An Observational Study Used to Illustrate Methodology
Paul Rosenbaum, Wharton, U of Pennsylvania

The Fisher Lecture was based on [8, 9, 12, 17] and [11, §6]. These differ in details documented in the articles but not emphasized in the presentation.

What is matching with fine balance? Constrains an optimal (i.e., minimum distance) match to exactly balance the marginal distributions of a nominal covariate, without restricting who is matched to whom. A tool in a toolbox, used with: propensity scores, covariate distances, directional penalties.

Optimal assignment [1] Pairs T rows to T distinct columns in a $T \times C$ distance matrix, $C \geq T$, so the total of the T within-pair distances is minimized. There are $C!/(C-T)!$ possible pairings, but the best can be found in $O(C^3)$ arithmetic steps.

Simple implementation of minimum-distance fine-balance. Add $C-T$ rows, making a $C \times C$ matrix, adding 0’s and $\infty$’s to remove required numbers from the control group, leaving behind marginal balance. Still requires $O(C^3)$ arithmetic steps. Network implementation makes more efficient use of space.

Fine balance: references [8], extensions [6, 19, 20, 22], R packages Pimentel’s rcbalance, Yu’s DiPs and bigmatch, Zubizarreta’s designmatch.

Notation Covariate $(x, u)$, with $x$ observed, $u$ unobserved. $I$ pairs, $i = 1, \ldots, I$, of two subjects, $j = 1, 2$, one treated, $Z_{ij} = 1$, one control, $Z_{ij} = 0$, matched so $x_{1i} = x_{12}$ but perhaps $u_{1i} \neq u_{12}$. Potential responses $(r_{T_{ij}}, r_{C_{ij}})$, $r_{T_{ij}}$ observed under treatment, $Z_{ij} = 1$, $r_{C_{ij}}$ observed under control, $Z_{ij} = 0$, so $r_{ij} = Z_{ij} r_{T_{ij}} + (1-Z_{ij}) r_{C_{ij}}$ is observed but the causal effect $r_{T_{ij}} - r_{C_{ij}}$ is not observed [5,16]. Write $F$ for $\{r_{T_{ij}}, r_{C_{ij}}, x_{ij}, u_{ij}\}$, $i = 1, \ldots, I, j = 1, 2$ and $Z$ for the event $\{Z_{i1} + Z_{i2} = 1, i = 1, \ldots, I\}$. Randomization [3] would ensure $\Pr(Z_{i1} = 1 \mid Z, F) = \frac{1}{2}$, $i = 1, \ldots, I$. Fisher’s hypothesis of no effect is $H_0: r_{T_{ij}} = r_{C_{ij}}, \forall i, j$. Treated-minus-control pair $i$ difference is $D_i = (2Z_{i1} - 1)(R_{i1} - R_{i2})$, so that $D_i = (2Z_{i1} - 1)(r_{C_{ij}} - r_{T_{ij}})$ if $H_0$ is true.

Two statistics Let $q_i$ be the rank of $|D_i|$, $q_i = 0$ if $|D_i| = 0$, $q_i = 1$ if $D_i > 0$, $q_i = 0$ otherwise. Wilcoxon’s statistic is $W = \sum s_i q_i$, and Stephenson’s is $S_m = \sum s_i \cdot \binom{a-1}{m-1}$, where $\binom{a}{b} = 0$ for $a < b$. $S_1$ is the sign test. $S_2$ is (almost) $W$. In an experiment under $H_0$, randomization creates the null distribution of $W$ and $S_m$. Invert for CIs and estimates.

Sensitivity to departures from randomization Model: Subjects with the same $x$ may differ in their odds of treatment by at most a factor of $\Gamma \geq 1$ due to differences in $u$. Yields $1/(1 + \Gamma) \leq \Pr(Z_{i1} = 1 \mid Z, F) \leq \Gamma/(1 + \Gamma)$, and then, for each $\Gamma$, sharp bounds on the null distribution of $W$ and $S_m$. For $W$, the upper bound is a random variable $\overline{W}$ which is the sum of $I$ independent random variables taking the value $i$ with probability $\Gamma/(1 + \Gamma)$ or 0 with probability $1/(1 + \Gamma)$, $i = 1, \ldots, I$. Invert for confidence intervals and point estimates.

Amplification: alternative interpretation of this analysis If unobserved bias led to a $\Delta$-fold increase in the odds of a positive response, $D_i > 0$, and a $\Lambda$-fold increase in the odds of treatment, $Z_{i1} - Z_{i2} = 1$, then this is the same as a bias of $\Gamma = (\Delta \Lambda + 1)/(\Delta + \Lambda)$; see [10]. For instance, $\Gamma = 1.25$ corresponds with $\Delta = 2$, $\Lambda = 2$, and $\Gamma = 1.5$ corresponds with $\Delta = 4, \Lambda = 2$.

Design sensitivity Consider a theoretical situation with a causal effect and no unmeasured biases; however, the investigator cannot know this. In this situation, there a number $\Gamma$, the design sensitivity, so as $I \to \infty$, the study is sensitive to bias $\Gamma > \Gamma$ and insensitive to bias $\Gamma < \Gamma$; see [7,12], [11, Chapter 14], and [15, Chapter 10]. Example, if $D_i \sim N(\frac{1}{2}, 1)$ and Wilcoxon’s $W$ is used, then $\Gamma = 3.17$; however, switch to a better statistic and $\Gamma = 4.2$; yet, that statistic has Pitman efficiency 0.98 relative to $W$ in a randomized experiment with Gaussian errors [12, Tables 1, 3]. Increase $\Gamma$ adaptively [13].

Mixture of large effects and nonresponders Conover and Salsburg [2] found the locally most powerful rank test for comparing $r_{C_{ij}} \sim iid F$ to $r_{T_{ij}} \sim iid (1-p) F + pF^m$ as $I \to \infty$ and $p \to 0$, where $F^m = F \times \cdots \times F$ is the maximum of $m$ iid observations from $F$. This is a Lehmann alternative [4] who discussed $m = 2$. Conover-Salsburg ranks are not easy to interpret, but become indistinguishable from Stephenson’s [18] ranks as $I \to \infty$. Stephenson’s ranks permit confidence statements for the proportion of extreme responses caused by the treatment [9]. Gaussian version: $r_{C_{ij}} \sim \Phi(\cdot)$ and $r_{T_{ij}} \sim (1-p) \Phi(\cdot) + p\Phi^m(\cdot)$ with $p = .25$. For $m = 5, W$ and $S_{1.0}$ are close, with $\Gamma = 1.6$ for $W$ and $\Gamma = 2.0$ for $S_{1.0}$. For $m = 500, \Gamma = 2.4$ for $W$ and $\Gamma = 8.9$ for $S_{1.0}$.

Sensitivity references, extensions, R packages References [11, Chapter 16], [15, Chapters 9-10], [9,10]. Extension [12]. Functions senWilcox and senD in R package DOS. Function amplify in package sensitivitymult.

[2] Conover WJ., Salsburg DS. Locally most powerful tests for detecting treatment effects when only a subset of patients can be expected to 'respond' to treatment. Biometrics 1988;44:189-96.


[18] Stephenson WR. A general class of one-sample nonparametric test statistics based on subsamples. JASA 1981;76:960-966. senU in DOS


