Title

Adaptive pre-specification in randomized trials with and without pair-matching

Permalink
https://escholarship.org/uc/item/3ft2w42z

Journal
Statistics in Medicine, 35(25)

ISSN
0277-6715

Authors
Balzer, Laura B
van der Laan, Mark J
Petersen, Maya L
et al.

Publication Date
2016-11-10

DOI
10.1002/sim.7023

Peer reviewed

Published in final edited form as:
Stat Med. 2016 November 10; 35(25): 4528–4545. doi:10.1002/sim.7023.

Adaptive Pre-specification in Randomized Trials With and Without Pair-Matching

Laura B. Balzer[a,*], **Mark J. van der Laan**[b], **Maya L. Petersen**[b], and **the SEARCH Collaboration**

[a]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

[b]Division of Biostatistics, University of California, Berkeley, CA 94110-7358, USA

Abstract

In randomized trials, adjustment for measured covariates during the analysis can reduce variance and increase power. To avoid misleading inference, the analysis plan must be pre-specified. However, it is often unclear *a priori* which baseline covariates (if any) should be adjusted for in the analysis. Consider, for example, the Sustainable East Africa Research in Community Health (SEARCH) trial for HIV prevention and treatment. There are 16 matched pairs of communities and many potential adjustment variables, including region, HIV prevalence, male circumcision coverage and measures of community-level viral load. In this paper, we propose a rigorous procedure to data-adaptively select the adjustment set, which maximizes the efficiency of the analysis. Specifically, we use cross-validation to select from a pre-specified library the candidate targeted maximum likelihood estimator (TMLE) that minimizes the estimated variance. For further gains in precision, we also propose a collaborative procedure for estimating the known exposure mechanism. Our small sample simulations demonstrate the promise of the methodology to maximize study power, while maintaining nominal confidence interval coverage. We show how our procedure can be tailored to the scientific question (intervention effect for the study sample vs. for the target population) and study design (pair-matched or not).

Keywords

Causal inference; Covariate selection; Data-adaptive; Pair-matched; Randomized trials; Targeted maximum likelihood estimation (TMLE)

1. Introduction

The objective of a randomized trial is to evaluate the effect of an intervention on the outcome of interest. In this setting, the difference in the average outcomes among the treated units and the average outcomes among the control units provides a simple and unbiased estimator of the intervention effect. Adjusting for measured covariates during the analysis can substantially reduce the estimator's variance and thereby increase study power (e.g. [1–5]). Nonetheless, recommendations on how and when to adjust in randomized trials have

[*]Correspondence to: Laura Balzer, Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA. lbbalzer@hsph.harvard.edu.

been conflicting [6–12]. The advice seems to depend on the study design, the unit of randomization, the application and the sample size. As a result, many researchers are left wondering how to adjust for baseline covariates, if at all.

Let n be the number of study units (e.g. patients or communities). Consider a trial where the treatment is randomly allocated to $n/2$ units and the remaining units are assigned to the control. There is a rich literature on locally efficient estimation in this setting (e.g. [4, 5, 13–15]). For example, parametric regression can be used to obtain an unbiased and more precise estimate of the intervention effect. Briefly, the outcome is regressed on the exposure and covariates according to a working model. Following Rosenblum and van der Laan [16], we use “working” to emphasize that the regression function need not be and often is not correctly specified. This working model can include interaction terms and can be linear or non-linear. The estimated coefficients are then used to obtain the predicted outcomes for all units under the treatment and the control. The difference or ratio of the average of the predicted outcomes provides an estimate of the intervention effect. For observational studies, this algorithm is sometimes referred to as the “parametric G-Computation” [17].

For continuous outcomes and linear working models without interaction terms, this procedure is known as analysis of covariance (ANCOVA) [2], and the coefficient for the exposure is equal to the estimate of the intervention effect. For binary outcomes, Moore and van der Laan [5] detailed the potential gains in precision from adjustment via logistic regression for estimating the treatment effect on the absolute or relative scale (i.e. risk difference, risk ratio or odds ratio). Furthermore, the authors showed that parametric maximum likelihood estimation (MLE) was equivalent to targeted maximum likelihood estimation (TMLE) in this setting [18, 19]. As a result, the asymptotic properties of the TMLE, including double robustness and asymptotic linearity, hold even if the working model for outcome regression is misspecified. Furthermore, this approach is locally efficient in that the TMLE will achieve the lowest possible variance among a large class of estimators if the working model is correctly specified. Rosenblum and van der Laan [16] expanded these results for a large class of general linear models. Indeed, the parametric MLE and TMLE can be considered special cases of the double robust estimators of Scharfstein *et al.* [20] and semiparametric approaches of Tsiatis *et al.* [4] and Zhang *et al.* [13]. For a recent and detailed review of these estimation approaches, we refer the reader to Colantuoni and Rosenblum [21].

Now consider a pair-matched trial, where the intervention is randomly allocated within the $n/2$ matched pairs. The proposed estimation strategies have been more limited in this setting. Indeed, the perceived “analytical limitations” of pair-matched trials have led some researchers to shy away from this design [11, 22, 23]. As with a completely randomized trial, the unadjusted difference in treatment-specific means provides an unbiased but inefficient estimate of the intervention effect. To include covariates in the analysis and to potentially increase power, Hayes and Moulton [8] suggested regressing the outcome on the covariates (but not on the exposure) and then contrasting the observed versus predicted outcomes within matched pairs. Alternatively, TMLE can provide an unbiased and locally efficient approach in pair-matched trials [24–26]. Specifically, the algorithm can be implemented as if the trial were completely randomized: (1) fit a working model for the mean outcome, given the exposure and covariates, (2) obtain predicted outcomes for all units under the treatment and control, and (3) contrast the average of the predicted outcomes on the relevant scale. Inference, however, must respect the pair-matching scheme [24–26].

A common challenge to both designs is the selection of covariates for inclusion in the analysis. Many variables are measured prior to implementation of the intervention, and it is difficult to *a priori* specify an appropriate working model. For a completely randomized trial, covariate adjustment will lead to gains in precision if (i) the covariates are predictive of the outcome and (ii) the covariates are imbalanced between treatment groups (e.g. [27]). Balance is guaranteed as sample size goes to infinity, but rarely seen in practice. Analogously in a pair-matched trial, covariate adjustment will improve precision if there is an imbalance on predictive covariates after matching.

Limited sample sizes pose an additional challenge to covariate selection. A recent review of randomized clinical trials reported that the median number of participants was 58 with an interquartile range of 27–161 [28]. Likewise, a recent review of cluster randomized trials reported that the median number of units was 31 with an interquartile range of 13–60 [29]. In small trials, adjusting for too many covariates can lead to overfitting and inflated Type I error rates (e.g. [15, 25, 27]). Finally, *ad hoc* selection of the adjustment set leads to concerns that researchers will go on a “fishing expedition” to find the covariates resulting in the most power and again risking inflation of Type I error rates (e.g. [4, 7, 30]).

In summary, covariate adjustment in randomized trials can provide meaningful improvements in precision and thereby statistical power. To avoid misleading statistical inference, the working model, including the adjustment variables, must be specified *a priori*. In practice, sample size often limits the size of the adjustment set, and best set is unclear before the trial’s conclusion. This results in an important challenge: the need to learn from the data to realize precision gains, but to do so in pre-specified and rigorous way to maintain valid statistical inference.

In this paper, we apply the principle of *empirical efficiency maximization* to data-adaptively select from a pre-specified library the candidate TMLE, which minimizes variance and thereby maximizes the precision of the analysis [14, 31]. We contribute to the existing methodology by modifying this strategy for pair-matched trials. To our knowledge, such a data-adaptive procedure has not been proposed or implemented for this study design. We further contribute to the literature by collaboratively estimating the exposure mechanism for additional gains in precision [32, 33]. We also generalize the results for estimation and inference to both the population and sample average treatment effects [26, 34]. Our finite sample simulations demonstrate the practical performance with limited numbers of independent units, as is common in early phase clinical trials and in cluster randomized trials. As a motivating example, we discuss the Sustainable East Africa Research in Community Health (SEARCH) study, an ongoing cluster randomized trial for HIV prevention and treatment (NCT01864603) [35]. The methodology proposed in this article will be used in the primary analysis of the SEARCH trial. Full R code is provided in the Supplementary Material [36].

2. Motivating Example and Causal Parameters

SEARCH is a community randomized trial to estimate the effect of immediate and streamlined antiretroviral therapy (ART) on HIV incidence as well as other health, economic and educational outcomes. The trial is being conducted in 32 rural communities in Uganda and Kenya. Extensive baseline characteristics were collected through ethnographic mapping and community-wide censuses. Examples include region, occupational mix, measures of mobility, HIV prevalence and community-level HIV RNA viral load. A subset of these characteristics was used to create the 16 best matched pairs of communities [25]. The intervention was then randomized within matched pairs. In treatment communities, HIV testing is expanded, and all individuals testing HIV+ are immediately eligible for ART with enhanced services for linkage, initiation, and retention in care. In control communities, all individuals testing HIV+ are eligible for ART, according to in-country guidelines. The primary outcome is the five-year cumulative incidence of HIV and will be measured through longitudinal follow-up. The observed data for a given SEARCH community can be denoted

$$O=(W,A,Y)$$

where W represents the vector of baseline covariates, A represents the intervention assignment, and Y denotes the outcome. Specifically, W includes region, HIV prevalence, male circumcision coverage and community-level HIV RNA viral load; A is a binary indicator equalling one if the community was randomized to the treatment and zero if the community was randomized to the control; and Y is the estimated five-year cumulative HIV incidence.

In this paper, we consider estimation and inference for the population average treatment effect (PATE) and the sample average treatment effect (SATE). Let $Y(a)$ denote the outcome if possibly contrary-to-fact the unit were assigned intervention-level $A = a$. The causal parameters are functions of the distribution of the full data, comprised of the baseline covariates and the counterfactual outcomes of interest: $(W, Y(1), Y(0))$ [34, 37]. Specifically, the PATE is the expected difference in the counterfactual outcomes if all members of the population were assigned the intervention and if all members of that population were assigned the control:

$$PATE=[Y(1)-Y(0)] \quad (1)$$

where the expectation is over the full data distribution. There is one true value of PATE for the target population. For the SEARCH trial, the population effect is the expected difference in the counterfactual cumulative incidence of HIV if all communities in the hypothetical target population implemented the test-and-treat strategy versus the counterfactual cumulative incidence of HIV if all communities in that target population maintained the standard of care.

The sample parameter is the average difference in the counterfactual outcomes for the study units [34]:

$$SATE = \frac{1}{n}\sum_{i=1}^{n}[Y_i(1) - Y_i(0)] \quad (2)$$

were $Y_i(a)$ denotes the outcome if possibly contrary-to-fact unit i were assigned intervention-level $A = a$. The SATE is data-adaptive; its true value depends on the n units in the sample. The SATE is easily interpretable and arguably the most relevant when the study units were not sampled from some super population of interest. For the SEARCH trial, the SATE is the average difference in the counterfactual cumulative incidence of HIV under the test-and-treat strategy and under the standard of care for the 32 study communities.

3. Targeted Estimation in a Randomized Trial Without Matching

In this section, we ignore the pair-matching scheme in the SEARCH trial and assume the observed data consist of n independent, identically distributed (i.i.d.) copies of $O = (W, A, Y)$ with some true, but unknown distribution P_0, which factorizes as

$$P_0(O) = P_0(W)P_0(A|W)P_0(Y|A,W).$$

We do not make any assumptions about the common covariate distribution $P_0(W)$ or about the common conditional distribution of the outcome, given the intervention and covariates $P_0(Y|A, W)$. By design, the intervention A is randomized with probability 0.5. Therefore, the exposure mechanism is known: $P_0(A = 1|W) \equiv g_0(1|W) = 0.5$. The statistical model $\mathcal{M}$, describing the set of possible observed data distributions, is semiparametric.

Since the intervention is randomized, we can easily identify the PATE (Eq. 1) from the observed data distribution. Our statistical estimand is the difference in the expected outcome given the treatment and covariates, and the expected outcome given the control and covariates, averaged (standardized) with respect to the covariate distribution in the population [17]:

$$\begin{aligned}\Psi(P_0) &= \mathbb{E}_0\left[\mathbb{E}_0(Y|A=1,W) - \mathbb{E}_0(Y|A=0,W)\right]\\ &= \mathbb{E}_0\left[\overline{Q}_0(1,W) - \overline{Q}_0(0,W)\right]\end{aligned}$$

where $\bar{Q}_0(A, W) \equiv \mathbb{E}_0(Y|A, W)$ denotes the true conditional mean outcome, given the intervention and covariates. As discussed in the introduction, there are many algorithms available for unbiased and locally efficient estimation of this statistical parameter in a randomized trial (e.g. [4, 5, 13–15]). Throughout, our focus is on TMLE, a general methodology for the construction of double robust, semiparametric, efficient substitution estimators [18, 19].

A TMLE for the population effect (Eq. 1) also serves as a consistent and asymptotically linear estimator of the sample effect (Eq. 2) [26]. The estimator can be implemented in three steps.

Step 1. Initial estimation: Estimate the expected outcome, given the exposure and covariates $\bar{Q}_0(A, W) \equiv \mathbb{E}_0(Y|A, W)$. We could rely on a pre-specified parametric working model or implement a more data-adaptive approach (as discussed below). The initial estimator is denoted $\bar{Q}_n(A, W)$.

Step 2. Targeting: Update the initial estimator $\bar{Q}_n(A, W)$.

- Calculate the "clever" covariate based on the known or estimated exposure mechanism $g_n(A| W)$:

$$H_n(A, W) = \left(\frac{(A=1)}{g_n(1|W)} - \frac{(A=0)}{g_n(0|W)} \right).$$

- If the outcome is continuous and unbounded, run linear regression of the outcome Y on the covariate $H_n(A, W)$ with the initial estimator as offset. Plug in the estimated coefficient ε_n to yield the targeted update: $\overline{Q}_n^*(A, W) = \overline{Q}_n(A, W) + \varepsilon_n H_n(A, W)$.
- If the outcome is binary or bounded in [0, 1][†], run logistic regression of the outcome Y on the covariate $H_n(A, W)$ with the logit(x) ≡ $log\{x/(1 - x)\}$ of the initial estimator as offset. Plug in the estimated coefficient ε_n to yield the targeted update: $\overline{Q}_n^*(A, W) = \text{logit}^{-1}\{\text{logit}[\overline{Q}_n(A, W)] + \varepsilon_n H_n(A, W)\}$.

Step 3. Parameter estimation: Obtain the predicted outcomes for all observations under the treatment $\overline{Q}_n^*(1, W)$ and control $\overline{Q}_n^*(0, W)$. Average the difference in predicted outcomes:

$$\Psi_n(\overline{Q}_n^*) = \frac{1}{n} \sum_{i=1}^{n} [\overline{Q}_n^*(1, W_i) - \overline{Q}_n^*(0, W_i)].$$

If the initial estimator for $\bar{Q}_0(A, W)$ is based on a working regression model with an intercept and a main term for the exposure and if the exposure mechanism is treated as known (i.e. not estimated), then the updating step can be skipped [16]. Further precision, however, can be attained by using a data-adaptive algorithm for initial estimation of the outcome regression $\bar{Q}_0(A, W)$ and/or by estimating the exposure mechanism $g_0(A| W)$ [39].

Under standard regularity conditions, the TMLE is an asymptotically linear estimator of both the population and sample effects [19, 26]. The estimator minus the true effect can be written as an empirical mean of an influence curve and a second order term going to 0 in probability. As a result, the TMLE is asymptotically normal with variance well-approximated by the variance of its influence curve, divided by sample size n. The influence curve for the TMLE of the population effect (PATE) is given by

[†] In greater generality, the logistic fluctuation can also be used for a continuous outcome that is bounded in [a, b] by first applying the following transformation to the outcome: $Y^* = (Y - a)/(b - a)$. Use of logistic regression over linear regression can provide stability under data sparsity and/or with rare outcomes (e.g. [25, 38]).

$$D^{\mathscr{P}}(g_0,\overline{Q})(O)=\left(\frac{(A=1)}{g_0(1|W)}-\frac{(A=0)}{g_0(0|W)}\right)(Y-\overline{Q}(A,W))+\overline{Q}(1,W)-\overline{Q}(0,W)-\Psi(\overline{Q})$$

where $\bar{Q}(A, W)$ denotes the limit of the targeted estimator of the conditional mean function $\bar{Q}_0(A, W)$ and where we are assuming the exposure mechanism $g_0(A| W)$ is known or consistently estimated, as will always be true in a randomized trial [19]. A plug-in estimator of this influence curve is given by

$$D_n^{\mathscr{P}}(g_n,\overline{Q}_n^*)(O)=\left(\frac{(A=1)}{g_n(1|W)}-\frac{(A=0)}{g_n(0|W)}\right)(Y-\overline{Q}_n^*(A,W))+\overline{Q}_n^*(1,W)-\overline{Q}_n^*(0,W)-\psi_n^* \quad (3)$$

where ψ_n^* denotes the point estimate. In finite samples, the variance of the TMLE for the PATE is well-approximated by the sample variance of this estimated influence curve, scaled by sample size:

$$\sigma_n^{2,\mathscr{P}}=\frac{Var_n\{D_n^{\mathscr{P}}(g_n,\overline{Q}_n^*)\}}{n}.$$

The influence curve for the TMLE of the sample effect (SATE) relies on non-identifiable quantities, specifically the counterfactual outcomes $Y_i(1)$ and $Y_i(0)$ [26]. Nonetheless, a conservative plug-in estimator of its influence curve is obtained by ignoring these non-identifiable quantities:

$$D_n^{\mathscr{S}}(g_n,\overline{Q}_n^*)(O)=\left(\frac{(A=1)}{g_n(1|W)}-\frac{(A=0)}{g_n(0|W)}\right)(Y-\overline{Q}_n^*(A,W)). \quad (4)$$

In finite samples, the variance of the TMLE for the SATE is conservatively approximated by the sample variance of this estimated influence curve, scaled by sample size:

$$\sigma_n^{2,\mathscr{S}}=\frac{Var_n\{D_n^{\mathscr{S}}(g_n,\overline{Q}_n^*)\}}{n}.$$

We refer the reader to Balzer *et al.* [26] for further details.

With an estimate of the standard error ($\sigma_n^{\mathscr{P}}$ for the population effect or $\sigma_n^{\mathscr{S}}$ for the sample effect), we construct Wald-Type 95% confidence intervals as $\psi_n^* \pm 1.96\sigma_n$. Analogously, we can test the null hypothesis of no average effect with the test statistic ψ_n^*/σ_n. For trials with a limited number of independent units, the Student's *t*-distribution is an appropriate alternative to the standard normal distribution. Randomization inference may not be appropriate however, because it is testing the sharp null of no treatment effect for any unit [40, 41], whereas our interest is in the null hypothesis of no treatment effect on average.

Comparing Eq. 3 and 4, we see that for the SATE there is no variance contribution from the covariate distribution, which is considered fixed. As a result, the sample effect will often be estimated with more precision than the population effect [34, 42, 43]. Indeed, the TMLE for the PATE and the TMLE for the SATE will only have the same efficiency bound if the conditional mean $\bar{Q}_0(A, W)$ is consistently estimated *and* if there is no variability in the intervention effect across units [26]. In many settings, there will be effect heterogeneity, and specifying the SATE as the target of inference can yield more power, especially in large trials. In small trials, the gains in precision from targeting the SATE can be attenuated, because this influence curve-based variance estimator is conservative (biased upwards).

3.1. Adaptive Pre-specified Approach for Step 1. Initial Estimation

Consider again the SEARCH trial for HIV prevention and treatment. Recall that the outcome Y is the five-year cumulative incidence of HIV and bounded between 0 and 1. The first step of the TMLE algorithm is to obtain an initial estimator of the expected outcome, given the exposure and measured covariates $\bar{Q}_0(A, W)$. Suppose that as a working model, we consider running logistic regression[‡] of the outcome Y on the treatment A and covariates W. It is unclear *a priori* which covariates should be included in the working model and in what form. For example, baseline HIV prevalence is a known predictor of the outcome and may be imbalanced between the treatment and control groups. Therefore, as initial estimator of $\bar{Q}_0(A, W)$, we could consider a logistic regression working model with an intercept and main terms for the treatment and HIV prevalence. Likewise, there might be substantial heterogeneity in the treatment effect by region and allowing for an interaction between region and the intervention may reduce the variance of the TMLE. Including all the covariates and the relevant interactions in the working model is likely to result in overfitting and misleading inference. To facilitate selection between candidate initial estimators and thereby candidate TMLEs, we propose the following cross-validation selector.

First, we propose a library of candidate working models for initial estimation of the conditional mean outcome $\bar{Q}_0(A, W)$. This library should be pre-specified in the protocol or the analysis plan. A possible library could consist of the following logistic regression working models:

$$\begin{gathered}\text{logit}[\overline{Q}^{(a)}(A, W)] = \beta_0 + \beta_1 A \\ \text{logit}[\overline{Q}^{(b)}(A, W)] = \beta_0 + \beta_1 A + \beta_2 W1 \\ \text{logit}[\overline{Q}^{(c)}(A, W)] = \beta_0 + \beta_1 A + \beta_2 W2 + \beta_3 A \times W2\end{gathered}$$

where, for example, $W1$ denotes baseline prevalence and $W2$ denotes region. Of course, there are many more candidate algorithms, and we are considering this simple set for pedagogic purposes. We also note that the first working model corresponds to the unadjusted estimator.

[‡] Logistic regression naturally respects the bounds on this continuous outcome. Prior work has suggested that use of the logistic regression over linear regression can provide stability when there are positivity violations and/or the outcome is rare [25, 38]. Full R code is available in the Supplementary Material.

Second, we need to pre-specify a loss function to measure the performance of the candidate estimators. Following the principle of empirical efficiency maximization [14, 31], we propose using the squared influence curve of the TMLE for the parameter of interest. The expectation of this loss function, called the "risk", is then the asymptotic variance of the TMLE. Thereby, our goal is to select the candidate estimator that maximizes precision. If the target of inference is the population effect, our loss function is

$$\mathscr{L}^{\mathscr{P}}(g_0,\overline{Q})(O)=\{D^{\mathscr{P}}(g_0,\overline{Q})(O)\}^2 \quad (5)$$

where we are not estimating the known exposure mechanism $g_0(A|W) = 0.5$. Since the true influence curve of the TMLE for the sample effect relies on non-identifiable quantities [26], our loss function for the SATE is the estimated influence curve-squared:

$$\mathscr{L}^{\mathscr{S}}(g_0,\overline{Q})(O)=\{D_n^{\mathscr{S}}(g_0,\overline{Q})(O)\}^2 \quad (6)$$

where again we are not estimating the known exposure mechanism $g_0(A|W) = 0.5$. In this case, the loss function for the SATE corresponds to the L2 squared error loss function: $\mathscr{L}^{\mathscr{S}}(g_0, \bar{Q}) = (Y - \bar{Q}(A,W))^2$.

Next, we need to pre-specify our cross-validation scheme, used to generate an estimate of the risk for each of the candidate estimators. For generality, we present V-fold cross-validation, where the data are randomly split into V partitions, called "folds", of size ≈ n/V. To respect the limited sample sizes common in early phase clinical trials and in cluster randomized trials, leave-one-out cross-validation is often appropriate. Leave-one-out cross-validation corresponds with $V = n$-fold cross-validation, where each fold corresponds to one observation. The cross-validation procedure for initial estimation of the conditional mean $\bar{Q}_0(A,W)$ can be implemented as follows.

A. For each fold $v = \{1, \ldots, V\}$ in turn,

  a. Set the observation(s) in fold v to be the validation set and the remaining observations to be the training set.

  b. Fit each algorithm for estimating $\bar{Q}_0(A,W)$ using only data in the training set. For the above library, we would run logistic regression of the outcome Y on the exposure A and covariates W, according to the working model. Denote the initial regression fits as $\overline{Q}_n^{(a)}(A,W), \overline{Q}_n^{(b)}(A,W)$ and $\overline{Q}_n^{(c)}(A,W)$, respectively.

  c. For each algorithm, use the estimated fit to predict the outcome(s) for the observation(s) in the validation set under the treatment and the control. For the first algorithm, for example, we would have $\overline{Q}_n^{(a)}(1, W_k)$ and $\overline{Q}_n^{(a)}(0, W_k)$ for observation O_k in the validation set.

d. For each algorithm, evaluate the loss function for the observation(s) in the validation set by plugging in the algorithm-specific predictions. For example, if our target of inference were the SATE, we would have for the first algorithm

$$\mathscr{L}^{\mathscr{S}}(g_0, \overline{Q}_n^{(a)})(O_k) = \left[\left(\frac{(A_k=1)}{g_0(1|W_k)} - \frac{(A_k=0)}{g_0(0|W_k)}\right)(Y_k - \overline{Q}_n^{(a)}(A_k, W_k))\right]^2$$

for observation O_k in the validation set. The exposure mechanism is known: $g_0(A|W) = 0.5$.

e. For each algorithm, obtain an estimate of the risk by averaging the estimated losses across the observations in validation set v. If our target of inference were the SATE, we would have for the first algorithm

$$Risk_v^{(a)} = \frac{1}{n_v}\sum_{k\in v}\mathscr{L}^{\mathscr{S}}(g_0, \overline{Q}_n^{(a)})(O_k)$$

where n_v denotes the number of observations in validation set v.

B. For each algorithm, average the estimated risks across the V folds.

C. Select the algorithm with the smallest cross-validated risk. This is the algorithm yielding the smallest cross-validated variance estimate.

The selected working model is then used for initial estimation of the conditional mean outcome $\bar{Q}_0(A, W)$ in Step 1 of the TMLE algorithm, described above (Sec. 3). Specifically, we would re-fit the selected algorithm using all the data. Since the exposure mechanism was treated as known and our library was limited to simple parametric working models with a main term for the exposure and an intercept, the updating step (Step 2) can be skipped. In other words, the chosen estimator is already targeted $\overline{Q}_n(A, W) = \overline{Q}_n^*(A, W)$ and can be used for Step 3 parameter estimation.

4. Targeted Estimation in a Randomized Trial With Matching

Recall the pair-matching scheme briefly described in Section 2 for the SEARCH trial. First, the potential study units were selected. Then the baseline covariates, such as region, occupational mix and measures of migration, were collected. A matching algorithm was applied to the baseline covariates of candidate units to create the best 16 matched pairs. The intervention was randomized within the resulting pairs, and the outcome will be measured with longitudinal follow-up. This pair-matching scheme is considered to be *adaptive*, because the resulting matched pairs are a function of the baseline covariates of all the candidate units [24–26]. This design has also been called "nonbipartite matching" and "optimal multivariate matching" [44–46].

The adaptive design creates a dependence in the data. Since the construction of the matched pairs is a function of the baseline covariates of all n study units, the observed data do not consist of $n/2$ i.i.d. paired observations, as current practice sometimes assumes (e.g. [8, 22, 47, 48]). Instead, we have n dependent copies of $O = (W, A, Y)$. Nonetheless, there remains substantial conditional independence in the data. Mainly, once we consider the baseline covariates of the study units as fixed, we recover $n/2$ conditionally independent units:

$$\overline{O}_j = (O_{j1}, O_{j2}) = ((W_{j1}, A_{j1}, Y_{j1}), (W_{j2}, A_{j2}, Y_{j2}))$$

where the index $j = 1, \ldots, n/2$ denotes the partitioning of the candidate units $\{1, \ldots n\}$ into matched pairs according to similarity in their baseline covariates $(W_1, \ldots, W_n)$. Throughout subscripts $j1$ and $j2$ index the observations within matched pair j. The conditional distribution of the observed data, given the baseline covariates of the study units, factorizes as

$$\begin{aligned} P_0(O_1, \ldots, O_n | W_1, \ldots W_n) &= \prod_{j=1}^{n/2} P_0(A_{j1}, A_{j2} | W_1, \ldots, W_n) P_0(Y_{j1} | A_{j1}, W_{j1}) P_0(Y_{j2} | A_{j2}, W_{j2}) \\ &= \prod_{j=1}^{n/2} 0.5 \times P_0(Y_{j1} | A_{j1}, W_{j1}) \times P_0(Y_{j2} | A_{j2}, W_{j2}) \end{aligned}$$

where the second line follows from randomization of the intervention within matched pairs. For estimation and inference of the population effect (PATE), we need to assume that each community's baseline covariates W_i are independently drawn from some common distribution $P_0(W)$. For estimation and inference of the sample effect (SATE), this assumption on the covariate distribution can be weakened [26].

Despite the dependence in the data, a TMLE for the population or sample effect can be implemented by ignoring the pair-matched design [24, 26]. In other words, a point estimate is obtained by following the procedure outlined in Section 3. In Step 1, we obtain an initial estimator of the conditional mean outcome with an *a priori*-specified parametric working model or with a more data-adaptive method (as detailed below). In Step 2, we target the initial estimator by using information in the known or estimated exposure mechanism. Finally in Step 3, we obtain the predicted outcomes for all observations under the treatment and the control, and then take the sample average of the difference in these targeted predictions.

In a trial with adaptive pair-matching, the TMLE is an asymptotically normal estimator of both the population and sample effects [24, 26]. For the PATE, we could estimate its variance with the sample variance of the estimated influence curve in the non-matched trial $Var_n\{D_n^{\mathscr{P}}(g_n, \overline{Q}_n^*)\}$ divided by n [24]. This variance estimator, however, ignores any gains in precision from pair-matching and will be conservative under reasonable assumptions. A less conservative variance estimator is obtained by accounting for the potential correlations of the residuals within matched pairs:

$$\rho_n(\overline{Q}_n^*)=\frac{1}{n/2}\sum_{j=1}^{n/2}(Y_{j1}-\overline{Q}_n^*(A_{j1},W_{j1}))(Y_{j2}-\overline{Q}_n^*(A_{j2},W_{j2})) \qquad (7)$$

[24]. In finite samples, we recommend estimating of the variance of the TMLE for the population effect under pair-matching with

$$\sigma_n^{2,\mathscr{P}*}=\frac{Var_n\{D_n^{\mathscr{P}}(g_n,\overline{Q}_n^*)\}-2\rho_n(\overline{Q}_n^*)}{n}.$$

In a pair-matched trial, the TMLE minus the sample effect (SATE) again behaves as an empirical mean of an influence curve, depending on non-identifiable quantities [26]. Nonetheless, a conservative plug-in estimator of its influence curve is given by

$$D_n^{\mathscr{S}*}(g_n,\overline{Q}_n^*)(\overline{O}_j)=\frac{1}{2}\left[D_n^{\mathscr{S}}(g_n,\overline{Q}_n^*)(O_{j1})+D_n^{\mathscr{S}}(g_n,\overline{Q}_n^*)(O_{j2})\right]$$

where $D_n^{\mathscr{S}}(g_n,\overline{Q}_n^*)(O)$ is the estimated influence curve for observation O in the non-matched trial (Eq. 4). In finite samples, we conservatively estimate the variance of the TMLE for the sample effect with the sample variance of the estimated (paired) influence curve divided by $n/2$:

$$\sigma_n^{2,\mathscr{S}*}=\frac{Var_n\{D_n^{\mathscr{S}*}(g_n,\overline{Q}_n^*)\}}{n/2}.$$

If we order observations within matched pairs such that first corresponds to the intervention (A_{j1} = 1) and the second to the control (A_{j2} = 0) and do not estimate the exposure mechanism $g_0(A|W)$ = 0.5, we have

$$D_n^{\mathscr{S}*}(g_0,\overline{Q}_n^*)(\overline{O}_j)=(Y_{j1}-\overline{Q}_n^*(1,W_{j1}))-(Y_{j2}-\overline{Q}_n^*(0,W_{j2})).$$

In this setting, the sample variance of the pairwise differences in residuals, divided by $n/2$, provides a conservative variance estimator. With an estimate of the standard error ($\sigma_n^{\mathscr{P}*}$ for the population effect or $\sigma_n^{\mathscr{S}*}$ for the sample effect), we can create 95% confidence intervals and conduct hypothesis tests, as described above.

4.1. Adaptive Pre-specified Approach for Step 1. Initial Estimation

By balancing intervention groups with respect to baseline determinants of the outcome, pair-matching increases the efficiency of the study (e.g. [24, 26, 49]). Nonetheless, residual imbalance on the baseline predictors often remains, and adjusting for these covariates during the analysis can further increase efficiency. In the SEARCH trial, for example, the matched pairs were created before baseline HIV prevalence was measured. As a result, there is likely to be variation across the pairs in baseline prevalence, a known driver of HIV incidence. Adjusting for baseline prevalence during the analysis is likely to increase power via two mechanisms: (1) reducing the variance of the TMLE for the point estimate, and (2) resulting in a less conservative variance estimator. Unfortunately, it is unclear *a priori* whether adjusting for prevalence will yield more power than adjusting for other covariates, such as male circumcision coverage or measures of community-level HIV RNA viral load. With only 16 (conditionally) independent units, we are limited as to the size of the adjustment set. Adjusting for too many covariates can result in over-fitting. As before, we want to data-adaptively select the candidate TMLE (i.e. working regression model), which maximizes the empirical efficiency.

The data-adaptive procedure for initial estimation of the conditional mean outcome $\bar{Q}_0(A, W)$ for a non-matched trial (Sec. 3.1) can be modified for a pair-matched trial. As before, we need to pre-specify our library of candidate estimators, our measure of performance, and the cross-validation scheme. We can use the same library of candidate working models for initial estimation of the conditional mean outcome $\bar{Q}_0(A, W)$. To measure performance, however, we want to use as risk the estimated variance of the TMLE under pair-matching. To elaborate, consider the loss function for the sample effect in a non-matched trial. Minimizing the sum of squared residuals (Eq. 6) targets the conditional mean outcome $\bar{Q}_0(A, W)$. As a result, the algorithm could select a working model adjusting for a covariate that is highly predictive of the outcome but on which we matched perfectly. In the SEARCH trial, for example, communities were paired within region, because HIV incidence is expected to be highly heterogeneous across regions. Therefore, minimizing the empirical variance of $D_n^{\mathscr{S}}(g_0, \overline{Q})$ might lead to selection of the candidate TMLE with main terms for the intervention and region. This selection would not improve the precision of the analysis over the unadjusted algorithm. (We already "controlled" for region in the design.) Instead, we want to select the candidate TMLE maximizing precision for the parameter of interest in a pair-matched trial. Thereby, our loss function for the PATE is

$$\mathscr{L}^{\mathscr{P}*}(g_0,\overline{Q})(\overline{O}_j)=\frac{1}{2}\{D_n^{\mathscr{P}}(g_0,\overline{Q})(O_{j1})\}^2+\frac{1}{2}\{D_n^{\mathscr{P}}(g_0,\overline{Q})(O_{j2})\}^2-2(Y_{j1}-\overline{Q}(A_{j1},W_{j1}))\,(Y_{j2}-\overline{Q}(A_{j2},W_{j2})), \tag{8}$$

and our loss function for the SATE is

$$\mathscr{L}^{\mathscr{I}*}(g_0,\overline{Q})(\overline{O}_j)=\{D_n^{\mathscr{I}*}(g_0,\overline{Q}_n^*)(\overline{O}_j)\}^2. \quad (9)$$

(For further details, see Section 1 of the Supplementary Material.) Again, we are treating the exposure mechanism as known: $g_0(A|W) = 0.5$.

Finally, in the cross-validation scheme, the pair should be treated as the unit of (conditional) independence. In other words, when the data are split into V-folds, the pairing should be preserved. In small trials, leave-one-pair-out cross-validation will often be appropriate. With these modifications, we can implement the cross-validation scheme, outlined in Section 3.1, to data-adaptively select the candidate working model, which minimizes the estimated variance of the TMLE in a pair-matched trial. As before, the selected working model would then be refit using all the data and used to estimate outcomes for all observations under the treatment and control. The average difference in the predicted outcomes would provide an estimate of the intervention effect.

5. Collaborative Estimation of the Exposure Mechanism

Even though the intervention A is randomized with balanced allocation, estimating the known exposure mechanism $g_0(A|W) = 0.5$ can increase the precision of the analysis [39]. As before, we want to respect the study design (i.e. pair-matched or not) as well as adjust for a covariate only if its inclusion improves the empirical efficiency. For example, we will generally not want to adjust for a covariate that is imbalanced between the intervention groups (i.e. predictive of A) but not predictive of the outcome. Likewise, if a given covariate (e.g. $W1$) was included in the working model for conditional mean outcome $\bar{Q}_0(A, W)$, further adjusting for this covariate when estimating the exposure mechanism may not increase precision. To this end, we incorporate the Collaborative TMLE (C-TMLE) approach into our algorithm [32, 33].

5.1. Adaptive Pre-specified Approach for Step 2. Targeting

First, we propose a library of candidate estimators of the exposure mechanism $g_0(A|W)$. As before, this library should be pre-specified in the protocol or analysis plan. A possible library could consist of the following logistic regression working models:

$$\begin{aligned}\text{logit}[g^{(a)}(W)]&=\beta_0\\ \text{logit}[g^{(b)}(W)]&=\beta_0+\beta_1W1\\ \text{logit}[g^{(c)}(W)]&=\beta_0+\beta_1W2\end{aligned}$$

where, for example, $W1$ is baseline prevalence and $W2$ is male circumcision coverage. Each algorithm would yield a different update to a given initial estimator of the conditional mean outcome $\bar{Q}_n(A, W)$, selected by the data-adaptive procedure for Step 1 (Sec. 3.1 for trials without matching and Sec. 4.1 for trials with matching). In other words, each candidate estimator of $g_0(A|W)$ results in a different targeted estimator $\overline{Q}_n^*(A,W)$. We also note that the first working model corresponds to the unadjusted estimator.

To choose between candidate algorithms, we need to pre-specify a measure of performance. As before, we propose using as risk the estimated asymptotic variance of the TMLE, appropriate for the study design (i.e. pair-matched or not) and the scientific question (i.e. population or sample effect). Therefore, our loss functions are

- Without matching and for the PATE: $\mathcal{L}^{\mathcal{P}}(g, \bar{Q}_n)$ as in Eq. 5
- Without matching and for the SATE: $\mathcal{L}^{\mathcal{S}}(g, \bar{Q}_n)$ as in Eq. 6
- With matching and for the PATE: $\mathcal{L}^{\mathcal{P}*}(g, \bar{Q}_n)$ as in Eq. 8
- With matching and for the SATE: $\mathcal{L}^{\mathcal{S}*}(g, \bar{Q}_n)$ as in Eq. 9

where g denotes a candidate estimator of the exposure mechanism and $\bar{Q}_n$ denotes our selected initial estimator of the outcome regression (Sec. 3.1 and 4.1).

Finally, we need to pre-specify our cross-validation scheme, used to obtain an honest measure of risk and to reduce the potential for over-fitting. As before, we present V-fold cross-validation, where the data are partitioned into V folds of size ≈ n/V. If matching was used, the partitioning should preserve the pairs. The cross-validation selector for collaborative estimation of the exposure mechanism can be implemented as follows.

A. For each fold $v = \{1, \ldots, V\}$ in turn,

  a. Set the observation(s) in fold v to be the validation set and the remaining observations to be the training set.

  b. Using only data in the training set, fit each algorithm for estimating the exposure mechanism. For the above library, we would run logistic regression of the exposure A on the covariates W, according to the working model. Denote the estimated exposure mechanisms as $g_n^{(a)}(A|W)$, $g_n^{(b)}(A|W)$ and $g_n^{(c)}(A|W)$, respectively.

  c. For each algorithm, use the estimated fit of the exposure mechanism to target the initial estimator $\bar{Q}_n(A, W)$, also fit with the training set. Denote the targeted regression fits as $\bar{Q}_n^{(a),*}(A, W)$, $\bar{Q}_n^{(b),*}(A, W)$ and $\bar{Q}_n^{(c),*}(A, W)$ where the superscript corresponds to the algorithm used to estimate the exposure mechanism.

  d. For each algorithm, obtain targeted predictions of the outcome(s) for the observation(s) in the validation set under the treatment and the control. For the first algorithm for fitting the exposure mechanism, for example, we would have $\overline{Q}_n^{(a),*}(1, W_k)$ and $\overline{Q}_n^{(a),*}(0, W_k)$ for observation O_k in the validation set.

  e. For each algorithm, evaluate the loss function for the observation(s) in the validation set by plugging in the algorithm-specific predictions. For example, if our target of inference were the SATE in a non-matched trial, we would have for the first algorithm

$$\mathscr{L}^{\mathscr{S}}(g_n^{(a)},\overline{Q}_n^{(a),*})(O_k)=\left[\left(\frac{(A_k=1)}{g_n^{(a)}(1|W_k)}-\frac{(A_k=0)}{g_n^{(a)}(0|W_k)}\right)(Y_k-\overline{Q}_n^{(a),*}(A_k,W_k))\right]^2$$

for observation O_k in the validation set.

f. For each algorithm for estimating the exposure mechanism, obtain an estimate of the risk by averaging the estimated losses across the observations in validation set v. If our target of inference were the SATE in a non-matched trial, we would have for the first algorithm for estimating the exposure mechanism

$$Risk_v^{(a)}=\frac{1}{n_v}\sum_{k\in v}\mathscr{L}^{\mathscr{S}}(g_n^{(a)},\overline{Q}_n^{(a),*})(O_k)$$

where n_v denotes the number of observations in validation set v.

B. For each algorithm, average the estimated risks across the V folds.

C. Select the algorithm with the smallest cross-validated risk. This is the algorithm yielding the smallest cross-validated variance estimate.

The chosen estimator for estimating the exposure mechanism is then used for targeting in Step 2 of the TMLE algorithm.

6. Obtaining Inference

In summary, we have proposed the following data-adaptive C-TMLE to maximize the precision and power of a randomized trial.

Step 1. Initial estimation of the conditional mean outcome with the working model $\bar{Q}_n(A,W)$, which was data-adaptively selected to maximize the empirical efficiency of the analysis (Sec. 3.1 for a non-matched trial and Sec. 4.1 for a matched trial).

Step 2. Targeting the initial estimator using the estimated exposure mechanism $g_n(A|W)$, which was data-adaptively selected to further maximize the empirical efficiency of the analysis (Sec. 5.1).

Step 3. Obtaining a point estimate by averaging the difference in the targeted predictions of the outcome under the treatment and under the control:

$$\Psi_n(\overline{Q}_n^*)=\frac{1}{n}\sum_{i=1}^{n}[\overline{Q}_n^*(1,W_i)-\overline{Q}_n^*(0,W_i)].$$

We now need a variance estimator that accounts for the selection process. For this, we propose using a cross-validated variance estimator. As before, the data are split into validation and training sets, respecting the unit of (conditional) independence. The selected TMLE is fit using the data in the training set and used to estimate the influence curve[§] for the observation(s) in the validation set. The sample variance of the cross-validated estimate of the influence curve can then be used for hypothesis testing and the construction of Wald-type confidence intervals. Step-by-step instructions are given in Section 2 of the Supplementary Material. We note that for a very small library (e.g. 2 candidate TMLEs), simulations support the use of the standard, as opposed to cross-validated, variance estimator for inference. For further details, see the Section 3 of the Supplementary Material.

7. Small Sample Simulations

We present the following simulation studies to demonstrate (1) implementation of the proposed methodology, (2) the potential gains in precision and power from data-adaptive estimation of the conditional mean outcome, (3) the additional gains in precision and power from collaborative estimation of the exposure mechanism, and (4) maintenance of nominal confidence interval coverage. All simulations were conducted in R v3.2.3 [36].

7.1. Study 1

For each unit $i = \{1, \dots, n\}$, we generated the nine baseline covariates by drawing from a multivariate normal with mean 0 and variance 1. The correlation between the first three covariates $\{W1, W2, W3\}$ and between the second three covariates $\{W4, W5, W6\}$ was 0.5, while the correlation between the remaining covariates $\{W7, W8, W9\}$ was 0. The exposure A was randomized such that the treatment allocation was balanced overall. For the non-matched trial, we randomly assigned the intervention to $n/2$ units and the control to the remaining $n/2$ units. For the pair-matched trial, we used the non-bipartite matching algorithm nbpMatch to pair units on covariates $\{W1, \dots, W6\}$ [50], and the exposure A was randomized within the resulting matched pairs. Recall A is a binary indicator, equalling 1 if the unit was assigned the treatment and 0 if the unit was assigned the control. For each unit, the outcome Y was then generated as

$$Y=0.4A+0.25(W1+W2+W4+W5+U_Y)+0.25A(W1+U_Y)$$

where U_Y was drawn from a standard normal. We also generated the counterfactual outcomes $Y(1)$ and $Y(0)$ by intervening to set $A = a$. To reflect the limited sample sizes common in early phase clinical trials and in cluster randomized trials, we selected a sample size of $n = 40$. This resulted in $n/2 = 20$ conditionally independent units in the pair-matched trial.

[§] For the TMLE of the population effect in a pair-matched trial, we also need a cross-validated estimate of the correction term ρ_n (Eq. 7). This term is a function of the residuals, which can be estimated for each pair in the validation set based on targeted estimator $\overline{Q}_n^*(A, W)$, fit with the training set.

For each study design (non-matched or matched), this data generating process was repeated 2,500 times. Recall that the sample effect (Eq. 2) is a data-adaptive parameter; its value changes with each new selection of units. Thereby, for each repetition, the SATE was calculated as the sample average of the difference in the counterfactual outcomes. The SATE ranged from 0.22 to 0.59 with a mean of 0.40. In contrast, the population effect (Eq. 1) is constant and was calculated by averaging the difference in the counterfactual outcomes over a population of 900,000 units. The true value of the PATE was 0.40.

We compared the performance of the unadjusted estimator to TMLE with various approaches to covariate adjustment. Specifically, we implemented the TMLE algorithm, where the initial estimation of the conditional mean outcome $\bar{Q}_0(A,W)$ was based on a linear working model with main terms for the intervention A and the irrelevant covariate $W9$ and where the exposure mechanism was treated as known: $g_0(A|W) = 0.5$. This approach was equivalent to standard maximum likelihood estimation (MLE) and represented the unfortunate scenario where the researcher pre-specified adjustment for a covariate that was not predictive of the outcome.

We also implemented a TMLE with the data-adaptive approach for Step 1 initial estimation of the conditional mean outcome (Sec. 3.1 and 4.1). Our library consisted of 10 working linear regression models, each with an intercept, a main term for the exposure A and a main term for one baseline covariate: $\{\varnothing, W1, \ldots, W9\}$, where $\varnothing$ corresponds to the unadjusted estimator. Our measure of performance (i.e. our risk function) was the estimated asymptotic variance of the TMLE, appropriate for the target parameter and study design. We chose the candidate working model based on leave-one-out cross-validation for the non-matched trial and leave-one-pair-out cross-validation for the matched trial. We also implemented Collaborative-TMLE (C-TMLE), which couples the data-adaptive approach for Step 1 initial estimation of the conditional mean outcome (Sec. 3.1 and 4.1) with the data-adaptive approach for Step 2 targeting (Sec. 5.1). For the latter, our library of candidates to estimate the exposure mechanism consisted of 10 working logistic regression models, each with an intercept and a main term for one baseline covariate: $\{\varnothing, W1, \ldots, W9\}$. The same loss function and cross-validation scheme were used for C-TMLE.

For the unadjusted estimator and the MLE, inference was based on the estimated influence curve. For the data-adaptive TMLEs, inference was based on the cross-validated estimate of the influence curve (Sec. 6). We assumed the standardized estimator followed the Student's t-distribution with $n - 2 = 38$ degrees of freedom for the non-matched trial and with $n/2 - 1 = 19$ degrees of freedom for the matched trial.

7.1.1. Results—Table 1 illustrates the performance of the estimators over the 2,500 simulated data sets. Specifically, we show the mean squared error (MSE), the relative MSE (rMSE), the average standard error estimate $\hat{\sigma}$, the attained power and the 95% confidence interval coverage. As expected, matching improved efficiency. The MSE of the unadjusted estimator, for example, was over 2 times larger in the non-matched trial than in the pair-matched trial. Furthermore, for the pair-matched trial, targeting the sample effect, as opposed to the population effect, resulted in substantial gains in attained power: 36% with the unadjusted estimator for the PATE and 53% with the same estimator for the SATE. For the trial without matching, targeting the sample parameter increased efficiency (smaller MSE), but did not directly translate into increased power due to the conservative variance estimator for the SATE.

In all scenarios, the MSE of the MLE, adjusting for the irrelevant covariate $W9$, was worse than the other estimators. This demonstrates the potential peril of relying on one pre-specified adjustment variable. Indeed, the TMLE with data-adaptive selection of the initial estimator of $\bar{Q}_0(A, W)$ improved precision over the unadjusted estimator and the MLE. Collaborative estimation of the exposure mechanism $g_0(A|W)$ led to further gains in precision. Consider, for example, estimation of the PATE in a trial without matching. The MSE of the unadjusted estimator was 1.49 times larger than the TMLE and 1.57 times larger than the C-TMLE. The attained power was 34%, 48% and 48%, respectively. As a second example, consider the attained power to detect that the SATE was different from zero in the pair-matched trial. We would have 53% power with the unadjusted estimator and with the MLE, adjusting for the irrelevant covariate $W9$. By incorporating the cross-validation selector for initial estimation of $\bar{Q}_0(A, W)$, the TMLE achieved 65% power. By further incorporating collaborative estimation of the exposure mechanism $g_0(A|W)$, the C-TMLE achieved 67% power.

Overall, the greatest efficiency was achieved with C-TMLE for the SATE in the pair-matched trial. Indeed, the MSE of the unadjusted estimator for the population parameter in the trial without matching was 3 times larger than the MSE of the C-TMLE for the sample effect in the pair-matched trial. Throughout, the confidence interval coverage was maintained near or above the nominal rate of 95%. Table 1 of the Supplementary Material provides the proportion of times each working model was selected with the TMLE and C-TMLE algorithms.

7.2. Study 2

For the second simulation study, we increased the complexity of the data-generating process and reduced the sample size to $n = 30$. As before, we generated nine baseline covariates from a multivariate normal with mean 0, variance 1 and the same correlation structure. We also generated a binary variable R, equalling 1 with probability 0.5 and equalling −1 with probability 0.5. The final covariate Z was generated as a function of these baseline covariates and random noise U_Z:

$$Z = R \times \text{logit}^{-1}(W1 + W4 + W7 + 0.5U_Z)$$

where U_Z was drawn independently from a standard normal. As before, the intervention A was randomized with balanced allocation. For the pair-matched trial, we used the non-bipartite matching algorithm nbpMatch to explore two matching sets [50]. In the first, units were matched on R, a baseline covariate strongly impacting Z. In the second, units were matched on $\{R, W2, W5, W8\}$. For each unit, the outcome Y was then generated as

$$Y = \text{logit}^{-1}[0.75A + 0.5(W2 + W5 + W8) + 1.5Z + 0.25U_Y + 0.75A(W2 - W5) + 0.5AZ]/7.5$$

where U_Y was drawn from a standard normal. Thereby, the outcome was a continuous variable bounded in [0, 1] (e.g. a proportion). We also generated the counterfactual outcomes $Y(1)$ and $Y(0)$ by intervening to set $A = a$. For each study design, this data generating process was repeated 2,500 times. The SATE and PATE were calculated as before. The SATE ranged from 0.2% to 3.3% with a mean of 1.6%. The true value of the PATE was 1.6%. Table 2 depicts the relationship between the baseline covariates and the outcome as well as the adaptive pair-matching schemes.

We compared the same algorithms: the unadjusted estimator, the MLE adjusting for the irrelevant covariate $W9$, the TMLE with data-adaptive estimation of the conditional mean outcome, and the C-TMLE pairing data-adaptive estimation of the conditional mean outcome with data-adaptive targeting. Our library for initial estimation of the conditional mean outcome $\bar{Q}_0(A, W)$ consisted of 12 working logistic regression models, each with an intercept and a main term for the exposure A and a main term for one candidate adjustment variable $\{\varnothing, R, W1, \ldots, W9, Z\}$. Our library for collaborative estimation of the exposure mechanism $g_0(A|W)$ included 12 working logistic regression models, each with an intercept and a main term for one candidate adjustment variable: $\{\varnothing, R, W1, \ldots, W9, Z\}$. We used the same measure of performance and cross-validation scheme. As before, inference was based on the estimated influence curve for the unadjusted estimator and the MLE and on the cross-validated estimate of the influence curve for the TMLEs (Sec. 6). We assumed the standardized estimator followed the Student's t-distribution with $n - 2 = 28$ degrees of freedom for the non-matched trial and with $n/2 - 1 = 14$ degrees of freedom for the matched trial.

7.2.1. Results—The results for the second simulation study are given in Table 3 and largely echoed the above findings. Pair-matching, even on a single covariate (i.e. matching set 1), improved the precision of the analysis. Targeting the sample effect instead of the population effect further improved efficiency. Allowing for data-adaptive selection of the working model for initial estimation of $\bar{Q}_0(A, W)$ yielded even greater precision, and the most efficient analysis was with C-TMLE. Indeed, the MSE of the unadjusted estimator for the PATE in the non-matched trial was nearly 4.5 times higher than the MSE of the C-TMLE for the SATE when matching on predictive covariates (i.e. matching set 2). This resulted in 29% more power to detect the intervention effect.

For these simulations, there was a notable impact of parameter specification on estimator performance. We first focus on the estimation of the PATE and then on estimation of the SATE. When the population effect was the target of inference, the gains in attained power from pair-matching were attenuated despite the gains in MSE. This was likely due to the slight underestimation of the standard error in the non-matched trial and overestimation in the pair-matched trial. Indeed, the 95% confidence interval coverage in the non-matched trial was slightly less than nominal (93–94%), while the coverage when matching well (i.e. set 2) approached 100%. For this set of simulations, the correction factor ρ_n (Eq. 7) used in variance estimation for the pair-matched design was approximately 0. As a result, the variance estimator in the pair-matched trial was quite conservative, and the cross-validation selection scheme was more optimized for the non-matched trial. The latter point is evidenced by Table 2 in the Supplementary Material, which shows the proportion of times each candidate working model was selected. The logistic regression model adjusting for R was selected for initial estimation of $\bar{Q}_0(A, W)$ in 10% of the studies without matching and in 7% of the studies when matching well on R (i.e. set 1). Furthermore, when matching on several covariates (i.e. set 2), the selection of working models for $\bar{Q}_0(A, W)$ was very similar to the selection in the non-matched trial.

In contrast, when estimating the SATE, smaller MSE translated to greater attained power, while maintaining nominal, if not conservative, confidence interval coverage. For example, the attained power of the TMLE was 33% in the non-matched trial, 40% when matching on a single covariate and 47% when matching on several covariates. Likewise, the attained power of the C-TMLE was 34% in the non-matched trial, 44% in the trial pair-matching on a single covariate and 53% in trial matching on several covariates. In Table 2 of the Supplementary Material, we see that the working model adjusting for R was selected for initial estimation of $\bar{Q}_0(A, W)$ in 10% of the studies without matching and only in 2% of the studies when matching well on R (i.e. set 1). In the latter, more weight was given to other predictive baseline covariates, such as $W2$ and Z.

8. Discussion

This paper builds on the rich history of covariate adjustment in randomized trials [1–4, 13, 15, 21, 27, 51]. In particular, Rubin and van der Laan [14] proposed the principle of *empirical efficiency maximization* as a strategy to select the estimator of conditional mean outcome $\bar{Q}_0(A, W)$ that minimized the empirical variance of the estimated efficient influence curve. Their procedure, however, relied on solving a weighted nonlinear least squares problem. Our approach only requires researchers to take the sample variance. More recently, van der Laan and Gruber [32] proposed collaborative estimation of the exposure mechanism to achieve the greatest bias reduction in the targeting step of TMLE in a observational study. In randomized trials, there is no risk of bias from regression model misspecification [16]. Thereby, the collaborative approach, implemented here, serves only to increase precision by estimating the known exposure-mechanism. To our knowledge, this is the first research into C-TMLE in a randomized trial setting. Most recently, van der Laan [31] suggested selection of the candidate (C-)TMLE based on minimizing the estimated variance of its influence curve. Our paper generalizes this scheme for estimation and inference of both the population and sample average treatment effects in randomized trials with and without pair-matching.

Our simulations illustrate the performance of the proposed procedure in realistically-sized (i.e. small) trials. In particular, with only 15 (conditionally) independent units, our procedure was able to identify the optimal working model for initial estimation of $\bar{Q}_0(A, W)$ from a library of 12 candidates as well as for collaborative estimation of $g_0(A|W)$ from a library of 12 candidates, while maintaining close to nominal confidence interval coverage. The simulations also indicated the most efficient combination (design, target parameter and adjustment approach) was estimating the sample effect with C-TMLE in pair-matched trial. Indeed, this approach was nearly 4.5 times more efficient than targeting the population effect with the unadjusted estimator in the non-matched trial. Thereby, our procedure dispels the common concern of “analytical limitations” to pair-matched trials (e.g. [11, 22, 23]).

There are several areas of future work. First, our library of candidate estimators was limited to simple parametric working models. This choice was made both for pedagogic purposes and to avoid over-fitting in small studies. Although not studied directly in simulations, it should be possible for larger trials to expand the library to include working models with multiple adjustment variables and interactions as well as selection procedures (e.g. stepwise regression) and other semiparametric algorithms. Future work will involve simulations to evaluate the methodology in larger trials. Simulations, such as those presented here, can inform the practitioner as to the optimal library size for his/her specific application. Future work will also evaluate using cross-validation to select the size of the candidate library. Second, this manuscript focused on randomized trials with and without pair-matching. The application to matched triplets, as opposed to matched pairs, should be straightforward. However, the impact of other designs (e.g. adaptive stratification, restricted randomization, and the minimization method) on estimation and inference merits additional consideration. Finally, we focused on two causal parameters: the population and sample average treatment effects. However, TMLE is a general methodology for the construction of double robust, semiparametric, efficient substitution estimators for a wide range of parameters. Our proposed strategy for covariate selection should extend to other causal parameters, such as the conditional average treatment effect (e.g. [25, 52]), the average treatment effect among the treated (e.g. [43]), and the natural direct effect (e.g. [53, 54]).

Overall, we proposed a general strategy to increase power in randomized trials. The proposed methodology is applicable to early and later phase clinical trials as well as cluster randomized trials. Specifically, we used cross-validation to select the candidate TMLE that optimized the efficiency of the analysis. Since the step-by-step algorithm (including the library definition) was pre-specified, there was no risk of bias or misleading inference from *ad hoc* analytic decisions. In other words, we have proposed a black box procedure to data-adaptively select the most powerful analysis. Furthermore, including the unadjusted estimator as a candidate obviates the need for guidelines on whether or not to adjust (e.g. [21, 27]). Finally, our procedure is tailored to the scientific question (population vs. sample effect) and study design (with or without pair-matching). Decisions about whether to adjust and how to adjust are made with a rigorous and principled approach, removing some of the “human art” from statistics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Contract/grant sponsor: NIH grant numbers R01AI074345, UM1AI069502 & U01AI099959

Research reported in this publication was supported by Division of AIDS, NIAID of the National Institutes of Health under award numbers R01-AI074345, UM1AI069502 and U01AI099959 and in part by the President’s Emergency Plan for AIDS Relief and Gilead Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, PEPFAR, or Gilead. The SEARCH project gratefully acknowledges the Ministries of Health of Uganda and Kenya, our research team, collaborators and advisory boards, and especially all communities and participants involved.

References

1. Fisher, R. Statistical methods for research workers. 4. Oliver and Boyd Ltd; Edinburgh: 1932.
2. Cochran W. Analysis of covariance: its nature and uses. Biometrics. 1957; 13:261–281. DOI: 10.2307/2527916
3. Cox D, McCullagh P. Some aspects of analysis of covariance. Biometrics. 1982; 38(3):541–561. DOI: 10.2307/2530040 [PubMed: 7171689]
4. Tsiatis A, Davidian M, Zhang M, Lu X. Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: A principled yet flexible approach. Statistics in Medicine. 2008; 27(23):4658–4677. DOI: 10.1002/sim.3113 [PubMed: 17960577]
5. Moore K, van der Laan M. Covariate adjustment in randomized trials with binary outcomes: Targeted maximum likelihood estimation. Statistics in Medicine. 2009; 28(1):39–64. DOI: 10.1002/sim.3445 [PubMed: 18985634]
6. Statistical principles for clinical trials E9. Feb. 1998 ICH Harmonised Tripartite Guideline.
7. Pocock S, Assmann S, Enos L, Kasten L. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Statistics in Medicine. 2002; 21(19):2917–2930. DOI: 10.1002/sim.1296 [PubMed: 12325108]
8. Hayes, R.; Moulton, L. Cluster Randomised Trials. Chapman & Hall/CRC; Boca Raton: 2009.
9. Austin P, Manca A, Zwarensteina M, Juurlinka D, Stanbrook M. A substantial and confusing variation exists in handling of baseline covariates in randomized controlled trials: a review of trials published in leading medical journals. Journal of Clinical Epidemiology. 2010; 63:142–153. DOI: 10.1016/j.jclinepi.2009.06.002 [PubMed: 19716262]
10. Kahn B, Jairath V, Doré C, Morris T. The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. Trials. 2014; 15(139):1–7. DOI: 10.1186/1745-6215-15-139 [PubMed: 24382030]
11. Campbell, M. Cluster randomized trials. In: Ahrens, W.; Pigeot, I., editors. Handbook of Epidemiology, 2nd edition. Springer; 2014.
12. European Medicines Agency. Guideline on adjustment for baseline covariates in clinical trials. London: Feb. 2015
13. Zhang M, Tsiatis A, Davidian M. Improving Efficiency of Inferences in Randomized Clinical Trials Using Auxiliary Covariates. Biometrics. 2008; 64(3):707–715. DOI: 10.1111/j.1541-0420.2007.00976.x [PubMed: 18190618]
14. Rubin DB, van der Laan M. Empirical efficiency maximization: improved locally efficient covariate adjustment in randomized experiments and survival analysis. The International Journal of Biostatistics. 2008; 4(1) Article 5. doi: 10.2202/1557-4679.1084
15. Shen C, Li X, Li L. Inverse probability weighting for covariate adjustment in randomized studies. Statistics in Medicine. 2014; 33:555–568. DOI: 10.1002/sim.5969 [PubMed: 24038458]
16. Rosenblum M, van der Laan M. Simple, efficient estimators of treatment effects in randomized trials using generalized linear models to leverage baseline variables. The International Journal of Biostatistics. 2010; 6(1) Article 13. doi: 10.2202/1557-4679.1138
17. Robins J. A new approach to causal inference in mortality studies with sustained exposure periods–application to control of the healthy worker survivor effect. Mathematical Modelling. 1986; 7:1393–1512. DOI: 10.1016/0270-0255(86)90088-6
18. van der Laan M, Rubin D. Targeted maximum likelihood learning. The International Journal of Biostatistics. 2006; 2(1) Article 11. doi: 10.2202/1557-4679.1043
19. van der Laan, M.; Rose, S. Targeted Learning: Causal Inference for Observational and Experimental Data. Springer; New York Dordrecht Heidelberg London: 2011.
20. Scharfstein D, Rotnitzky A, Robins J. Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models (with Rejoiner). Journal of the American Statistical Association. 1999; 94(448):1096–1120. 1135–1146. DOI: 10.2307/2669930
21. Colantuoni, E.; Rosenblum, M. Technical Report. Vol. 263. Johns Hopkins University, Dept. of Biostatistics Working Papers; Feb. 2015 Leveraging prognostic baseline variables to gain precision in randomized trials. http://biostats.bepress.com/jhubiostat/paper263

22. Klar N, Donner A. The merits of matching in community intervention trials: a cautionary tale. Statistics in Medicine. 1997; 16(15):1753–1764. DOI: 10.1002/(SICI)1097-0258(19970815)16:15〈1753::AID-SIM597〉3.0.CO;2-E [PubMed: 9265698]

23. Imbens, G. Technical Report. 2011. Experimental design for unit and cluster randomized trials. NBER Technical Working Paper

24. van der Laan M, Balzer L, Petersen M. Adaptive Matching in Randomized Trials and Observational Studies. Journal of Statistical Research. 2012; 46(2):113–156. [PubMed: 25097298]

25. Balzer L, Petersen M, van der Laan M. the SEARCH Consortium. Adaptive pair-matching in randomized trials with unbiased and efficient effect estimation. Statistics in Medicine. 2015; 34(6): 999–1011. DOI: 10.1002/sim.6380 [PubMed: 25421503]

26. Balzer L, Petersen M, van der Laan M. Targeted estimation and inference of the sample average treatment effect in trials with and without pair-matching. Statistics in Medicine. 2016; Early View. doi: 10.1002/sim.6965

27. Moore K, Neugebauer R, Valappil T, van der Laan M. Robust extraction of covariate information to improve estimation efficiency in randomized trials. Statistics in Medicine. 2011; 30(19):2389–2408. DOI: 10.1002/sim.4301 [PubMed: 21751231]

28. Califf R, Zarin D, Kramer J, Sherman R, Aberle L, Tasneem A. Characteristics of clinical trials registered in ClinicalTrials.gov, 2007–2010. JAMA. 2012; 307(17):1838–1847. DOI: 10.1001/jama.2012.3424 [PubMed: 22550198]

29. Selvaraj S, Prasad V. Characteristics of cluster randomized trials: Are they living up to the randomized trial? JAMA Internal Medicine. 2013; 173(23):313.doi: 10.1001/jamainternmed.2013.1638 [PubMed: 23337957]

30. Olken B. Pre-analysis plans in economics. Technical Report. Massachusetts Institute of Technology Department of Economics; 2015. http://economics.mit.edu/files/10399

31. van der Laan, M. Appendix A.19: Efficiency maximization and TMLE. In: van der Laan, M.; Rose, S., editors. Targeted Learning: Causal Inference for Observational and Experimental Data. Springer; New York Dordrecht Heidelberg London: 2011.

32. van der Laan M, Gruber S. Collaborative double robust targeted maximum likelihood estimation. The International Journal of Biostatistics. 2010; 6(1)doi: 10.2202/1557-4679.1181

33. Gruber, S.; van der Laan, M. C-TMLE of an Additive Point Treatment Effect. In: van der Laan, M.; Rose, S., editors. Targeted Learning: Causal Inference for Observational and Experimental Data. Springer; New York Dordrecht Heidelberg London: 2011.

34. Neyman J. 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. 1923; 5:465–480.

35. University of California, San Francisco. Sustainable East Africa Research in Community Health (SEARCH). ClinicalTrials.gov. 2013. http://clinicaltrials.gov/show/NCT01864603

36. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2015. http://www.R-project.org

37. Rubin D. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology. 1974; 66(5):688–701. DOI: 10.1037/h0037350

38. Gruber S, van der Laan M. A targeted maximum likelihood estimator of a causal effect on a bounded continuous outcome. The International Journal of Biostatistics. 2010; 6(1) Article 26. doi: 10.2202/1557-4679.1260

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

40. Small D, Ten Have T, Rosenbaum P. Randomization inference in a group–randomized trial of treatments for depression: Covariate adjustment, noncompliance, and quantile effects. Journal of the American Statistical Association. 2008; 103(481):271–279. DOI: 10.1198/016214507000000897

41. Zhang K, Traskin M, Small D. A powerful and robust test statistic for randomization inference in group-randomized trials with matched pairs of groups. Biometrics. 2012; 68:75–84. DOI: 10.1111/j.1541-0420.2011.01622.x [PubMed: 21732926]

42. Rubin DB. Comment: Neyman (1923) and causal inference in experiments and observational studies. Statistical Science. 1990; 5(4):472–480.

43. Imbens G. Nonparametric estimation of average treatment effects under exogeneity: a review. Review of Economics and Statistics. 2004; 86(1):4–29. DOI: 10.1162/003465304323023651

44. Greevy R, Lu B, Silber J, Rosenbaum P. Optimal multivariate matching before randomization. Biostatistics. 2004; 5(2):263–275. DOI: 10.1093/biostatistics/5.2.263 [PubMed: 15054030]

45. Zhang K, Small D. Comment: The essential role of pair matching in cluster-randomized experiments, with application to the Mexican universal health insurance evaluation. Statistical Science. 2009; 25(1):59–64. DOI: 10.1214/09-STS274B

46. Lu B, Greevy R, Xu X, Beck C. Optimal Nonbipartite Matching and its Statistical Applications. American Statistician. 2011; 65(1):21–30. DOI: 10.1198/tast.2011.08294 [PubMed: 23175567]

47. Freedman L, Gail M, Green S, Corle D. The COMMIT Research Group. The Efficiency of the Matched-Pairs Design of the Community Intervention Trial for Smoking Cessation (COMMIT). Controlled Clinical Trials. 1997; 18(2):131–139. DOI: 10.1016/S0197-2456(96)00115-8 [PubMed: 9129857]

48. Campbell M, Donner A, Klar N. Developments in cluster randomized trials and *Statistics in Medicine*. Statistics in Medicine. 2007; 26(1):2–19. DOI: 10.1002/sim.2731 [PubMed: 17136746]

49. Imai K, King G, Nall C. The essential role of pair matching in cluster-randomized experiments, with application to the Mexican Universal Health Insurance Evaluation. Statistical Science. 2009; 24(1):29–53. DOI: 10.1214/08-STS274

50. Beck, C.; Lu, B.; Greevy, R. nbpMatching: functions for optimal non-bipartite optimal matching. 2016. https://CRAN.R-project.org/package=nbpMatching, R package version 1.5.0

51. Yuan S, Zhang H, Davidian M. Variable selection for covariate-adjusted semiparametric inference in randomized clinical trials. Statistics in Medicine. 2012; 31:3789–3804. DOI: 10.1002/sim.5433 [PubMed: 22733628]

52. Abadie A, Imbens G. Simple and bias-corrected matching estimators for average treatment effects. Technical Report. 2002; 283 NBER technical working paper.

53. Robins J, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology. 1992; 3:143–155. DOI: 10.1097/00001648-199203000-00013 [PubMed: 1576220]

54. Pearl, J. Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann; San Francisco: 2001. Direct and indirect effects; p. 411-420.

Table 1

Summary of estimator performance for Simulation 1. The rows denote the study design and the estimator: unadjusted, MLE adjusting for $W9$, TMLE with data-adaptive selection of the initial estimator, and Collaborative-TMLE (C-TMLE) with data-adaptive selection of the initial estimator paired with data-adaptive estimation of the exposure mechanism.

	PATE					SATE				
	MSE[a]	**rMSE**[b]	$\hat{\sigma}^c$	**Pow**[d]	**Cov**[e]	**MSE**[a]	**rMSE**[b]	$\hat{\sigma}^c$	**Pow**[d]	**Cov**[e]
Non-Matched										
Unadj	6.8E-2	1.00	0.25	0.34	0.94	6.4E-2	1.06	0.25	0.34	0.94
MLE	6.9E-2	0.98	0.25	0.35	0.94	6.5E-2	1.04	0.25	0.35	0.94
TMLE	4.5E-2	1.49	0.20	0.48	0.94	4.2E-2	1.62	0.20	0.48	0.95
C-TMLE	4.3E-2	1.57	0.20	0.48	0.95	4.0E-2	1.70	0.20	0.48	0.96
Matched										
Unadj	3.2E-2	2.10	0.22	0.36	0.99	2.9E-2	2.31	0.18	0.53	0.97
MLE	3.4E-2	2.01	0.22	0.37	0.98	3.1E-2	2.19	0.18	0.53	0.96
TMLE	2.6E-2	2.64	0.19	0.51	0.98	2.3E-2	2.93	0.16	0.65	0.96
C-TMLE	2.5E-2	2.71	0.18	0.53	0.98	2.2E-2	3.03	0.15	0.67	0.96

[a] Mean squared error: the bias (average deviation between the point estimate and sample-specific true value) - squared plus the variance

[b] Relative MSE: the MSE of the unadjusted estimator for the PATE in a non-matched trial relative to (divided by) the MSE of another estimator

[c] Average standard error estimate, based on the estimated influence curve

[d] Attained power: proportion of times the false null hypothesis was rejected

[e] Confidence interval (CI) coverage: proportion of times the true value was contained in the 95% CI

Table 2

For Simulation 2, the relationships between baseline covariates and the outcome as well as the adaptive pair-matching schemes.

		correlation 0.5			**correlation 0.5**			**correlation 0**			
	R	***W*1**	***W*2**	***W*3**	***W*4**	***W*5**	***W*6**	***W*7**	***W*8**	***W*9**	***Z***
Parents of covariate *Z*	✓	✓			✓			✓			
Parents of the outcome *Y*			✓			✓			✓		✓
Matching set 1	✓										
Matching set 2	✓		✓			✓			✓		

Table 3

Summary of estimator performance for Simulation 2. The rows denote the study design and the estimator: unadjusted, MLE adjusting for $W9$, TMLE with data-adaptive selection of the initial estimator, and Collaborative-TMLE (C-TMLE) with data-adaptive selection of the initial estimator paired with data-adaptive estimation of the exposure mechanism.

	PATE					SATE				
	MSE[a]	rMSE[b]	$\hat{\sigma}^c$	Pow[d]	Cov[e]	MSE[a]	rMSE[b]	$\hat{\sigma}^c$	Pow[d]	Cov[e]
Non-matched										
Unadj	1.8E-4	1.00	0.013	0.24	0.94	1.6E-4	1.12	0.013	0.24	0.95
MLE	1.8E-4	0.95	0.012	0.25	0.93	1.7E-4	1.06	0.012	0.25	0.94
TMLE	1.2E-4	1.50	0.010	0.33	0.94	9.8E-5	1.79	0.010	0.33	0.96
C-TMLE	1.1E-4	1.54	0.010	0.34	0.93	9.5E-5	1.85	0.010	0.34	0.96
Match Set 1										
Unadj	1.1E-4	1.54	0.012	0.21	0.98	9.2E-5	1.90	0.011	0.28	0.97
MLE	1.2E-4	1.48	0.012	0.23	0.97	9.7E-5	1.81	0.011	0.29	0.97
TMLE	9.2E-5	1.91	0.010	0.31	0.97	6.9E-5	2.52	0.009	0.40	0.96
C-TMLE	9.0E-5	1.95	0.010	0.33	0.96	6.9E-5	2.53	0.008	0.44	0.95
Match Set 2										
Unadj	6.5E-5	2.70	0.011	0.17	0.99	4.6E-5	3.79	0.009	0.37	0.98
MLE	7.3E-5	2.41	0.011	0.20	0.99	5.4E-5	3.27	0.009	0.37	0.98
TMLE	5.3E-5	3.30	0.009	0.28	0.99	3.8E-5	4.66	0.008	0.47	0.98
C-TMLE	5.3E-5	3.28	0.009	0.32	0.99	3.9E-5	4.44	0.007	0.53	0.97

[a] Mean squared error: the bias (average deviation between the point estimate and sample-specific true value) - squared plus the variance

[b] Relative MSE: the MSE of the unadjusted estimator for the PATE in a non-matched trial relative to (divided by) the MSE of another estimator

[c] Average standard error estimate, based on the estimated influence curve

[d] Attained power: proportion of times the false null hypothesis was rejected

[e] Confidence interval (CI) coverage: proportion of times the true value was contained in the 95% CI