Published in final edited form as:
Stat Med. 2015 March 15; 34(6): 999–1011. doi:10.1002/sim.6380.

Adaptive pair-matching in randomized trials with unbiased and efficient effect estimation

Laura B. Balzer[a,*], **Maya L. Petersen**[a], **Mark J. van der Laan**[a], and **the SEARCH Consortium**
[a] Division of Biostatistics, University of California, Berkeley, CA 94110-7358, USA

Abstract

In randomized trials, pair-matching is an intuitive design strategy to protect study validity and to potentially increase study power. In a common design, candidate units are identified, and their baseline characteristics used to create the best $n/2$ matched pairs. Within the resulting pairs, the intervention is randomized, and the outcomes measured at the end of follow-up. We consider this design to be adaptive, because the construction of the matched pairs depends on the baseline covariates of all candidate units. As a consequence, the observed data cannot be considered as $n/2$ independent, identically distributed (i.i.d.) pairs of units, as common practice assumes. Instead, the observed data consist of n dependent units. This paper explores the consequences of adaptive pair-matching in randomized trials for estimation of the average treatment effect, conditional the baseline covariates of the n study units. By avoiding estimation of the covariate distribution, estimators of this conditional effect will often be more precise than estimators of the marginal effect. We contrast the unadjusted estimator with targeted minimum loss-based estimation (TMLE) and show substantial efficiency gains from matching and further gains with adjustment. This work is motivated by the Sustainable East Africa Research in Community Health (SEARCH) study, an ongoing community randomized trial to evaluate the impact of immediate and streamlined antiretroviral therapy on HIV incidence in rural East Africa.

Keywords

adaptive designs; causal inference; efficiency; pair-matching; randomized trials; targeted minimum loss-based estimation (TMLE)

1. Introduction

Pair-matching helps balance treatment groups with respect to important determinants of the outcome at baseline [1, 2]. In observational studies, matching can help control for confounding. In randomized trials, there is no confounding; the probability of receiving the intervention or the control is a known constant. Nonetheless, covariate imbalance is common in small trials, and data sparsity may limit our ability to adjust for these characteristics during the analysis. Thereby, matching is sometimes implemented in

[*] Correspondence to: Laura Balzer, Division of Biostatistics, University of California, Berkeley, CA 94110-7358, USA. lbbalzer@berkeley.edu.

randomized trials to protect study credibility. For example, the “face validity” [3] of a randomized trial for violence prevention could be compromised if neighborhoods with highest baseline violence were all randomized, by chance, to the control level of the intervention. Matching is also implemented to improve study power. By decreasing variation in the outcome within pairs, matching may, but is not guaranteed to, increase study efficiency. The conflicting recommendations on pair-matching have inspired a heated debate in the literature for over sixty years [3–17].

Much of the work in the design and the analysis of pair-matched trials has assumed that the observed data consist of $n/2$ independent and identically distributed (i.i.d.) units (e.g. [3, 17–21]). Such a data structure could arise by randomly sampling $n/2$ matched pairs from some target population of pre-existing matched units. Often, however, there may be substantial logistical or financial barriers to practical implementation of this design. Alternatively, this data structure could arise by (i) sampling a unit from an infinite target population, (ii) measuring its baseline covariates, (iii) repeatedly sampling units until the baseline covariates of the second were sufficiently close to the first, (iv) randomizing the intervention within the matched pair, (v) measuring the outcomes, and (vi) repeating this process $n/2$ times. This pair-matching scheme may also be impractical and is likely to be resource intensive. Theoretically, this design also yields less information for estimating the (population) average treatment effect than a design randomly pairing two sampled units [22].

A different pair-matching scheme was implemented in the Sustainable East Africa Research in Community Health (SEARCH) trial [23, 24]. SEARCH is a multinational, multidisciplinary consortium to evaluate the health, economic and educational impacts of a community-based strategy for immediate and streamlined antiretroviral therapy (ART) for all HIV-positive persons. In the trial, 54 candidate communities were identified from rural Uganda and Kenya. These clusters satisfied the study's inclusion criteria, which included community size, health care infrastructure and sufficient distance from other potential study units. Thirty-two communities were then pair-matched within region and on baseline predictors of HIV transmission and health care delivery. The intervention has been randomized within the resulting 16 matched pairs and the 5-year cumulative incidence of HIV will be measured at the conclusion of the trial. We consider this design to be *adaptive*, because partitioning of the study communities into matched pairs was a function of the baseline covariates of all candidates. Thereby, the observed data do not consist of $n = 32$ i.i.d. random variables or of $n/2 = 16$ i.i.d. paired random variables. Instead, the observed data consist of n dependent units. For examples of other types of adaptive designs, see [25–28].

To the best of our understanding, adaptive pair-matching has been implemented in several other cluster randomized trials. Examples include the Mwanza trial to prevent HIV [29], the PRISM trial to prevent postpartum depression [30], and the SPACE study to promote physical activity [31]. The process of selecting $n/2$ pairs based on the covariates of n candidates is also known in other literature as “nonbipartite matching” [14, 32] and has motivated the development of “optimal multivariate matching” algorithms to pair units based on several covariates simultaneously [33–36]. Previously, van der Laan *et al.* [37] explored the consequences of adaptive pair-matching for estimation of the population

average treatment effect. This paper explores the consequences of adaptive pair-matching for estimation of the average treatment effect, conditional on the baseline covariates of the *n* study units. For brevity, we will refer to this causal parameter as the conditional average treatment effect (CATE). This parameter was initially proposed in Abadie and Imbens [38], can be interpreted as the intervention effect, given the covariates of the sample at hand, and often leads to more precise estimators [16, 39, 40].

Adjustment for baseline covariates during the analysis can help control for chance imbalances in important determinants of the outcome and can also increase study efficiency [41–43]. Nonetheless, the recommendations on whether and how to adjust in pair-matched trials have been conflicting (e.g. [3, 15–17, 21, 44, 45]). The intervention effect can be estimated with the average of the differences in the outcomes within matched pairs. Alternatively, one could take a multi-step approach of first fitting a regression model with terms for the pairs and covariates (but not the intervention) and then contrasting the observed versus predicted outcomes within matched pairs [17, 29, 46]. In all cases, the estimation approach should be tailored to the parameter of interest (i.e population vs. conditional average treatment effect). To the best of our knowledge, this is the first paper to propose targeted minimum loss-based estimation (TMLE) for the CATE in a randomized trial. Without risking bias due to regression misspecification [41–43], TMLE allows for further adjustment for baseline characteristics (beyond that attained by matching alone) and thereby can provide an efficient estimate of the intervention effect.

The remaining article is outlined as follows. We first describe the adaptive design and the resulting data structure. Second, we motivate the use of the CATE as the causal parameter of interest. Third, we discuss two estimators of the corresponding statistical parameter: the unadjusted difference in outcomes within matched pairs and targeted minimum loss-based estimation (TMLE). The latter estimator allows for further adjustment of important baseline covariates, beyond that attained with matching, and is thereby more powerful under reasonable scenarios. We also provide asymptotically conservative variance estimators and finite sample simulations. We conclude with some practical recommendations. While the SEARCH trial serves as the motivating example, our conclusions are applicable to other randomized trials and also general to other study outcomes beyond incidence. Moreover, we focus on data at the level of the experimental unit (i.e. the unit of randomization). Thereby, our results are applicable to both individually randomized trials as well as cluster randomized trials. Detailed proofs are given in the Supplementary Material.

2. The Estimation Problem

The SEARCH consortium will estimate the impact of immediate antiretroviral therapy (ART), initiated at all CD4+ T cell counts and delivered by a streamlined care system, on the 5-year cumulative incidence of HIV [24]. The trial began enrolling communities in 2013, and data collection is ongoing. In communities randomized to the intervention, all individuals testing positive for HIV will be immediately eligible for ART with streamlined delivery, which includes enhanced services for initiation, linkage and retention in care. In communities randomized to the control, all individuals testing positive for HIV will be offered ART according to in-country guidelines, which are primarily based on CD4+ T cell

counts. HIV incidence, as well as other health, economic and educational outcomes, will be measured among approximately 320,000 individuals, followed longitudinally for the 5 years of the trial. The SEARCH study aims to understand the impact this community-based "test-and-treat" program on both HIV-positive individuals and their greater communities [47–54].

For the purposes of understanding the adaptive design, we focus on the cluster-level data. Let N denote the number of candidate communities considered for inclusion in the study, n denote the number of communities selected for the SEARCH trial, and $n/2$ denote the number of matched pairs. Let W represent the pre-intervention community-level covariates, which include region, proximity to trucking routes, occupational mix and baseline population HIV RNA levels [55]. A subset of these baseline covariates were used to select the $n/2$ best matched pairs of communities from the N possible candidates. Within the resulting pairs, the intervention was randomized. The treatment variable A is a binary indicator, equalling one if the community was assigned to the intervention (all individuals testing positive for HIV are immediately offered ART with streamlined care delivery) and equalling zero if the community was assigned to the control (all individuals testing positive for HIV are offered ART according to in-country guidelines). Finally, the outcome Y is the 5-year cumulative incidence of HIV, which will be measured through longitudinal follow-up. Thereby, the data structure for a SEARCH community is $O = (W, A, Y)$.

The adaptive design has important implications for estimation and inference [37]. Mainly, the partitioning of the sample into $n/2$ pairs is a function of the baseline covariates of all N candidates. Adaptive pair-matching results in n dependent copies of O. Nonetheless, given the covariates of all candidate communities $W^N = (W_1, \ldots, W_N)$, the observed data can be represented as $n/2$ conditionally independent random variables:

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

where the index $j = 1, \ldots, n/2$ denotes the partitioning of the candidates $\{1, \ldots, N\}$ into matched pairs according to similarity on their baseline covariates W^N. Throughout the subscripts $j1$ and $j2$ denote the first and second communities within matched pair j. We place no assumptions on the joint distribution of covariates $P_0(W^N)$, where subscript 0 denotes the true but unknown distribution. The treatment assignment mechanism is known; with probability 0.5, the first unit is randomized to the intervention and the second to the control:

$$P_0\left(A_{j1}=1, A_{j2}=0 | W^N\right) = P_0\left(A_{j1}=0, A_{j2}=1 | W^N\right) = 0.5$$

Study communities are assumed to be causally independent (i.e. no contamination or spillover effects). In other words, we assume that the baseline covariates and intervention assignment of one community do not affect the outcome of another study community. We further assume that there is no spatial correlation between study communities. For further discussion of the impact of spatial dependence in cluster randomized trials, we refer the reader to [56–60]. Recent work, relaxing these assumptions and considering a network of interacting units, is elaborated in van der Laan [61]. Under these assumptions, the

conditional distribution of the observed data, given the baseline covariates of the candidate units, factorizes as

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

Throughout, P_0^n denotes the true conditional distribution of the observed data, given the baseline covariates of the n study units $W^n = (W_1, \ldots, W_n)$. There are no other restrictions on the set of possible observed data distributions, and the resulting statistical model $\mathscr{M}$ is semiparametric.

2.1. The Conditional Average Treatment Effect (CATE)

The goal of the SEARCH trial is to estimate the effect of a strategy for immediate and streamlined ART for all HIV diagnosed persons on the 5-year cumulative HIV incidence in rural East African communities. A common target of inference is the population average treatment effect $E[Y(1)] - E[Y(0)]$ or its relative counterpart $E[Y(1)]/E[Y(0)]$, where $Y(a)$ denotes the counterfactual cumulative incidence under treatment level $A = a$. This causal parameter is the difference in the expected outcomes if all communities (in some hypothetical target population) were to receive the intervention and if all communities (in some hypothetical target population) were to receive the control.

An alternative estimand involves conditioning on the baseline covariates of the study communities [16, 38–40]:

$$\psi^F = \frac{1}{n}\sum_{i=1}^{n} E\left[Y_i(1) - Y_i(0)\,|W^n\right]$$

where $Y_i(a)$ denotes the counterfactual cumulative incidence under treatment level $A = a$ for unit i. This parameter is the difference in the expected counterfactual outcomes, treating the baseline covariates of the study communities as fixed. As a result, the parameter is data-adaptive; its value changes with the sample of study units. Nonetheless, ψ^F can be interpreted as the intervention effect, given the covariates the sample units. Greater generalizability is up to the reader and not implicitly assumed in the parameter specification. Furthermore, by obviating estimation of the covariate distribution, estimators of the conditional parameter will also often be more precise than those of the population parameter [16, 38–40].

2.2. Estimation

Since the intervention is randomized within matched pairs, the causal parameter is readily identifiable from the conditional distribution of the observed data. The statistical estimand is

$$\begin{aligned}\Psi\left(P_0^n\right) &= \frac{1}{n}\sum_{i=1}^{n}\left[E_0\left(Y_i|A_i{=}1,W_i\right)-E_0\left(Y_i|A_i{=}0,W_i\right)\right]\\ &= \frac{1}{n}\sum_{i=1}^{n}\left[\bar{Q}_0\left(1,W_i\right)-\bar{Q}_0\left(0,W_i\right)\right]\end{aligned}$$

where $\bar{Q}_0(A, W)$ denotes the conditional mean outcome, given the intervention A and covariates W. In other words, the target parameter is the average difference in the strata-specific expected HIV incidence under the intervention and control for the n study communities. This estimand is still random through the vector of covariates $W^n = (W_1, \ldots, W_n)$. The true value ψ_0 depends on the sample of n units.

An intuitive estimator of ψ_0 is the average difference in outcomes within matched pairs:

$$\hat{\psi}_{unadj}=\frac{1}{n/2}\sum_{j=1}^{n/2}\left(Y_{j1}-Y_{j2}\right)$$

where the observations within matched pair j have been ordered such that the first corresponds to the intervention, $A_{j1} = 1$, and the second the control, $A_{j2} = 0$. This estimator is equivalent to taking the difference in the average outcomes among intervention units $\bar{Q}_n(1) = E_n(Y|A = 1)$ and the average outcomes among control units $\bar{Q}_n(0) = E_n(Y|A = 0)$. Since the intervention is randomized, the unadjusted estimator is unbiased for the parameter of interest, given the vector of covariates W^n. (See Appendix A in the Supplementary Material for the accompanying proof.) When the measured covariates are predictive of the outcome, this simple difference-in-means estimator tends to be *inefficient* as it fails to adjust for measured covariates. Despite recent advances in matching algorithms [14, 35, 36], there is likely to be some residual imbalance on pre-intervention determinants of the outcome within matched pairs. Furthermore, even if we succeeded in matching well on all available characteristics, there might be additional baseline covariates that are predictive of the outcome, but were unavailable during the matching process. In the SEARCH trial, for example, baseline population HIV RNA levels are thought to be a major driver of incidence but were unavailable during matching.

An alternative approach is to use TMLE, which can provide an unbiased and efficient estimate of the intervention effect. The TMLE for $\Psi\left(P_0^n\right)$ is given by the following substitution estimator:

$$\hat{\psi}_{adj}=\frac{1}{n}\sum_{i=1}^{n}\left[\bar{Q}_n^*\left(1,W_i\right)-\bar{Q}_n^*\left(0,W_i\right)\right]$$

where $\bar{Q}_n^*\left(A,W\right)$ denotes a targeted estimate of the conditional mean function $E_0(Y|A, W)$. In general, this targeting step is used to achieve the optimal bias-variance trade-off for the parameter of interest and to solve the efficient score equation [62]. We refer the reader to

van der Laan and Rose [63] for a detailed discussion and worked examples of TMLE. In an adaptive pair-matched trial, TMLE for $\Psi\left(P_0^n\right)$ can be implemented as follows.

1. Estimate the conditional mean function $\bar{Q}_0(A, W)$ by regressing the outcome Y on the treatment A and covariates W, while ignoring the dependence in the data.
   - For a binary outcome or a bounded continuous outcome, the negative log likelihood is a valid loss function and provides stability in the context of sparsity [64]. Specifically, the boundedness property of the logistic function guarantees the predicted outcomes are within the appropriate range (e.g. [0,1] for a proportion). For a continuous outcome, initial estimation of the conditional mean $\bar{Q}_0(A, W)$ can also be based on linear regression, which can yield more power than non-linear (logistic) regression in randomized trials. In particular, Rubin and van der Laan [41] detail the use of least squares regression to optimize the fit of $\bar{Q}_0(A, W)$ to achieve the lowest possible variance. Initial estimation can also be based on an *a priori* specified data-adaptive method, such as Super Learner [65]. In all cases, there is no risk of bias due to model misspecification [41–43].
2. If the initial regression model included an intercept and a main term for the exposure, the estimator of the conditional mean outcome $\bar{Q}_n(A, W)$ is already targeted. Skip to step 3. Otherwise, update the initial estimator as follows.
   - If logistic regression was used for initial estimation, then the following fluctuation sub-model is appropriate:

$$logit\left[\bar{Q}_n\left(A,W\right)\left(\varepsilon\right)\right]=logit\left[\bar{Q}_n\left(A,W\right)\right]+\varepsilon H\left(A\right),\quad \text{where}\quad H\left(A\right)=\left(\frac{\mathbb{I}\left(A=1\right)}{P_0\left(A=1\right)}-\frac{\mathbb{I}\left(A=0\right)}{P_0\left(A=0\right)}\right)$$

     and ε is the univariate parameter. If linear regression was used, then the following fluctuation sub-model is appropriate:

$$\bar{Q}_n\left(A,W\right)\left(\varepsilon\right)=\bar{Q}_n\left(A,W\right)+\varepsilon H\left(A\right)$$

     with ε and $H(A)$ are defined as above. In practice, run logistic (linear) regression of the outcome Y on the covariate $H(A)$, using the initial estimate as offset. Then plug the estimated coefficient ε_n into the fluctuation model to yield the targeted estimates $\bar{Q}_n^*\left(A,W\right)=\bar{Q}_n\left(A,W\right)\left(\varepsilon_n\right)$.
3. Take the sample average of the differences in the expected outcomes:

$$\hat{\psi}_{adj}=\frac{1}{n}\sum_{i=1}^{n}\left[\bar{Q}_n^*\left(1,W_i\right)-\bar{Q}_n^*\left(0,W_i\right)\right]$$

where $\bar{Q}_n^*(1, W_i)$ denotes the expected outcome for unit *i* under the intervention and $\bar{Q}_n^*(0, W_i)$ denotes the expected outcome for unit *i* under the control. It is worth emphasizing that empirical mean, here, is part of target parameter mapping; we are not estimating the covariate distribution as would be required for the population average treatment effect.

In practice, many cluster randomized trials have a limited number of (conditionally) independent units. For example, there are only 16 conditionally independent pairs in the SEARCH trial. As a result, the number of parameters in the regression model for $\bar{Q}_0(A, W)$ can quickly approach the number of observations. Therefore, the curse of dimensionality can prevent adjustment for all the measured covariates *W* or the inclusion of multiple interaction terms. Nonetheless, it is often possible to adjust for a single or few covariates and obtain efficiency gains without risk [42, 43]. Furthermore, when the regression model for $\bar{Q}_0(A, W)$ includes an intercept and the exposure *A* as a main term, the initial estimator is already targeted. Thus, we can obtain an unbiased and more efficient estimator in two steps: estimate $\bar{Q}_0(A, W)$ with main terms linear or logistic regression, and take the sample average of the differences in the expected outcomes under the treatment and control.

2.3. Statistical Inference

As established in Appendix B of the Supplementary Material, both the unadjusted estimator and the TMLE are asymptotically linear and normally distributed. Briefly, an estimator is asymptotically linear if the difference between the estimator and the estimand behaves (in first order) as an empirical mean of a function, known as the influence curve, of the unit data [63]. Then the limit distribution of the standardized estimator is normal with mean 0 and variance given by the variance of its influence curve. With an estimate of the influence curve and thereby an estimate of the variance, the standard normal distribution can be used for confidence interval construction and hypothesis testing in large studies. For trials with limited numbers of (conditionally) independent units, the Student's *t*-distribution with *n*/2-1 degrees of freedom is an appropriate alternative to the standard normal distribution. Randomization inference, in contrast, may not be appropriate, as it is testing a different null hypothesis of a constant treatment effect (e.g. $Y_i(0) = Y_i(1)\ \forall i$) [66, 67]. The causal and statistical estimands, considered here, are in terms of a sample average effect over the study units.

The influence curve for the TMLE of $\Psi(P_0^n)$ in a trial with adaptive pair-matching is the following function of the paired data (proof in Appendix B of the Supplementary Material):

$$\begin{aligned} IC\left(\bar{O}_j\right) &= \bar{D}^*\left(\bar{O}_j\right) - E_0\left[\bar{D}^*\left(\bar{O}_j\right)|W^n\right] \\ \bar{D}^*\left(\bar{O}_j\right) &= \tfrac{1}{2}\left\{D^*(O_{j1}) + D^*(O_{j2})\right\} \\ D^*(O_i) &= \left(\tfrac{(A_i=1)}{P_0(A_i)} - \tfrac{(A_i=0)}{P_0(A_i)}\right)\left(Y_i - \bar{Q}(A_i, W_i)\right) \end{aligned}$$

where $Q(A, W)$ denotes the limit of the targeted estimator of the conditional mean function $Q_0(A, W)$ and where the marginal probability of being assigned the treatment or the control is known: $P_0(A) = 0.5$. Through the conditional expectation of $D^*(\bar{O}_j)$, given the vector of covariates W^n, the influence curve relies on the true but unknown conditional mean outcome $Q_0(A, W)$:

$$E_0\left[\bar{D}^*\left(\bar{O}_j\right)|W^n\right] = \frac{1}{2}\left\{\left(\bar{Q}_0(1, W_{j1}) - \bar{Q}(1, W_{j1})\right) - \left(\bar{Q}_0(0, W_{j1}) - \bar{Q}(0, W_{j1})\right) + \left(\bar{Q}_0(1, W_{j2}) - \bar{Q}(1, W_{j2})\right) - \left(\bar{Q}_0(0, W_{j2}) - \bar{Q}(0, W_{j2})\right)\right\}$$

This term captures deviations between the true and estimated mean outcomes for observations within a matched pair. The influence curve for the unadjusted estimator $\hat{\psi}_{unadj}$ is analogous, but with $Q(A, W)$ replaced with the limit of the treatment-specific mean $Q_{n(A)} = E_n(Y|A)$. For either estimator, there is no contribution from the covariate distribution, which is considered fixed.

The asymptotic variance of the unadjusted estimator or the TMLE is then given by the variance of its influence curve, divided by $n/2$. Improved estimation of the conditional mean outcome $Q_0(A, W)$ leads to more precise estimators of intervention effect $\Psi(P_0^n)$. Specifically, if this conditional mean is consistently estimated (i.e. if $Q(A, W) = Q_0(A, W)$), then the term, involving deviations between the true and estimated means, is zero, and the estimator of $\Psi(P_0^n)$ is asymptotically efficient. In other words, the estimator's influence curve equals the efficient influence curve, and the estimator has lowest possible variance among a large class of estimators [63]. Otherwise, the estimator is still be unbiased, but does not achieve the efficiency bound. When the baseline covariates W impact the outcome, the targeted estimator of the conditional mean outcome $\bar{Q}_n^*(A, W)$ is expected to be closer to the true mean $Q_0(A, W)$ than the unadjusted estimator $Q_n(A)$. As a result, the asymptotic variance of the TMLE $\hat{\psi}_{adj}$ is often smaller than that of the unadjusted estimator $\hat{\psi}_{unadj}$. Thus, for both individual and cluster randomized trials, TMLE is often a more efficient estimator of the CATE than the unadjusted.

Consistent estimation of the influence curve and thereby the asymptotic variance rely on consistent estimation of this conditional mean $Q_0(A, W)$, which might be particularly challenging when n is small, as common in cluster randomized trials. Nonetheless, we can conservatively approximate the influence curve of the unadjusted estimator $\hat{\psi}_{unadj}$ or the TMLE $\hat{\psi}_{adj}$ by the difference in residuals within matched pairs (proof in Appendix B.1 of the Supplementary Material):

$$\hat{IC}_{unadj}\left(\bar{O}_j\right) = \left(Y_{j1} - \bar{Q}_n(1)\right) - \left(Y_{j2} - \bar{Q}_n(0)\right)$$
$$\hat{IC}_{adj}\left(\bar{O}_j\right) = \left(Y_{j1} - \bar{Q}_n^*(1, W_{j1})\right) - \left(Y_{j2} - \bar{Q}_n^*(0, W_{j2})\right)$$

respectively. Again, $\bar{Q}_n(A)$ denotes an unadjusted estimate of the treatment-specific mean, $\bar{Q}_n^*(A, W)$ denotes a targeted estimate of the conditional mean outcome, and observations in matched pair *j* have been ordered such that the first corresponds to intervention ($A_{j1} = 1$) and the second to the control ($A_{j2} = 0$). An asymptotically conservative variance estimator is then given by the sample variance of the estimated influence curve, divided by $n/2$. For $\hat{\psi}_{unadj}$, this is equivalent to the sample variance of the within pair differences, divided by $n/2$, and is commonly recommended for pair-matched randomized trials [17] even though it is known to be conservative if the conditional parameter is the target of inference [16, 38–40]. To obtain a less conservative variance estimator for $\hat{\psi}_{unadj}$, Abadie and Imbens [40] proposed a matching estimator, involving the variance of pairs-of-pairs with similar covariates. Our approach to reduce the true variance of the estimator and obtain a less conservative variance estimate is through adjustment with TMLE. In most practical settings, the sum of squared adjusted residuals is smaller than the sum of squared unadjusted residuals. Thereby, the estimated variance of the TMLE is often smaller than the estimated variance of the unadjusted algorithm. In summary, this implies that covariate adjustment with TMLE results in a more precise estimator (i.e. smaller true variance) and a less conservative variance estimator.

We also briefly note that a randomized trial with adaptive pair-matching will often be more efficient for estimation of the CATE than a randomized trial without matching. The designs will only have the same efficiency bound if the conditional mean outcome is consistently estimated (i.e. $\bar{Q}(A, W) = \bar{Q}_0(A, W)$). In practice, we expect there to be some deviations between the true and estimated means. If these deviations are positively correlated within matched pairs, the asymptotic variance of the TMLE will be smaller in the adaptive trial than in the completely randomized trial. In finite samples, we also expect there to be an efficiency gain from adaptive pair-matching. Mainly, if we succeed in matching pairs on predictive covariates, then the sample covariance of the residuals within matched pairs will be positive and the adaptive design will yield more power. We refer the reader to Appendix C in the Supplementary Material for further details and associated proofs.

3. Simulations

We present the following set of simulations to demonstrate (1) implementation of the above estimators, (2) the potential gain in efficiency with adaptive pair-matching, and (3) the further gain with adjustment during the analysis due to having a more precise estimator and a less conservative variance estimator. These simulations are not intended to represent the full complexities of a cluster randomized trial. (To be clear, these simulations were not the ones used when developing the design and analysis of the SEARCH trial.) Nonetheless, they explore some of the challenges faced, such as rare outcomes, the inability to match on all baseline covariates, and limited numbers of conditionally independent units. All simulations were done in R v3.0.1 [68].

3.1. Data Generating Process & Estimators

For $n = 32$ units, three baseline covariates $W = (W1, W2, W3)$ were independently drawn from a normal distribution with mean 0 and standard deviation 1. A fourth covariate Z was generated as a function of these baseline covariates and random noise U_Z:

$$Z = expit\left[-0.25 + 0.5^{*}W1 + W2 + 2^{*}W3 + 0.5^{*}U_{Z}\right]/4$$

where the *expit* function is the inverse of the logit function and U_Z was drawn independently from a normal with mean 0 and standard deviation 1. To imitate adaptive pair-matching, the nonbipartite matching algorithm (nbpMatching v1.3.6 [33]) was applied to the set of n covariates $W^n = (W_1, \ldots, W_n)$ with $W_i = (W1_i, W2_i, W3_i)$. Within the resulting 16 matched pairs, the exposure A was randomized. As before, A is binary indicator, equaling 1 if the unit was randomized to the intervention and 0 otherwise. Finally, the outcome Y was generated as

$$Y = expit\left[\beta_0 + 0.5^{*}W1 + 0.5^{*}W2 + 0.5^{*}W3 + 7^{*}Z - A + 0.25^{*}A^{*}Z\right]/15 + U_{Y}$$

where random noise U_Y was drawn independently from a uniform distribution with minimum 0 and maximum 0.025. Dividing by 15 was done to scale the outcome Y, representing a proportion, to be within plausible ranges for the cumulative incidence of HIV. The term β_0 was set to either -2 or 0.5 to examine the performance of the estimators when the outcome was rare ("Simulation A") or more common ("Simulation B"). To simulate the null scenario, the treatment was randomly assigned within pairs but the outcomes generated as if all communities received the control ($A = 0$). For comparison, we also simulated equivalent data for a non-matched randomized trial with balanced allocation of the treatment.

Over 5000 data sets, we examined the performance of the unadjusted estimator and TMLE. For the latter, we compared linear to logistic main terms regression with various adjustment sets. Linear regression can result in more efficient estimation, by minimizing the empirical variance of the influence curve [41]. With rare outcomes, however, logistic regression can provide stability, by guaranteeing the predicted outcomes respect the model bounds (i.e. are in [0,1]). Therefore, we expected the TMLE with logistic regression to result in better performance when the outcome was rare (Simulation A) and the TMLE with linear regression to result in better performance when the outcome was more common (Simulation B). In terms of adjustment sets, we compared regression models with main terms for the exposure A and the covariate Z as well as regression models with main terms for the exposure A, the matching covariates W and the remaining covariate Z. Recall Z was an important determinant of the outcome but not used in matching. We expected that the fully adjusted estimator (TMLE with main terms for (A, W, Z)) would suffer from over-fitting. Since main terms regression models were used, the fluctuation step of the TMLE algorithm did not provide an update. In all cases, there was no risk of bias due to regression model misspecification [42, 43]. Inference was based on the sample variance of the estimated influence curve and the Student's t-distribution with 15 degrees of freedom. The

corresponding TMLE implementation and proof of statistical inference for the non-matched randomized trial are given in Appendix C of the Supplementary Material.

3.2. Results

Recall the true value of the statistical estimand depends on the $n = 32$ communities in the sample. Table 1 shows the minimum, mean and maximum value of the intervention effect ψ_0 over the 5,000 simulated data sets. For comparison, the table also gives the corresponding summaries of the exposure-specific effects:

$\psi_0(a) = \frac{1}{n}\sum_{i=1}^{n} E_0(Y|A_i=a, W_i, Z_i)$. This estimand is the sample average of the conditional mean outcome, setting the exposure $A = a$ and given the covariates (W, Z). For Simulation A, representing a rare outcome, the average values of the effect under the exposure $\psi_0(1)$ and the control $\psi_0(0)$ were 0.024 and 0.032, respectively. The corresponding mean value of the intervention effect ψ_0 was -0.009, translating to 26.41% reduction in the incidence of the outcome (on average). For Simulation B, representing a more common outcome, the average values of the conditional effect under the exposure $\psi_0(1)$ and the control $\psi_0(0)$ were 0.05 and 0.061, respectively. The corresponding average value of the target parameter ψ_0 was -0.011, translating to a 17.90% reduction in the incidence of the outcome (on average).

For Simulation A, Table 2 illustrates the performance of the estimators over 5,000 simulated data sets. All estimators were unbiased. As expected, there was an efficiency gain with matching. The standard deviation (square root of the variance of the point estimates) of the unadjusted estimator was 1.58 times higher without matching than with matching. Likewise, the attained power (proportion of simulated trials where the null hypothesis was correctly rejected) jumped from 34% to 64% with matching. As expected, adaptive pair-matching on the three covariates W reduced variability in the outcomes within matched pairs. The coefficient of variation, measuring of the variability in outcomes between units in the absence of the intervention, was $k = 0.53$, while the matched-pair coefficient of variation, measuring of the variability in outcomes within matched pairs in the absence of the intervention, was $k_m = 0.29$ [17].

There was also an efficiency gain from adjustment. For the non-matched design, the standard deviation of the unadjusted estimator was 1.58 times higher than the standard deviation of the TMLE, using linear regression to adjust for Z. The corresponding power increased from 34% to 72%. For the adaptive design, the standard deviation of the unadjusted estimator was 1.13 times higher than the standard deviation of the TMLE, using linear regression to adjust for Z. The corresponding attained power increased from 64% to 74%. For both designs, there was a further precision gain by using logistic regression to adjust for Z. Under sparsity, logistic regression can be more stable than linear regression and is guaranteed to yield parameter estimates within the appropriate range (i.e. [0,1] for proportions) [64]. While there was some power gain from adjusting for all four covariates (W, Z), there was also a risk in over-fitting the regression model and under-estimating the variance. Recall the variance estimators in Section 2.3 are asymptotically conservative, and the simulations represent finite samples. Indeed, with a main terms regression model for the conditional mean outcome, there were 5 parameters with only 16 conditionally independent

units. As a result, the confidence interval coverage (proportion of studies containing the true parameter value) was less than the nominal rate of 95% for the fully adjusted estimator. Likewise, the type I error rate (proportion of studies falsely rejecting the null hypothesis) was greater than $\alpha = 0.05$ for the fully adjusted estimator, as shown in Table 1 of Appendix D in the Supplementary Material. Conversely, for both the unadjusted estimator and the TMLE only adjusting for *Z*, there was good confidence interval coverage and control of type I error rates. Indeed, there was some evidence of over-coverage of confidence intervals and conservative Type I error rates for the unadjusted estimator in both designs, as predicted by theory.

The results for Simulation B, representing a more common outcome, are also given in Table 2 and largely echoed the above findings. Because the exposure was randomized, all estimators were unbiased. As before, there was a substantial efficiency gain with matching. Adaptive matching on the three covariates $W = (W1, W2, W3)$ reduced variability in the outcomes within pairs. The coefficient of variation was $k = 0.27$, while the matched-pair coefficient of variation was $k_m = 0.14$. Again, there was also a substantial precision gain from adjustment. With a more common outcome, however, there was a greater gain in power from adjusting for *Z* with linear regression than logistic regression for both designs. Here, minimizing the sum of squared residuals helped to minimize the empirical variance of the influence curve and thereby maximize the empirical efficiency [41]. With the fully adjusted estimator, again there was some risk of over-fitting and inference was optimistic. In contrast, for both the unadjusted estimator and the TMLE adjusting only for *Z*, there was good confidence interval coverage as well as Type I error control (Table 1 of Appendix D in the Supplementary Material). In summary, our finite sample simulations support our theoretical results: adaptive pair-matching yields more power than complete randomization, and further efficiency gains can be attained through adjustment during the analysis.

4. Discussion

To our knowledge, this is the first paper to study and articulate the consequences of adaptive pair-matching for estimation of the average treatment effect, given the baseline covariates of the *n* study units. This work was motivated by SEARCH trial, which aims to estimate the effect of immediate ART, delivered in a streamlined fashion, on the five-year cumulative incidence of HIV. The decision to pair-match communities in the trial was motivated by a desire to protect study credibility and by the potential to increase study power. Through careful definition of the data generating experiment, we recognized that the design would not yield *n*/2 i.i.d. paired units, as current practice assumes. Instead, by constructing the matched pairs as a function of the baseline covariates of all candidate communities, the adaptive design results in *n* dependent units and *n*/2 conditionally independent units, given the baseline covariates of the study communities.

To the best of our understanding, adaptive pair-matching is a common design and has been implemented in other cluster randomized trials (e.g. [29–31]). In practice, adaptive pair-matching (a.k.a. “nonbipartite matching”) can be carried out with standard software. For example, the nbpMatching package [33] in R and the corresponding web application will generate the set of optimal matched pairs as function of a user-supplied matrix of covariates

[35, 36]. These tools allow the user to weight covariates differently (e.g. on importance or relevance to the outcome) and to specify the maximum number of matches - choices, which should be driven by subject matter knowledge as well as resource constraints.

We focused on estimation of the CATE. By obviating estimation of the covariate distribution, estimators of the conditional parameter will often be less variable than estimators of the population parameter [16, 38–40]. We contrasted the unadjusted estimator with TMLE adjusting for baseline covariates. We provided a step-by-step implementation of the latter estimator and detailed proofs of inference. Both estimators can be implemented ignoring the dependence in the data and with standard software, such as the tmle [69] and ltmle [70] packages in R. Asymptotically conservative inference can obtained with the sample variance of the pairwise differences in residuals, divided by $n/2$. When the baseline covariates are predictive of the outcome, the unadjusted estimator will be less efficient than the TMLE. Furthermore, the estimated variance of the TMLE will often be less conservative than that of the unadjusted estimator.

Finite sample simulations were used to evaluate estimator performance and verify our theoretical results. Since the intervention was randomized, all estimators were unbiased [41–43]. There was an efficiency gain with matching and a further gain with adjustment. When the outcome was quite rare, adjusting for a single baseline covariate with logistic regression yielded more power than adjustment with linear regression. When the outcome was more common, the converse was observed. While the variance estimators are asymptotically conservative, there was some risk of over-adjusting in small trials. Indeed, with only 16 (conditionally) independent units, adjusting for all 4 baseline covariates resulted in under-coverage of the confidence intervals and higher than nominal Type I error rates.

Previously, Imai *et al.* [15] suggested, “randomization by cluster without prior construction of matched pairs, when pairing is feasible, is an exercise in self-destruction.” Our work also suggests that asymptotically and in finite samples, a randomized trial with adaptive pair-matching will often be more efficient for estimation of the CATE than its completely randomized counterpart. The trials will only have the same efficiency bound when the conditional mean outcome, given the exposure and covariates, is consistently estimated. In practice, we expect there to be some deviations between the true and estimated means. When these deviations are positively correlated within matched pairs, the design with adaptive pair-matching will be more efficient (Appendix C.3). In finite samples, pair-matching will also often result in a positive covariance of the residuals (deviations between the observed and predicted outcomes) within matched pairs and thereby smaller finite sample variance.

Overall, adaptive pair-matching is an intuitive strategy to group candidate units on similarity in their baseline covariates. Pair-matching will protect study credibility. Combining subject matter knowledge with modern matching algorithms (e.g. nbpMatching [33]) is likely to result in studies, where pair-matching substantially improves study power. We recommend specifying the intervention effect in terms of the conditional parameter, which considers the covariate distribution as fixed and obviates its estimation, resulting in less variable estimators. We also recommend adjusting for baseline variables as the data allow.

Simulations, such as those presented here, can help inform the practitioner as to the optimal adjustment set. Future work will involve the use of cross-validation to data-adaptively select for the adjustment set. We also plan to formally study the asymptotic and finite sample properties of analysis approaches based on covariate-adjusted residuals for estimation and inference of the CATE [17, 29, 46]. We will also investigate the impact of adaptive stratification on estimation and inference for both the population and conditional average treatment effect. While our work was motivated by a cluster randomized trial with the outcome of cumulative incidence, the results are generally applicable to other trials with binary or continuous outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors would like to acknowledge and thank the entire SEARCH consortium for their helpful comments and discussion. The authors would also like to thank the anonymous reviewers for their insightful comments and questions. This work was supported, in part, by the National Institute Of Allergy And Infectious Diseases of the National Institutes of Health under award numbers R01AI074345 and U01AI99959. The content is solely the responsibility of the authors and does not necessarily represent the officials views of the National Institutes of Health. Maya Petersen is a recipient of a Doris Duke Clinical Scientist Development Award.

References

1. Costanza M. Matching. Preventive Medicine. 1995; 24(5):425–433. doi:10.1006/pmed.1995.1069. [PubMed: 8524715]
2. Rothman, K.; Greenland, S.; Lash, T. Modern Epidemiology. Lippincott Williams & Wilkins; Phildelphia: 2008.
3. 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]
4. Cochran W. Matching in Analytical Studies. American Journal of Public Health and the Nations Health. 1953; 43(6):684–691. doi:10.2105/AJPH.43.6_Pt_1.684.
5. Billewicz W. Matched samples in medical investigations. British Journal of Preventive & Social Medicine. 1964; 18:167–173. doi:10.1136/jech.18.4.167. [PubMed: 14216076]
6. Youkeles L. 184 Note: Loss of Power through Ineffective Pairing of Observations in Small Two-Treatment All-or-None Experiments. Biometrics. 1963; 19(1):175–180. doi:10.2307/2527582.
7. Mathen K. Matching in comparative studies in public health. Indian Journal of Public Health. 1963; 8(4):161–169. [PubMed: 14073639]
8. Billewicz W. The Efficiency of Matched Samples: An Empirical Investigation. Biometrics. 1965; 21(3):623–644. doi:10.2307/2528546. [PubMed: 5858095]
9. Miettinen O. The matched pairs design in the case of all-or-none responses. Biometrics. 1968; 24(2):339–352. doi: 10.2307/2528039. [PubMed: 5683874]
10. McKinlay S. Pair-Matching–A Reappraisal of a Popular Technique. Biometrics. 1977; 33(4):725–735. doi:10.2307/2529471. [PubMed: 588658]
11. Wacholder S, Weinberg C. Paired versus Two-Sample Design for a Clinical Trial of Treatments with Dichotomous Outcome: Power considerations. Biometrics. 1982; 38(3):801–812. doi: 10.2307/2530059. [PubMed: 7171700]
12. Martin D, Diehr P, Perrin E, Koepsell T. The effect of matching on the power of randomized community intervention studies. Statistics in Medicine. 1993; 12(3-4):329–338. doi:10.1002/sim.4780120315. [PubMed: 8456215]

13. Diehr P, Martin D, Koepsell T, Cheadle A. Breaking the matches in a paired t-test for community interventions when the number of pairs is small. Statistics in Medicine. 1995; 14(13):1491–1504. doi:10.1002/sim.4780141309. [PubMed: 7481187]
14. 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]
15. 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.
16. Imbens, G. Experimental design for unit and cluster randomized trials. 2011. Technical Report, NBER technical working paper
17. Hayes, R.; Moulton, L. Cluster Randomised Trials. Chapman & Hall/CRC; Boca Raton: 2009.
18. 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]
19. Murray, D. Design and Analysis of Group-Randomized Trials. Oxford University Press; New York: 1998.
20. Donner, A.; Klar, N. Design and Analysis of Cluster Randomization Trials in Health Research. Arnold; London: 2000.
21. 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]
22. Balzer, L.; Petersen, M.; van der Laan, M. Why match in individually and cluster randomized trials?. University of California; Berkeley: 2012. Technical Report 294, Division of Biostatistics. Available at http://www.bepress.com/ucbbiostat/paper294
23. Granich R, Gupta S, B Suthar A, Smyth C, Hoos D, Vitoria M, Simao M, Hankins C, Schwartlander B, Ridzon R, et al. Antiretroviral therapy in prevention of HIV and TB: update on current research efforts. Current HIV research. 2011; 9(6):446–469. doi: 10.2174/157016211798038597. [PubMed: 21999779]
24. Sustainable East Africa Research in Community Health (SEARCH). University of California; San Francisco: 2013. ClinicalTrials.govURL http://clinicaltrials.gov/show/NCT01864603
25. Wei L, Smythe R, Lin D, Park T. Statistical inference with data-dependent treatment allocation rules. Journal of the American Statistical Association. 1990; 85(156-162) doi: 10.1080/01621459.1990.10475319.
26. Tamura R, Faries D, Andersen J, Heiligenstein J. A case study of an adaptive clinical trial in the treatment of out-patients with depressive disorder. Journal of the American Statistical Association. 1994; 89(427):768–776. doi: 10.2307/2290902.
27. Collins L, Murphy S, Bierman K. A conceptual framework for adaptive preventive interventions. Prevention Science. 2004; 5(3):185–196. doi:10.1023/B:PREV.0000037641.26017.00. [PubMed: 15470938]
28. Chambaz A, van der Laan M. Targeting the optimal design in randomized clinical trials with binary outcomes and no covariate: Theoretical study. The International Journal of Biostatistics. 2011; 7(1) Article 10, doi:10.2202/1557-4679.1247.
29. Grosskurth H, Mosha F, Todd J, Mwijarubi E, Klokke A, Senkoro K, Mayaud P, Changalucha J, Nicoll A, ka-Gina G, et al. Impact of improved treatment of sexually transmitted diseases on HIV infection in rural Tanzania: randomised controlled trial. The Lancet. 1995; 346(8974):530–536. doi:10.1016/S0140-6736(95)91380-7.
30. Watson L, Small R, Brown S, Dawson W, Lumley J. Mounting a community-randomized trial: sample size, matching, selection, and randomization issues in PRISM. Controlled Clinical Trials. 2004; 25(3):235–250. doi: 10.1016/j.cct.2003.12.002. [PubMed: 15157727]
31. Toftager M, Christiansen L, Kristensen P, Troelsen J. Space for physical activity-a multicomponent intervention study: study design and baseline findings from a cluster randomized controlled trial. BMC Public Health. 2011; 11:777. doi:10.1186/1471-2458-11-777. [PubMed: 21985278]

32. King G, Gakidou E, Ravishankar N, Moore R, Lakin J, Vargas M, Tellez-Rojo M, Hernandez Avila J, Hernandez Avila M, Hernandez Llamas H. A “politically robust” experimental design for public policy evaluation, with application to the Mexican Universal Health Insurance Program. Journal of Policy Analysis and Management. 2007; 26(3):479–506. doi:10.1002/pam.20279. [PubMed: 17633445]
33. Lu, B.; Greevy, R.; Beck, C. nbpMatching: functions for non-bipartite optimal matching. R package version 1.3.6. 2012. URL http://CRAN.R-project.org/package=nbpMatching
34. 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.
35. 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]
36. Greevy RA, Grijalva CG, Roumie CL, Beck C, Hung AM, Murff HJ, Liu X, Griffin MR. Reweighted Mahalanobis distance matching for cluster-randomized trials with missing data. Pharmacoepidemiology and Drug Safety. 2012; 21:148–154. doi:10.1002/pds.3260. [PubMed: 22552990]
37. 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]
38. Abadie, A.; Imbens, G. Simple and bias-corrected matching estimators for average treatment effects. 2002. Technical Report 283, NBER technical working paper
39. 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.
40. Abadie A, Imbens G. Estimation of the conditional variance in paired experiments. Annales d'Economie et de Statistique. 2008; (91/92):175–187.
41. 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.
42. 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]
43. 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.
44. Imai K. Variance identification and efficiency analysis in randomized experiments under the matched-pair design. Statistics in Medicine. 2008; 27(24):4857–4873. doi:10.1002/sim.3337. [PubMed: 18618425]
45. Hill J, Scott M. Comment: The essential role of pair matching. Statistical Science. 2009; 24(1):54–58. doi: 10.1214/09-STS274A.
46. The COMMIT Research Group. Community intervention trial for smoking cessation (COMMIT): I. cohort results from a four-year community intervention. American Journal of Public Health. 1995; 85:183–192. doi:10.2105/ajph.85.2.183. [PubMed: 7856777]
47. Thirumurthy H, Goldstein M, Graff Zivin J. The economic impact of AIDS treatment: labor supply in Western Kenya. Journal of Human Resources. 2008; 43:511–552. doi:10.1353/jhr.2008.0009. [PubMed: 22180664]
48. Zivin GJ, Thirumurthy H, Goldstein M. AIDS Treatment and Intrahousehold Resource Allocation: Children's Nutrition and Schooling in Kenya. Journal of Public Economics. 2009; 93(7-8):1008–1015. doi:10.1016/j.jpubeco.2009.03.003. [PubMed: 22180689]
49. Cohen M, Chen YQ, McCauley M, Gamble T, Hosseinipour M, Kumarasamy N, et al. Prevention of HIV-1 infection with early antiretroviral therapy. New England Journal of Medicine. 2011; 365(6):493–505. doi:10.1056/NEJMoa1105243. [PubMed: 21767103]
50. Lawson S, Harries A. Reducing tuberculosis-associated early mortality in antiretroviral treatment programmes in sub-Saharan Africa. AIDS. 2011; 25(12):1554–1555. doi:10.1097/QAD.0b013e328348fb61. [PubMed: 21747238]

51. Middelkoop K, Bekker L, Myer L, Johnson L, Kloos M, Morrow C, Wood R. Antiretroviral therapy and TB notification rates in a high HIV prevalence South African community. Journal of Acquired Immune Deficiency Syndromes. 2011; 56(3):263–269. doi:10.1097/QAI.0b013e31820413b3. [PubMed: 21317585]
52. Flateau C, Le Loup G, Pialoux G. Consequences of HIV infection on malaria and therapeutic implications: a systematic review. The Lancet Infectious Diseases. 2011; 11(7):541–556. doi:10.1016/S1473-3099(11)70031-7. [PubMed: 21700241]
53. Fowler M, Gable A, Lampe M, Etima M, Owor M. Perinatal HIV and its prevention: progress toward an HIV-free generation. Clinics in Perinatology. 2012; 37(4):699–719. vii. doi:10.1016/j.clp.2010.09.002. [PubMed: 21078445]
54. Schouten E, Jahn A, Midiani D, Makombe S, Mnthambala A, Chirwa Z, Harries A, van Oosterhout J, Meguid T, Ben-Smith A, et al. Prevention of mother-to-child transmission of HIV and the health related Millennium Development Goals: time for a public health approach. The Lancet. 2011; 378(9787):282–284. doi:10.1016/S0140-6736(10)62303-3.
55. Jain V, Liegler T, Kabami J, Chamie G, Clark T, Black D, Geng E, Kwarisiima D, Wong J, Abdel-Mohsen M, et al. Assessment of population-based HIV RNA levels in a rural East African setting using a fingerprick-based blood collection method. Clinical Infectious Diseases. 2013; 56(4):598–605. doi:10.1093/cid/cis881. [PubMed: 23243180]
56. Barrios T, Diamond R, Imbens GW, Kolesar M. Clustering, spatial correlations, and randomization inference. Journal of the American Statistical Association. 2012; 107(498):578–591. doi:10.3386/w15760.
57. Bandyopadhyay D, Reich BJ, Slate EH. A spatial beta-binomial model for clustered count data on dental caries. Statistical methods in medical research. 2011; 20(2):85–102. doi:10.1177/0962280210372453. [PubMed: 20511359]
58. Cressie, N.; Cassie, N. Statistics for Spatial Data. John Wiley and Sons; 1993.
59. Reich BJ, Bandyopadhyay D, Bondell HD. A nonparametric spatial model for periodontal data with nonrandom missingness. Journal of the American Statistical Association. 2013; 108(503):820–831. doi:10.1080/01621459.2013.795487.
60. Silcocks P, Kendrick D. Spatial effects should be allowed for in primary care and other community-based cluster RCTS. Trials. 2010; 11(1):55. doi:10.1186/1745-6215-11-55. [PubMed: 20470402]
61. van der Laan M. Causal inference for a population of causally connected units. Journal of Causal Inference. 2014; 0(0):1–62. doi:10.1515/jci-2013-0002.
62. 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.
63. van der Laan, M.; Rose, S. Targeted Learning: Causal Inference for Observational and Experimental Data. Springer; Berlin Heidelberg New York: 2011.
64. 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.
65. van der Laan M, Polley E, Hubbard A. Super learner. Statistical Applications in Genetics and Molecular Biology. 2007; 6(1):25.
66. 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.
67. 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]
68. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2014. URL http://www.R-project.org
69. Gruber, S.; van der Laan, M. Targeted Maximum Likelihood Estimation. R package version 1.2.0-1. 2012. Available at http://CRAN.R-project.org/package=tmle

70. Schwab, J.; Lendle, S.; Petersen, M.; van der Laan, M. ltmle: Longitudinal Targeted Maximum Likelihood Estimation. R package version 0.9.3-1. 2014. Available at http://CRAN.R-project.org/package=ltmle

Table 1

Summary of the true value of the exposure-specific effects $\psi_0(a) = 1/n \sum_i E_0(Y_i \mid A_i = a, W_i, Z_i)$ and the target parameter ψ_0 over 5,000 simulations of $n = 32$ communities. The rows indicate the setting with Simulation A corresponding to a rare outcome and Simulation B corresponding to a more common outcome. Recall the true value is dependent on the sample.

	$\psi_0(1)$			$\psi_0(0)$			$\psi_0 = \psi_0(1) - \psi_0(0)$		
	min	**mean**	**max**	**min**	**mean**	**max**	**min**	**mean**	**max**
Simulation A	0.018	0.024	0.031	0.023	0.032	0.043	−0.012	−0.009	−0.005
Simulation B	0.038	0.050	0.061	0.050	0.061	0.069	−0.013	−0.011	−0.007

Table 2

For Simulation A (rare outcome) and Simulation B (more common outcome), summary of the estimator performance over 5,000 simulations of $n = 32$ communities. The rows indicate the estimator and the columns the performance metric.

	Bias[a]	Std. Dev.[b]	Std. Error[c]	*t*-stat[d]	CI Cov.[e]	Power[f]
Simulation A			No Matching			
Unadj.	−0.00011	0.0054	0.0053	−1.6	96	34
TMLE linear for *Z*	−0.00008	0.0034	0.0032	−2.7	94	72
TMLE logit for *Z*	−0.00008	0.0033	0.0030	−2.9	94	78
TMLE linear for (W, *Z*)	−0.00011	0.0033	0.0027	−3.2	91	82
TMLE logit for (W, *Z*)	−0.00013	0.0031	0.0024	−3.6	90	88
			Adaptive Pair-Matching			
Unadj.	−0.00004	0.0034	0.0035	−2.5	96	64
TMLE linear for *Z*	−0.00004	0.0030	0.0030	−2.9	96	74
TMLE logit for *Z*	−0.00006	0.0030	0.0028	−3.2	94	80
TMLE linear for (W, *Z*)	−0.00004	0.0030	0.0029	−3.1	95	79
TMLE logit for (W, *Z*)	−0.00008	0.0029	0.0026	−3.5	93	84
Simulation B			No Matching			
Unadj.	−0.00007	0.0063	0.0062	−1.8	95	38
TMLE linear for *Z*	−0.00009	0.0035	0.0033	−3.4	94	88
TMLE logit for *Z*	−0.00013	0.0037	0.0036	−3.1	95	84
TMLE linear for (W, *Z*)	−0.00015	0.0032	0.0026	−4.2	91	96
TMLE logit for (W, *Z*)	−0.00037	0.0036	0.0031	−3.7	91	91
			Adaptive Pair-Matching			
Unadj.	−0.00007	0.0036	0.0036	−3.1	96	80
TMLE linear for *Z*	−0.00008	0.0030	0.0030	−3.8	96	92
TMLE logit for *Z*	−0.00011	0.0031	0.0033	−3.4	97	89
TMLE linear for (W, *Z*)	−0.00010	0.0029	0.0028	−4.1	95	95
TMLE logit for (W, *Z*)	−0.00023	0.0031	0.0032	−3.6	96	90

[a] average deviation between the point estimate & sample-specific true value

[b] square root of the variance of the point estimates

[c] average standard error estimate based on the influence curve

[d] average value of the test statistic (point estimate divided by standard error estimate)

[e] proportion of intervals containing the true parameter value (in percent)

[f] proportion of studies correctly rejecting the null hypothesis (in percent)