Research Article

Statistics in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

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

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

Appendix A: The unadjusted estimator is unbiased

Recall the unadjusted estimator is the average difference in the outcomes within matched pairs:

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

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:

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

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

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

Contract/grant sponsor: NIH grant numbers R01AI074345 & U01AI99959

Appendix B: Statistical inference for the TMLE

In this subsection, we establish that the proposed TMLE is an asymptotically linear estimator of the conditional average treatment effect (CATE) 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 $\hat{\psi}_{unadj}$, 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 $O = (W, A, Y)$:

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

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 $\bar{Q}_n^*$:

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

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

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

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

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

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:

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

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

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

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 $\bar{O}_j = (O_{j1}, O_{j2})$:

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

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 $P(\mid \sum_j Y_j \mid > x) \leq 2\exp\left(-\frac{1}{2}\frac{x^2}{v + Mx/3}\right)$, where $v \geq$ 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 [1]: for each $\delta > 0$ and $\eta > 0$, we now have

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

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 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

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

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

Uniform bound: *Assume $\sup_{\bar{Q} \in \mathcal{F}} \sup_O \mid D^*(\bar{Q}, g_0) \mid < M < \infty$, 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 $\bar{Q} \in \mathcal{F}$, $\frac{1}{n/2} \sum_{j=1}^{n/2} P_0^n IC_j(\bar{Q}, \bar{Q}_0, g_0)^2 \to \sigma_0^2(\bar{Q})$ 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*

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

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,*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 average treatment effect, conditional on the baseline covariates of the n study units:

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

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.* [2], 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 complete randomization

In this section, we consider estimation and inference for the conditional average treatment effect (CATE) 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 within adaptive pairs and a trial with complete (independent) randomization.

Appendix C.1: TMLE for the CATE under complete 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:

$$\text{logit}\big[\bar{Q}_n(A, W)(\epsilon)\big] = \text{logit}\big[\bar{Q}_n(A, W)\big] + \epsilon H(A)$$

$$\text{where } H(A) = \left(\frac{\mathbb{I}(A = 1)}{g_0(A)} - \frac{\mathbb{I}(A = 0)}{g_0(A)} \right)$$

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

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

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

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

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:

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

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

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

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

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

where $P_0^n f = E\big[f(O^n)|W^n\big]$ denotes the conditional expectation of the function f of the data O^n, given the covariate 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:

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

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

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

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.

For a completely randomized trial, 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 $P_0^n D^*(\bar{Q}, g_0)$ 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 $v \geq$ 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 [1]: for each $\delta > 0$ and $\eta > 0$, we now have

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

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 complete (i.e. 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*

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

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* $\parallel f\parallel=\sigma_n(f)$*. 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,*

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

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*

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

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 variance estimator, 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 complete randomization can be expressed as

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

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

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

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

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

We can consistently estimate the upper bound Σ_0^u with

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

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

Appendix C.3: Comparison of asymptotic variances of the TMLEs in the independent design (i.e. under complete randomization) 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:

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

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

where

$$\bar{D}^*(\bar{Q}, g_0)(\bar{O}_j) = \frac{1}{2}\left\{ D^*(\bar{Q}, g_0)(O_{j1}) + D^*(\bar{Q}, g_0)(O_{j2}) \right\}$$

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

The corresponding asymptotic variances are

$$\Sigma_{0,I} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} P_0^n \left[D^*(\bar{Q}, g_0)^2 \right] - \left[P_0^n D^*(\bar{Q}, g_0) \right]^2$$

$$\Sigma_{0,M} = \lim_{n\to\infty} \frac{1}{n/2} \sum_{j=1}^{n/2} P_0^n \left[\bar{D}^*(\bar{Q}^*, g_0)^2 \right] - \left[P_0^n \bar{D}^*(\bar{Q}^*, g_0) \right]^2,$$

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

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

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

$$\begin{aligned}\Sigma_{0,M} &= \lim_{n\to\infty} \frac{1}{2n} \sum_{i=1}^{n} \left\{ 2E_0 \left[\left(Y_i - \bar{Q}_0(1, W_i) \right)^2 \middle| A_i = 1, W^n \right] + 2E_0 \left[\left(Y_i - \bar{Q}_0(0, W_i) \right)^2 \middle| A_i = 0, W^n \right] \right. \\ &\quad \left. + \left[\bar{Q}_0(1, W_i) - \bar{Q}(1, W_i) + \bar{Q}_0(0, W_i) - \bar{Q}(0, W_i) \right]^2 \right\} - \rho_0 \\ &= 0.5\Sigma_{0,I} - \rho_0\end{aligned}$$

where ρ_0 is the following pairwise product

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

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

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

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

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

Then, the difference 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.

Appendix D: Simulation results under the null

The following tables give the simulation results when there is no effect. The null scenario was simulated by randomly assigning the intervention, but generating the outcomes under the control ($A = 0$). Recall Simulation A represents a rare outcome and Simulation B represents a more common outcome.

References

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

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

Table 1. For Simulation A (rare outcome) and Simulation B (more common outcome) with no treatment effect, summary of the estimator performance over 5,000 simulations of $n = 32$ communities. The rows indicate the estimator and the columns the performance metric.

	Bias[a]	Std. Dev.[b]	Std. Error[c]	t-stat[d]	CI Cov.[e]	α[f]
Simulation A			No Matching			
Unadj.	0.00015	0.0061	0.0060	0.0	95	5
TMLE linear for Z	0.00001	0.0033	0.0032	0.0	94	6
TMLE logit for Z	0.00003	0.0032	0.0030	0.0	94	6
TMLE linear for (W, Z)	0.00003	0.0030	0.0026	0.0	91	9
TMLE logit for (W, Z)	0.00005	0.0029	0.0024	0.0	90	10
			Adaptive Pair-Matching			
Unadj.	0.00002	0.0034	0.0034	0.0	96	4
TMLE linear for Z	0.00005	0.0028	0.0028	0.0	95	5
TMLE logit for Z	0.00005	0.0027	0.0027	0.0	95	5
TMLE linear for (W, Z)	0.00005	0.0027	0.0026	0.0	94	6
TMLE logit for (W, Z)	0.00005	0.0027	0.0025	0.0	94	6
Simulation B			No Matching			
Unadj.	0.00017	0.0058	0.0057	0.0	95	5
TMLE linear for Z	0.00006	0.0035	0.0033	0.0	94	6
TMLE logit for Z	0.00007	0.0036	0.0035	0.0	94	6
TMLE linear for (W, Z)	0.00007	0.0031	0.0027	0.0	91	9
TMLE logit for (W, Z)	0.00007	0.0035	0.0030	0.0	91	9
			Adaptive Pair-Matching			
Unadj.	0.00002	0.0034	0.0033	0.0	95	5
TMLE linear for Z	0.00006	0.0029	0.0028	0.0	95	5
TMLE logit for Z	0.00005	0.0030	0.0029	0.0	95	5
TMLE linear for (W, Z)	0.00004	0.0028	0.0026	0.0	94	6
TMLE logit for (W, Z)	0.00004	0.0030	0.0028	0.0	94	6

[a] average deviation between the point estimate and sample-specific true value
[b] square root of the variance of the point estimates
[c] average standard error estimate based on the influence curve
[d] average value of the test statistic (point estimate divided by standard error estimate)
[e] proportion of intervals containing the true parameter value (in percent)
[f] proportion of studies falsely rejecting the null hypothesis (in percent)

