Supplementary Materials for “Adaptive Selection of the Optimal Strategy to Improve Precision and Power in Randomized Trials” by Laura B. Balzer, Erica Cai, Lucas Godoy Garraza, and Pracheta Amaranath.

SUMMARY: The Supporting Information is organized as follows. In Web Appendix A, we provide a detailed discussion of targeted minimum loss-based estimation (TMLE) with and without Adaptive Prespecification (APS). In Web Appendix B, we provide additional simulation studies. In Web Appendix C, we provide more information and results for the application to real data from ACTG Study 175.
Web Appendix A. More on TMLE and Adaptive Pre-specification (APS)

Recall $W$ denotes the baseline covariates; $A$ is an indicator of being randomized to the intervention, and $Y$ is the outcome. Let $Y_i(a)$ denote the counterfactual outcome for randomized unit (i.e., participant) $i = \{1, \ldots, N\}$ under treatment-level $A = a$. When defining the causal effects, we take contrasts of “treatment-specific means”, defined at the population-level $\psi^p(a) = \mathbb{E}[Y(a)]$, conditional on covariates $\psi^c(a) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}[Y_i(a)|W_i]$, or for the study sample $\psi^s(a) = \frac{1}{N} \sum_{i=1}^{N} Y_i(a)$ (Neyman, 1923; Imbens, 2004; Imai, 2008; Balzer et al., 2015, 2016). For example, the population average treatment effect (PATE) is $\psi^p(1) - \psi^p(0)$; the conditional average treatment effect (CATE) is $\psi^c(1) - \psi^c(0)$, and the sample average treatment effect (SATE) is $\psi^s(1) - \psi^s(0)$. Of course, these treatment-specific means can also be contrasted on the relative or odds ratio scale.

A.1 Step-by-step implementation of TMLE in randomized trials

For a review of TMLE and its relation to other effect estimators in randomized trials, we refer to Colantuoni and Rosenblum (2015). For demonstration, we first focus on a binary outcome $Y \in \{0, 1\}$. In this setting and as described by Benkeser et al. (2021), we can adjust for covariates to improve precision when estimating the population risk difference (RD)$=\psi^p(1) - \psi^p(0)$ or risk ratio (RR)$=\frac{\psi^p(1)}{\psi^p(0)}$, as follows (Scharfstein et al., 1999; Moore and van der Laan, 2009; Rosenblum and van der Laan, 2010):

1. Use a working logistic regression model to estimate the conditional probability of the outcome, given the intervention indicator and selected covariates: $\mathbb{P}(Y = 1|A, W)$.
2. Using the fit from Step #1, predict the outcome under the intervention $\hat{\mathbb{P}}(Y = 1|A = 1, W_i)$ and under the control $\hat{\mathbb{P}}(Y = 1|A = 0, W_i)$ for $i = \{1, \ldots, N\}$.
3. Contrast the average arm-specific predictions to estimate the risk difference (RD) or risk ratio (RR), respectively:
\[
\hat{RD} = \frac{1}{N} \sum_{i=1}^{N} \hat{P}(Y = 1|A = 1, W_i) - \frac{1}{N} \sum_{i=1}^{N} \hat{P}(Y = 1|A = 0, W_i)
\]

\[
\hat{RR} = \frac{1}{N} \sum_{i=1}^{N} \hat{P}(Y = 1|A = 1, W_i) \div \frac{1}{N} \sum_{i=1}^{N} \hat{P}(Y = 1|A = 0, W_i)
\]

The same implementation is used for point estimation of conditional effects defined as contrasts of \(\psi^c(a)\) and for sample effects defined as contrasts of \(\psi^s(a)\) (Balzer et al., 2015, 2016).

Importantly, the working regression model in Step 1 is robust to misspecification. It can include interactions and other user-specified terms. Furthermore, the number of parameters (i.e., coefficients) to be fit in Step #1 would be equivalent to if we had, instead, used a linear working model. An important benefit of using logistic regression over linear regression is that the logit-link guarantees the bounds on the binary outcome will be respected (Gruber and van der Laan, 2010); this can be particularly important under data sparsity (e.g., due to rare outcomes). That said, when using Adaptive Prespecification (APS), we can consider candidate outcome regression estimators using both linear and logit-links, and let the data decide which candidate maximizes empirical efficiency, as detailed below.

Doubly robust methods incorporate information in the propensity score when solving the efficient score equation. In TMLE this is done during the targeting step. In a randomized trial, the propensity score \(\mathbb{P}(A = 1|W)\) is known and does not need to be estimated. For example, in a two-armed trial with equal allocation probability: \(\mathbb{P}(A = 1|W) = \mathbb{P}(A = 1) = 0.5\). When the propensity score is not estimated, TMLE reduces to the above approach (Rosenblum and van der Laan, 2010) and, therefore, requires estimation of the same number of regression parameters. However, collaborative estimation of the known propensity score has been shown repeatedly to increase efficiency (e.g., van der Laan and Robins (2003); van der Laan and Rose (2011); Balzer et al. (2016)).
For a binary outcome or, more generally, a bounded continuous outcome, we now present the full TMLE algorithm, including estimation of the propensity score and subsequent targeting. Steps 1-2 are the same as before. We use * to denote targeted estimates.

1. Use a working logistic regression model to estimate the conditional expectation of the outcome, given the intervention indicator and selected covariates: $E(Y|A,W)$.
2. Using the fit from Step #1, predict the outcome under the intervention $\hat{E}(Y|A = 1, W_i)$ and under the control $\hat{E}(Y|A = 0, W_i)$ for $i = \{1, \ldots, N\}$.
3. Target the initial outcome predictions $\hat{E}(Y|A,W)$ using information in the estimated propensity score $\hat{P}(A = 1|W)$. The targeting approach is driven by the efficient influence curve and, thus, depends on the statistical estimand of interest (van der Laan and Rose, 2011). One approach to simultaneously target effects defined on the difference, ratio, and odds ratio scales is to use a two-dimensional “clever covariate” (Moore and van der Laan, 2009):
   
   a. Use a working logistic regression model to estimate the conditional probability of the intervention given the selected covariates: $\mathbb{P}(A = 1|W)$.
   
   b. Using the propensity score fit from the previous step, calculate the clever covariates $\hat{H}(1, W_i) = \frac{1(A_i=1)}{\hat{P}(A=1|W_i)}$ and $\hat{H}(0, W_i) = \frac{1(A_i=0)}{\hat{P}(A=0|W_i)}$ for $i = \{1, \ldots, N\}$
   
   c. Estimate the fluctuation coefficients $\epsilon_1$ and $\epsilon_0$ in the following working logistic regression model:

   $$\text{logit}[\mathbb{E}^*(Y|A,W)] = \text{logit}[\mathbb{E}(Y|A,W)] + \epsilon_1 \hat{H}(1,W) + \epsilon_0 \hat{H}(0,W)$$

   Note there is no intercept and the coefficient for the logit of the initial estimates is set to 1. In practice, we run logistic regression of the observed outcome $Y$ on the the clever covariates $\hat{H}(1,W)$ and $\hat{H}(0,W)$ with the logit of the initial estimates $\mathbb{E}(Y|A,W)$ as offset. Using maximum likelihood estimation corresponds to solving the relevant component of the efficient score equation (a.k.a., the efficient influence
curve equation) (van der Laan and Rose, 2011). Denote the resulting estimates as \( \hat{\epsilon}_1 \) and \( \hat{\epsilon}_0 \).

(d) For \( i = \{1, \ldots, N\} \), obtain targeted outcome predictions under the intervention and control, calculated as

\[
\hat{E}^*(Y|A = 1, W_i) = \text{logit}^{-1}\left[ \text{logit}\{\hat{E}(Y|A = 1, W_i)\} + \frac{\hat{\epsilon}_1}{\text{P}(A = 1|W_i)} \right]
\]

\[
\hat{E}^*(Y|A = 0, W_i) = \text{logit}^{-1}\left[ \text{logit}\{\hat{E}(Y|A = 0, W_i)\} + \frac{\hat{\epsilon}_0}{\text{P}(A = 0|W_i)} \right]
\]

(4) Contrast the average arm-specific, targeted predictions to estimate the risk difference or risk ratio, respectively:

\[
\hat{RD}^* = \frac{1}{N} \sum_{i=1}^{N} \hat{E}^*(Y|A = 1, W_i) - \frac{1}{N} \sum_{i=1}^{N} \hat{E}^*(Y|A = 0, W_i)
\]

\[
\hat{RR}^* = \frac{1}{N} \sum_{i=1}^{N} \hat{E}^*(Y|A = 1, W_i) \div \frac{1}{N} \sum_{i=1}^{N} \hat{E}^*(Y|A = 0, W_i)
\]

If the effect on the difference scale is of primary interest, the targeting step can be simplified by using a one-dimensional clever covariate: \( \hat{H}(A, W) = \frac{1(A_i = 1)}{\text{P}(A = 1|W_i)} - \frac{1(A_i = 0)}{\text{P}(A = 0|W_i)} \). We could even define candidate TMLEs using alternative targeting approaches and select the optimal using APS. In our extension of APS for trials with many randomized units, we consider more flexible candidate estimators of the outcome regression in Step #1 and the propensity score in Step #3.

A.2. Statistical inference & influence curves (functions)

Under standard regularity conditions, which are weak in randomized trials, TMLE is an asymptotically linear estimator of the population, conditional, and sample effects (Moore and van der Laan, 2009; Rosenblum and van der Laan, 2010; van der Laan and Rose, 2011; Balzer et al., 2015, 2016). Therefore, the estimator minus the estimand can be written as an empirical mean of its influence curve plus a remainder term going to 0 in probability. For
example, for the TMLE $\hat{\psi}^p(a)$ of the population parameter $\psi^p(a) = \mathbb{E}[Y(a)]$, we have

$$\hat{\psi}^p(a) - \psi^p(a) = \frac{1}{N} \sum_{i=1}^{N} D^p(a, O_i) + o_p(1/\sqrt{N})$$

where $D^p(a, O_i)$ is the influence curve for observation $O_i$. (The form of the influence curve is given explicitly below.) Asymptotically linear estimators enjoy properties following from the Central Limit Theorem. In particular, the standardized estimator is normally distributed in the large-data limit with variance given by the variance of its influence curve. In randomized trials, TMLE is locally efficient in that it will have the lowest possible asymptotic variance if the working model for the outcome regression is correctly specified; then its influence curve will equal the efficient influence curve.

The variance of asymptotically linear estimators is conveniently estimated by the sample variance of the estimated influence curve divided by sample size $N$. With a variance estimate, we can create Wald-Type 95% confidence intervals and test the null hypothesis. Additionally, building on the principle of empirical efficiency maximization (Rubin and van der Laan, 2008), we can use the estimated influence curve-squared as our measure of performance in APS, as outlined below (Balzer et al., 2016).

For the population parameter $\psi^p(a)$, the estimated influence curve for TMLE for observation $O_i$ is given by

$$\hat{D}^p(a; O_i) = \left( \frac{1}{\mathbb{P}(A = a|W_i)} \right) \left( Y_i - \mathbb{E}(Y|A, W_i) \right) + \mathbb{E}(Y|A = a, W_i) - \hat{\psi}^p(a) \quad (1)$$

when the propensity score $\mathbb{P}(A = a|W)$ is treated as known (i.e., no targeting is done) and

$$\hat{D}^{ps}(a; O_i) = \left( \frac{1}{\hat{\mathbb{P}}(A = a|W_i)} \right) \left( Y_i - \hat{\mathbb{E}}^*(Y|A, W_i) \right) + \hat{\mathbb{E}}^*(Y|A = a, W_i) - \hat{\psi}^{ps}(a) \quad (2)$$

when we have estimated the propensity score $\hat{\mathbb{P}}(A = a|W)$ and obtained a targeted estimator. The latter provides a conservative approximation (pp. 96 of van der Laan and Rose (2011)). As detailed in Balzer et al. (2015, 2016), the influence curve of TMLE for both the conditional
parameter $\psi^c(a)$ and the sample parameter $\psi^s(a)$ is conservatively approximated by
\[
\hat{D}^c(a; O_i) = \hat{D}^s(a; O_i) = \left( \mathbb{1}(A_i = a) \right) \left( Y_i - \hat{E}(Y | A, W_i) \right)
\] (3)
when the propensity score $\mathbb{P}(A = a | W)$ is treated as known (i.e., no targeting is done) and
\[
\hat{D}^{c*}(a; O_i) = \hat{D}^{s*}(a; O_i) = \left( \mathbb{1}(A_i = a) \right) \left( Y_i - \hat{E}^*(Y | A, W_i) \right)
\] (4)
when we have estimated the propensity score $\hat{P}(A = a | W)$ and obtained a targeted estimator.

By applying the Delta Method, we can obtain inference for any contrast of treatment-specific means (e.g., the difference, ratio, or odds ratio). For example, the estimated influence curve for TMLE for the $\text{PATE}=\psi^p(1) - \psi^p(0)$ is given by $\hat{D}^{p*}(1; O) - \hat{D}^{p*}(0; O)$. For relative effects, it is easier to obtain inference on logarithmic scale and then transform back (Moore and van der Laan, 2009). For example, the estimated influence curve for TMLE of the sample risk ratio on the logarithmic scale $\log\left\{ \psi^s(1) / \psi^s(0) \right\}$ is given by $\hat{D}^{s*}(1; O) / \hat{\psi}^{s*}(1) - \hat{D}^{s*}(0; O) / \hat{\psi}^{s*}(0)$.

A.3. Maximizing empirical efficiency with Adaptive Prespecification (APS)

The goal of APS is to have a fully automated and data-adaptive approach to select, from a prespecified set, the candidate estimation approach that maximizes precision. A schematic of APS to select the optimal TMLE is given in Figure 1 of the main text. We provide additional mathematical details here.

We consider $V$-fold cross-validation, and let $K_v$ denote the set of indices for the observations in fold $v$. Additionally, let superscript $-v$ denote estimators fit on data excluding observations in fold $v$. First, we select the candidate TMLE to minimize the cross-validated variance estimate when treating the propensity score as known:
\[
\text{CV-Risk} = \frac{1}{V} \sum_{v=1}^{V} \left[ \frac{1}{n_v} \sum_{k \in K_v} \left\{ \hat{D}^{-v}(O_k) \right\}^2 \right]
\]
where $n_v = |K_v|$ denotes the number of observations in fold $v$ and $\hat{D}^{-v}(O_k)$ denotes the cross-validated estimate of influence curve for the TMLE without targeting (Eqs. 1 and 3).
For example, for the PATE=$\psi^p(1) - \psi^p(0)$, we would choose the candidate estimator of the outcome regression that minimized $\frac{1}{V} \sum_{v=1}^{V} \frac{1}{n_v} \sum_{k \in K_v} \left\{ \hat{D}^{\psi,-v}(1, O_k) - \hat{D}^{\psi,-v}(0, O_k) \right\}^2$. For the SATE=$\psi^s(1) - \psi^s(0)$ in a two-armed trial with equal allocation probability $\mathbb{P}(A = 1) = 0.5$, this corresponds to selecting the candidate outcome regression estimator that minimizes the L2 loss (Balzer et al., 2016).

Second, given the selected outcome regression estimator, we select the candidate TMLE, which further minimizes the cross-validated variance estimate after targeting based on the candidate estimator of the propensity score:

$$\text{CV-Risk} = \frac{1}{V} \sum_{v=1}^{V} \left[ \frac{1}{n_v} \sum_{k \in K_v} \left\{ \hat{D}^{\psi,-v}(O_k) \right\}^2 \right]$$

where $\hat{D}^{\psi,-v}(O_k)$ denotes the cross-validated estimate of influence curve for the TMLE after targeting (Eqs. 2 and 4). For example, for the PATE=$\psi^p(1) - \psi^p(0)$, we would chose the candidate TMLE minimizing $\frac{1}{V} \sum_{v=1}^{V} \frac{1}{n_v} \sum_{k \in K_v} \left\{ \hat{D}^{\psi,-v}(1, O_k) - \hat{D}^{\psi,-v}(0, O_k) \right\}^2$. Importantly, the unadjusted estimator of the known propensity score is always included. Therefore, if adjusting during propensity score estimation does not further improve precision, the algorithm defaults to the TMLE only adjusting with the optimal estimator for the outcome regression. Equally important, the unadjusted estimator of the outcome regression is always included. Therefore, if adjusting when fitting the outcome regression does not improve precision, the algorithm defaults to unadjusted effect estimator.

### A.4. Alternative designs

Throughout, we have focused on two-armed trials with simple randomization. For discussion of TMLE applied in trials with covariate-adaptive randomization schemes, such as pair-matching and stratification, we refer the reader to Balzer et al. (2015, 2016); Wang et al. (2021). For extensions to cluster randomized trials, including inference for effects defined at the cluster- or individual-level and inference with arm-specific dependence structures, we refer the reader to Benitez et al. (2023); Nugent (2022).
Web Appendix B. More details on the simulation

B.1. Data generating processes for the continuous outcome simulations

To study performance with a continuous outcome and on the difference scale, we conducted the following simulation study. For 5000 simulated trials each with 500 participants, we generated 5 measured covariates and 2 unmeasured covariates:

\[ W_1 \sim \text{Bern}(p = 0.5), \quad W_2 \sim \text{Bern}(p = 0.2 + W_1), \quad W_3 \sim \text{Unif}(0, 5), \quad W_4 = \text{logit}^{-1}(-2 + W_1 + W_2 + \text{Unif}(0, 2)), \quad W_5 = 1 + \text{Binom}(3, p = 0.3), \quad U_1 \sim \text{Unif}(0, 0.5), \quad \text{and} \quad U_2 \sim \text{Unif}(0, 1), \]

respectively. We then generated the continuous, counterfactual outcomes \( Y(a) \) in 3 settings of varying complexity:

1. “Linear”: \( Y(a) = 90 + 0.07a + 0.7W_1 + 0.3W_2 + 0.1W_3 + 0.3W_4 + 0.4W_5 + 0.25aW_1 + 5U_1 + U_2 \)
2. “Interactive”: \( Y(a) = 150 + 0.05a + 0.33W_1 - 0.25W_2 + 0.5W_3 - 0.2W_4 + 0.05W_5 + 0.1aW_1 + 0.02aW_3 + 0.3aU_1 + 5.8U_1 + U_2 \)
3. “Polynomial”: \( Y(a) = 90 + 0.17a + 0.33(W_1 + W_2 + W_3 + W_4 + W_5) - 0.2W_1W_3 + 0.5W_1(0.8 - 0.6W_4)W_3 + 0.25(1 - W_1)(-0.2 + 0.15W_4) + 4.7U_1 + U_2 \).

We additionally considered a “Treatment only” scenario where none of the measured covariates influences the outcome: \( Y(a) = 90 + 0.1a + 3U_1 + U_2 \). As before, we generated the observed treatment \( A \) using simple randomization and randomization within strata defined by \( \mathbb{1}(W_1 > 0) \). Again, we set the observed outcome \( Y \) equal to the counterfactual outcome \( Y(a) \) when \( A = a \).

B.2. Additional simulation results

In Tables 1-4, we provide additional simulation results. Throughout “DGP” denotes the data generating process, and “Design” refers to whether the trial used simple randomization or stratified randomization.

In Tables 1 and 2, we summarize the selection of candidate estimators for the outcome regression and for the propensity score in the large-trial APS implementation for the 5000
simulated trials when there was an effect. The candidate estimators for the outcome regression were the unadjusted estimator ("Unadjust"), a working generalized linear model adjusting for a single baseline covariate ("GLM"), main terms regression adjusting for all covariates ("Main"), stepwise regression ("Step"), stepwise regression with all possible pairwise interactions ("StepInt"), least absolute shrinkage and selection operator ("LASSO"), and to multivariate adaptive regression splines ("MARS"). For the propensity score, we used a similar set of candidates, but excluded stepwise regression with interactions and MARS. We see that selection is adaptive to the data generating process for both outcome types.

[Table 1 about here.]

[Table 2 about here.]

In Tables 3 and 4, we provide simulation results when there is no effect (i.e., under the null). "Cover." denotes the 95% confidence interval coverage; "Type-I" denotes the proportion of trials where the true null hypothesis was rejected; "MSE" denotes the mean squared error; "Var." denotes the variance of the point estimates, and "Rel.Eff." denotes relative efficiency, approximated by the ratio of the MSE of a given estimator to that of the unadjusted estimator. "Unadjusted" refers to the unadjusted effect estimator, "Static" to forced adjustment for \( W_1 \) in the outcome regression, "Small APS" to TMLE with the small-trial implementation of APS, and "Large APS" to TMLE with the large-trial implementation of APS.

[Table 3 about here.]

[Table 4 about here.]
Web Appendix C. Additional information and results for the real data application

In Table 5, we provide the baseline characteristics, used as candidate adjustment variables, in the application to ACTG 175 Study (Hammer et al., 1996), whose data are publicly available in the `speff2trial` R package (Juraska et al., 2022). While there was an imbalance in the number randomized to the intervention ($N = 1587$) and to the control ($N = 526$), the baseline covariates were well-balanced between arms.

[Table 5 about here.]

In Tables 6 and 7, we provide additional results for the effect on average CD4 count (continuous outcome) and the probability of having a CD4 count $>$ 350 c/mm$^3$ (binary outcome), overall and within subgroups defined by age (18-29 years and 30+ years) and gender. A coding vignette to reproduce these and additional results is available at https://github.com/LauraBalzer/AdaptivePrespec.

“Small TMLE” and “Large TMLE” refer to using APS to only select the outcome regression estimator in the small-trial and large-trial implementation, respectively. “Small CTMLE” and “Large CTMLE” refer to using APS for selection of the outcome regression estimator and collaborative selection of the known propensity score estimator in the small-trial and large-trial implementation, respectively. “Out.Reg.” and “PScore” refer to the fixed or adaptively selected estimator of the outcome regression and propensity score, respectively. For estimation of the outcome regression and the propensity score, the large-trial APS considered the unadjusted estimator, adjustment for a single covariate a working generalized linear model (“GLM”), a main terms regression adjusting for all covariates (“Main”), step-wise regression, LASSO, MARS, and MARS after screening based on pairwise correlations. As before, the small-trial implementation was limited to the unadjusted estimator and adjustment for a single covariate a working generalized linear model.
References

Balzer, L., Petersen, M., and van der Laan, M. (2016). Targeted estimation and inference of the sample average treatment effect in trials with and without pair-matching. *Stat Med* **35**, 3717–3732.

Balzer, L., Petersen, M., van der Laan, M., and the SEARCH Consortium (2015). Adaptive pair-matching in randomized trials with unbiased and efficient effect estimation. *Stat Med* **34**, 999–1011.

Balzer, L., van der Laan, M., Petersen, M., and SEARCH (2016). Adaptive pre-specification in randomized trials with and without pair-matching. *Stat Med* **35**, 4528–4545.

Benitez, A., Petersen, M., van der Laan, M., Santos, N., et al. (2023). Comparative methods for the analysis of cluster randomized trials. *Stat Med* **42**, 3443–3466.

Benkeser, D., Díaz, I., Luedtke, A., Segal, J., Scharfstein, D., and Rosenblum, M. (2021). Improving precision and power in randomized trials for COVID-19 treatments using covariate adjustment for binary, ordinal, and time-to-event outcomes (with rejoinder). *Biometrics* **77**, 1467–1481.

Colantuoni, E. and Rosenblum, M. (2015). Leveraging prognostic baseline variables to gain precision in randomized trials. *Stat Med* **34**, 2602–2617.

Gruber, S. and van der Laan, M. (2010). A targeted maximum likelihood estimator of a causal effect on a bounded continuous outcome. *The International Journal of Biostatistics* **6**, Article 26.

Hammer, S., Katzenstein, D., Hughes, M., Gundacker, H., Schooley, R., Haubrich, R., et al.
(1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per mm$^3$. *NEJM* **335**, 1081–1090.

Imai, K. (2008). Variance identification and efficiency analysis in randomized experiments under the matched-pair design. *Statistics in Medicine* **27**, 4857–4873.

Imbens, G. (2004). Nonparametric estimation of average treatment effects under exogeneity: a review. *Review of Economics and Statistics* **86**, 4–29.

Juraska, M., with contributions from P.B. Gilbert, Lu, X., Zhang, M., Davidian, M., and Tsiatis, A. (2022). `speff2trial`: Semiparametric Efficient Estimation for a Two-Sample Treatment Effect. R package version 1.0.5.

Moore, K. and van der Laan, M. (2009). Covariate adjustment in randomized trials with binary outcomes: Targeted maximum likelihood estimation. *Stat Med* **28**, 39–64.

Neyman, J. (1923). Sur les applications de la theorie des probabilites aux experiences agricoles: Essai des principes (In Polish). English translation by D.M. Dabrowska and T.P. Speed (1990). *Statistical Science* **5**, 465–480.

Nugent, J. (2022). Estimation of causal effects in complex clustered data. [https://scholarworks.umass.edu/dissertations_2/2706/](https://scholarworks.umass.edu/dissertations_2/2706/).

Rosenblum, M. and van der Laan, M. (2010). Simple, efficient estimators of treatment effects in randomized trials using generalized linear models to leverage baseline variables. *Int J Biostat* **6**, Article 13.

Rubin, D. and van der Laan, M. (2008). Empirical efficiency maximization. *Int J Biostat* **4**, Article 5.

Scharfstein, D., Rotnitzky, A., and Robins, J. (1999). Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models (with Rejoinder). *Journal of the American Statistical Association* **94**, 1096–1120 (1135–1146).

van der Laan, M. and Robins, J. (2003). *Unified Methods for Censored Longitudinal Data*
and Causality. Springer-Verlag, New York Berlin Heidelberg.

van der Laan, M. and Rose, S. (2011). Targeted Learning: Causal Inference for Observational and Experimental Data. Springer, New York Dordrecht Heidelberg London.

Wang, B., Susukida, R., Mojtabai, R., Amin-Esmaeili, M., and Rosenblum, M. (2021). Model-robust inference for clinical trials that improve precision by stratified randomization and covariate adjustment. J Am Stat Assoc 0, 1–12.
Table 1

Proportion of times (%) a candidate algorithm was selected for estimation of the outcome regression ("OutReg") and the propensity score ("PScore") in TMLE with the large-trial APS in the **binary outcome** setting when there was an effect. Stepwise regression with all possible pairwise interactions and MARS were not considered as candidates for the propensity score.

| Target | DGP     | Design      | Unadj | GLM  | Main | Step | StepInt | LASSO | MARS |
|--------|---------|-------------|-------|------|------|------|---------|-------|------|
| OutReg | Linear  | Simple      | 0%    | 0%   | 0.2% | 0%   | 98.1%   | 0.1%  | 1.6% |
|        | Linear  | Stratified  | 0%    | 0%   | 0.1% | 0%   | 98%     | 0.1%  | 1.8% |
|        | Interactive | Simple | 0%    | 0%   | 25.5%| 0%   | 17.3%   | 56.9% | 0.3% |
|        | Interactive | Stratified | 0%    | 0%   | 25.3%| 0%   | 17%     | 57.2% | 0.4% |
|        | Polynomial | Simple      | 0%    | 0%   | 31.7%| 0%   | 12.1%   | 55.7% | 0.5% |
|        | Polynomial | Stratified  | 0%    | 0%   | 31.3%| 0%   | 12.1%   | 55.9% | 0.7% |
| PScore | Linear  | Simple      | 34.9% | 64.3%| 0.2% | 0%   | -       | 0.6%  | -    |
|        | Linear  | Stratified  | 37.3% | 62.2%| 0.1% | 0%   | -       | 0.4%  | -    |
|        | Interactive | Simple | 35%   | 64.2%| 0.2% | 0%   | -       | 0.6%  | -    |
|        | Interactive | Stratified | 36%   | 63%  | 0.3% | 0%   | -       | 0.7%  | -    |
|        | Polynomial | Simple      | 40%   | 59.4%| 0.2% | 0%   | -       | 0.4%  | -    |
|        | Polynomial | Stratified  | 39.9% | 59.3%| 0.2% | 0%   | -       | 0.6%  | -    |
Table 2
Proportion of times (%) a candidate algorithm was selected for estimation of the outcome regression (“OutReg”) and the propensity score (“PScore”) in TMLE with the large-trial APS in the continuous outcome setting when there was an effect. Stepwise regression with all possible pairwise interactions and MARS were not considered as candidates for the propensity score.

| Target | DGP     | Design | Unadj | GLM | Main | Step | StepInt | LASSO | MARS |
|--------|---------|--------|-------|-----|------|------|---------|-------|------|
| OutReg | Linear  | Simple | 0%    | 0%  | 6.7% | 0%   | 0%      | 91.9% | 1.4% |
|        | Linear  | Stratified | 0% | 0% | 5.8% | 0% | 0% | 92.8% | 1.3% |
|        | Interactive | Simple | 0% | 0.5% | 9.4% | 0% | 0% | 89.4% | 0.7% |
|        | Interactive | Stratified | 0% | 0.6% | 8.3% | 0% | 0% | 90.4% | 0.7% |
|        | Polynomial | Simple | 0% | 0% | 13% | 0% | 0% | 86.5% | 0.5% |
|        | Polynomial | Stratified | 0% | 0% | 12% | 0% | 0% | 87.6% | 0.5% |
| PScore | Linear  | Simple | 68.8% | 31.2% | 0% | 0% | - | 0% | - |
|        | Linear  | Stratified | 72.5% | 27.5% | 0% | 0% | - | 0% | - |
|        | Interactive | Simple | 70.5% | 29.5% | 0% | 0% | - | 0% | - |
|        | Interactive | Stratified | 73.5% | 26.5% | 0% | 0% | - | 0% | - |
|        | Polynomial | Simple | 68.2% | 31.8% | 0% | 0% | - | 0% | - |
|        | Polynomial | Stratified | 71.9% | 28.1% | 0% | 0% | - | 0% | - |
Table 3

Simulation results for the binary outcome under the null; the true value of the sample risk ratio is 1. “Txt only” refers to the setting when there are no prognostic covariates.

| DGP        | Design  | Estimator | Cover. | Type-I | MSE | Bias | Var. | Rel.Eff. |
|------------|---------|-----------|--------|--------|-----|------|------|----------|
| Txt only   | Simple  | Unadjusted| 0.952  | 0.048  | 0.003| 0.001| 0.003| 1.000    |
|            |         | Static    | 0.951  | 0.049  | 0.003| 0.001| 0.003| 1.001    |
|            |         | Small APS | 0.952  | 0.048  | 0.003| 0.001| 0.003| 0.988    |
|            |         | Large APS | 0.953  | 0.047  | 0.003| 0.001| 0.003| 0.983    |
| Stratified | Unadjusted| Static  | 0.951  | 0.049  | 0.003| 0.001| 0.003| 1.000    |
|            |         | Small APS | 0.951  | 0.049  | 0.003| 0.001| 0.003| 0.988    |
|            |         | Large APS | 0.952  | 0.048  | 0.003| 0.001| 0.003| 0.988    |
| Linear     | Simple  | Unadjusted| 0.953  | 0.047  | 0.006| 0.003| 0.006| 1.000    |
|            |         | Static    | 0.952  | 0.048  | 0.005| 0.003| 0.005| 0.900    |
|            |         | Small APS | 0.963  | 0.037  | 0.005| 0.003| 0.005| 0.783    |
|            |         | Large APS | 0.950  | 0.050  | 0.004| 0.002| 0.004| 0.659    |
| Stratified | Unadjusted| Static  | 0.956  | 0.044  | 0.006| 0.002| 0.006| 1.000    |
|            |         | Small APS | 0.949  | 0.051  | 0.005| 0.002| 0.005| 0.958    |
|            |         | Large APS | 0.962  | 0.038  | 0.005| 0.002| 0.005| 0.824    |
|            |         | Large APS | 0.949  | 0.051  | 0.004| 0.001| 0.004| 0.698    |
| Interactive| Simple  | Unadjusted| 0.954  | 0.046  | 0.003| 0.002| 0.003| 1.000    |
|            |         | Static    | 0.953  | 0.047  | 0.003| 0.001| 0.003| 0.898    |
|            |         | Small APS | 0.961  | 0.039  | 0.002| 0.001| 0.002| 0.793    |
|            |         | Large APS | 0.945  | 0.055  | 0.002| 0.001| 0.002| 0.677    |
| Stratified | Unadjusted| Static  | 0.955  | 0.045  | 0.003| 0.001| 0.003| 1.000    |
|            |         | Small APS | 0.946  | 0.054  | 0.003| 0.001| 0.003| 0.962    |
|            |         | Large APS | 0.959  | 0.041  | 0.002| 0.001| 0.002| 0.822    |
|            |         | Large APS | 0.939  | 0.061  | 0.002| 0.001| 0.002| 0.730    |
| Polynomial | Simple  | Unadjusted| 0.951  | 0.049  | 0.005| 0.004| 0.005| 1.000    |
|            |         | Static    | 0.945  | 0.055  | 0.004| 0.003| 0.004| 0.838    |
|            |         | Small APS | 0.954  | 0.046  | 0.004| 0.003| 0.004| 0.771    |
|            |         | Large APS | 0.943  | 0.057  | 0.003| 0.002| 0.003| 0.633    |
| Stratified | Unadjusted| Static  | 0.964  | 0.036  | 0.004| 0.003| 0.004| 1.000    |
|            |         | Small APS | 0.951  | 0.049  | 0.004| 0.003| 0.004| 0.927    |
|            |         | Large APS | 0.959  | 0.041  | 0.003| 0.002| 0.003| 0.847    |
Table 4
Estimator performance with the *continuous outcome under the null*, the true value of the SATE is 0. “Txt only” refers to the setting when there are no prognostic covariates.

| DGP    | Design | Estimator | Cover. | Type-I | MSE | Bias | Var. | Rel.Eff. |
|--------|--------|-----------|--------|--------|-----|------|------|----------|
|_txt only_ | Simple | Unadjusted | 0.950  | 0.050  | 0.002 | -0.001 | 0.002 | 1.000    |
|        |        | Static    | 0.949  | 0.051  | 0.002 | -0.001 | 0.002 | 1.003    |
|        |        | Small APS | 0.949  | 0.051  | 0.002 | -0.001 | 0.002 | 1.006    |
|        |        | Large APS | 0.950  | 0.050  | 0.002 | -0.001 | 0.002 | 1.005    |
| Stratified |        | Unadjusted | 0.945  | 0.055  | 0.002 | 0.000  | 0.002 | 1.000    |
|        |        | Static    | 0.945  | 0.055  | 0.002 | 0.000  | 0.002 | 1.000    |
|        |        | Small APS | 0.946  | 0.054  | 0.002 | 0.000  | 0.002 | 1.003    |
|        |        | Large APS | 0.945  | 0.055  | 0.002 | 0.000  | 0.002 | 1.005    |
| Linear | Simple | Unadjusted | 0.953  | 0.047  | 0.007 | -0.001 | 0.007 | 1.000    |
|        |        | Static    | 0.947  | 0.053  | 0.006 | -0.002 | 0.006 | 0.836    |
|        |        | Small APS | 0.948  | 0.052  | 0.006 | -0.002 | 0.006 | 0.825    |
|        |        | Large APS | 0.947  | 0.053  | 0.005 | -0.001 | 0.005 | 0.664    |
| Stratified |        | Unadjusted | 0.966  | 0.034  | 0.006 | 0.001  | 0.006 | 1.000    |
|        |        | Static    | 0.945  | 0.055  | 0.006 | 0.001  | 0.006 | 1.000    |
|        |        | Small APS | 0.948  | 0.052  | 0.006 | 0.001  | 0.006 | 0.969    |
|        |        | Large APS | 0.944  | 0.056  | 0.005 | 0.001  | 0.005 | 0.799    |
| Interactive | Simple | Unadjusted | 0.948  | 0.052  | 0.011 | -0.001 | 0.011 | 1.000    |
|        |        | Static    | 0.946  | 0.054  | 0.011 | -0.001 | 0.011 | 0.990    |
|        |        | Small APS | 0.948  | 0.052  | 0.007 | -0.001 | 0.007 | 0.611    |
|        |        | Large APS | 0.946  | 0.054  | 0.006 | -0.001 | 0.006 | 0.585    |
| Stratified |        | Unadjusted | 0.942  | 0.058  | 0.011 | 0.003  | 0.011 | 1.000    |
|        |        | Static    | 0.939  | 0.061  | 0.011 | 0.003  | 0.011 | 1.000    |
|        |        | Small APS | 0.947  | 0.053  | 0.007 | 0.001  | 0.007 | 0.601    |
|        |        | Large APS | 0.943  | 0.057  | 0.006 | 0.001  | 0.006 | 0.583    |
| Polynomial | Simple | Unadjusted | 0.948  | 0.052  | 0.008 | -0.001 | 0.008 | 1.000    |
|        |        | Static    | 0.943  | 0.057  | 0.007 | -0.001 | 0.007 | 0.938    |
|        |        | Small APS | 0.952  | 0.048  | 0.006 | -0.002 | 0.006 | 0.729    |
|        |        | Large APS | 0.948  | 0.052  | 0.004 | -0.001 | 0.004 | 0.585    |
| Stratified |        | Unadjusted | 0.948  | 0.052  | 0.007 | 0.002  | 0.007 | 1.000    |
|        |        | Static    | 0.940  | 0.060  | 0.007 | 0.002  | 0.007 | 1.000    |
|        |        | Small APS | 0.957  | 0.043  | 0.005 | 0.001  | 0.005 | 0.712    |
|        |        | Large APS | 0.942  | 0.058  | 0.005 | 0.001  | 0.005 | 0.620    |
Table 5
Baseline characteristics and candidate adjustment variables in the real data application to the ACTG 175 Study, by arm and overall (Hammer et al., 1996). Continuous variables are shown in median [Q1, Q3], and binary variables are shown in N (%).

|                                | Intervention | Control   | Overall   |
|--------------------------------|--------------|-----------|-----------|
|                                | N=1587       | N=526     | N=2113    |
| Age (years)                    | 34 [29,40]   | 34 [29,40]| 34 [29,40]|
| Aged 18-29 years               | 399 (25%)    | 137 (26%) | 536 (25%) |
| Male                           | 1319 (83%)   | 427 (81%) | 1746 (83%)|
| Non-white race                 | 456 (29%)    | 154 (29%) | 610 (29%) |
| Weight (kg)                    | 74 [67,82]   | 75 [68,84]| 74 [67,83]|
| Has hemophilia                 | 118 (7%)     | 37 (7%)   | 155 (7%)  |
| Karnofsky score (scale 0-100)  | 100 [90,100] | 100 [90,100]| 100 [90,100]|
| Symptomatic                    | 279 (18%)    | 88 (17%)  | 367 (17%) |
| ART experienced                | 926 (58%)    | 307 (58%) | 1233 (58%)|
| Time on ART (days)             | 139 [0,739]  | 138 [0,731]| 139 [0,735]|
| Started ART 1-52wks prior      | 312 (20%)    | 96 (18%)  | 408 (19%) |
| Non-zidovudine prior to baseline| 31 (2%)      | 16 (3%)   | 47 (2%)   |
| Baseline CD4 count (cells/mm³) | 339 [260,423]| 346 [271,422]| 340 [263,423]|
| Baseline CD4>350               | 724 (46%)    | 252 (48%) | 976 (46%) |
| Baseline CD8 count (cells/mm³) | 897 [655,1212]| 880 [656,1190]| 894 [655,1210]|
| Baseline CD48>350              | 1540 (97%)   | 520 (99%) | 2060 (97%)|
Table 6
Additional results from the real data application for the additive effect on the continuous outcome: CD4 count at week 20. “Younger” refers to age 18-29 years, while “older” refers to age ≥30+ years.

| Estimator     | Effect (95%CI)          | Rel.Var. | Out.Reg. | PScore |
|---------------|-------------------------|----------|----------|--------|
| Overall (N=2113) |                         |          |          |        |
| Unadjusted    | 46.4 (33.0, 59.7)       | 1.000    | Unadj.   | Unadj. |
| Static        | 46.8 (33.5, 60.0)       | 0.991    | Fixed    | Fixed  |
| Small TMLE    | 48.4 (37.9, 58.9)       | 0.621    | GLM      | Unadj. |
| Small CTMLE   | 48.5 (38.0, 59.0)       | 0.617    | GLM      | GLM    |
| Large TMLE    | 47.7 (37.9, 57.6)       | 0.546    | MARS     | Unadj. |
| Large CTMLE   | 47.8 (38.0, 57.6)       | 0.542    | MARS     | GLM    |
| Older women (N=258) |                       |          |          |        |
| Unadjusted    | 53.6 (18.4, 88.7)       | 1.000    | Unadj.   | Unadj. |
| Static        | 53.3 (18.5, 88.1)       | 0.979    | Fixed    | Fixed  |
| Small TMLE    | 62.5 (32.5, 92.4)       | 0.723    | GLM      | Unadj. |
| Small CTMLE   | 64.3 (35.3, 93.4)       | 0.682    | GLM      | GLM    |
| Large TMLE    | 62.5 (32.5, 92.4)       | 0.723    | GLM      | Unadj. |
| Large CTMLE   | 64.3 (35.3, 93.4)       | 0.682    | GLM      | GLM    |
| Younger women (N=109) |                  |          |          |        |
| Unadjusted    | -13.6 (-81.4, 54.2)     | 1.000    | Unadj.   | Unadj. |
| Static        | -14.0 (-81.8, 53.9)     | 1.002    | Fixed    | Fixed  |
| Small TMLE    | 29.6 (-15.8, 75.1)      | 0.449    | GLM      | Unadj. |
| Small CTMLE   | 37.3 (-4.9, 79.4)       | 0.386    | GLM      | GLM    |
| Large TMLE    | 29.6 (-15.8, 75.1)      | 0.449    | GLM      | Unadj. |
| Large CTMLE   | 37.3 (-4.9, 79.4)       | 0.386    | GLM      | GLM    |
| Older men (N=1319) |                    |          |          |        |
| Unadjusted    | 50.8 (34.3, 67.4)       | 1.000    | Unadj.   | Unadj. |
| Static        | 50.8 (34.2, 67.3)       | 0.999    | Fixed    | Fixed  |
| Small TMLE    | 50.2 (37.0, 63.4)       | 0.637    | GLM      | Unadj. |
| Small CTMLE   | 50.3 (37.1, 63.4)       | 0.636    | GLM      | GLM    |
| Large TMLE    | 48.9 (36.2, 61.6)       | 0.591    | LASSO    | Unadj. |
| Large CTMLE   | 48.9 (36.2, 61.6)       | 0.591    | LASSO    | GLM    |
| Younger men (N=427) |                   |          |          |        |
| Unadjusted    | 47.4 (17.7, 77.1)       | 1.000    | Unadj.   | Unadj. |
| Static        | 46.2 (16.6, 75.8)       | 0.994    | Fixed    | Fixed  |
| Small TMLE    | 42.0 (19.1, 65.0)       | 0.599    | GLM      | Unadj. |
| Small CTMLE   | 41.2 (18.3, 64.2)       | 0.598    | GLM      | GLM    |
| Large TMLE    | 45.6 (23.9, 67.4)       | 0.538    | Main     | Unadj. |
| Large CTMLE   | 45.0 (23.3, 66.7)       | 0.534    | Main     | GLM    |
Table 7
Additional results from the real data application for the relative effect on the binary outcome: having a CD4 count > 350 c/mm$^3$ at week 20. Young" refers to age 18-29 years, while "old" refers to age 30+ years.

| Estimator      | Effect (95%CI) | Rel.Var. | Out.Reg. | PScore |
|----------------|----------------|----------|----------|--------|
| Overall (N=2113) | Unadjusted     | 1.23 (1.10, 1.37) | 1.000 | Unadj. | Unadj.  |
|                | Static         | 1.23 (1.11, 1.37) | 1.001 | Fixed  | Fixed   |
|                | Small TMLE     | 1.26 (1.15, 1.38) | 0.707 | GLM    | Unadj.  |
|                | Small CTMLE    | 1.26 (1.15, 1.38) | 0.702 | GLM    | GLM     |
|                | Large TMLE     | 1.25 (1.15, 1.37) | 0.678 | LASSO  | Unadj.  |
|                | Large CTMLE    | 1.26 (1.15, 1.37) | 0.672 | LASSO  | GLM     |
| Older women (N=258) | Unadjusted | 1.21 (0.89, 1.65) | 1.000 | Unadj. | Unadj.  |
|                | Static         | 1.21 (0.89, 1.65) | 0.984 | Fixed  | Fixed   |
|                | Small TMLE     | 1.28 (0.96, 1.71) | 0.877 | GLM    | Unadj.  |
|                | Small CTMLE    | 1.27 (0.96, 1.68) | 0.847 | GLM    | GLM     |
|                | Large TMLE     | 1.28 (0.96, 1.71) | 0.877 | GLM    | Unadj.  |
|                | Large CTMLE    | 1.27 (0.96, 1.69) | 0.841 | GLM    | GLM     |
| Younger women (N=109) | Unadjusted | 1.05 (0.71, 1.55) | 1.000 | Unadj. | Unadj.  |
|                | Static         | 1.04 (0.71, 1.53) | 1.000 | Fixed  | Fixed   |
|                | Small TMLE     | 1.35 (1.00, 1.83) | 0.606 | GLM    | Unadj.  |
|                | Small CTMLE    | 1.35 (1.01, 1.81) | 0.571 | GLM    | GLM     |
|                | Large TMLE     | 1.35 (1.00, 1.83) | 0.606 | GLM    | Unadj.  |
|                | Large CTMLE    | 1.35 (1.01, 1.81) | 0.571 | GLM    | GLM     |
| Older men (N=1319) | Unadjusted    | 1.25 (1.09, 1.45) | 1.000 | Unadj. | Unadj.  |
|                | Static         | 1.25 (1.09, 1.44) | 1.000 | Fixed  | Fixed   |
|                | Small TMLE     | 1.25 (1.11, 1.41) | 0.702 | GLM    | Unadj.  |
|                | Small CTMLE    | 1.25 (1.11, 1.41) | 0.702 | GLM    | GLM     |
|                | Large TMLE     | 1.25 (1.11, 1.40) | 0.672 | LASSO  | Unadj.  |
|                | Large CTMLE    | 1.25 (1.11, 1.40) | 0.669 | LASSO  | GLM     |
| Younger men (N=427) | Unadjusted | 1.23 (0.98, 1.56) | 1.000 | Unadj. | Unadj.  |
|                | Static         | 1.23 (0.97, 1.55) | 0.988 | Fixed  | Fixed   |
|                | Small TMLE     | 1.24 (1.03, 1.49) | 0.616 | GLM    | Unadj.  |
|                | Small CTMLE    | 1.24 (1.03, 1.49) | 0.612 | GLM    | GLM     |
|                | Large TMLE     | 1.23 (1.03, 1.46) | 0.549 | LASSO  | Unadj.  |
|                | Large CTMLE    | 1.22 (1.03, 1.45) | 0.528 | LASSO  | GLM     |