Biometrika (2016), **103**, 1, *pp*. 1–13

Printed in Great Britain
Advance Access publication on 31 July 2016

Supplementary material for: 'Randomization tests of causal effects under interference'

BY G. W. BASSE
Department of Statistics, University of California, 418 Evans Hall, Berkeley, California 94720, U.S.A.
gbasse@berkeley.edu

A. FELLER
Goldman School of Public Policy, University of California, 2607 Hearst Avenue, Berkeley, California 94720, U.S.A.
afeller@berkeley.edu

AND P. TOULIS
Booth School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, Illinois 60637, U.S.A.
panos.toulis@chicagobooth.edu

1. PROOFS OF THEOREMS AND STATEMENTS

1·1. *Proof of validity of classical Fisher test*

We reproduce the proof of Hennessy et al. (2016) with slight modifications. This proof will provide an introduction to the proof of the validity of the conditional test that follows.

Proof. We need to show that:

$$\mathrm{pr}(p \leq \alpha \mid H_0) \leq \alpha, \text{ for all } \alpha \in [0,1],$$

where the probability is with respect to $\mathrm{pr}(Z^{\mathrm{obs}})$, and $p = \mathrm{pval}(Z^{\mathrm{obs}})$ is defined as

$$p = \mathrm{pr}\{T(Z \mid Y^{\mathrm{obs}}) \geq T(Z^{\mathrm{obs}} \mid Y^{\mathrm{obs}})\}.$$

Let U be a random variable with the same distribution as $T(Z \mid Y^{\mathrm{obs}})$, as induced by $\mathrm{pr}(Z)$ and let F_{U} be its cumulative distribution function. We can then write

$$p = 1 - F_{\mathrm{U}}\{T(Z^{\mathrm{obs}} \mid Y^{\mathrm{obs}})\}.$$

By definition, under H_0 we have $Y(Z) = Y(Z^{\mathrm{obs}})$ for all Z, and so $T(Z \mid Y^{\mathrm{obs}}) = T\{Z \mid Y(Z)\}$. It follows that, under H_0, U has the same distribution as $T(Z \mid Y^{\mathrm{obs}})$. The randomness in $T(Z^{\mathrm{obs}} \mid Y^{\mathrm{obs}})$ is induced by the randomness in Z^{obs}. In the testing procedure, $Z^{\mathrm{obs}} \sim \mathrm{pr}(Z^{\mathrm{obs}})$. Combining with the above, we see that the distribution of $T(Z^{\mathrm{obs}} \mid Y^{\mathrm{obs}})$ induced by $\mathrm{pr}(Z^{\mathrm{obs}})$ is the same as that of U under H_0. We thus have

$$p = 1 - F_{\mathrm{U}}(U).$$

By the probability integral transform theorem, p is uniform, and so $\mathrm{pr}(p \leq \alpha \mid H_0) \leq \alpha$. □

1·2. *Proof of Theorem* 1

The proof of Theorem 1 follows that of the classical Fisher test, with some important modifications.

THEOREM 1. *Let H_0 be a null hypothesis and $T(Z \mid Y, \mathcal{C})$ a test statistic, such that T is imputable with respect to H_0 under some conditioning mechanism $m(\mathcal{C} \mid Z)$; that is, under H_0, it holds that*

$$T\{Z' \mid Y(Z'), \mathcal{C}\} = T\{Z' \mid Y(Z), \mathcal{C}\}, \quad (1)$$

for all $Z, Z', \mathcal{C}$, for which $\mathrm{pr}(Z, \mathcal{C}; m) > 0$ and $\mathrm{pr}(Z', \mathcal{C}; m) > 0$. Consider the procedure where we first draw $\mathcal{C} \sim m(\mathcal{C} \mid Z^{\mathrm{obs}})$, and then compute the conditional p-value,

$$\mathrm{pval}(Z^{\mathrm{obs}}; \mathcal{C}) = E_Z[\mathbb{I}\{T(Z \mid Y^{\mathrm{obs}}, \mathcal{C}) > T^{\mathrm{obs}}\} \mid \mathcal{C}], \quad (2)$$

where $T^{\mathrm{obs}} = T(Z^{\mathrm{obs}} \mid Y^{\mathrm{obs}}, \mathcal{C})$, and the expectation is with respect to $\mathrm{pr}(Z \mid \mathcal{C}) = \mathrm{pr}(Z, \mathcal{C}; m)/\mathrm{pr}(\mathcal{C})$. This procedure is valid at any level, that is, $\mathrm{pr}\{\mathrm{pval}(Z^{\mathrm{obs}}; \mathcal{C}) \leq \alpha \mid \mathcal{C}\} \leq \alpha$, for any $\alpha \in [0, 1]$.

Proof. We need to show that

$$\mathrm{pr}(p_{\mathcal{C}} \leq \alpha \mid H_0, \mathcal{C}) \leq \alpha$$

for all $\mathcal{C}$ such that $\mathrm{pr}(\mathcal{C} \mid Z^{\mathrm{obs}}) > 0$, where the probability is with respect to $\mathrm{pr}(Z^{\mathrm{obs}} \mid \mathcal{C})$, and $p_{\mathcal{C}}$ is defined as

$$p_{\mathcal{C}} = \mathrm{pr}\{T(Z \mid Y^{\mathrm{obs}}, \mathcal{C}) \geq T(Z^{\mathrm{obs}} \mid Y^{\mathrm{obs}}, \mathcal{C}) \mid \mathcal{C}\}.$$

Fix $\mathcal{C}$. Let U be a random variable with the same distribution as $T(Z \mid Y^{\mathrm{obs}}, \mathcal{C})$ as induced by $\mathrm{pr}(Z \mid \mathcal{C})$ and let F_{U} be its cumulative distribution function. We can then write:

$$p_{\mathcal{C}} = 1 - F_{\mathrm{U}}\{T(Z^{\mathrm{obs}} \mid Y^{\mathrm{obs}}, \mathcal{C})\}.$$

In the procedure, we have $Z^{\mathrm{obs}} \sim \mathrm{pr}(Z^{\mathrm{obs}})$ and $\mathcal{C} \sim \mathrm{pr}(\mathcal{C} \mid Z^{\mathrm{obs}})$, implying that $\mathrm{pr}(Z^{\mathrm{obs}}, \mathcal{C}) > 0$. So, by imputatability of the test statistic in Equation (1) under H_0,

$$T\{Z \mid Y(Z), \mathcal{C}\} = T(Z \mid Y^{\mathrm{obs}}, \mathcal{C})$$

for all $Z \sim \mathrm{pr}(Z \mid \mathcal{C})$, since this guarantees $\mathrm{pr}(Z, \mathcal{C}) > 0$. This means that under H_0, U has the same distribution as $T(Z \mid Y^{\mathrm{obs}}, \mathcal{C})$. The randomness in $T(Z^{\mathrm{obs}} \mid Y^{\mathrm{obs}}, \mathcal{C})$ is induced by the randomness in Z^{obs} conditional on $\mathcal{C}$. Combining with the above, we see that the distribution of $T(Z^{\mathrm{obs}} \mid Y^{\mathrm{obs}}, \mathcal{C})$ induced by $\mathrm{pr}(Z^{\mathrm{obs}} \mid \mathcal{C})$ is the same as that of U under H_0. We thus have:

$$p_{\mathcal{C}} = 1 - F_{\mathrm{U}}(U).$$

By the probability integral transform theorem, $p_{\mathcal{C}}$ is uniform and so $\mathrm{pr}(p_{\mathcal{C}} \leq \alpha \mid H_0, \mathcal{C}) \leq \alpha$. □

1·3. *Proof of Theorem* 2

For the reader's convenience we repeat the definitions of the contrast null hypothesis, conditioning mechanism, and test statistic, which are used in Theorem 2:

$$H_0: \ Y_i(Z) = Y_i(Z'), i = 1, \ldots, N, \text{ for all } Z, Z' \text{ for which } h_i(Z), h_i(Z') \in \{a, b\}, \quad (3)$$

$$m(\mathcal{C} \mid Z) = f(\mathcal{U} \mid Z) g(\mathcal{Z} \mid \mathcal{U}, Z), \quad (4)$$

$$T(Z \mid Y, \mathcal{C}) = \mathrm{AVE}\{Y_i \mid i \in \mathcal{U}, h_i(Z) = a\} - \mathrm{AVE}\{Y_i \mid i \in \mathcal{U}, h_i(Z) = b\}, \quad (5)$$

where $\mathcal{C} = (\mathcal{U}, \mathcal{Z})$, and $\mathcal{U}, \mathcal{Z}$ are any subsets of units and assignment vectors, respectively and AVE denotes the average. The main challenge is to prove that the conditions of the theorem ensure that the test statistic in Equation (5) is imputable under H_0.

THEOREM 2. *Let H_0 be a null hypothesis as in Equation* (3)*, $m(\mathcal{C} \mid Z)$ be a conditioning mechanism as in Equation* (4)*, and T be a test statistic defined only on focal units, as in Equation* (5)*. Then, T is imputable under H_0 if $m(\mathcal{C} \mid Z) > 0$ implies that $Z \in \mathcal{Z}$, and for every $i \in \mathcal{U}$ and $Z' \in \mathcal{Z}$ that*

$$h_i(Z') \in \{a, b\}, \text{ if } h_i(Z) \in \{a, b\}, \quad (6)$$

$$h_i(Z') = h_i(Z), \text{ if } h_i(Z) \notin \{a, b\}. \quad (7)$$

If T is imputable the randomization test for H_0 as described in Theorem 1 *is valid at any level α.*

Proof. For a conditioning event $\mathcal{C} = (\mathcal{U}, \mathcal{Z})$, suppose that $m(\mathcal{C} \mid Z) > 0$ implies that $Z \in \mathcal{Z}$ and that:

$$\text{for all } i \in \mathcal{U}, Z' \in \mathcal{Z}, \quad \begin{cases} h_i(Z') \in \{a, b\} & \text{if } h_i(Z) \in \{a, b\}, \\ h_i(Z') = h_i(Z) & \text{if } h_i(Z) \notin \{a, b\}. \end{cases}$$

Now let $Z, Z', \mathcal{C}$ be such that $\mathrm{pr}(Z, \mathcal{C}; m) > 0$ and $\mathrm{pr}(Z', \mathcal{C}; m) > 0$. By definition of a conditioning mechanism, this implies that $m(\mathcal{C} \mid Z) > 0$ and $m(C \mid Z') > 0$. It follows that $Z \in \mathcal{Z}$ and $Z' \in \mathcal{Z}$. Now take $i \in \mathcal{U}$. If $h_i(Z') \notin \{a, b\}$, then, by assumption, $h_i(Z) = h_i(Z')$ since $Z, Z' \in \mathcal{Z}$. And so by Equation (5) of the main paper, we have that $Y_i(Z') = Y_i(Z)$. If instead $h_i(Z') \in \{a, b\}$, then $h_i(Z) \in \{a, b\}$ and so under the null hypothesis $Y_i(Z') = Y_i(Z)$, as well. Therefore, we proved that $Y_{\mathcal{U}}(Z') = Y_{\mathcal{U}}(Z)$, where $Y_{\mathcal{U}}(Z)$ denotes the subvector of outcomes of units in $\mathcal{U}$ under assignment vector Z. Since the test statistic, $T(Z \mid Y, \mathcal{C})$, is defined only on $Y_{\mathcal{U}}$, the subvector of outcomes of units in $\mathcal{U}$, it follows that $T\{Z' \mid Y(Z'), \mathcal{C}\} = T\{Z' \mid Y(Z), \mathcal{C}\}$, and so T is imputable. □

1·4. *Proof of Proposition 1*

PROPOSITION 1. *Consider the following testing procedure:*

1. *In control households ($W_j = 0$), choose one unit at random. In treated households ($W_j = 1$), choose one unit at random among the non-treated units ($Z_i = 0$).*
2. *Compute the distribution of the test statistic in Equation* (5) *induced by all permutations of exposures on the chosen focal units, using $a = (0, 0)$ and $b = (0, 1)$ as the contrasted exposures.*
3. *Compute the p-value.*

Steps 1-3 outline a procedure that is valid for testing the null hypothesis of no spillover effect, H_0^s.

Proof. Define

$$\mathbb{U}(Z) = \{\mathcal{U} \in \mathbb{U} \,:\, Z_i \mathbb{I}(i \in \mathcal{U}) = 0, \; i = 1, \ldots, N, \text{ and } \sum_i \mathbb{I}(i \in \mathcal{U}) R_{ij} = 1, \text{for every household } j\}.$$

In words, $\mathbb{U}(Z)$ is the set of all subsets of units for which no unit in the subset is treated under Z, and each household has exactly one unit in the subset. Step 1 of the procedure in Proposition 1 chooses focals according to conditioning mechanism $m(\mathcal{C} \mid Z) = f(\mathcal{U} \mid Z) g(\mathcal{Z} \mid \mathcal{U}, Z)$, where we define

$$f(\mathcal{U} \mid Z) = \mathrm{Unif}\{\mathbb{U}(Z)\}, \tag{8}$$

$$g(\mathcal{Z} \mid \mathcal{U}, Z) = \mathbb{I}[\mathcal{Z} = \{Z' : h_i(Z') \in \{(0, 0), (0, 1)\} \text{ for all } i \in \mathcal{U}\}]. \tag{9}$$

That is, $f(\mathcal{U} \mid Z)$ is uniform on $\mathbb{U}(Z)$ and g is degenerate on the set of assignments for which all units in $\mathcal{U}$ are either in control or exposed to spillovers. In what follows, we fix a conditioning event $\mathcal{C} = (\mathcal{U}, \mathcal{Z})$.

Let $H = H(Z) \in \{0, 1\}^K$ denote the exposure of focal units under Z, where we use 0 for control and 1 for spillovers. Also, let $W = W(Z) \in \{0, 1\}^K$ denote the household assignment under assignment vector Z. Since there is one focal per household and household assignment determines the exposure of a focal, H and W are equal almost surely:

$$H(Z) = W(Z), \text{ for all } Z, \text{ and so we can write } H = W, \text{ almost surely.}$$

For any $Z, Z' \in \mathcal{Z}$, it holds that

$$g(\mathcal{Z} \mid \mathcal{U}, Z) = g(\mathcal{Z} \mid \mathcal{U}, Z').$$

This follows from definition of g in Equation (9) since $g(\mathcal{Z} \mid \mathcal{U}, Z) \equiv g(\mathcal{Z} \mid \mathcal{U})$ does not depend on Z given a fixed $\mathcal{U}$; note that $\mathcal{U}$ depends on Z itself, but still g does not depend on Z if $\mathcal{U}$ is given.

For any $w \in \{0, 1\}^K$, it holds that:

$$\sum_{Z : W(Z) = w} f(\mathcal{U} \mid Z) \mathrm{pr}(Z \mid W = w) = \mathrm{const}.$$

To see this, first note that $\mathrm{pr}(Z \mid W) = \prod_{k:W_k=1}^{K} 1/n_k$, where n_k is the number of units in the household. Furthermore,

$$\sum_{Z:W(Z)=w} f(\mathcal{U} \mid Z) = \prod_{k:W_k=0} 1/n_k.$$

Therefore,

$$\sum_{Z:W(Z)=w} f(\mathcal{U} \mid Z)\mathrm{pr}(Z \mid W = w) = \prod_{k} 1/n_k = \text{const.}$$

Actually this is equal to the marginal probability of the focal set, $\mathrm{pr}(\mathcal{U})$.

We now put things together and prove that the conditioning mechanism yields a randomization distribution that is uniform in its support. Fix a conditioning event $\mathcal{C} = (\mathcal{U}, \mathcal{Z})$. Then,

$$\begin{aligned}
\mathrm{pr}(H \mid \mathcal{C}) &= \mathrm{pr}(W \mid \mathcal{C}) \text{ [\textit{from Step 1}]} \\
&\propto \mathrm{pr}(\mathcal{C} \mid W)\mathrm{pr}(W) \\
&\propto \sum_{Z} \mathrm{pr}(\mathcal{C}, Z \mid W)\mathrm{pr}(W) \\
&\propto \sum_{Z:W(Z)=W} \mathrm{pr}(\mathcal{C} \mid Z)\mathrm{pr}(Z \mid W)\mathrm{pr}(W) \\
&\propto \sum_{Z:W(Z)=W} f(\mathcal{U} \mid Z)g(\mathcal{Z} \mid \mathcal{U}, Z)\mathrm{pr}(Z \mid W)\mathrm{pr}(W) \qquad (10) \\
&\propto g(\mathcal{Z} \mid \mathcal{U})\mathrm{pr}(W) \sum_{Z:W(Z)=W} f(\mathcal{U} \mid Z)\mathrm{pr}(Z \mid W) \\
&\propto \mathrm{pr}(W) \\
&= \binom{N}{N_1}^{-1}. \qquad (11)
\end{aligned}$$

From the definition of the test statistic:

$$T(Z \mid Y, \mathcal{C}) = T(Z' \mid Y, \mathcal{C}) \quad \text{if} \quad H(Z) = H(Z').$$

Therefore, we can write $T(Z \mid Y, \mathcal{C}) \equiv T(H \mid Y, \mathcal{C})$. From the above, we know that the conditional distribution of the focals' exposure under the particular conditioning mechanism is a permutation of their exposures under Z^{obs}, as prescribed by the testing procedure of Proposition 1. This is sufficient for validity since the test statistic is in fact a function of H.

2. ADDITIONAL DISCUSSION OF ALTERNATIVE METHODS

2·1. *Equivalence of tests from Athey et al. (2017) and Aronow (2012) for two-stage designs*

The tests described by Athey et al. (2017) and Aronow (2012) coincide for testing spillover effects, H_0^s, in our two-stage randomized setting. We will show that the method of Aronow (2012) is equivalent to our procedure, with $f(\mathcal{U} \mid Z) = f(\mathcal{U})$. Briefly, the method of Aronow (2012) can be summarized as follows:

1. Draw a set of units $\mathcal{U} \subset \mathbb{U}$, uniformly at random, as in Athey et al. (2017).
2. Compute the p-value by using the conditional randomization distribution $\mathrm{pr}(Z \mid \mathcal{U}, Z_{\mathcal{U}} = Z_{\mathcal{U}}^{\mathrm{obs}})$, where $Z_{\mathcal{U}}$ is the subvector of Z that is restricted to the units in $\mathcal{U}$.

The conditional randomization distribution is therefore equal to:

$$\begin{aligned}
\mathrm{pr}(Z \mid \mathcal{U}, Z_{\mathcal{U}} = Z_{\mathcal{U}}^{\mathrm{obs}}) &\propto \mathrm{pr}(\mathcal{U}, Z_{\mathcal{U}} = Z_{\mathcal{U}}^{\mathrm{obs}} \mid Z)\mathrm{pr}(Z) \propto \mathrm{pr}(Z_{\mathcal{U}} = Z_{\mathcal{U}}^{\mathrm{obs}} \mid \mathcal{U}, Z)\mathrm{pr}(\mathcal{U} \mid Z)\mathrm{pr}(Z) \\
&= \mathbb{I}(Z_{\mathcal{U}} = Z_{\mathcal{U}}^{\mathrm{obs}})\mathrm{pr}(\mathcal{U})\mathrm{pr}(Z).
\end{aligned}$$

Now, consider a conditioning event $\mathcal{C} = (\mathcal{U}, \mathcal{Z})$ from a mechanism $m_f(\mathcal{C} \mid Z) = f(\mathcal{U})g(\mathcal{Z} \mid \mathcal{U}, Z)$, where according to Equation (11) in the main paper is degenerate on the set:

$$\mathcal{Z} = [Z' : h_i(Z') = (1,1) \text{ if } h_i(Z^{\text{obs}}) = (1,1) \text{ and } h_i(Z') \in \{(0,0),(0,1)\} \text{ otherwise, for all } i \in \mathcal{U}]. \quad (12)$$

Under this definition and the setting of spillover effects, for every unit $i \in \mathcal{U}$ in the focal set and every assignment vector $Z' \in \mathcal{Z}$ in the test, we will have either $Z'_i = 0$ if $Z^{\text{obs}}_i = 0$ or $Z'_i = 1$ if $Z^{\text{obs}}_i = 1$. Thus, if $Z' \in \mathcal{Z}$ it follows that $Z'_{\mathcal{U}} = Z^{\text{obs}}_{\mathcal{U}}$. Suppose the reverse is true, that is, $Z'_{\mathcal{U}} = Z^{\text{obs}}_{\mathcal{U}}$. Consider unit i in the focal set for which $Z^{\text{obs}}_i = 1$. Then, $Z'_i = 1$ as well, and so $h_i(Z') = h_i(Z^{\text{obs}})$ for such units. Consider unit i in the focal set for which $Z^{\text{obs}}_i = 0$. Then, $Z'_i = 0$ as well, and so $h_i(Z') =\in \{(0,0),(0,1)\}$, by definition of exposures. Thus, if $Z'_{\mathcal{U}} = Z^{\text{obs}}_{\mathcal{U}}$ it follows that $Z' \in \mathcal{Z}$. Therefore, the two statements are equivalent, and the conditioning mechanism with $f(\mathcal{U} \mid Z) = f(\mathcal{U})$ will yield the same test as in Athey et al. (2017) and Aronow (2012).

2·2. *When the test of Athey et al. (2017) is a permutation test*

The method of Athey et al. can be cast in our framework, where $f(\mathcal{U} \mid Z) = f(\mathcal{U})$, i.e., the selection of focals does not depend on the observed assignment, and where the randomization distribution, $\text{pr}(Z \mid \mathcal{C})$, is uniform over the set $\mathcal{Z}$ defined in Equation (12). We denote by $\mathcal{U}^{\text{eff}}(Z) = \{i \in \mathcal{U} : Z_i = 0\}$. We denote by $H_i = h_i(Z)$ the exposure of unit i under assignment vector Z.

First, notice that $\mathcal{U}^{\text{eff}}(Z) = \mathcal{U}^{\text{eff}}(Z^{\text{obs}})$, for every $Z \in \mathcal{Z}$. Now, consider unit $i \in \mathcal{U}^{\text{eff}}(Z^{\text{obs}})$. We have:

$$\begin{aligned}
\text{pr}(H_i = (1,0) \mid Z \in \mathcal{Z}) &= \text{pr}(Z_i = 0, W_{[i]} = 1 \mid Z \in \mathcal{Z}) \\
&= \frac{\text{pr}(Z_i = 0 \mid W_{[i]} = 1)\text{pr}(W_{[i]} = 1)}{\text{pr}(Z \in \mathcal{Z})} \\
&= \frac{(n_i - 1)/n_i\binom{N}{N_1}}{\text{pr}(Z \in \mathcal{Z})} \\
&\propto \frac{n_i - 1}{n_i}.
\end{aligned}$$

We thus have the constraint that for all $Z \in \mathcal{Z}$:

$$\sum_{i \in \mathcal{U}^{\text{eff}}(Z)} \mathbb{I}\{H_i(Z) = (1,0)\} = N^{\text{eff}}_1(Z^{\text{obs}}).$$

In words, the number of exposed units is constant for all $Z \in \mathcal{Z}$. Putting it all together, we see that $P(H \mid Z \in \mathcal{Z})$ is such that:

1. $\sum_{i \in \mathcal{U}^{\text{eff}}} \mathbb{I}\{H_i = (1,0)\} = N^{\text{eff}}_1$.
2. For all $i \in \mathcal{U}^{\text{eff}}$, $\text{pr}\{H_i = (1,0) \mid Z \in \mathcal{Z}(\mathcal{U}, Z^{\text{obs}})\} \propto (n_i - 1)/n_i$. □

This result implies that the method of Athey et al. (2017) can be implemented as a permutation test only when the households are of equal sizes. This is not true in our application, and not expected to be true more generally, and thus poses computational challenges in implementing the test of Athey et al. (2017).

3. Simulations and analysis details

3·1. *Simulations*

We compare the power of the test we proposed in the previous section, which chooses the focal units conditionally on Z^{obs}, to that of the test in Athey et al. (2017) which chooses the focals unconditionally of Z^{obs}. We use the term "unconditional focals" to describe that approach, but we note that this could encompass selection of focals based on existing covariate information, such as a network between units. For example, Athey et al. (2017) propose an approach where after a unit is selected as focal subsequent focal units are selected beyond a certain distance to the initial focal unit; this is known as the ϵ-net approach. Such approaches are still unconditional to the observed treatment assignment, Z^{obs}.

Figure 1 illustrates the potential power gains, by considering the extreme case of $K = 500$ households of equal size $n = 50$ with $K_1 = 250$ treated households, and focusing on the power of the test of no primary effect H_0^p. If we are interested in testing the no spillover effect hypothesis H_0^s, the expected difference in the number of effective focal units between our test and the test of Athey et al. (2017) decreases with n. In the case of the no primary effect hypothesis H_0^p, the difference increases with n. This phenomenon is illustrated in Figure 2.

3·2. *Details of analysis: covariate adjustment*

In all the analyses in the paper, covariates where taken into account via the same model-assisted approach used in Section 7 and Section 9.2 of Basse & Feller (2017). Briefly, we use a holdout set to estimate the parameter of a regression, then we use those estimators parameters to obtain predicted values $\{\hat{Y}_i\}_i$ for the outcomes in our sample and compute the residuals $\hat{e}_i = Y_i^{\text{obs}} - \hat{Y}_i$. We then apply the conditional testing methodology to the residuals, instead of the original potential outcomes; in that way, the residuals can be thought of as transformed outcomes. Note that this approach is similar to that used by Rosenbaum et al. (2002).

3·3. *Details of analysis: confidence intervals*

We ran an additional analysis comparing the size of confidence intervals for our method and for that of Athey et al. (2017). Specifically, for each of H_0^s and H_0^p, we drew 100 focal sets using our method, and 100 using the method of Athey et al. (2017), and computed the associated confidence intervals, obtained by inverting sequences of Fisher randomization tests (Rosenbaum et al., 2002). Figure 3 summarizes the results. We see that our method leads to smaller confidence intervals compared to the method of Athey et al. (2017), and that the difference is larger for the primary effect than for the spillover effect.

3·4. *Details of analysis: point estimates*

Point estimates are obtained using a variant of the Hodges-Lehmann estimator (Hodges Jr & Lehmann, 1963). Specifically, for a conditioning event $\mathcal{C}$, we numerically solve the equation $E(T \mid \mathcal{C}, H_\tau^P) = T^{\text{obs}}$, where H_τ^P is the null hypothesis $Y_i(1,1) = Y_i(0,0) + \tau$, by considering a grid of values for τ, and computing the expectation of the null distribution of T under the hypothesis H_τ^P and keeping the value $\hat{\tau}$ of τ that is closest to T^{obs}.

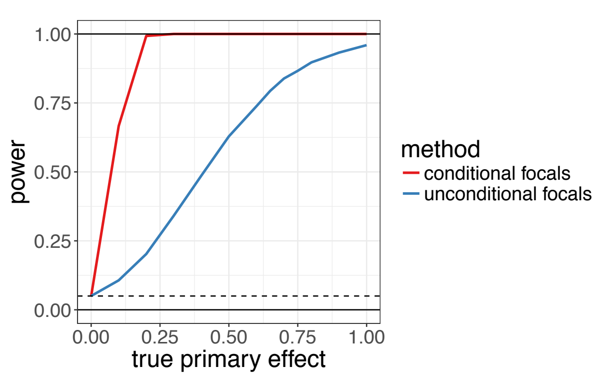

Fig. 1: Power of the test of no primary effect obtained with choice of focals unconditional to the observed assignment (Athey et al. (2017)) versus conditional choice, for different true values of the true primary effect.

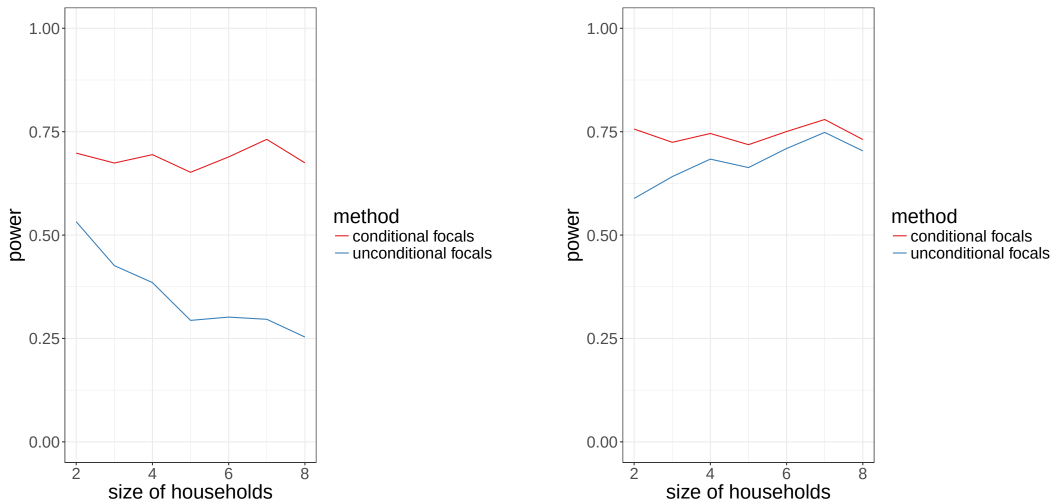

Fig. 2: Power of the two methods for testing the null hypotheses of no primary effect, on the left, and no spillover effect, on the right, as a function of household size n_i.

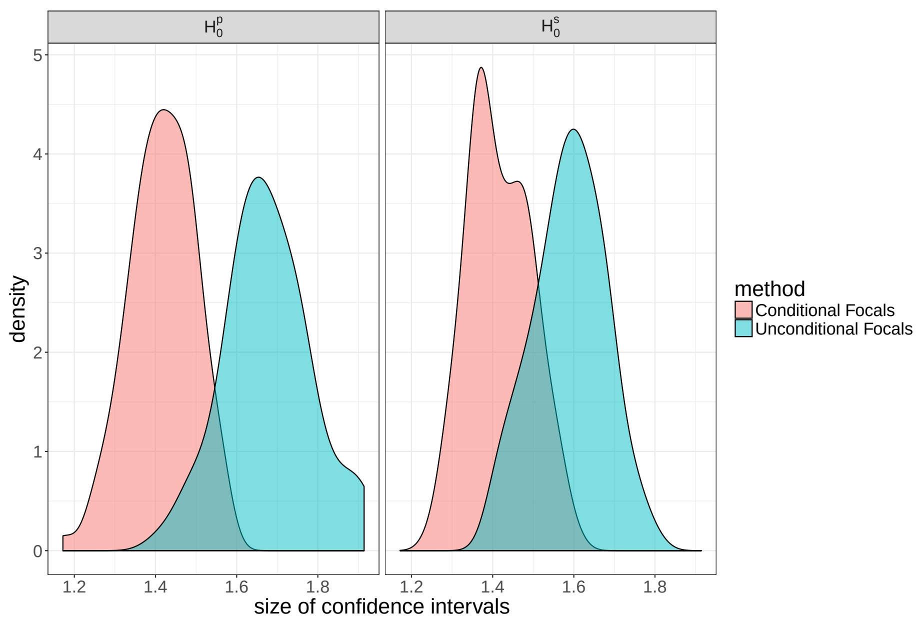

Fig. 3: Size of the confidence intervals for the primary and spillover effects obtained by the two methods.

3·5. *Details of analysis: results for testing H_0^P* 195

The median value of the Hodges-Lehmann for the primary effect is approximately equal for both choices of functions f and is approximately equal to -1.5 days, with associated confidence interval $[-2.2, 0.75]$ for our method, and $[-2.3, -0.8]$ for the method of Athey et al. (2017). The average length of confidence intervals obtained with our method is 1.4 days, versus 1.6 days for the method of Athey et al. (2017). The fraction of focals leading to a p-value below 0.05 is 100% in our case, based on a Monte-Carlo estimate from 100 replications, versus 92% for the method in Athey et al. (2017).

4. COMPARISON OF POWERS OF TESTS

4·1. *Model, p-values and power*

In this section, we make an approximate theoretical analysis of the power of our test and the power of the test by Athey et al. (2017). Our analysis is performed under two approximations. First, in the context of classical Fisher randomization tests, we argue that, in general, tests that are balanced and use more units are more powerful. So, balance and size of treatment arms can be used as a proxy for the power of the test. Second, we argue that since in the two-stage randomization case, our test and the test in Athey et al. (2017) can be conceived as classical Fisher randomization tests run on the focal units, the aforementioned power approximation for the classical Fisher randomization test applies.

Consider a classical Fisher randomization test, with complete randomization where N_1 out of N units are treated. Let $p = N_1/N$. Suppose that that the true effect is constant additive τ, and that we test for the null of no effect H_0. In order to give concrete analytical heuristics, we consider a model for the potential outcomes and focus on asymptotics; see also Lehmann & Romano (2006) for this approach:

$$Y_i(Z_i) \sim \tau Z_i + \mathcal{N}(\mu, \sigma^2).$$

As mentioned, we will focus our argument on asymptotic heuristics. Denote by $V = var(T \mid Y^{\text{obs}}, H_0)$ the randomization variance of the test statistics conditional on Y^{obs}, and assuming H_0 is true. We have, for large N:

$$V = \frac{1}{N}\left[\frac{\sigma^2}{p(1-p)} + \tau^2\right].$$

Denote by V^{obs} the variance of the test statistic $V^{\text{obs}} = var(T)$. We have, for large N,

$$V^{\text{obs}} = V - \frac{\tau^2}{N},$$

and so by applying the appropriate CLT's, we have:

$$\frac{T}{V^{1/2}} \approx \mathcal{N}(0,1), \qquad \frac{T^{\text{obs}} - \tau}{(V^{\text{obs}})^{1/2}} \approx \mathcal{N}(0,1).$$

Note the application of the CLT is heuristic here, and some regularity conditions are required. We can then obtain an approximation of the distribution of a one-sided p-value for large N:

$$\begin{aligned} \text{pval} &= \text{pr}(T \geq T^{\text{obs}}) \\ &\approx 1 - \Phi\left(\frac{T^{\text{obs}}}{V^{1/2}}\right), \end{aligned}$$

using the asymptotics from above. We can then verify that:

$$\begin{aligned} \frac{T^{\text{obs}}}{V^{1/2}} &= \frac{T^{\text{obs}} - \tau}{(V^{\text{obs}})^{1/2}}(1-C)^{1/2} + (NC)^{1/2} \\ &\approx W(1-C)^{1/2} + (NC)^{1/2}, \end{aligned}$$

where $W \sim \mathcal{N}(0,1)$ and $C = \tau^2[\sigma^2/\{p(1-p)\} + \tau^2]^{-1}$ and so :

$$\text{pval} = 1 - \Phi(W(1-C)^{1/2} + (NC)^{1/2}). \tag{13}$$

We can use the approximation of Equation (13) to deal with the power. For $\alpha \in [0,1]$, the power of the test at level α will be

$$\beta_\alpha = \text{pr}(\text{pval} \leq \alpha),$$

but we verify that

$$\text{pval} \le \alpha \quad \Leftrightarrow \quad W \ge \frac{\Phi^{-1}(1-\alpha), -(NC)^{1/2}}{(1-C)^{1/2}}$$

and so the power of the test will be approximately:

$$\beta_\alpha = 1 - \Phi\left(\frac{\Phi^{-1}(1-\alpha) - (NC)^{1/2}}{(1-C)^{1/2}}\right). \tag{14}$$

4·2. *Comparing classical tests*

We are interested in comparing tests with different proportions p of treated units, and with different numbers N of units. We will denote these quantities by $N^{(1)}$ and $N^{(2)}$ for the number of units, and $p^{(1)}$ and $p^{(2)}$ for the proportions. Let $\beta^{(1)}$ and $\beta^{(2)}$ be the associated powers. Finally, notice that:

$$\begin{aligned}\beta^{(1)} \le \beta^{(2)} &\Leftrightarrow \frac{\Phi^{-1}(1-\alpha) - (N^{(1)}C^{(1)})^{1/2}}{(1-C^{(1)})^{1/2}} \ge \frac{\Phi^{-1}(1-\alpha) - (N^{(2)}C^{(2)})^{1/2}}{(1-C^{(2)})^{1/2}} \\ &\Leftrightarrow \gamma^{(1)} \ge \gamma^{(2)}\end{aligned}$$

where $\gamma^{(1)} = \{\Phi^{-1}(1-\alpha) - (N^{(1)}C^{(1)})^{1/2}\}/\{(1-C^{(1)})^{1/2}\}$

Suppose that both tests have the same number of units $N^{(1)} = N^{(2)} = N$, but different fractions of treated units $p^{(1)} \ne p^{(2)}$. We have

$$\begin{aligned}\gamma^{(1)} - \gamma^{(2)} &= N^{1/2}\left(\frac{(C^{(2)})^{1/2}}{1-(C^{(2)})^{1/2}} - \frac{(C^{(1)})^{1/2}}{1-(C^{(1)})^{1/2}}\right) + \left(\frac{\Phi^{-1}(1-\alpha)}{(1-C^{(1)})^{1/2}} - \frac{\Phi^{-1}(1-\alpha)}{(1-C^{(2)})^{1/2}}\right) \\ &\to N^{1/2}\left(\frac{(C^{(2)})^{1/2}}{1-(C^{(2)})^{1/2}} - \frac{(C^{(1)})^{1/2}}{1-(C^{(1)})^{1/2}}\right)\end{aligned}$$

and so for large N,

$$\begin{aligned}\gamma^{(1)} - \gamma^{(2)} \ge 0 &\Leftrightarrow \frac{(C^{(2)})^{1/2}}{1-(C^{(2)})^{1/2}} - \frac{(C^{(1)})^{1/2}}{1-(C^{(1)})^{1/2}} \ge 0 \\ &\Leftrightarrow p^{(1)}(1-p^{(1)}) \le p^{(2)}(1-p^{(2)}) \\ &\Leftrightarrow |p^{(1)} - \frac{1}{2}| \ge |p^{(2)} - \frac{1}{2}|.\end{aligned}$$

So in conclusion:

$$\beta^{(1)} \le \beta^{(2)} \quad \Leftrightarrow \quad |p^{(1)} - \frac{1}{2}| \ge |p^{(2)} - \frac{1}{2}|$$

which, in words, means that the balanced test has more power asymptotically.

Suppose that $N^{(1)} \ne N^{(2)}$ but that the fractions of treated units in each test is identical. That is, $p^{(1)} = p^{(2)} = p$. The immediate consequence is that $C^{(1)} = C^{(2)} = C$, and so:

$$\gamma^{(1)} - \gamma^{(2)} = \frac{C^{1/2}}{1-C^{1/2}}\left((N^{(2)})^{1/2} - (N^{(1)})^{1/2}\right),$$

and so:

$$\beta^{(1)} \le \beta^{(2)} \quad \Leftrightarrow N^{(1)} \le N^{(2)},$$

which in words means that the test with more units has more power asymptotically.

4·3. *Comparing the power of our test with that of Athey et al. (2017)*

If we restrict our attention to the special case where all households have equal size $n_i = n$, then both our method and the method of Athey et al. (2017) can be seen as classical Fisher randomization tests applied on a set of ”effective” focal units, where the set of ”effective focals” is always at least as large with our method as in the method of Athey et al. (2017), and is always balanced if the initial assignment $\mathrm{pr}(Z)$ is balanced. We can then leverage the result of Section 4·2 to argue heuristically that for classical Fisher randomization tests, larger and more balanced is generally better, and so we expect our method to lead to more powerful test. This has been confirmed in the simulations of Section 3·1 and in the analysis.

4·4. *Comparison with unconditional focal selection under a different design*

In this section, we perform an analysis outside of the two-stage design setting to illustrate the generality of our framework. We assume there is a network between units such that $\mathcal{N}_i$ denotes the neighborhood of unit i. As in the two-stage setting, we will show that being able to condition on the observed treatment assignment, which is possible in our framework, can lead to better randomization tests.

We consider a network between units and the following exposure functions:

$$h_i(Z) = \begin{cases} a & \text{if} \quad Z_i = 1, \\ b & \text{if} \quad Z_i = 0, \ \sum_{j \in \mathcal{N}_i} Z_j < d, \\ c & \text{if} \quad Z_i = 0, \ \sum_{j \in \mathcal{N}_i} Z_j > d, \end{cases}$$

and assume that N_1 units are treated completely at random in the network, and that we wish to test the null hypothesis:

$$H_0 : Y_i(Z) = Y_i(Z'), i = 1, \ldots, N, \quad \text{for all } Z, Z' : h_i(Z), h_i(Z') \in \{b, c\}.$$

This example is very different from the two-stage randomization setting considered in the main text, but there is one commonality: the units who received treatment are useless for testing H_0, and so it is wasteful to include them in the focal set. It is easy to verify that if focals are chosen completely at random, the distribution of the effective number of focals is $|\text{EFF}(\mathcal{U})| \sim M - \text{Hypergeom}(N, N_1, M)$, and so the expected number of focal units is $E\{|\text{EFF}(\mathcal{U})|\} = M - M(N_1/N)$. In the case where half the units are treated, that is $N_1 = N/2$, we have:

$$E\{|\text{EFF}(\mathcal{U})|\} = \frac{M}{2},$$

so in effect we lose half of the focal units. Choosing focals unconditionally but based on ϵ-nets would be better than choosing the focals completely at random but would not solve the fundamental reason why power is lost. Moreover, if choosing focals based on ϵ-nets is helpful, then it could always be combined with conditioning on the observed assignment to yield an even more powerful test.

To illustrate our framework in this setting, we could use following procedure:

1. Draw Z, completely at random with N_1 treated units, and N_0 control units.
2. Choose M focal units at random among the N_0 units with $Z_i = 0$. Let $\mathcal{U}$ be the set of focal units.
3. Draw $Z' \sim \mathrm{pr}(Z' \mid \mathcal{U})$ as follows. Set $Z'_i = 0$ for all $i \in \mathcal{U}$. Then choose $N_0 - M$ units at random among the $N - M$ non-focal units, and set $Z'_i = 0$ for these units. Finally, set $Z_i = 1$ for the remaining N_1 units.

We claim that the abovementioned procedure in Step 3 samples indeed from the correct conditional randomization distribution.

Proof. By definition of the procedure in Steps 2 and 3, it holds that $\mathrm{pr}(Z') \propto 1$ if $\sum_i Z'_i = N_1$, and also $\mathrm{pr}(\mathcal{U} \mid Z') = \text{const.}$, if $|\mathcal{U}| = M$ and $Z'_i = 0$ for every $i \in \mathcal{U}$. Therefore, $\mathrm{pr}(Z' \mid \mathcal{U}) = \text{Unif}(\mathcal{Z}(\mathcal{U}))$, where $\mathcal{Z}(\mathcal{U}) = \{Z : \text{for all } i \in \mathcal{U}, Z_i = 0 \ \text{ and } \ \sum_i Z_i = N_1\}$. Which is what step 3 does. □

Note that in this case our approach does not lead to a permutation test; and neither does the method of Athey et al. Nevertheless, it leads to a procedure that is easily implementable and that uses more information than that of Athey et al.

5. TESTING THE NULL HYPOTHESIS OF NO PRIMARY EFFECT

The paper focused on testing the null hypothesis of no spillover effects H_0^S. In this section, we briefly give equivalent results for testing the null hypothesis of no primary effect H_0^P. We omit the proofs, since they follow exactly the same outlines as the proof for H_0^S. A simple choice of f function for testing the null hypothesis of no primary effect is

$$f(\mathcal{U} \mid Z) = \text{Unif}\{\mathbb{U}^{(P)}(Z)\},$$

where

$$\mathbb{U}^{(P)}(Z) = \{\mathcal{U} \in \mathbb{U} : Z_i = 1 \Rightarrow i \in \mathcal{U}, \text{ for all } i \in \mathcal{U} \quad \text{and} \quad \sum_i \mathbb{I}(i \in \mathcal{U}) R_{ij} = 1, \text{for every household } j\}.$$

If applied to Theorem 2, this choice of f leads to the following procedure, which mirrors that of Proposition 1:

1. In control households, $W_j = 0$, choose one unit at random. In treated households, $W_j = 1$, choose the treated unit as focal.
2. Compute the distribution of the test statistic Equation (5) induced by all permutations of exposures on focal units, using $a = (0, 0)$ and $b = (1, 1)$ as the contrasted exposures.
3. Compute the p-value.

This procedure is valid conditionally and marginally for testing H_0^P.

6. ADDITIONAL NOTES ON THE CHOICE OF EXPOSURE MAPPING $h()$

6·1. *More complex exposure mappings*

The class of null hypotheses that our method is designed to test is summarized in Equation (7) of our manuscript, reproduced below for convenience:

$$H_0 : Y_i(Z) = Y_i(Z'), (i = 1, \ldots, N) \text{ for all } Z, Z' \text{ for which } h_i(Z), h_i(Z') \in \{a, b\}, \qquad (15)$$

for some exposure function h, the choice of which is limited by a few theoretical and practical considerations. The only strong theoretical constraint implicit in Equation (7) of the manuscript is that the two exposures a and b being contrasted must be well defined for all units under consideration. For instance, in the test of no spillovers H_0^s, the two exposures contrasted are the spillover exposure $(1, 0)$, and the control exposure $(0, 0)$, which are well defined for all units. If we had households with a single individual, then the exposure $(1, 0)$ would not be defined for that unit and the null hypothesis of Equation (7) would consequently be ill-posed if it included that unit.

Still, the formulation in Equation (15) provides enough flexibility to test a wide variety of null hypotheses. Here, we illustrate with a couple of short but representative examples on network interference. Similar to Athey et al. (2017), let $G_{ij} = 1$ if units i and j are neighbors in the network, and $G_{ij} = 0$ otherwise. By convention, $G_{ii} = 0$ for all i.

Suppose we want to test spillovers on control units from first-order neighbors. Then, we could define:

$$h_i(Z) = \begin{cases} a & \text{if } Z_i = 1, \\ b & \text{if } Z_i = 0, \ \sum_j G_{ij} Z_j > 0, \\ c & \text{if } Z_i = 0, \ \sum_j G_{ij} Z_j = 0. \end{cases}$$

Now testing the hypothesis in Equation (15) contrasting the exposures b and c defined above will test whether there are spillovers on control units.

As another example, suppose we want to test spillovers on control units from up to second-order neighbors. Let $H_{ij} = 1$ if i and j are second-order neighbors but not first-order neighbors, so $G_{ij} = 0$. Then, we could define:

$$h_i(Z) = \begin{cases} a & \text{if } Z_i = 1, \\ b & \text{if } Z_i = 0, \ \sum_j (G_{ij} + H_{ij}) Z_j > 0, \\ c & \text{if } Z_i = 0, \ \sum_j (G_{ij} + H_{ij}) Z_j = 0. \end{cases}$$

Now testing the hypothesis in Equation (15) contrasting the exposures b and c defined above will test whether there are spillovers on control units from first-order or second-order neighbors. We could also test the hypothesis that there are no second-order spillovers without putting constraints on first-order spillovers. For that test, we could define:

$$h_i(Z) = \begin{cases} a & \text{if } Z_i = 1, \\ b & \text{if } Z_i = 0, \ \sum_j G_{ij} Z_j > 0, \\ c & \text{if } Z_i = 0, \ \sum_j G_{ij} Z_j = 0, \ \sum_j H_{ij} Z_j > 0, \\ d & \text{if } Z_i = 0, \ \sum_j (G_{ij} + H_{ij}) Z_j = 0. \end{cases}$$

Now testing the hypothesis in Equation (15) contrasting the exposures c and d defined above will test whether there are spillovers on control units from second-order neighbors only. We can follow similar approaches for testing higher than second-order spillovers.

We now consider an example closer to the scenario of our application. Consider the same design as in our manuscript, but assume that all households have $n = 3$ units. We are interested in testing whether an untreated unit in a treated household receives a different spillover if the eldest of its two siblings is treated compared to the spillover received if the youngest of its two siblings is treated.

In order to test this null hypothesis, we need to consider a more complex exposure mapping than the one in our manuscript. Let $E_i \in \{0, 1\}$ be the treatment assignment of the eldest of unit i's two siblings, and consider the exposure mapping:

$$h_i(Z) = (H_i, Z_i, E_i).$$

Each unit now has four potential outcomes:

$$Y_i(Z) \in \{Y_i(1,1,0), Y_i(0,0,0), Y_i(1,0,1), Y_i(1,0,0)\},$$

the other combinations being impossible. With this exposure mapping the null hypothesis of no differential spillover effect from the eldest sibling can be written as:

$$H_0 : Y_i(1,0,1) = Y_i(1,0,0) \ (i = 1, \ldots, N).$$

6·2. *Exposure mappings and the choice of test statistic*

The choice of test statistic T is related to the choice of exposure mapping h to the extent that it provides a good estimate of the differential effect between exposures a and b in Equation (15). Furthermore, if we have some prior belief about the potential outcomes and the interference structure, it can be incorporated in the test statistic. Athey et al. (2017) have a nice and insightful discussion about possible test statistics in Section 5.3 of their paper, which is applicable in our setting as well.

REFERENCES

ARONOW, P. M. (2012). A general method for detecting interference between units in randomized experiments. *Sociological Methods & Research* **41**, 3–16.

ATHEY, S., ECKLES, D. & IMBENS, G. W. (2017). Exact p-values for network interference. *Journal of the American Statistical Association* , Forthcoming.

BASSE, G. & FELLER, A. (2017). Analyzing two-stage experiments in the presence of interference. *Journal of the American Statistical Association* , Forthcoming.

HENNESSY, J., DASGUPTA, T., MIRATRIX, L., PATTANAYAK, C. & SARKAR, P. (2016). A conditional randomization test to account for covariate imbalance in randomized experiments. *Journal of Causal Inference* **4**, 61–80.

Hodges Jr, J. L. & Lehmann, E. L. (1963). Estimates of location based on rank tests. *The Annals of Mathematical Statistics* , 598–611.

Lehmann, E. L. & Romano, J. P. (2006). *Testing statistical hypotheses*. Springer Science & Business Media.

Rosenbaum, P. R. et al. (2002). Covariance adjustment in randomized experiments and observational studies. *Statistical Science* **17**, 286–327.

[*Received July* 2017. *Revised July* 2017]