University of California, Berkeley
U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2014 *Paper* 320

Adaptive Pair-Matching in the SEARCH Trial and Estimation of the Intervention Effect

Laura Balzer* Maya L. Petersen†
Mark J. van der Laan‡

*Division of Biostatistics, University of California, Berkeley - the SEARCH Collaboration, lbbalzer@hsph.harvard.edu

†Division of Biostatistics, University of California, Berkeley - the SEARCH Collaboration, mayaliv@berkeley.edu

‡Division of Biostatistics, University of California, Berkeley - the SEARCH Collaboration, laan@berkeley.edu

http://biostats.bepress.com/ucbbiostat/paper320

Adaptive Pair-Matching in the SEARCH Trial and Estimation of the Intervention Effect

Laura Balzer, Maya L. Petersen, and Mark J. van der Laan

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 consequence, the observed data cannot be considered as n/2 independent, identically distributed (i.i.d.) pairs of units, as current 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, given the baseline covariates of the n study units. 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, a community randomized trial to evaluate the impact of immediate and streamlined antiretroviral therapy on HIV incidence in rural East Africa.

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 control for these characteristics during the analysis. Thereby, matching is sometimes implemented in 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 ([3–15]).

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. [15–20]). Such a data structure could arise by randomly sampling $n/2$ matched pairs from some target population of pre-existing matched units. In practice, however, the existence of such a target population is rare. Alternatively, this data structure could arise by (i) sampling a unit from the 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 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 [21].

A different pair-matching scheme was implemented in the Sustainable East Africa Research in Community Health (SEARCH) trial. SEARCH is an multinational, interdisciplinary consortium to evaluate the health, economic and educational impact 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. This process only occurred once. 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. 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. For examples of other types of adaptive designs, see Wei et al. [22], Tamura et al. [23], Collins et al. [24], Chambaz and van der Laan [25].

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 [26], the PRISM trial to prevent postpartum depression [27] and the SPACE study to promote physical activity [28]. Indeed, this design has motivated the development of so-called optimal matching algorithms to pair units based on similarity in their measured covariates [29, 30]. Previously, van der Laan et al. [31] 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 sample average conditional treatment effect. Specifically, we focus on the average treatment

effect for the n study units, given their measured covariates. This parameter avoids assuming some hypothetical infinite target population of units, is readily interpretable as the intervention effect for the sample at hand, and often leads to more precise estimates. The sample average treatment effect was originally proposed in Neyman [32] for a completely randomized trial and later considered in Imai et al. [14], Imai [20], Abadie and Imbens [33] for a pair-matched trial. The sample average treatment effect, given the covariate distribution, was initially proposed in Abadie and Imbens [34] and detailed in Abadie and Imbens [33], Imbens [35].

In pair-matched trials, the intervention effect is commonly estimated with the unadjusted pair-wise difference in outcomes with possible weighting for cluster size. The recommendations of whether and how to adjust for baseline covariates in pair-matched trials have been conflicting [3, 14, 15, 19, 20, 36]. For example, Hayes and Moulton [15] discuss a two-stage approach of first fitting a regression model with terms for the pairs and covariates, but not the intervention, and then contrasting the observed versus expected outcomes within matched pairs. Alternatively, random effects models and generalized estimating equations allow for a single regression with possible adjustment for individual and cluster-level covariates. Such estimation techniques require substantive modeling assumptions and may be unreliable in small trials, when the number of model parameters quickly approaches the number of units. As a result, some researchers, such as Imai et al. [14], advocate design-based estimation, which relies only on the unadjusted difference in weighted means. To the best of our knowledge, this is the first paper to propose targeted minimum loss-based estimation (TMLE) for the sample average treatment effect, given the vector of measured covariates, in an adaptive pair-matched trial. Without risking bias due to regression misspecification [37–39], TMLE allows for further adjustment of baseline covariates (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 sample average treatment effect, given the vector of measured covariates, as the causal parameter of interest. For brevity, we will refer to this causal parameter as the conditional treatment effect. Third, we present 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. We also provide conservative variance estimators and finite sample simulations. We conclude with some practical recommendations. Throughout, the SEARCH trial is used as the motivating example and detailed proofs are relegated to the Appendix.

2 The Estimation Problem

The SEARCH consortium will estimate the impact of immediate ART, initiated at all CD4 counts and delivered by a streamlined care system, on the 5-year cumulative incidence of HIV. In each study community, an annual health campaign will offer multi-disease prevention and treatment services, including HIV testing and referral [40]. Community members, failing to attend the campaign, will be offered home-based HIV testing. In communities randomized to Arm A, all individuals testing positive for HIV will be immediately eligible for ART with streamlined delivery. Elements of streamlined care include enhanced services for linkage, adherence and retention. In communities randomized to Arm B, all individuals testing positive for HIV will be offered ART according to in-country guidelines. 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 [41–48].

For the purposes of understanding the adaptive design, we focus on the cluster-level data and a single-time point setting. 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 [49]. A subset of the 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 Arm A (all individuals testing positive for HIV are immediately offered ART with streamlined care delivery) and equalling zero if the community was assigned to Arm B (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, 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 [31]. Mainly, the partitioning of the sample into $n/2$ pairs was a function of the baseline covariates of all N candidates. Adaptive pair-matching induces a statistical dependence between units, and the data generating process occurs only once. Nonetheless, given the covariates of all candidate communities $W^N = (W_1, \ldots, W_N)$, the observed data can be represented as $n/2$ independent random variables:

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

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 and vice versa:

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

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. Recent work, relaxing this assumption and considering a network of interacting units, is elaborated in van der Laan [50]. The conditional distribution of the observed data, given the baseline covariates of the candidate units, factorizes as

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

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 $\mathcal{M}$ is semiparametric.

2.1 The Conditional Treatment Effect

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 with annual HIV testing. The corresponding causal effect is often taken to be 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.

For the SEARCH trial, this parameter is not the most appropriate. In particular, it relies on the notion of some hypothetical target population of clusters. Instead, the SEARCH team identified 54 candidates communities from Western Uganda (Mbarara region), Eastern Uganda (Tororo region) and the Southern Nyanza Province in Kenya. These communities were selected to satisfy the study's inclusion criteria and to minimize possible contamination or spillover effects. From these candidates, the 16 best matched pairs were formed. Therefore, we focus on estimation of the sample average treatment effect, given the vector of baseline covariates:

$$\psi^F = \frac{1}{n}\sum_{i=1}^{n} E[Y_i^1 - Y_i^0 \mid W^n]$$

where Y_i^a denotes the counterfactual cumulative incidence under intervention level $A = a$ for unit i. This parameter is the average treatment effect for n SEARCH communities, given their measured baseline covariates. It avoids the notion of an infinite target population and is readily interpretable as the intervention effect for the study communities. Greater generalizability is up to the reader and not implicitly assumed in the parameter specification. By obviating estimation of the covariate distribution, estimators of the sample parameter will also often be more precise than those of the population parameter [20, 32–35].

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

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

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 is the difference in the average outcomes among units receiving the intervention and among units receiving the control:

$$\hat{\Psi}_{unadj}(P_n) = \frac{\sum_{i=1}^n A_i Y_i}{\sum_{i=1}^n A_i} - \frac{\sum_{i=1}^n (1 - A_i) Y_i}{\sum_{i=1}^n (1 - A_i)} = \bar{Q}_n(1) - \bar{Q}_n(0)$$

where P_n denotes the empirical distribution and $\bar{Q}_n(A) = E_n(Y|A)$ denotes an unadjusted estimate of the treatment-specific mean. This estimator is equivalent to the average difference in outcomes within matched pairs:

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

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, $A_{j2} = 0$. 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 for the accompanying proof.) This estimator is also unbiased for the population average treatment effect as well as for the sample average treatment effect, given the vector of counterfactual outcomes $(Y_i^1, Y_i^1 : \ i = 1, \ldots n)$ [20, 32]. Unfortunately, this simple difference-in-means estimator is *inefficient* as it fails to control for measured covariates. Despite recent advances in matching algorithms, 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 an additional baseline covariate that is predictive of the outcome, but was 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.

To the best of our knowledge, this is the first paper to propose targeted minimum loss-based estimation (TMLE) for the conditional treatment effect in an adaptive pair-matched trial. TMLE can provide an unbiased and efficient estimate of the intervention effect. The TMLE for $\Psi(P_0^n)$ is given by the following substitution estimator:

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

where $\bar{Q}_n^*(A, W)$ denotes a targeted estimate of the conditional mean function $E_0(Y|A, W)$. Briefly, 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 [51]. We refer the reader to van der Laan and Rose [52] for a detailed discussion and worked examples of TMLE. In an adaptive pair-matched trial, TMLE for $\Psi(P_0^n)$ 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 [53, 54].
2. Update the initial estimator $\bar{Q}_n(A, W)$ with a logistic fluctuation sub-model:

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

   In practice, run logistic regression of the outcome Y on the covariate $H(A)$, using the initial estimate as offset. Plugging the estimated coefficient ϵ_n into the fluctuation model yields targeted estimates $\bar{Q}_n^*(A, W)$.
3. Obtain the expected outcome for each unit under the intervention $\bar{Q}_n^*(1, W_i)$ and under the control $\bar{Q}_n^*(0, W_i)$.
4. Take the sample average of the differences in the expected outcomes:

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

Initial estimation of $\bar{Q}_0(A,W)$ 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 [37] detail the use of least squares regression to optimize the fit of $\bar{Q}_0(A,W)$ to have the lowest possible variance. If linear regression is used, then the following fluctuation sub-model is appropriate for Step 2:

$$\bar{Q}_n(\epsilon) = \bar{Q}_n + \epsilon H(A)$$

In practice, many cluster randomized trials, such as SEARCH, have a limited number of (conditionally) independent units. As a result, the number of parameters in the regression model 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, adjustment for a single or few covariates can safely yield efficiency gains without risk. 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 [39]. In other words, one can obtain an unbiased and more efficient estimate by simply fitting $\bar{Q}_0(A,W)$ with main terms linear or logistic regression and then taking the sample average of the differences in the expected outcomes under the treatment and control [37–39, 55].

2.3 Statistical Inference

As established in Appendix B, 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 [52]. Then the limit distribution of the standardized estimator is normal with mean 0 and variance given by the variance of its influence curve. Given 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.

The influence curve of the unadjusted estimator $\hat{\Psi}_{unadj}(P_n)$ of $\Psi(P_0^n)$ is conservatively approximated by the difference in residuals within matched pairs:

$$\hat{IC}_{unadj}(\bar{O}_j) = Y_{j1} - \bar{Q}_n(1) - \big(Y_{j2} - \bar{Q}_n(0)\big)$$

where $\bar{Q}_n(A)$ denotes an unadjusted estimate of the treatment-specific mean and again observations in matched pair j have been ordered such that the first corresponds to intervention ($A_{j1} = 1$) and the second to the control ($A_{j2} = 0$). A conservative variance estimator is then given by the sample variance of the estimated influence curve $\hat{\sigma}^2_{unadj}$, divided by $n/2$. This is equivalent to the sample variance of the within pair differences, divided by $n/2$. The latter is commonly recommended for pair-matched randomized trials [15]. As others have noted [20, 33–35] and we prove in Appendix B, this variance estimator will often be conservative. The true variance relies on the unknown conditional mean function $\bar{Q}_0(A,W)$. If this mean is consistently estimated, then this variance estimator will also be consistent; otherwise, the variance estimator will be conservative. In practice, the baseline covariates W, used for matching and possibly adjustment, are predictive of the outcome, and thereby the unadjusted treatment-specific mean $\bar{Q}_n(A)$ will not be a consistent for $\bar{Q}_0(A,W)$ and the variance will be overestimated on average. To obtain a less conservative variance estimator for $\hat{\Psi}_{unadj}(P_n)$, Abadie and Imbens [33] proposed a matching estimator, involving the variance of pairs-of-pairs with similar covariates. Additionally, Imai [20] derived upper and lower bounds on the conditional variance, given the vector of paired counterfactuals.

Our approach to reduce the true variance of the estimator and obtain a less conservative variance estimate is through adjustment for measured covariates with TMLE. The influence curve of the

proposed TMLE $\hat{\Psi}_{adj}(P_n)$ is also conservatively approximated by the difference in the residuals within matched pairs:

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

where $\bar{Q}^*_n(A, W)$ denotes a targeted estimate of the conditional mean function and again observations have again been ordered within matched pairs. A conservative variance estimator is then given by the sample variance of the estimated influence curve $\hat{\sigma}^2_{adj}$, divided by $n/2$. Assuming the baseline covariates W are predictive of the outcome, the estimated conditional mean function, adjusting for both the exposure and covariates $\bar{Q}^*_n(A, W)$, should be closer to the true mean $\bar{Q}_0(A, W)$ than the unadjusted estimate $\bar{Q}_n(A)$. As a result the asymptotic variance of the TMLE will often be smaller than that of the unadjusted estimator. Furthermore, in most practical settings, the variance of the unadjusted residuals $\hat{\sigma}^2_{unadj}$ will be larger than the variance of the adjusted residuals $\hat{\sigma}^2_{adj}$. Intuitively, the sum of squared unadjusted residuals is often greater than the sum of squared adjusted residuals. Thereby, both theoretically and in finite samples, TMLE is expected yield more precise and powerful estimates of the conditional treatment effect than the unadjusted estimator.

For both estimators, asymptotically conservative 95% confidence intervals can be constructed as $\hat{\psi} \pm 1.96\ \hat{\sigma}/\sqrt{n/2}$ where $\hat{\sigma}$ is the standard deviation of the estimated influence curve. Likewise, we can test the null hypothesis of no effect with test statistic $T = \sqrt{n/2}\hat{\psi}/\hat{\sigma}$. For trials with limited numbers of (conditionally) independent units, appropriate alternatives to the standard normal distribution include the Student's t-distribution with $n/2$-1 degrees of freedom and the permutation distribution. In an adaptive pair-matched trial, the permutation distribution is approximated by randomizing the intervention assignment within the fixed matched pairs and recalculating the test statistic. As noted by Imbens [35], the variance of estimators of the sample average treatment effect cannot be estimated with the bootstrap, because the statistical estimand depends on the sample and thereby changes over the bootstrapped samples.

We also briefly note that a randomized trial with adaptive pair-matching will often be more efficient for estimation of the conditional treatment effect than a non-matched trial (i.e. a trial with complete randomization). The designs will only have the same efficiency bound if the conditional mean outcome is consistently estimated: $\bar{Q}^*_n(A, W) = \bar{Q}_0(A, W)$. 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 design. 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 residuals within matched pairs will be positive and the adaptive design will yield more precise and powerful estimates. We refer the reader to Appendices C-D for further details and associated proofs.

3 Simulations

We present the following set of simulations to demonstrate implementation of the above estimators, the potential gain in efficiency with adaptive pair-matching and the further gain with adjustment during the analysis. While these simulations are not intended to represent the full complexities of the SEARCH trial, they explore some of the challenges faced, such as rare outcomes and the inability to match on all baseline covariates. All simulations were done in `R` v3.0.1 [56].

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:

$$Z = \text{expit}\big[-0.25 + 0.5^*W1 + W2 + 2^*W3 + 0.5^*U_Z\big]/4$$

where the expit function is the inverse of the logistic function and U_Z was drawn independently from a normal with mean 0 and standard deviation 1. The covariate Z represents an important pre-intervention determinant of the outcome and was not available to the matching process. To imitate adaptive pair-matching, the nonbipartite matching algorithm (`nbpMatching` v1.3.6 [30]) 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 $n/2$ matched pairs, the exposure A was randomized. Finally, the outcome Y was generated as

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

where random noise U_Y was drawn independently from a uniform distribution with minimum 0 and maximum 0.025. The intercept β_0 was set to either -2 or 0.5 to examine the performance of the estimators when the outcome was quite 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 intervention.

Over 5000 data sets, we examined the performance of the unadjusted estimator and TMLE with various adjustment sets. Specifically, we estimated the conditional mean outcome adjusting for only the exposure A and the covariate Z, which is an important determinant of the outcome but not used in matching. We also estimated the conditional mean function adjusting for the exposure A, the matching covariates W and the remaining covariate Z. Linear and logistic main terms regression models were used. Thereby, the fluctuation step of the TMLE algorithm did not provide an update [39]. Inference was based on the sample variance of the estimated influence curve and the Student's t-distribution with $n/2 - 1$ degrees of freedom. The corresponding TMLE implementation and proof of statistical inference for the non-matched randomized trial is given in Appendix C.

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 conditional treatment 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, given exposure $A = a$ and covariates (W, Z). For Simulation A, representing a quite 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).

	$\psi_0(1)$			$\psi_0(0)$			$\psi_0 = \psi_0(1) - \psi_0(0)$		
	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 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 quite rare outcome and Simulation B corresponding to a more common outcome. Recall the true value is dependent on the sample.

For Simulation A, Table 2 illustrates the performance of the estimators over 5,000 simulated data sets. All estimators are unbiased. Nonetheless, the average deviation between the point estimate and the sample-specific true value tended to be smaller with adaptive pair-matching than without. 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 1.57 times higher without matching than with matching. Likewise, the attained power (proportion of simulated trials where the null hypothesis was correctly rejected) jumped from 31% 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 $k_m = 0.29$ [15].

There was also an efficiency gain from adjustment. For the non-matched design, the standard deviation of the unadjusted estimator was 1.57 times higher than the standard deviation of the TMLE, using linear regression to adjust for Z. The corresponding power increased from 31% to 69%. 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 gained by using logistic regression to adjust for Z. With a very sparse outcome, logistic regression can be more stable than linear regression and is guaranteed to yield parameter estimates that respect the statistical model [53, 54]. 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. 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. Conversely, for both the unadjusted estimator and the TMLE only adjusting for Z, there is good confidence interval coverage and 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.

The results for Simulation B, representing a more common outcome, are given in Table 3 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

Simulation A	Bias	Std. Dev.	Std. Error	t-stat	CI Cov.	Power	α
	No Matching						
Unadj.	-1.13E-4	5.39E-3	5.34E-3	-1.64	0.96	0.31	0.04
TMLE linear for Z	-7.50E-5	3.44E-3	3.23E-3	-2.71	0.95	0.69	0.05
TMLE logit for Z	-8.01E-5	3.29E-3	2.98E-3	-2.94	0.94	0.75	0.05
TMLE linear for (W, Z)	-1.15E-4	3.28E-3	2.74E-3	-3.21	0.92	0.80	0.08
TMLE logit for (W, Z)	-1.35E-4	3.12E-3	2.43E-3	-3.62	0.91	0.87	0.09
	Adaptive Pair-Matching						
Unadj.	-4.37E-5	3.44E-3	3.46E-3	-2.54	0.96	0.64	0.04
TMLE linear for Z	-4.18E-5	3.04E-3	3.03E-3	-2.92	0.96	0.74	0.05
TMLE logit for Z	-5.68E-5	2.97E-3	2.78E-3	-3.20	0.94	0.80	0.05
TMLE linear for (W, Z)	-3.66E-5	2.98E-3	2.85E-3	-3.11	0.95	0.79	0.06
TMLE logit for (W, Z)	-7.80E-5	2.92E-3	2.58E-3	-3.45	0.93	0.84	0.06

Table 2: For Simulation A, summary of the estimator performance over 5,000 simulations of $n = 32$ communities: The rows indicate the estimator and the columns the performance metric with bias as the average deviation between the point estimate and sample-specific true value, standard deviation ("Std. Dev.") as the square root of the variance of the point estimates, standard error estimates ("Std. Error") as the average standard error estimate, t-stat as the average value of the test statistic, 95% confidence interval coverage ("CI Cov.") as the proportion of intervals containing the true parameter value, the attained power as the proportion of studies correctly rejecting the null hypothesis, and the Type I error rate α as the proportion of studies falsely rejecting the null hypothesis. The null scenario was simulated by randomly assigning the intervention, but generating the outcomes under the control ($A = 0$).

the empirical variance of the influence curve and thereby maximize the empirical efficiency [37]. For both the unadjusted estimator and the estimator adjusting only for Z, the standard error estimates were closely aligned with the true standard deviation. Again, there was a risk of over-fitting with the fully adjusted regression model for the conditional mean outcome. Due to the curse of dimensionality, inference was slightly optimistic. As a result, the confidence interval coverage was less than nominal for the non-matched design, and Type I error rates were too high for both designs.

4 Discussion

To our knowledge, this is the first paper to study and articulate the consequences of adaptive pair-matching for estimation of the sample average treatment effect, given the baseline covariates. 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 in communities with annual testing. 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. To the best of our understanding, adaptive pair-matching is a common design and has been implemented in other cluster randomized trials (e.g. [26–28]). It has also motivated the creation of `R` packages, such as `nbpMatching` [30], to match units on several covariates simultaneously.

We focused on estimation of the average treatment effect for the sample given the baseline covariates. This causal parameter avoids the assumption of a hypothetical target population of

Simulation B	Bias	Std. Dev.	Std. Error	t-stat	CI Cov.	Power	α
	No Matching						
Unadj.	-6.66E-5	6.34E-3	6.20E-3	-1.78	0.96	0.36	0.04
TMLE linear for Z	-9.02E-5	3.54E-3	3.28E-3	-3.36	0.95	0.86	0.05
TMLE logit for Z	-1.27E-4	3.74E-3	3.58E-3	-3.08	0.96	0.81	0.05
TMLE linear for (W, Z)	-1.53E-4	3.15E-3	2.62E-3	-4.25	0.92	0.95	0.08
TMLE logit for (W, Z)	-3.70E-4	3.63E-3	3.07E-3	-3.71	0.93	0.90	0.08
	Adaptive Pair-Matching						
Unadj.	-7.45E-5	3.59E-3	3.61E-3	-3.10	0.96	0.80	0.05
TMLE linear for Z	-8.37E-5	3.00E-3	2.98E-3	-3.77	0.96	0.92	0.05
TMLE logit for Z	-1.10E-4	3.10E-3	3.32E-3	-3.37	0.97	0.89	0.05
TMLE linear for (W, Z)	-9.70E-5	2.91E-3	2.76E-3	-4.07	0.95	0.95	0.06
TMLE logit for (W, Z)	-2.30E-4	3.13E-3	3.19E-3	-3.56	0.96	0.90	0.06

Table 3: For Simulation B, summary of the estimator performance over 5,000 simulations of $n = 32$ communities: The rows indicate the estimator and the columns the performance metric with bias as the average deviation between the point estimate and sample-specific true value, standard deviation ("Std. Dev.") as the square root of the variance of the point estimates, standard error estimates ("Std. Error") as the average standard error estimate, t-stat as the average value of the test statistic, 95% confidence interval coverage ("CI Cov.") as the proportion of intervals containing the true parameter value, the attained power as the proportion of studies correctly rejecting the null hypothesis, and the Type I error rate α as the proportion of studies falsely rejecting the null hypothesis. The null scenario was simulated by randomly assigning the intervention, but generating the outcomes under the control ($A = 0$).

clusters and is simply the intervention effect for the n study units. Also, by obviating estimation of the covariate distribution, estimators of the conditional parameter will often be less variable than that of the population parameter. We contrasted the unadjusted estimator, commonly used, with TMLE adjusting for baseline covariates. We provided a step-by-step implementation of the latter estimator and detailed proofs of inference. In fact, both estimators can be implemented by ignoring the dependence in the data, and asymptotically conservative inference obtained with the sample variance of the pairwise differences in residuals. We showed that the unadjusted estimator was less efficient both in theory and in practice. We further showed that adaptive pair-matching would often yield a more precise estimates of the conditional effect than a completely randomized trial.

Finite sample simulations were used to evaluate estimator performance and verify our theoretical results. Since the intervention was randomized, all estimators were unbiased. There was an efficiency gain with matching and a further gain with adjustment. When the outcome was quite rare, adjusting for the 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 was risky and resulted in under-coverage of the confidence intervals and higher than nominal Type I error rates.

For unbiased and powerful estimation of the conditional treatment effect in randomized trials, we recommend adaptive pair-matching with TMLE to adjust for a single or few pre-exposure covariates. Adaptive pair-matching is an intuitive strategy to group candidate units on similarity in their baseline covariates and will often yield more precise estimates and thereby power than the non-matched trial. Adjusting for an important measured covariate during the analysis can lead to further gains in efficiency with little risk to optimistic inference. Depending on the number of pairs,

it may also be possible to adjust for other covariates, including those used in matching. Future work will involve investigating the impact of adaptive stratification on estimation and inference for both the population and sample average treatment effect.

Appendix A: The unadjusted estimator is unbiased

Recall the unadjusted estimator, commonly implemented, is the difference in the outcomes within matched pairs:

$$\hat{\Psi}_{unadj}(P_n) = \frac{1}{n/2}\sum_{j=1}^{n/2}\left[A_{j1}Y_{j1} - (1-A_{j1})Y_{j1} + A_{j2}Y_{j2} - (1-A_{j2})Y_{j2}\right]$$

If observations within matched pairs have been ordered such that the first corresponds to intervention and the second to the control, the estimator can be expressed $\frac{1}{n/2}\sum_{j=1}^{n/2}(Y_{j1} - Y_{j2})$. Given the vector of covariates $W^n = (W_1, \ldots, W_n)$, the unadjusted estimator is unbiased for the statistical estimand:

$$\begin{aligned}
E_0\left[\hat{\Psi}_{unadj}(P_n) \mid W^n\right] =& \frac{1}{n/2}\sum_{j=1}^{n/2}\Big[E_0[A_{j1}Y_{j1} \mid W^n] - E_0[(1-A_{j1})Y_{j1} \mid W^n] \\
&+ E_0[A_{j2}Y_{j2} \mid W^n] - E_0[(1-A_{j2})Y_{j2} \mid W^n]\Big] \\
=& \frac{1}{n/2}\sum_{j=1}^{n/2}\Big[\bar{Q}_0(1,W_{j1})E_0(A_{j1} \mid W^n) - \bar{Q}_0(0,W_{j1})E_0((1-A_{j1}) \mid W^n) \\
&+ \bar{Q}_0(1,W_{j2})E_0(A_{j2} \mid W^n) - \bar{Q}_0(0,W_{j2})E_0((1-A_{j2}) \mid W^n)\Big] \\
=& \frac{1}{n/2}\sum_{j=1}^{n/2}\frac{1}{2}\Big[\bar{Q}_0(1,W_{j1}) - \bar{Q}_0(0,W_{j2}) + \bar{Q}_0(1,W_{j2}) - \bar{Q}_0(0,W_{j2})\Big] \\
=& \frac{1}{n}\sum_{i=1}^{n}\bar{Q}_0(1,W_i) - \bar{Q}_0(0,W_i) = \Psi(P_0^n).
\end{aligned}$$

Thus, $\hat{\Psi}_{unadj}(P_n)$ is an unbiased estimator of $\Psi(P_0^n)$, conditional on W^n.

Appendix B: Statistical Inference for the proposed TMLE

In this subsection, we establish that the proposed TMLE is an asymptotically linear estimator of the conditional average treatment effect in an adaptive pair-matched trial, where $n/2$ matched pairs are created as a function of baseline covariates of n candidate units. We then consider the adaptive design, where $n/2$ matched pairs are created as function of the baseline covariates of $N > n$ candidate units and the remaining $(N-n)$ units discarded. The latter adaptive design is a generalization of the first and the derived theorems are applicable. The theoretical results also apply to the unadjusted estimator, which can be considered a special case.

Let P_0^n denote the conditional distribution of $O^n = (O_1, \ldots, O_n)$, given the vector of covariates

$W^n = (W_1, \ldots, W_n)$. The statistical estimand is a function of this conditional distribution:

$$\Psi(P_0^n) = \frac{1}{n}\sum_{i=1}^{n} \bar{Q}_0(1, W_i) - \bar{Q}_0(0, W_i),$$

where $\bar{Q}_0(A, W) = E_0(Y|A, W)$ denotes the conditional expectation of the outcome, given the exposure A and the covariates W. The TMLE for $\Psi(P_0^n)$ is defined by following plug-in estimator:

$$\Psi(\bar{Q}_n^*) = \frac{1}{n}\sum_{i=1}^{n} \bar{Q}_n^*(1, W_i) - \bar{Q}_n^*(0, W_i),$$

where $\bar{Q}_n^*(A, W)$ denotes targeted estimates of the conditional mean function $\bar{Q}_0(A, W)$. Let ψ_0 denote the true parameter value and ψ_n^* denote the estimate.

Let us define the following function of the observed data structure $O = (W, A, Y)$:

$$D^*(\bar{Q}, g_0)(O) \equiv \left(\frac{\mathbb{I}(A=1)}{g_0(A)} - \frac{\mathbb{I}(A=0)}{g_0(A)}\right)\big(Y - \bar{Q}(A, W)\big),$$

where the marginal probability of receiving the intervention or the control is $g_0(A) = P_0(A) = 0.5$ in a randomized trial with two arms. By construction, TMLE solves $D^*(\bar{Q}, g_0)(O)$ at the targeted update $\bar{Q}_n^*$:

$$P_n D^*(\bar{Q}_n^*, g_0)(O_i) = \frac{1}{n}\sum_{i=1}^{n} D^*(\bar{Q}_n^*, g_0)(O_i) = 0,$$

where P_n denotes the empirical distribution, placing mass $(1/n)$ on each O_i, $i = 1, \ldots, n$. It is of interest to note that this equality can be rewritten as

$$\frac{1}{n/2}\sum_{j=1}^{n/2} \left\{\bar{Q}_n^*(1, W_{j1}) - \bar{Q}_n^*(0, W_{j2})\right\} = \frac{1}{n/2}\sum_{j=1}^{n/2} \left\{Y_{j1} - Y_{j2}\right\},$$

where observations in pair j have again been ordered such that the first corresponds to the intervention $A_{j1} = 1$ and the second to the control $A_{j2} = 0$. Thus, the TMLE has the interesting property that if it is used to predict the counterfactual effect $Y^1 - Y^0$ for each pair j, then the average of these j-specific effects equals the unadjusted estimator.

Let $P_0^n f = E[f(O^n) \mid W^n]$ denote the conditional expectation of a function f of the data O^n, given the covariate vector W^n. For all $\bar{Q}(A, W)$, we have

$$P_0^n D^*(\bar{Q}, g_0)(O_i) = \bar{Q}_0(1, W_i) - \bar{Q}_0(0, W_i) - \big(\bar{Q}(1, W_i) - \bar{Q}(0, W_i)\big).$$

Therefore, the statistical estimand $\Psi(P_0^n)$ minus the TMLE $\Psi(\bar{Q}_n^*)$ can be written as the empirical mean of the above conditional expectation:

$$\Psi(P_0^n) - \Psi(\bar{Q}_n^*) = \frac{1}{n}\sum_{i=1}^{n} P_0^n D^*(\bar{Q}_n^*, g_0)(O_i).$$

Combining the latter equality with $P_n D^*(\bar{Q}_n^*, g_0)(O_i) = 0$ yields

$$\begin{aligned}(\psi_n^* - \psi_0) &= P_n\left\{D^*(\bar{Q}_n^*, g_0)(O_i) - P_0^n D^*(\bar{Q}_n^*, g_0)(O_i)\right\} \\ &= \frac{1}{n}\sum_{i=1}^{n}\left\{D^*(\bar{Q}_n^*, g_0)(O_i) - P_0^n D^*(\bar{Q}_n^*, g_0)(O_i)\right\}.\end{aligned}$$

We can re-write this equality in terms of the empirical distribution $P_{n/2}$, which puts mass $1/(n/2)$ on each paired data point $\bar{O}_j = (O_{j1}, O_{j2})$:

$$(\psi_n^* - \psi_0) = P_{n/2}\left\{\bar{D}^*(\bar{Q}_n^*, g_0) - P_0^n \bar{D}^*(\bar{Q}_n^*, g_0)\right\}$$
$$= \frac{1}{n/2}\sum_{j=1}^{n/2}\left\{\bar{D}^*(\bar{Q}_n^*, g_0)(\bar{O}_j) - P_0^n \bar{D}^*(\bar{Q}_n^*, g_0)\right\}$$
$$\text{where } \bar{D}^*(\bar{Q}, g_0)(\bar{O}_j) = \frac{1}{2}\{D^*(\bar{Q}, g_0)(O_{j1}) + D^*(\bar{Q}, g_0)(O_{j2})\}$$

Now let $\mathcal{F}$ be a set of multivariate real valued functions so that $\bar{Q}_n^*(A, W)$ is an element of $\mathcal{F}$ with probability 1. Define the process $(Z_n(\bar{Q}) : \bar{Q} \in \mathcal{F})$ by

$$Z_n(\bar{Q}) = \frac{1}{\sqrt{n/2}}\sum_{j=1}^{n/2}\left\{\bar{D}^*(\bar{Q}, g_0)(\bar{O}_j) - P_0^n \bar{D}^*(\bar{Q}, g_0)\right\}$$

Conditional on the covariate vector $W^n = (W_1, \ldots, W_n)$, $Z_n(\bar{Q})$ is a sum of $n/2$ independent mean zero random variables $\bar{D}^*(\bar{Q}, g_0)(\bar{O}_j) - P_0^n \bar{D}^*(\bar{Q}, g_0)$, $j = 1, \ldots, n/2$. Below we establish asymptotic equicontinuity of $(Z_n(\bar{Q}) : \bar{Q} \in \mathcal{F})$ so that $Z_n(\bar{Q}_n^*) - Z_n(\bar{Q}) \to 0$ in probability. Then, we can conclude that

$$\sqrt{n/2}(\psi_n^* - \psi_0) = \frac{1}{\sqrt{n/2}}\sum_{j=1}^{n/2}\left\{\bar{D}^*(\bar{Q}, g_0)(\bar{O}_j) - P_0^n \bar{D}^*(\bar{Q}, g_0)\right\} + o_P(1).$$

Since the main term on the right-hand side, conditional on W^n, is a sum of independent mean zero random variables, we can apply the central limit theorem for sums of independent random variables.

Let us define the following function of the paired data $\bar{O}_j = (O_{j1}, O_{j2})$:

$$IC_j(\bar{Q}, \bar{Q}_0, g_0) \equiv \bar{D}^*(\bar{Q}, g_0)(\bar{O}_j) - P_0^n \bar{D}^*(\bar{Q}, g_0),$$

where the notation recognizes that $P_0^n \bar{D}^*(\bar{Q}, g_0)$ also depends on the true conditional mean $\bar{Q}_0(A, W) = E_0(Y|A, W)$. We assume that

$$\Sigma_0 = \lim_{n\to\infty} \frac{1}{n/2}\sum_{j=1}^{n/2} P_0^n IC_j(\bar{Q}, \bar{Q}_0, g_0)^2$$

exists as a limit. Then, we have shown $\sqrt{n/2}(\psi_n^* - \psi_0) \Rightarrow_d N(0, \Sigma_0)$.

To establish the asymptotic equicontinuity result, we use a few fundamental building blocks. Let $\mathcal{F}_d = \{f_1 - f_2 : f_1, f_2 \in \mathcal{F}\}$. Let $\sigma_n^2(f) = P_0^n Z_n(f)^2$ be the conditional variance. Note that $Z_n(f)/\sigma_n(f)$ is a sum of $n/2$ independent mean zero bounded random variables and the variance of this sum equals 1. Bernstein's inequality states that $P(|\sum_j Y_j| > x) \leq 2\exp\left(-\frac{1}{2}\frac{x^2}{v + Mx/3}\right)$, where $v \geq \text{VAR}\sum_j Y_j$. Thus, by Bernstein's inequality, conditional on W^n, we have

$$P\left(\frac{|Z_n(f)|}{\sigma_n(f)} > x\right) \leq 2\exp\left(-\frac{1}{2}\frac{x^2}{1 + Mx/3}\right) \leq K\exp(-Cx^2),$$

for a universal K and C. This implies $\| Z_n(f)/\sigma_n(f) \|_{\psi_2} \leq (1 + K/C)^{0.5}$, where for a given convex function ψ with $\psi(0) = 0$, $\| X \|_{\psi} \equiv \inf\{C > 0 : E\psi(| X | /C) \leq 1\}$ is the so called Orlics norm, and $\psi_2(x) = \exp(x^2) - 1$. Thus $\| Z_n(f) \|_{\psi_2} \leq C_1 \sigma_n(f)$ for $f \in \mathcal{F}^d$. This result allows us to apply Theorem 2.2.4 in van der Vaart and Wellner [57]: for each $\delta > 0$ and $\eta > 0$, we now have

$$\| \sup_{\sigma_n(f) \leq \delta} | Z_n(f) | \|_{\psi_2} \leq K \left\{ \int_0^{\eta} \psi_2^{-1}(N(\epsilon, \sigma_n, \mathcal{F}_d) d\epsilon + \delta \psi_2^{-1}(N^2(\eta, \sigma_n, \mathcal{F}_d)), \right\}, \tag{1}$$

where $N(\epsilon, \sigma_n, \mathcal{F}_d)$ is the number of balls of size ϵ w.r.t. norm $\| f \| = \sigma_n(f)$ to cover $\mathcal{F}_d$.

Convergence of a sequence of random variables to zero with respect to ψ_2-orlics norm implies convergence in expectation to zero and thereby convergence of that sequence of random variables to zero in probability. Let δ_n be a sequence converging to zero, and let η_n also converge to zero but slowly enough so that the term $\delta_n \psi_2^{-1}(N^2(\eta_n, \sigma_n, \mathcal{F}^d))$ converges to zero as $n \to \infty$. By assumption, $\int_0^{\delta_n} \psi_2^{-1}(N(\epsilon, \sigma_n, \mathcal{F}^d) d\epsilon$ converges to zero. Thus,

$$\lim_{\delta_n \to 0} \left\{ \int_0^{\delta_n} \psi_2^{-1}(N(\epsilon, \sigma_n, \mathcal{F}^d) d\epsilon + \delta_n \psi_2^{-1}(N^2(\eta_n, \sigma_n, \mathcal{F}^d)) \right\} = 0.$$

This proves that

$$E \left(\sup_{\{f : \sigma_n(f) \leq \delta_n\}} | Z_n(f) | \right) \to 0.$$

Thus, if $\sigma_n(\bar{Q}_n^* - \bar{Q}) \to 0$ in probability, then $Z_n(\bar{Q}_n^* - \bar{Q}) \to 0$ in probability. This proves the following theorem.

Theorem 1 *Consider the TMLE $\Psi(\bar{Q}_n^*)$ of the statistical estimand $\Psi(P_0^n) = 1/n \sum_{i=1}^n \{\bar{Q}_0(1, W_i) - \bar{Q}_0(0, W_i)\}$. Let $P_0^n f$ represent the conditional expectation of a function f of O^n, given the vector of covariates W^n. This conditional expectation, $P_0^n f$, is thus still random through W^n. Let $\mathcal{F}$ be a set of multivariate real valued functions so that $\bar{Q}_n^*$ is an element of $\mathcal{F}$ with probability 1. Define*

$$Z_n(\bar{Q}) = \frac{1}{\sqrt{n/2}} \sum_{j=1}^{n/2} IC_j(\bar{Q}, \bar{Q}_0, g_0)$$

where

$$\begin{aligned} IC_j(\bar{Q}, \bar{Q}_0, g_0) &\equiv \bar{D}^*(\bar{Q}, g_0)(\bar{O}_j) - P_0^n \bar{D}^*(\bar{Q}, g_0) \\ \bar{D}^*(\bar{Q}, g_0)(\bar{O}_j) &= \frac{1}{2} \left\{ D^*(\bar{Q}, g_0)(O_{j1}) + D^*(\bar{Q}, g_0)(O_{j2}) \right\} \\ D^*(\bar{Q}, g_0)(O_i) &= \left(\frac{\mathbb{I}(A_i = 1)}{g_0(A_i)} - \frac{\mathbb{I}(A_i = 0)}{g_0(A_i)} \right) \big(Y_i - \bar{Q}(A_i, W_i) \big). \end{aligned}$$

where $g_0(A) = P_0(A)$ is known. We make the following assumptions.

Uniform bound: *Assume $\sup_{\bar{Q} \in \mathcal{F}} \sup_O | D^*(\bar{Q}, g_0) | < M < \infty$, where the second supremum is over a set that contains the support of each O_i.*

Convergence of variances: *Assume that for a specified $\{\sigma_0^2(\bar{Q}) : \bar{Q} \in \mathcal{F}\}$, for any $\bar{Q} \in \mathcal{F}$, $\frac{1}{n/2} \sum_{j=1}^{n/2} P_0^n IC_j(\bar{Q}, \bar{Q}_0, g_0)^2 \to \sigma_0^2(\bar{Q})$ a.s (i.e, for almost every $(W^n, n \geq 1)$).*

Convergence of $\bar{Q}_n^*$ to some limit: *For any $\bar{Q}_1, \bar{Q}_2 \in \mathcal{F}$, we define*

$$\sigma_n^2(\bar{Q}_1 - \bar{Q}_2) = \frac{1}{n/2} \sum_{j=1}^{n/2} P_0^n \{IC_j(\bar{Q}_1, \bar{Q}_0, g_0) - IC_j(\bar{Q}_2, \bar{Q}_0, g_0)\}^2,$$

where we note that the right-hand side indeed only depends on $\bar{Q}_1, \bar{Q}_2$ through its difference $\bar{Q}_1 - \bar{Q}_2$.

Assume that for a particular $\bar{Q}^* \in \mathcal{F}$, $\sigma_n^2(\bar{Q}_n^* - \bar{Q}^*) \to 0$ in probability as $n \to \infty$.

Entropy condition: *Let $\mathcal{F}^d = \{f_1 - f_2 : f_1, f_2 \in \mathcal{F}\}$. Let $N(\epsilon, \sigma_n, \mathcal{F}^d)$ be the covering number of the class $\mathcal{F}^d$ w.r.t norm/dissimilarity $\| f \| = \sigma_n(f)$. Assume that the class $\mathcal{F}$ satisfies*

$$\lim_{\delta_n \to 0} \int_0^{\delta_n} \sqrt{\log N(\epsilon, \sigma_n, \mathcal{F}^d)} d\epsilon = 0$$

Asymptotic equicontinuity of process: *Then,*

$$Z_n(\bar{Q}_n^*) - Z_n(\bar{Q}^*) \text{ converges to zero in probability, as } n \to \infty.$$

First order linear approximation: *As a consequence,*

$$\sqrt{n/2}(\psi_n^* - \psi_0) = Z_n(\bar{Q}^*) + o_P(1).$$

Asymptotic normality: *In addition, $Z_n(\bar{Q}^*)$ converges to $N(0, \sigma_0^2(\bar{Q}^*))$, so that*

$$\sqrt{n/2}(\psi_n^* - \psi_0) \text{ converges in distribution to } N(0, \sigma_0^2(\bar{Q}^*)).$$

The asymptotic variance $\sigma_0^2(\bar{Q}^*)$ equals the limit of

$$\sigma_{0,n}^2 = \frac{1}{n/2} \sum_{j=1}^{n/2} P_0^n \left\{ IC_j(\bar{Q}_n^*, \bar{Q}_0, g_0)(\bar{O}_j) \right\}^2$$

If Y_i is d-dimensional outcome, then the application of the above theorem to each component of ψ_n^* yields the desired asymptotic linearity for the d-dimensional ψ_n^* and thereby the asymptotic normality as well.

Appendix B.1: Conservative variance estimation

The above result suggests the following estimator of the asymptotic variance of the standardized TMLE:

$$\hat{\Sigma} = \frac{1}{n/2} \sum_{j=1}^{n/2} \left\{ IC_j(\bar{Q}_n^*, \bar{Q}_{n,np}, g_0)(\bar{O}_j) \right\}^2$$

where $\bar{Q}_{n,np}$ is a consistent estimator of $\bar{Q}_0$. Unfortunately, such a variance estimator relies upon consistent estimation of the conditional mean function $\bar{Q}_0$, which is particular concerning when n is small. However, we will now show that one can obtain a conservative variance estimate, which does not rely on a consistent estimator of the conditional mean function $\bar{Q}_0$.

The asymptotic variance of the standardized estimator $\sqrt{n/2}(\psi_n^* - \psi_0)$ can be expressed as

$$\begin{aligned}\Sigma_0 &= \lim_{n\to\infty} \frac{1}{n/2} \sum_{j=1}^{n/2} P_0^n \left[IC_j(\bar{Q}^*, \bar{Q}_0, g_0)(\bar{O}_j) \right]^2 \\ &= \lim_{n\to\infty} \frac{1}{n/2} \sum_{j=1}^{n/2} P_0^n \left[\bar{D}^*(\bar{Q}^*, g_0)(\bar{O}_j) \right]^2 - \left[P_0^n \bar{D}^*(\bar{Q}^*, g_0)(\bar{O}_j) \right]^2.\end{aligned}$$

The latter term is zero when $\bar{Q}^*(A, W) = \bar{Q}_0(A, W)$:

$$\begin{aligned}P_0^n \bar{D}^*(\bar{Q}^*, g_0)(\bar{O}_j) =& \frac{1}{2} \Big\{ \bar{Q}_0(1, W_{j1}) - \bar{Q}_0(0, W_{j1}) - \left(\bar{Q}^*(1, W_{j1}) - \bar{Q}^*(0, W_{j1}) \right) \\ &+ \bar{Q}_0(1, W_{j2}) - \bar{Q}_0(0, W_{j2}) - \left(\bar{Q}^*(1, W_{j2}) - \bar{Q}^*(0, W_{j2}) \right) \Big\}\end{aligned}$$

Thus, the true variance Σ_0 is always less than or equal to an upper bound Σ_0^u, where

$$\Sigma_0^u = \lim_{n\to\infty} \frac{1}{n/2} \sum_{j=1}^{n/2} P_0^n \left\{ \bar{D}^*(\bar{Q}^*, g_0)(\bar{O}_j) \right\}^2$$

Again, if the conditional mean is consistently estimated $\bar{Q}^*(A, W) = \bar{Q}_0(A, W)$, $\Sigma_0^u = \Sigma_0$.

We can consistently estimate the upper bound Σ_0^u with

$$\hat{\Sigma}^u = \frac{1}{n/2} \sum_{j=1}^{n/2} \left\{ \bar{D}^*(\bar{Q}_n^*, g_0)(\bar{O}_j) \right\}^2$$

Recall

$$\begin{aligned}\bar{D}^*(\bar{Q}, g_0)(\bar{O}_j) =& \frac{1}{2} \left\{ D^*(\bar{Q}, g_0)(O_{j1}) + D^*(\bar{Q}, g_0)(O_{j2}) \right\} \\ =& \frac{1}{2} \Bigg[\left(\frac{\mathbb{I}(A_{j1} = 1)}{g_0(A_{j1})} - \frac{\mathbb{I}(A_{j1} = 0)}{g_0(A_{j1})} \right) \left(Y_{j1} - \bar{Q}(A_{j1}, W_{j1}) \right) \\ &+ \left(\frac{\mathbb{I}(A_{j2} = 1)}{g_0(A_{j2})} - \frac{\mathbb{I}(A_{j2} = 0)}{g_0(A_{j2})} \right) \left(Y_{j2} - \bar{Q}(A_{j2}, W_{j2}) \right) \Bigg]\end{aligned}$$

Ordering the observations within pairs, such that index $j1$ corresponds to the unit randomized to the intervention ($A_{j1} = 1$) and $j2$ corresponds to the unit randomized to the control ($A_{j2} = 0$), it follows that

$$\bar{D}^*(\bar{Q}, g_0)(\bar{O}_j) = Y_{j_1} - \bar{Q}(1, W_{j1}) - \left(Y_{j_2} - \bar{Q}(0, W_{j2}) \right),$$

allowing us to represent the conservative variance estimator $\hat{\Sigma}^u$ as the difference in residuals within matched pairs:

$$\hat{\Sigma}^u = \frac{1}{n/2} \sum_{j=1}^{n/2} \left\{ Y_{j_1} - \bar{Q}_n^*(1, W_{j1}) - \left(Y_{j_2} - \bar{Q}_n^*(0, W_{j2}) \right) \right\}^2.$$

Appendix B.2: Generalization to $N > n$ candidate units

Now consider the common adaptive design, where first N candidate units are selected, the best $n/2$ matched pairs selected as a function of the covariate vector $W^N = (W_1, \ldots, W_n, \ldots, W^N)$, and the remaining $N - n$ units discarded. In the SEARCH trial, for example, 16 matched pairs were formed as a function of the baseline covariates of 54 candidate communities. As a result of this adaptive design, the treatment assignment mechanism depends on the N candidate communities. Nonetheless, in a randomized trial, the conditional likelihood of the observed data factorizes as

$$\begin{aligned} P_0^N(O_1, \ldots, O_n | W_1, \ldots, W_N) &= \prod_{j=1}^{n} g_0(A_{j1}, A_{j2} | W_1, \ldots, W_N) P_0(Y_{j1} | A_{j1}, W_{j1}) P_0(Y_{j2} | A_{j2}, W_{j2}) \\ &= 0.5 \prod_{j=1}^{n} P_0(Y_{j1} | A_{j1}, W_{j1}) P_0(Y_{j2} | A_{j2}, W_{j2}) \\ &= P_0(O_1, \ldots, O_n \mid W_1, \ldots, W_n) = P_0^n(O^n | W^n) \end{aligned}$$

Therefore, given the baseline covariates of the n study units $W^n = (W_1, \ldots W_n)$, we still have $n/2$ conditionally independent observations. Furthermore, recall that the statistical estimand corresponds to the sample average treatment effect for the n units included in the study:

$$\Psi(P_0^n) = \frac{1}{n} \sum_{i=1}^{n} \bar{Q}_0(1, W_i) - \bar{Q}_0(0, W_i)$$

Since we condition on $W^n = (W_1, \ldots, W_n)$ in the target parameter and corresponding TMLE, the actual distribution that generated these n covariates is not important. Recall we make no assumptions about the joint distribution of $P_0(W^N)$. We only need to assume that the conditional variance still converges. As a result, we can apply the same TMLE and asymptotics. As detailed in van der Laan et al. [31], this is a much different result than when the target parameter is the marginal (population) average treatment effect. In the latter case, the so-called adaptive missingness has important implications for estimation and inference to a target population of units.

Appendix C: Comparison with independent randomization

In this section, we consider estimation and inference for the sample average conditional treatment effect in a trial where the intervention is completely randomized. We consider implementation of the TMLE and the corresponding asymptotics. We conclude with an efficiency comparison between a trial randomizing the intervention with adaptive pairs and a trial with independent randomization.

Appendix C.1: TMLE for the conditional effect with independent randomization

Let $\bar{Q}_n(A, W)$ be an initial estimator of $\bar{Q}_0(A, W)$, which can be obtained by regressing the outcome Y_i on exposure A_i and covariates W_i, $i = 1, \ldots, n$. For a binary or bounded continuous outcome, the negative log-likelihood is a valid loss function:

$$-L(\bar{Q})(O) = Y \log \bar{Q}(A, W) + (1 - Y) \log(1 - \bar{Q}(A, W))$$

Now consider the logistic fluctuation submodel:

$$\begin{aligned} \text{Logit}\left[\bar{Q}_n(\epsilon)\right] &= \text{Logit}\left[\bar{Q}_n\right] + \epsilon H(A) \\ \text{where } H(A) &= \left(\frac{\mathbb{I}(A = 1)}{g_0(A)} - \frac{\mathbb{I}(A = 0)}{g_0(A)} \right) \end{aligned}$$

In a randomized trial with two arms, the probability of receiving the intervention or control $g_0(A = a) = P_0(A = a) = 0.5$. Let ϵ_n be the minimizer of the empirical mean of the loss function:

$$\epsilon_n = \arg\min_{\epsilon} P_n L(\bar{Q}_n(\epsilon)) = \frac{1}{n}\sum_{i=1}^{n} L(\bar{Q}_n(\epsilon))(O_i)$$

The TMLE of the conditional mean function $\bar{Q}_0$ is defined by plugging in the estimated coefficient ϵ_n into the fluctuation model $\bar{Q}_n^* = \bar{Q}_n(\epsilon_n)$. The TMLE of $\Psi(P_0^n)$ is defined as the corresponding plug-in estimator:

$$\Psi(\bar{Q}_n^*) = \frac{1}{n}\sum_{i=1}^{n} \left\{\bar{Q}_n^*(1, W_i) - \bar{Q}_n^*(0, W_i)\right\}$$

As before, initial estimation of the conditional mean function $\bar{Q}_0(A, W)$ can also be based on least squares regression and targeting achieved with the following fluctuation submodel:

$$\bar{Q}_n(\epsilon) = \bar{Q}_n + \epsilon H(A)$$

Recall the definition of $D^*(\bar{Q}, g_0)(O)$ as a function of the observed data structure $O = (W, A, Y)$:

$$D^*(\bar{Q}, g_0)(O) = \left(\frac{\mathbb{I}(A = 1)}{g_0(A)} - \frac{\mathbb{I}(A = 0)}{g_0(A)}\right)(Y - \bar{Q}(A, W)),$$

where the probability of receiving the intervention or the control is $g_0(A) = P_0(A) = 0.5$ in a randomized trial. By construction, TMLE solves $D^*(\bar{Q}, g_0)(O)$ at the targeted update $\bar{Q}_n^*$:

$$P_n D^*(\bar{Q}_n^*, g_0) = \frac{1}{n}\sum_{i=1}^{n} D^*(\bar{Q}_n^*, g_0)(O_i) = 0$$

where P_n denotes the empirical distribution, placing mass $(1/n)$ on each $O_i,\ i = 1, \ldots, n$. For all $\bar{Q}(A, W)$, we also have

$$P_0^n D^*(\bar{Q}, g_0)(O_i) = \bar{Q}_0(1, W_i) - \bar{Q}_0(0, W_i) - \big(\bar{Q}(1, W_i) - \bar{Q}(0, W_i)\big),$$

where $P_0^n f = E\big[f(O^n)|W^n\big]$ denotes the conditional expectation of f of the data O^n, given the covariates vector W^n. Therefore, the statistical estimand $\Psi(P_0^n)$ minus the TMLE $\Psi(\bar{Q}_n^*)$ can be written as the empirical mean of the above conditional expectation:

$$\Psi(P_0^n) - \Psi(\bar{Q}_n^*) = \frac{1}{n}\sum_{i=1}^{n} P_0^n D^*(\bar{Q}_n^*, g_0)(O_i).$$

Combining the latter equality with $P_n D^*(\bar{Q}_n^*, g_0) = 0$ yields

$$\sqrt{n}(\psi_n^* - \psi_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left\{D^*(\bar{Q}_n^*, g_0)(O_i) - P_0^n D^*(\bar{Q}_n^*, g_0)\right\}.$$

Recall $\mathcal{F}$ is the set of multivariate real-valued functions such that $\bar{Q}_n^*(A, W)$ is an element of $\mathcal{F}$ with probability 1. Define the process $(Z_n(\bar{Q}) : \bar{Q} \in \mathcal{F})$ by

$$Z_n(\bar{Q}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left\{D^*(\bar{Q}, g_0)(O_i) - P_0^n D^*(\bar{Q}, g_0)\right\},$$

Conditional on the covariate vector $W^n = (W_1, \ldots, W_n)$, $Z_n(\bar{Q})$ is a sum of n independent mean zero random variables $D^*(\bar{Q}, g_0)(O_i) - P_0^n D^*(\bar{Q}, g_0)$, $i = 1, \ldots, n$. Below we establish asymptotic equicontinuity of $(Z_n(\bar{Q}) : \bar{Q} \in \mathcal{F})$ so that $Z_n(\bar{Q}_n^*) - Z_n(\bar{Q}) \to 0$ in probability. Then, we can conclude that

$$\sqrt{n}(\psi_n^* - \psi_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ D^*(\bar{Q}, g_0)(O_i) - P_0^n D_i^*(\bar{Q}, g_0) \right\} + o_P(1).$$

Since the main term on the right-hand side, conditional on W^n, is a sum of independent mean zero random variables, we can apply the central limit theorem for sums of independent random variables.

Let us define the following function of the unit data $O_i = (W_i, A_i, Y_i)$:

$$IC_i(\bar{Q}, \bar{Q}_0, g_0) \equiv D^*(\bar{Q}, g_0)(O_i) - P_0^n D^*(\bar{Q}, g_0),$$

where the notation recognizes that $P_0^n \bar{D}^*(\bar{Q}, g_0)$ also depends on the true conditional mean $\bar{Q}_0(A, W) = E_0(Y|A, W)$. We assume that

$$\Sigma_0 = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} P_0^n IC_i(\bar{Q}, \bar{Q}_0, g_0)^2$$

exists as a limit. Then, we have shown $\sqrt{n}(\psi_n^* - \psi_0) \Rightarrow_d N(0, \Sigma_0)$.

To establish the asymptotic equicontinuity result, we use a few fundamental building blocks. Let $\mathcal{F}_d = \{f_1 - f_2 : f_1, f_2 \in \mathcal{F}\}$. Let $\sigma_n^2(f) = P_0^n Z_n(f)^2$ be the conditional variance. Note that $Z_n(f)/\sigma_n(f)$ is a sum of n independent mean zero bounded random variables and the variance of this sum equals 1. Bernstein's inequality states that $P(\mid \sum_j Y_j \mid > x) \leq 2 \exp\left(-\frac{1}{2} \frac{x^2}{v + Mx/3}\right)$, where $v \geq \mathrm{VAR} \sum_j Y_j$. Thus, by Bernstein's inequality, conditional on W^n, we have

$$P\left(\frac{\mid Z_n(f) \mid}{\sigma_n(f)} > x \right) \leq 2 \exp\left(-\frac{1}{2} \frac{x^2}{1 + Mx/3} \right) \leq K \exp(-Cx^2),$$

for a universal K and C. This implies $\parallel Z_n(f)/\sigma_n(f) \parallel_{\psi_2} \leq (1 + K/C)^{0.5}$, where for a given convex function ψ with $\psi(0) = 0$, $\parallel X \parallel_\psi \equiv \inf\{C > 0 : E\psi(\mid X \mid /C) \leq 1\}$ is the so called Orlics norm, and $\psi_2(x) = \exp(x^2) - 1$. Thus $\parallel Z_n(f) \parallel_{\psi_2} \leq C_1 \sigma_n(f)$ for $f \in \mathcal{F}^d$. This result allows us to apply Theorem 2.2.4 in van der Vaart and Wellner [57]: for each $\delta > 0$ and $\eta > 0$, we now have

$$\parallel \sup_{\sigma_n(f) \leq \delta} \mid Z_n(f) \mid \parallel_{\psi_2} \leq K \left\{ \int_0^\eta \psi_2^{-1}(N(\epsilon, \sigma_n, \mathcal{F}_d) d\epsilon + \delta \psi_2^{-1}(N^2(\eta, \sigma_n, \mathcal{F}_d)), \right\}, \tag{2}$$

where $N(\epsilon, \sigma_n, \mathcal{F}_d)$ is the number of balls of size ϵ w.r.t. norm $\parallel f \parallel = \sigma_n(f)$ to cover $\mathcal{F}_d$.

Convergence of a sequence of random variables to zero with respect to ψ_2-orlics norm implies convergence in expectation to zero and thereby convergence of that sequence of random variables to zero in probability. Let δ_n be a sequence converging to zero, and let η_n also converge to zero but slowly enough so that the term $\delta_n \psi_2^{-1}(N^2(\eta_n, \sigma_n, \mathcal{F}^d))$ converges to zero as $n \to \infty$. By assumption, $\int_0^{\delta_n} \psi_2^{-1}(N(\epsilon, \sigma_n, \mathcal{F}^d) d\epsilon$ converges to zero. Thus,

$$\lim_{\delta_n \to 0} \left\{ \int_0^{\delta_n} \psi_2^{-1}(N(\epsilon, \sigma_n, \mathcal{F}^d) d\epsilon + \delta_n \psi_2^{-1}(N^2(\eta_n, \sigma_n, \mathcal{F}^d)) \right\} = 0.$$

This proves that

$$E\left(\sup_{\{f:\sigma_n(f)\leq\delta_n\}} \mid Z_n(f) \mid\right) \to 0.$$

Thus, if $\sigma_n(\bar{Q}_n^* - \bar{Q}) \to 0$ in probability, then $Z_n(\bar{Q}_n^* - \bar{Q}) \to 0$ in probability. This proves the following theorem.

Theorem 2 *Consider the TMLE $\Psi(\bar{Q}_n^*)$ for the statistical estimand $\Psi(P_0^n) = 1/n\sum_{i=1}^n\{\bar{Q}_0(1,W_i) - \bar{Q}_0(0,W_i)\}$ defined above for a trial with independent randomization. Let $P_0^n f$ represents a conditional expectation of a function f of O^n, given W^n. This conditional expectation is thus still random through W^n. Let $\mathcal{F}$ be a set of multivariate real valued functions so that $\bar{Q}_n^*$ is an element of $\mathcal{F}$ with probability 1. Define*

$$Z_n(\bar{Q}) = \frac{1}{\sqrt{n}}\sum_{i=1}^n IC_i(\bar{Q}, \bar{Q}_0, g_0),$$

where

$$IC_i(\bar{Q}, \bar{Q}_0, g_0) \equiv D^*(\bar{Q}, g_0)(O_i) - P_0^n D^*(\bar{Q}, g_0)$$
$$D^*(\bar{Q}, g_0)(O_i) = \left(\frac{\mathbb{I}(A_i = 1)}{g_0(A_i)} - \frac{\mathbb{I}(A_i = 0)}{g_0(A_i)}\right)(Y_i - \bar{Q}(A_i, W_i)).$$

We make the following assumptions.

Uniform bound: *Assume $\sup_{\bar{Q}\in\mathcal{F}}\sup_O \mid D^*(\bar{Q}, g_0) \mid < M < \infty$, where the second supremum is over a set that contains the support of each O_i.*

Convergence of variances: *Assume that for a specified $\{\sigma_0^2(\bar{Q}) : \bar{Q} \in \mathcal{F}\}$, for any $\bar{Q} \in \mathcal{F}$, $\frac{1}{n}\sum_{i=1}^n P_0^n IC_i(\bar{Q}, \bar{Q}_0, g_0)^2 \to \sigma_0^2(\bar{Q})$ a.s (i.e, for almost every $(W^n, n \geq 1)$).*

Convergence of $\bar{Q}_n^*$ to some limit: *For any $\bar{Q}_1, \bar{Q}_2 \in \mathcal{F}$, we define*

$$\sigma_n^2(\bar{Q}_1 - \bar{Q}_2) = \frac{1}{n}\sum_{i=1}^n P_0^n\left\{IC_i(\bar{Q}_1, \bar{Q}_0, g_0) - IC_i(\bar{Q}_2, \bar{Q}_0, g_0)\right\}^2,$$

where we note that the right-hand side indeed only depends on $\bar{Q}_1, \bar{Q}_2$ through its difference $\bar{Q}_1 - \bar{Q}_2$.

Assume that for a particular $\bar{Q}^* \in \mathcal{F}$, $\sigma_n^2(\bar{Q}_n^* - \bar{Q}^*) \to 0$ in probability as $n \to \infty$.

Entropy condition: *Let $\mathcal{F}^d = \{f_1 - f_2 : f_1, f_2 \in \mathcal{F}\}$. Let $N(\epsilon, \sigma_n, \mathcal{F}^d)$ be the covering number of the class $\mathcal{F}^d$ w.r.t norm/dissimilarity $\| f \| = \sigma_n(f)$. Assume that the class $\mathcal{F}$ satisfies*

$$\lim_{\delta_n\to 0}\int_0^{\delta_n}\sqrt{\log N(\epsilon, \sigma_n, \mathcal{F}^d)}d\epsilon = 0$$

Asymptotic equicontinuity of process: *Then,*

$Z_n(\bar{Q}_n^*) - Z_n(\bar{Q}^*)$ *converges to zero in probability, as $n \to \infty$.*

First order linear approximation: *As a consequence,*

$$\sqrt{n}(\psi_n^* - \psi_0) = Z_n(\bar{Q}^*) + o_P(1).$$

Asymptotic normality: *In addition, $Z_n(\bar{Q}^*)$ converges to $N(0, \sigma_0^2(\bar{Q}^*))$, so that*

$$\sqrt{n}(\psi_n^* - \psi_0) \text{ converges in distribution to } N(0, \sigma_0^2(\bar{Q}^*)).$$

The asymptotic variance $\sigma_0^2(\bar{Q}^*)$ equals the limit of

$$\sigma_{0,n}^2 = \frac{1}{n}\sum_{i=1}^{n} P_0^n \left\{ IC_i(\bar{Q}^*, \bar{Q}_0, g_0)(O_i) \right\}^2. \tag{3}$$

If Y_i is a d-dimensional outcome, then application of the above theorem to each component of ψ_n^* yields the desired asymptotic linearity for the d-dimensional ψ_n^* and thereby the asymptotic normality as well.

Appendix C.2: Conservative variance estimation

As before, we can obtain a conservative estimate of Σ_0, which does not rely on a consistent estimator of the conditional mean function $\bar{Q}_0(A, W)$. The asymptotic variance of the standardized estimator in the design with independent randomization can be expressed as

$$\Sigma_0 = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} P_0^n \left\{ D^*(\bar{Q}^*, g_0)(O_i) \right\}^2 - \left\{ P_0^n D^*(\bar{Q}^*, g_0)(O_i) \right\}^2$$

The latter term is zero when $\bar{Q}^*(A, W) = \bar{Q}_0(A, W)$:

$$P_0^n D^*(\bar{Q}^*, g_0) = \bar{Q}_0(1, W_i) - \bar{Q}_0(0, W_i) - \left(\bar{Q}^*(1, W_i) - \bar{Q}^*(0, W_i) \right)$$

Thus, the true variance Σ_0 is always less than or equal to an upper bound Σ_0^u, where

$$\Sigma_0^u = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} P_0^n \left\{ D^*(\bar{Q}^*, g_0)(O_i) \right\}^2$$

We can consistently estimate the upper bound Σ_0^u with

$$\begin{aligned} \hat{\Sigma}^u &= \frac{1}{n}\sum_{i=1}^{n} \left\{ D^*(\bar{Q}_n^*, g_0)(O_i) \right\}^2 \\ &= \frac{4}{n}\sum_{i=1}^{n} \left(Y_i - \bar{Q}_n^*(A_i, W_i) \right)^2. \end{aligned}$$

where we have used that the treatment assignment mechanism $g_0(A) = P_0(A) = 0.5$ in a randomized trial.

Appendix D: Comparison of asymptotic variances of the TMLEs in the independent design and the adaptive pair-matched design

The above two theorems give us the following approximations for the TMLEs $\psi_{n,I}^*$ under independent randomization and $\psi_{n,M}^*$ under adaptive pair-matching:

$$\sqrt{n}(\psi_{n,I}^* - \psi_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left\{ D^*(\bar{Q}, g_0)(O_i) - P_0^n D^*(\bar{Q}, g_0)(O_i) \right\} + o_P(1)$$

$$\sqrt{n/2}(\psi_{n,M}^* - \psi_0) = \frac{1}{\sqrt{n/2}}\sum_{j=1}^{n/2} \left\{ \bar{D}^*(\bar{Q}, g_0)(\bar{O}_j) - P_0^n \bar{D}^*(\bar{Q}, g_0)(\bar{O}_j) \right\} + o_P(1)$$

where

$$\bar{D}^*(\bar{Q}, g_0)(\bar{O}_j) = \frac{1}{2}\Big\{ D^*(\bar{Q}, g_0)(O_{j1}) + D^*(\bar{Q}, g_0)(O_{j2}) \Big\}$$
$$D^*(\bar{Q}, g_0)(O_i) = \left(\frac{\mathbb{I}(A_i = 1)}{g_0(A_i)} - \frac{\mathbb{I}(A_i = 0)}{g_0(A_i)} \right) \big(Y_i - \bar{Q}(A_i, W_i) \big).$$

The corresponding asymptotic variances are

$$\Sigma_{0,I} = \lim_{n\to\infty} \frac{1}{n} \sum_i P_0^n \Big[D^*(\bar{Q}, g_0)(O_i)^2 \Big] - \Big[P_0^n D^*(\bar{Q}, g_0)(O_i) \Big]^2$$
$$\Sigma_{0,M} = \lim_{n\to\infty} \frac{1}{n/2} \sum_{j=1}^{n/2} P_0^n \Big[\bar{D}^*(\bar{Q}^*, g_0)(\bar{O}_j)^2 \Big] - \Big[P_0^n \bar{D}^*(\bar{Q}^*, g_0)(\bar{O}_j) \Big]^2,$$

respectively. Expanding out the squared terms and simplifying, the asymptotic variance of the standardized estimator in the independent design is

$$\Sigma_{0,I} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \Big\{ 2E_0 \Big[\big(Y_i - \bar{Q}_0(1, W_i) \big)^2 \Big| A_i = 1, W^n \Big] + 2E_0 \Big[\big(Y_i - \bar{Q}_0(0, W_i) \big)^2 \Big| A_i = 0, W^n \Big]$$
$$+ \big[\bar{Q}_0(1, W_i) - \bar{Q}(1, W_i) + \bar{Q}_0(0, W_i) - \bar{Q}(0, W_i) \big]^2 \Big\}$$

Likewise, the asymptotic variance of the standardized estimator in the adaptive design is

$$\Sigma_{0,M} = \lim_{n\to\infty} \frac{1}{2n} \sum_{i=1}^{n} \Big\{ 2E_0 \Big[\big(Y_i - \bar{Q}_0(1, W_i) \big)^2 \Big| A_i = 1, W^n \Big] + 2E_0 \Big[\big(Y_i - \bar{Q}_0(0, W_{j1}) \big)^2 \Big| A_i = 0, W^n \Big]$$
$$+ \big[\bar{Q}_0(1, W_i) - \bar{Q}(1, W_i) + \bar{Q}_0(0, W_i) - \bar{Q}(0, W_i) \big]^2 \Big\} - \rho_0$$
$$= 0.5\Sigma_{0,I} - \rho_0$$

where ρ_0 is the following pairwise product

$$\rho_0 = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n/2} \Big\{ \big[\bar{Q}_0(1, W_{j1}) - \bar{Q}(1, W_{j1}) + \bar{Q}_0(0, W_{j1}) - \bar{Q}(0, W_{j1}) \big] \times$$
$$\big[\bar{Q}_0(1, W_{j2}) - \bar{Q}(1, W_{j2}) + \bar{Q}_0(0, W_{j2}) - \bar{Q}(0, W_{j2}) \big] \Big\}$$

The proof is omitted here, but readily available upon request from the authors.

Thus, the asymptotic variance of the TMLE in the independent design is $\Sigma_{0,I}/n$ whereas the asymptotic variance of the TMLE in the adaptive design is $\Sigma_{0,M}/(n/2) = \Sigma_{0,I}/n - 2\rho_0/n$. When we match well on measured and unmeasured factors, the product of the deviations between the true conditional means and the estimated means within matched pairs is expected to be positive:

$$\rho_0 \geq 0$$

Under this condition, the adaptive design will be more efficient than the completely randomized trial. As an example, consider the unadjusted estimator and suppose we match perfectly on W, which is predictive of the outcome. Then the relevant term is

$$\big[\bar{Q}_0(1, W_j) - \bar{Q}_n(1) + \bar{Q}_0(0, W_j) - \bar{Q}_n(0) \big]^2 > 0$$

If we consistently estimate $\bar{Q}_0(A,W)$, then the cross-term ρ_0 is zero and the efficiency bound of the two designs is the same:

$$\Sigma_{0,M}/(n/2) = \Sigma_{0,I}/n$$

In finite samples, we also expect there to be an efficiency gain from pair-matching. Comparing the proposed variance estimators, we have

$$\begin{aligned}
\hat{\Sigma}^u_M &= \frac{1}{n/2}\sum_{j=1}^{n/2}\left[\big(Y_{j1} - \bar{Q}^*_n(1,W_{j1})\big) - \big(Y_{j2} - \bar{Q}^*_n(0,W_{j2})\big)\right]^2 \\
&= \frac{1}{n/2}\sum_{j=1}^{n/2}\Big[\big(Y_{j1} - \bar{Q}^*_n(1,W_{j1})\big)^2 + \big(Y_{j2} - \bar{Q}^*_n(0,W_{j2})\big)^2 \\
&\qquad - 2\big(Y_{j1} - \bar{Q}^*_n(1,W_{j1})\big)\big(Y_{j2} - \bar{Q}^*_n(0,W_{j2})\big)\Big] \\
\hat{\Sigma}^u_I &= \frac{4}{n}\sum_{i=1}^{n}(Y_i - \bar{Q}^*_n(A_i,W_i))^2 \\
&= \frac{4}{n}\sum_{j=1}^{n/2}\big(Y_{j1} - \bar{Q}^*_n(1,W_{j1})\big)^2 + \big(Y_{j2} - \bar{Q}^*_n(1,W_{j2})\big)^2
\end{aligned}$$

As consequence, the difference in the variance estimators is

$$\frac{\hat{\Sigma}^u_I}{n} - \frac{\hat{\Sigma}^u_M}{n/2} = \frac{2}{(n/2)^2}\sum_{j=1}^{n/2}\big(Y_{j1} - \bar{Q}^*_n(1,W_{j1})\big)\big(Y_{j2} - \bar{Q}^*_n(0,W_{j2})\big)$$

If we succeed in matching pairs on predictive covariates W, then the sample covariance of residuals within matched pairs will be positive. Under this condition (expected to hold in practice), adaptive pair-matching will yield more precise estimates in finite samples.

5 Acknowledgements

The authors would like to acknowledge and thank the entire SEARCH collaboration for their helpful comments and discussion. This work was supported in part by the National Institute Of Allergy And Infectious Diseases of the National Institutes of Health under award number UM1AI069502. Maya Petersen is a recipient of a Doris Duke Clinical Scientist Development Award.

References

[1] M.C. Costanza. Matching. *Preventive Medicine*, 24(5):425–433, 1995. doi: 10.1006/pmed.1995.1069.

[2] K.J. Rothman, S. Greenland, and T.L. Lash. *Modern Epidemiology*. Lippincott Williams & Wilkins, Phildelphia, 2008.

[3] N. Klar and A. Donner. The merits of matching in community intervention trials: a cautionary tale. *Statistics in Medicine*, 16(15):1753–1764, 1997. doi: 10.1002/(SICI)1097-0258(19970815)16:15⟨1753::AID-SIM597⟩3.0.CO;2-E.

[4] W.G. Cochran. Matching in Analytical Studies. *American Journal of Public Health and the Nations Health*, 43(6):684–691, 1953. doi: 10.2105/AJPH.43.6_Pt_1.684.

[5] W.Z. Billewicz. Matched samples in medical investigations. *British Journal of Preventive & Social Medicine*, 18:167–173, 1964. doi: 10.1136/jech.18.4.167.

[6] L.H. Youkeles. 184 Note: Loss of Power through Ineffective Pairing of Observations in Small Two-Treatment All-or-None Experiments. *Biometrics*, 19(1):175–180, 1963. doi: 10.2307/2527582.

[7] K.K. Mathen. Matching in comparative studies in public health. *Indian Journal of Public Health*, 8(4):161–169, 1963.

[8] W.Z. Billewicz. The Efficiency of Matched Samples: An Empirical Investigation. *Biometrics*, 21(3):623–644, 1965. doi: 10.2307/2528546.

[9] O.S. Miettinen. The matched pairs design in the case of all-or-none responses. *Biometrics*, 24 (2):339–352, 1968. doi: 10.2307/2528039.

[10] S.M. McKinlay. Pair-Matching–A Reappraisal of a Popular Technique. *Biometrics*, 33(4): 725–735, 1977. doi: 10.2307/2529471.

[11] S. Wacholder and C.R. Weinberg. Paired versus Two-Sample Design for a Clinical Trial of Treatments with Dichotomous Outcome: Power considerations. *Biometrics*, 38(3):801–812, 1982. doi: 10.2307/2530059.

[12] D.C. Martin, P. Diehr, E.B. Perrin, and T.D. Koepsell. The effect of matching on the power of randomized community intervention studies. *Statistics in Medicine*, 12(3-4):329–338, 1993. doi: 10.1002/sim.4780120315.

[13] P. Diehr, D.C. Martin, T. Koepsell, and A. Cheadle. Breaking the matches in a paired t-test for community interventions when the number of pairs is small. *Statistics in Medicine*, 14(13): 1491–1504, 1995. doi: 10.1002/sim.4780141309.

[14] K. Imai, G. King, and C. Nall. The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation. *Statistical Science*, 24(1):29–53, 2009. doi: 10.1214/08-STS274.

[15] R.J. Hayes and L.H. Moulton. *Cluster Randomised Trials*. Chapman & Hall/CRC, Boca Raton, 2009.

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

[17] D.M. Murray. *Design and Analysis of Group-Randomized Trials*. Oxford University Press, New York, 1998.

[18] A. Donner and N. Klar. *Design and Analysis of Cluster Randomization Trials in Health Research*. Arnold, London, 2000.

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

[20] K. Imai. Variance identification and efficiency analysis in randomized experiments under the matched-pair design. *Statistics in Medicine*, 27(24):4857–4873, 2008. doi: 10.1002/sim.3337.

[21] L.B. Balzer, M.L. Petersen, and M.J. van der Laan. Why Match in Individually and Cluster Randomized Trials? Technical Report 294, Division of Biostatistics, University of California, Berkeley, 2012. Available at http://www.bepress.com/ucbbiostat/paper294.

[22] L.J. Wei, R.T. Smythe, D.Y. Lin, and T.S. Park. Statistical inference with data-dependent treatment allocation rules. *Journal of the American Statistical Association*, 85(156-162), 1990. doi: 10.1080/01621459.1990.10475319.

[23] R.N. Tamura, D.E. Faries, J.S. Andersen, and J.H. Heiligenstein. A case study of an adaptive clinical trial in the treatment of out-patients with depressive disorder. *Journal of the American Statistical Association*, 89(427):768–776, 1994. doi: 10.2307/2290902.

[24] L.M. Collins, S.A. Murphy, and K.L. Bierman. A conceptual framework for adaptive preventive interventions. *Prevention Science*, 5(3):185–196, 2004. doi: 10.1023/B:PREV.0000037641.26017.00.

[25] A. Chambaz and M.J. van der Laan. Targeting the optimal design in randomized clinical trials with binary outcomes and no covariate: Theoretical study. *The International Journal of Biostatistics*, 7(1):Article 10, 2011. doi: 10.2202/1557-4679.1247.

[26] H. Grosskurth, F. Mosha, J. Todd, E. Mwijarubi, A. Klokke, K. Senkoro, P. Mayaud, J. Changalucha, A. Nicoll, G. ka-Gina, J. Newell, K. Mugeye, D. Mabey, and R. Hayes. Impact of improved treatment of sexually transmitted diseases on HIV infection in rural Tanzania: randomised controlled trial. *The Lancet*, 346(8974):530–536, 1995. doi: 10.1016/S0140-6736(95)91380-7.

[27] L. Watson, R. Small, S. Brown, W. Dawson, and J. Lumley. Mounting a community-randomized trial: sample size, matching, selection, and randomization issues in PRISM. *Controlled Clinical Trials*, 25(3):235–250, 2004. doi: 10.1016/j.cct.2003.12.002.

[28] M. Toftager, L.B. Christiansen, P.L. Kristensen, and J. Troelsen. Space for physical activity-a multicomponent intervention study: study design and baseline findings from a cluster randomized controlled trial. *BMC Public Health*, 11:777, 2011. doi: 10.1186/1471-2458-11-777.

[29] R. Greevy, B. Lu, J.H. Silber, and P. Rosenbaum. Optimal multivariate matching before randomization. *Biostatistics*, 5(2):263–275, 2004. doi: 10.1093/biostatistics/5.2.263.

[30] B. Lu, R. Greevy, and C. Beck. *nbpMatching: functions for non-bipartite optimal matching*, 2012. URL http://CRAN.R-project.org/package=nbpMatching. R package version 1.3.6.

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

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

[33] A. Abadie and G.W. Imbens. Estimation of the conditional variance in paired experiments. *Annales d'Economie et de Statistique*, (91/92):175–187, 2008.

[34] A. Abadie and G. Imbens. Simple and bias-corrected matching esti- mators for average treatment effects. Technical Report 283, NBER technical working paper, 2002.

[35] G.W. Imbens. Nonparametric estimation of average treatment effects under exogeneity: a review. *Review of Economics and Statistics*, 86(1):4–29, 2004. doi: 10.1162/003465304323023651.

[36] J. Hill and M. Scott. Comment: The essential role of pair matching. *Statistical Science*, 24 (1):54–58, 2009. doi: 10.1214/09-STS274A.

[37] Daniel B. Rubin and M.J. van der Laan. Empirical efficiency maximization: improved locally efficient covariate adjustment in randomized experiments and survival analysis. *The International Journal of Biostatistics*, 4(1):Article 5, 2008. doi: 10.2202/1557-4679.1084.

[38] K.L. Moore and M.J. van der Laan. Covariate adjustment in randomized trials with binary outcomes: Targeted maximum likelihood estimation. *Statistics in Medicine*, 28(1):39–64, 2009. doi: 10.1002/sim.3445.

[39] M. Rosenblum and M.J. van der Laan. Simple, efficient estimators of treatment effects in randomized trials using generalized linear models to leverage baseline variables. *The International Journal of Biostatistics*, 6(1):Article 13, 2010. doi: 10.2202/1557-4679.1138.

[40] G. Chamie, D. Kwarisiima, T.D. Clark, J. Kabami, V. Jain, E. Geng, M.L. Petersen, H. Thirumurthy, M.R. Kamya, D.V. Havlir, E.D. Charlebois, and the SEARCH Collaboration. Leveraging rapid community-based HIV testing campaigns for non-communicable diseases in rural Uganda. *PLoS One*, 7(8):e43400, 2012. doi: 10.1371/journal.pone.0043400.

[41] H. Thirumurthy, M. Goldstein, and J. Graff Zivin. The economic impact of AIDS treatment: labor supply in Western Kenya. *Journal of Human Resources*, 43:511–552, 2008. doi: 10.1353/jhr.2008.0009.

[42] G. J. Zivin, H. Thirumurthy, and M. Goldstein. AIDS Treatment and Intrahousehold Resource Allocation: Children's Nutrition and Schooling in Kenya. *Journal of Public Economics*, 93 (7-8):1008–1015, 2009. doi: 10.1016/j.jpubeco.2009.03.003.

[43] M.S. Cohen, Y. Q. Chen, M. McCauley, T. Gamble, M.C. Hosseinipour, N. Kumarasamy, et al. Prevention of HIV-1 infection with early antiretroviral therapy. *New England Journal of Medicine*, 365(6):493–505, 2011. doi: 10.1056/NEJMoa1105243.

[44] S.D. Lawson and A.D. Harries. Reducing tuberculosis-associated early mortality in antiretroviral treatment programmes in sub-Saharan Africa. *AIDS*, 25(12):1554–1555, 2011. doi: 10.1097/QAD.0b013e328348fb61.

[45] K. Middelkoop, L. Bekker, L. Myer, L.F. Johnson, M. Kloos, C. Morrow, and R. Wood. Antiretroviral therapy and TB notification rates in a high HIV prevalence South African community. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 56(3):263–269, 2011. doi: 10.1097/QAI.0b013e31820413b3.

[46] C. Flateau, G. Le Loup, and G. Pialoux. Consequences of HIV infection on malaria and therapeutic implications: a systematic review. *The Lancet Infectious Diseases*, 11(7):541–556, 2011. doi: 10.1016/S1473-3099(11)70031-7.

[47] M.G. Fowler, A.R. Gable, M.A. Lampe, M. Etima, and M. Owor. Perinatal HIV and its prevention: progress toward an HIV-free generation. *Clinics in Perinatology*, 37(4):699–719, vii, 2012. doi: 10.1016/j.clp.2010.09.002.

[48] E.J. Schouten, A. Jahn, D. Midiani, S.D. Makombe, A. Mnthambala, Z. Chirwa, A.D. Harries, J.J. van Oosterhout, T. Meguid, A. Ben-Smith, R. Zachariah, L. Lynen, M. Zolfo, W. Van Damme, C.F. Gilks, R. Atun, M. Shawa, and F. Chimbwandira. Prevention of mother-to-child transmission of HIV and the health related Millennium Development Goals: time for a public health approach. *The Lancet*, 378(9787):282–284, 2011. doi: 10.1016/S0140-6736(10)62303-3.

[49] V. Jain, T. Liegler, J. Kabami, G. Chamie, T.D. Clark, D. Black, E.H. Geng, D. Kwarisiima, J.K. Wong, M. Abdel-Mohsen, N. Sonawane, F.T. Aweeka, H. Thirumurthy, M.L. Petersen, E.D. Charlebois, M.R. Kamya, D.V. Havlir, and for the SEARCH Collaboration. Assessment of population-based HIV RNA levels in a rural East African setting using a fingerprick-based blood collection method. *Clinical Infectious Diseases*, 56(4):598–605, 2013. doi: 10.1093/cid/cis881.

[50] M.J. van der Laan. Causal inference for a population of causally connected units. *Journal of Causal Inference*, 0(0):1–62, 2014. doi: 10.1515/jci-2013-0002.

[51] M.J. van der Laan and D.B. Rubin. Targeted maximum likelihood learning. *The International Journal of Biostatistics*, 2(1):Article 11, 2006. doi: 10.2202/1557-4679.1043.

[52] M. van der Laan and S. Rose. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer, Berlin Heidelberg New York, 2011.

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

[54] L. Balzer and M.J. van der Laan. Estimating Effects on Rare Outcomes: Knowledge is Power. Technical Report 310, Division of Biostatistics, University of California, Berkeley, 2013. Available at http://biostats.bepress.com/ucbbiostat/paper310/.

[55] D.O. Scharfstein, A. Rotnitzky, and J.M. Robins. Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models (with Rejoiner). *Journal of the American Statistical Association*, 94(448):1096–1120 (1135–1146), 1999. doi: 10.2307/2669930.

[56] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. URL http://www.R-project.org. ISBN 3-900051-07-0.

[57] A.W. van der Vaart and J.A. Wellner. *Weak convergence and empirical processes*. Springer, Berlin Heidelberg New York, 1996.