

Simultaneous sensitivity analysis for observational studies using full matching or matching with multiple controls

DYLAN SMALL, JOSEPH L. GASTWIRTH, ABBA M. KRIEGER, AND PAUL R. ROSENBAUM*

In a matched observational study, a sensitivity analysis asks how the conclusions of the study would change if the matching had failed to adjust for an unobserved covariate with particular attributes. A ‘simultaneous’ sensitivity analysis characterizes the unobserved covariate u in terms of two sensitivity parameters which relate u to treatment and to response. In a previous paper, we proposed a simultaneous sensitivity analysis for matched pairs. The current paper extends this to matching with multiple controls and to full matching, using the technique of asymptotic separability. An example concerns the possibility that the military caused an increase in smoking by soldiers by various programs that subsidized cigarettes.

1. INTRODUCTION

In an observational study of treatment effects, subjects are not randomly assigned to treatment or control, as would happen in a randomized trial (Cochran 1965). In consequence, differing outcomes after treatment may not be effects caused by the treatment, and may instead reflect some form of bias in treatment assignment. Recorded differences in pretreatment covariates may often be removed by adjustments, for instance, by matching, but there is invariably concern that some important covariate was not measured, and hence not controlled by the adjustments. A sensitivity analysis asks: What would this unobserved covariate have to be like in order to alter the conclusions of the study? For various discussions of sensitivity analysis, see Cornfield, et al. (1959), Rosenbaum and Rubin (1983), Rosenbaum (1987), Gastwirth (1992), Imbens (2003), Diprete and Gangl (2004), and Wang and Krieger (2006).

In the current paper, we extend a method of sensitivity analysis for matched pairs proposed by Gastwirth, Krieger and Rosenbaum (1998) to the case of matching with sets larger than pairs. The proposed model and methodology are discussed in §2, and an example is developed in detail in §3. Unlike our previous work, we restrict attention to a binary

unobserved covariate u , and in §4, some limited results and a conjecture are given concerning a covariate that is not binary.

2. MODEL AND METHODOLOGY

2.1 Full matching and matching with multiple controls

There are I matched sets, $i = 1, \dots, I$, matched for observed covariates, \mathbf{x} , and within set i there are $n_i \geq 2$ subjects, $j = 1, \dots, n_i$, and $N = \sum n_i$ subjects in total. If the j^{th} subject in set i received the treatment, write $Z_{ij} = 1$; otherwise this subject received the control, and we write $Z_{ij} = 0$. Write $m_i = \sum_{j=1}^{n_i} Z_{ij}$ for the number of treated subjects in matched set i . In a *full matching*, each set contains either one treated subject and $n_i - 1$ controls or $n_i - 1$ treated subjects and one control, so $m_i = 1$ or $m_i = n_i - 1$ for each i . In *pair matching*, $n_i = 2$ and $m_i = 1$ for $i = 1, \dots, I$. In *matching with multiple controls*, $n_i \geq 2$ and $m_i = 1$ for $i = 1, \dots, I$. In *matching with a fixed number of controls*, $m_i = 1$ and n_i is the same for all i , while in *matching with variable controls*, n_i varies with i . For instance, one of the studies of aspirin as a cause of Reyes syndrome in children was matched with multiple controls (Hurwitz, et al. 1985).

There is a sense in which full matching is the optimal form of stratification on covariates, specifically in making subjects in the same stratum as comparable as possible. It is possible to show that, under very mild conditions: (i) among all divisions of treated and control subjects into strata containing at least one treated subject and one control, there is a full matching which minimizes the average distance between treated and control subjects in the same stratum, and (ii) for continuous covariates, with probability one, a stratification that minimizes the average distance within strata is necessarily a full matching; see Rosenbaum (1991). Intuitively, if a stratification is not a full matching, if $\min(m_i, n_i - m_i) > 1$ for some i , then stratum i can be subdivided without increasing the average distance within strata, and in the continuous case, there is a subdivision which decreases the average distance except on a set of probability zero. An optimal full matching may be

*Supported by a grant from NSF.

Table 1. Variance multiplier for a treated-minus-average control matched set difference when matching with $n_i - 1$ controls. With 2 controls rather than matched pairs, the reduction in variance is the same as that obtained by using infinitely many controls rather than 2 controls.

$n_i - 1$	1	2	4	10	∞
$1 + \frac{1}{n_i - 1}$	2	1.5	1.25	1.1	1

found using algorithms for minimum cost flow in a network (Rosenbaum 1991, Hansen and Klopfer 2006), and Hansen (2007) has implemented the algorithm in the `optmatch` package of the R statistical package (R Development Core Team 2007). Hansen (2004) discusses an example of full matching. Full matching is particularly useful when there are certain covariate values such that most subjects are treated and other covariate values such that most subjects are controls.

When there are comparatively few treated subjects and many potential controls, matching with multiple controls ($m_i = 1$, $n_i \geq 2$) can have greater efficiency than pair matching ($m_i = 1$, $n_i = 2$). Under the simple model associated with a paired t-test, with additive pair effects, a constant treatment effect, and independent errors with constant variance σ^2 , the treated-minus-average-control difference in pair i has variance $\sigma^2(1 + \frac{1}{n_i - 1})$. Table 1 shows the multiplier for $n_i - 1 = 1, 2, 4, 10$, and ∞ controls matched to a treated subject. Clearly, there is large gain from using two controls rather than matched pairs, a meaningful gain from using four controls rather than two, but the returns to additional controls diminish quickly after that. Smith (1997) presents a detailed discussion of the choice of n_i in an example of matching with a fixed number of controls; see also Haviland, et al. (2008) for another example. Ming and Rosenbaum (2000) show that matching with a variable number of controls can remove substantially more bias than matching with a fixed number of controls. For an example of matching with a variable number of controls, together with detailed efficiency results, see Haviland, et al. (2007).

Each subject ij has two potential responses, (r_{Tij}, r_{Cij}) , where subject (i, j) would exhibit response r_{Tij} under treatment ($Z_{ij} = 1$) or response r_{Cij} under control ($Z_{ij} = 0$), and the effect caused by the treatment, $r_{Tij} - r_{Cij}$, is not observed for any subject; see Neyman (1923) and Rubin (1974). The response actually observed from (i, j) is the response under the treatment (i, j) actually received, namely $R_{ij} = Z_{ij} r_{Tij} + (1 - Z_{ij}) r_{Cij}$. Fisher's (1935) sharp null hypothesis of no treatment effect asserts $H_0 : r_{Tij} - r_{Cij} = 0$, $\forall ij$. The treatment has a constant effect τ if $r_{Tij} - r_{Cij} = \tau$, $\forall ij$, in which case $R_{ij} = \tau Z_{ij} + r_{Cij}$. If there is a test for no effect, then the hypothesis $H_{\tau_0} : r_{Tij} - r_{Cij} = \tau_0$, $\forall ij$, is tested by computing, from the data and the null hypothesis, the adjusted responses, $R_{ij} - \tau_0 Z_{ij}$, which equal r_{Cij} if the hypothesis is true, and testing for no effect on these adjusted

responses; by inversion, this yields a confidence interval for a constant effect τ . In complete generality, a hypothesis about treatment effects is $H_{\tau_0} : r_{Tij} - r_{Cij} = \tau_{0ij}$, $\forall ij$, where $\tau_0 = (\tau_{011}, \dots, \tau_{0I, n_I})^T$ is a specified N -dimensional vector. This general hypothesis H_{τ_0} is tested by computing adjusted responses, $R_{ij} - \tau_{0ij} Z_{ij}$, which equal r_{Cij} if the hypothesis is true, and testing for no effect on these adjusted responses. In this way, any hypothesis about treatment effects may be converted into a test of no effect. The general hypothesis, H_{τ_0} , may be inverted to yield confidence intervals for attributable effects which summarize the plausible values τ_0 ; see Rosenbaum (2002, 2003). In light of these considerations, our discussion focuses on tests of the null hypothesis of no treatment effect, $H_0 : r_{Tij} - r_{Cij} = 0$, $\forall ij$, because tests of other hypotheses and confidence intervals are immediately available once this problem is solved.

In the discussion here, we assume that matching has controlled the observed covariates \mathbf{x}_{ij} , so that \mathbf{x} varies between but not within matched sets, $\mathbf{x}_{ij} = \mathbf{x}_{ik}$ for $\forall i, j, k$. Our concern is with an attribute or binary covariate, u_{ij} , $u_{ij} = 0$ or $u_{ij} = 1$, which was not observed and hence not controlled by matching, so that possibly $u_{ij} \neq u_{ik}$ for $j \neq k$. Write $\mathbf{u}_i = (u_{i1}, \dots, u_{i, n_i})^T$ for the n_i dimensional vector for matched set i , $i = 1, \dots, I$, and $\mathbf{u} = (u_{11}, \dots, u_{I, n_I})^T = (\mathbf{u}_1^T, \dots, \mathbf{u}_I^T)^T$ for the N -dimensional vector. In parallel, write $\mathbf{r}_C = (r_{C11}, r_{C12}, \dots, r_{CI, n_I})^T = (\mathbf{r}_{C1}^T, \dots, \mathbf{r}_{CI}^T)^T$ and $\mathbf{Z} = (Z_{11}, Z_{12}, \dots, Z_{I, n_I})^T = (\mathbf{Z}_1^T, \dots, \mathbf{Z}_I^T)^T$.

2.2 Model in the population before matching

In the population before matching, the model states that distinct subjects are mutually independent, that

$$(1) \quad Z_{ij} \perp\!\!\!\perp r_{Cij} \mid \mathbf{x}_{ij}, u_{ij},$$

where $A \perp\!\!\!\perp B \mid C$ is Dawid's (1979) notation for conditional independence of A and B given C ,

$$(2) \quad \Pr(Z_{ij} = 1 \mid \mathbf{x}_{ij}, u_{ij}) = \frac{\exp(\beta_i + \gamma u_{ij})}{1 + \exp(\beta_i + \gamma u_{ij})},$$

$$\Pr(Z_{ij} = 0 \mid u_{ij}) = 1 - \Pr(Z_{ij} = 1 \mid u_{ij}), \text{ and}$$

$$(3) \quad \Pr(r_{Cij} = r \mid \mathbf{x}_{ij}, u_{ij}) = \exp\{\zeta_i(u_{ij}) + \kappa_i(r) + \delta r u_{ij}\}$$

where β_i , γ , and δ are unknown numbers, $\kappa_i(\cdot)$ is an unknown function, and $\zeta_i(u_{ij})$ is a normalizing constant ensuring that (3) integrates to one, $\zeta_i(u_{ij}) = -\log \int \exp\{\kappa_i(r) + \delta r u_{ij}\} dr$. In (2) and (3), β_i and $\kappa_i(\cdot)$ vary with i to reflect the covariates \mathbf{x}_{ij} that are controlled in forming matched set i , where $\mathbf{x}_{ij} = \mathbf{x}_{ik}$. The sensitivity parameters γ and δ determine the strength of the relationship between the unobserved covariate u_{ij} and, respectively, the

treatment assignment Z_{ij} and response under control, r_{Cij} . Condition (1) says that u_{ij} is the relevant unobserved covariate, in the sense that treatment Z_{ij} and response under control r_{Cij} are dependent only because of their dependence on u_{ij} . Because β_i and $\kappa_i(r)$ are not restricted in any way, (2) and (3) are models only in specifying the form of the relationship between u_{ij} and Z_{ij} and between u_{ij} and r_{Cij} ; that is, if u_{ij} were irrelevant, in the sense that $\gamma = \delta = 0$, then (2) and (3) would permit any possible relationship between the matching characteristics \mathbf{x}_{ij} and Z_{ij} and between \mathbf{x}_{ij} and r_{Cij} . Motivation for a focus on binary u_{ij} is provided by Wang and Krieger (2006).

2.3 Permutation distributions in the matched sample

From the population model (1)–(3), matched samples are constructed by sampling in such a way that $\mathbf{x}_{ij} = \mathbf{x}_{ik}$ for $\forall i, j, k$ with $m_i = \sum_{j=1}^{n_i} Z_{ij}$ fixed. Write \mathbf{X} for the matrix with N rows containing the \mathbf{x}_{ij}^T . Because attention is restricted to full matching, in the discussion here, it is always the case that $m_i = 1$ or $m_i = n_i - 1$ for $i = 1, \dots, I$. Write $\mathbf{m} = (m_1, \dots, m_I)^T$. Let Π_{n_i} be the set of the $n_i!$ permutation matrices of size $n_i \times n_i$; that is, Π_{n_i} contains matrices $\boldsymbol{\pi}_i$ of size $n_i \times n_i$ in which each row and column contains a single 1 and $n_i - 1$ zeros. Then $\boldsymbol{\pi}_i \mathbf{r}_{Ci}$ permutes the n_i responses to control within matched set i . Let Π be the set containing the $n_1! \times \dots \times n_I!$ block diagonal $N \times N$ matrices whose i^{th} diagonal block $\boldsymbol{\pi}_i$ is in Π_{n_i} for $i = 1, \dots, I$, so a matrix $\boldsymbol{\pi} \in \Pi$ acts on \mathbf{r}_C through $\boldsymbol{\pi} \mathbf{r}_C$ to permute responses within the I matched sets. Write $|A|$ for the number of elements in a finite set A , so $|\Pi_{n_i}| = n_i!$.

For a vector \mathbf{w}_i of dimension n_i , let $\text{Orb}(\mathbf{w}_i) = \{\boldsymbol{\pi}_i \mathbf{w}_i : \boldsymbol{\pi}_i \in \Pi_{n_i}\}$, and for a vector \mathbf{w} of dimension N , let $\text{Orb}(\mathbf{w}) = \{\boldsymbol{\pi} \mathbf{w} : \boldsymbol{\pi} \in \Pi\}$. If the coordinates of \mathbf{r}_{Ci} were all different, then $|\text{Orb}(\mathbf{r}_{Ci})| = n_i!$, but $|\text{Orb}(\mathbf{r}_{Ci})|$ would be smaller if some coordinates were tied. Also, because $m_i = 1$ or $m_i = n_i - 1$, the coordinates of \mathbf{Z}_i are always tied in such a way that $|\text{Orb}(\mathbf{Z}_i)| = n_i$. Notice that \mathbf{m} alone determines $\text{Orb}(\mathbf{Z})$. Write $r_{Ci(j)}$ for the order statistics of the r_{Cij} within matched set i , so $r_{Ci(1)} \leq \dots \leq r_{Ci(n_i)}$ for $i = 1, \dots, I$, and write $\vec{\mathbf{r}}_{Ci} = \{r_{Ci(1)}, r_{Ci(2)}, \dots, r_{Ci(n_i)}\}$ for the n_i -dimensional vector, $i = 1, \dots, I$, and $\vec{\mathbf{r}}_C = \{r_{C1(1)}, r_{C1(2)}, \dots, r_{CI(n_I)}\}^T = (\vec{\mathbf{r}}_{C1}^T, \dots, \vec{\mathbf{r}}_{CI}^T)^T$ for the N -dimensional vector. Notice that $\vec{\mathbf{r}}_C$ determines $\text{Orb}(\mathbf{r}_C)$.

Using (1)–(3)

$$(4) \quad \begin{aligned} & \Pr(\mathbf{Z} = \mathbf{z}, \mathbf{r}_C = \mathbf{r} \mid \vec{\mathbf{r}}_C, \mathbf{m}, \mathbf{X}, \mathbf{u}) \\ &= \frac{\exp(\gamma \mathbf{u}^T \mathbf{z})}{\sum_{\mathbf{b} \in \text{Orb}(\mathbf{Z})} \exp(\gamma \mathbf{u}^T \mathbf{b})} \frac{\exp(\delta \mathbf{u}^T \mathbf{r})}{\sum_{\mathbf{w} \in \text{Orb}(\vec{\mathbf{r}}_C)} \exp(\delta \mathbf{u}^T \mathbf{w})} \\ &= \prod_{i=1}^I \frac{\exp(\gamma \mathbf{u}_i^T \mathbf{z}_i)}{\sum_{\mathbf{b}_i \in \text{Orb}(\mathbf{Z}_i)} \exp(\gamma \mathbf{u}_i^T \mathbf{b}_i)} \end{aligned}$$

$$\times \frac{\exp(\delta \mathbf{u}_i^T \mathbf{r}_i)}{\sum_{\mathbf{w}_i \in \text{Orb}(\vec{\mathbf{r}}_{Ci})} \exp(\delta \mathbf{u}_i^T \mathbf{w}_i)},$$

so for any function $t(\mathbf{Z}, \mathbf{r}_C)$,

$$(5) \quad \begin{aligned} & \Pr\{t(\mathbf{Z}, \mathbf{r}_C) \geq k \mid \vec{\mathbf{r}}_C, \mathbf{m}, \mathbf{X}, \mathbf{u}\} \\ &= \sum_{\mathbf{z} \in \text{Orb}(\mathbf{Z})} \sum_{\mathbf{r} \in \text{Orb}(\vec{\mathbf{r}}_C)} \chi\{t(\mathbf{z}, \mathbf{r}) \geq k\} \\ & \times \frac{\exp(\gamma \mathbf{u}^T \mathbf{z})}{\sum_{\mathbf{b} \in \text{Orb}(\mathbf{Z})} \exp(\gamma \mathbf{u}^T \mathbf{b})} \frac{\exp(\delta \mathbf{u}^T \mathbf{r})}{\sum_{\mathbf{w} \in \text{Orb}(\vec{\mathbf{r}}_C)} \exp(\delta \mathbf{u}^T \mathbf{w})}, \end{aligned}$$

where $\chi(E) = 1$ if the event E occurs and $\chi(E) = 0$ otherwise.

Let \mathcal{U} be the set of possible values of \mathbf{u} . Because $u_{ij} = 0$ or $u_{ij} = 1$, there are $|\mathcal{U}| = 2^N$ possible values of \mathbf{u} ; however, it is not necessary to consider all of these. Because of the sum over all $\mathbf{z} \in \text{Orb}(\mathbf{Z})$ and $\mathbf{r} \in \text{Orb}(\vec{\mathbf{r}}_C)$, the value of (5) is unchanged by replacing \mathbf{u} by $\boldsymbol{\pi} \mathbf{u}$ for $\boldsymbol{\pi} \in \Pi$. In light of this, it suffices to consider the $n_i + 1$ possible values of \mathbf{u}_i with $u_{i1} \leq u_{i2} \leq \dots \leq u_{i, n_i}$.

We are particularly interested in statistics of the form $t(\mathbf{Z}, \mathbf{r}_C) = \mathbf{Z}^T \mathbf{q}$ where $\mathbf{q} = \mathbf{q}(\mathbf{r}_C, \mathbf{m})$ with the invariance property $\boldsymbol{\pi} \mathbf{q}(\mathbf{r}_C, \mathbf{m}) = \mathbf{q}(\boldsymbol{\pi} \mathbf{r}_C, \mathbf{m})$ for $\boldsymbol{\pi} \in \Pi$. For instance, this invariance property holds with: (i) $\mathbf{q}(\mathbf{r}_C, \mathbf{m}) = \mathbf{r}_C$, (ii) $q_{ij}(\mathbf{r}_C, \mathbf{m})$ equal to the rank of r_{Cij} among $r_{Ci1}, \dots, r_{Ci, n_i}$, and (iii) $q_{ij}(\mathbf{r}_C, \mathbf{m})$ equal to the rank of the aligned response, $r_{Cij} - (1/n_i) \sum_{\ell=1}^{n_i} r_{Ci\ell}$, among all N aligned responses. In matched set i , for $k = 0, \dots, n_i$, define $\tilde{\mathbf{u}}_k = (0, 0, \dots, 0, 1, 1, \dots, 1)^T$ as the vector with k zeros followed by $n_i - k$ ones,

$$\begin{aligned} \mu_{ik} &= \sum_{\mathbf{z}_i \in \text{Orb}(\mathbf{Z}_i)} \sum_{\mathbf{r}_i \in \text{Orb}(\vec{\mathbf{r}}_{Ci})} \mathbf{z}_i^T \mathbf{q}_i(\mathbf{r}, \mathbf{m}) \\ & \times \frac{\exp(\gamma \tilde{\mathbf{u}}_k^T \mathbf{z}_i)}{\sum_{\mathbf{b}_i \in \text{Orb}(\mathbf{Z}_i)} \exp(\gamma \tilde{\mathbf{u}}_k^T \mathbf{b}_i)} \\ & \times \frac{\exp(\delta \tilde{\mathbf{u}}_k^T \mathbf{r}_i)}{\sum_{\mathbf{w}_i \in \text{Orb}(\vec{\mathbf{r}}_{Ci})} \exp(\delta \tilde{\mathbf{u}}_k^T \mathbf{w}_i)} \end{aligned}$$

and

$$\begin{aligned} \sigma_{ik}^2 &= \sum_{\mathbf{z}_i \in \text{Orb}(\mathbf{Z}_i)} \sum_{\mathbf{r}_i \in \text{Orb}(\vec{\mathbf{r}}_{Ci})} \{\mathbf{z}_i^T \mathbf{q}_i(\mathbf{r}, \mathbf{m}) - \mu_{ik}\}^2 \\ & \times \frac{\exp(\gamma \tilde{\mathbf{u}}_k^T \mathbf{z}_i)}{\sum_{\mathbf{b}_i \in \text{Orb}(\mathbf{Z}_i)} \exp(\gamma \tilde{\mathbf{u}}_k^T \mathbf{b}_i)} \\ & \times \frac{\exp(\delta \tilde{\mathbf{u}}_k^T \mathbf{r}_i)}{\sum_{\mathbf{w}_i \in \text{Orb}(\vec{\mathbf{r}}_{Ci})} \exp(\delta \tilde{\mathbf{u}}_k^T \mathbf{w}_i)} \end{aligned}$$

with \mathbf{r} differing from $\vec{\mathbf{r}}_C$ only in set i ,

$$\mathbf{r} = \begin{bmatrix} \vec{\mathbf{r}}_{C1} \\ \vdots \\ \vec{\mathbf{r}}_{C,i-1} \\ \mathbf{r}_i \\ \vec{\mathbf{r}}_{C,i+1} \\ \vdots \\ \vec{\mathbf{r}}_{CI} \end{bmatrix}.$$

Also, let

$$\begin{aligned} \mu_{i \max} &= \max_{k \in \{0,1,\dots,k\}} \mu_{ik} \\ A_i &= \{k : \mu_{ik} = \mu_{i \max}\} \\ \sigma_{i \max}^2 &= \max_{k \in A_i} \sigma_{ik}^2. \end{aligned}$$

By Proposition 1 in Gastwirth, J. L., Krieger, A. M. and Rosenbaum, P. R. (2000),

$$(6) \quad \begin{aligned} &\max_{\mathbf{u} \in \mathcal{U}} \Pr \{t(\mathbf{Z}, \mathbf{r}_C) \geq k \mid \vec{\mathbf{r}}_C, \mathbf{m}, \mathbf{X}, \mathbf{u}\} \\ &\approx 1 - \Phi \left(\frac{k - \sum_{i=1}^I \mu_{i \max}}{\sqrt{\sum_{i=1}^I \sigma_{i \max}^2}} \right), \end{aligned}$$

in the sense that the difference between the two sides of (6) tends to zero as $I \rightarrow \infty$ for the well behaved statistics $t(\mathbf{Z}, \mathbf{r}_C) = \mathbf{Z}^T \mathbf{q}$.

3. AN OBSERVATIONAL STUDY OF THE EFFECT OF MILITARY SERVICE ON SMOKING

3.1 Background

Several studies have found that among U.S. men, military veterans smoke more than nonveterans (Bray, Marsden and Peterson, 1991; Feigelman, 1994; Kroutil, Bray and Marsden, 1994; Klevens, Giovino, Peddicord, Nelson, Mowery and Grummer-Strawn, 1995; McKinney, McIntire, Carmody and Joseph, 1997). Reasons why military service might *cause* an increase in smoking include cigarettes being sold at reduced prices in U.S. military commissaries; cigarettes being distributed freely in rations to personnel stationed overseas (this was discontinued in 1975); the stress and boredom of military life; and the targeting of military personnel by the cigarette industry through advertising linking patriotism, strength and toughness with smoking (McIntire, Carmody and Joseph, 1997). However, the association between military service and smoking might not be causal. The group of veterans might differ from nonveterans in family background, mental and physical aptitude and many other characteristics, and these differences might fully explain the association between smoking and military service rather than

any causal effect of military service on smoking. To study the effect of smoking on military service, we undertook an observational study using the Wisconsin Longitudinal Study (WLS), an especially detailed longitudinal study.

3.2 Data

The WLS began its data collection with high school seniors in 1957 in Wisconsin. The respondents have been followed up to the present day with periodic surveys. For two of the many empirical studies based on the WLS, see Singer, Ryff, Carr and Magee (1998) and Warren, Sheridan and Hauser (2002).

We define veteran status using the question asked in 1975, ‘‘Have you ever been in the military’’? We remove all men for whom the answer to the veteran status question is missing and also remove men who served in the military but were reserves. Our outcome is pack years smoked based on data from the 1992–1993 survey (when respondents were approximately 53 years old). Specifically, pack years is defined as the multiplication of the answers to the following two questions asked in 1992–1993: (1) For how many years did/have you smoked regularly; and (2) About how many packs did/do you usually smoke per day now. We remove all men for whom some of the information needed to compute pack years is missing. This leaves 1273 veterans and 1320 nonveterans.

The mean pack years for veterans is 20.6 and for nonveterans is 15.9. A two sample t-test of the null hypothesis that veterans and nonveterans’ have the same mean pack years gives a p -value of < 0.0001 and a 95% confidence interval for the mean difference between veterans and nonveterans is (2.7, 6.7).

To investigate whether the positive association between military service and pack years is causal, we consider the following measured confounders: 1957 population of town in which graduate attended high school (pop57), father’s years of schooling (edfa57q), mother’s years of schooling (edmo57q), family’s income when graduate was a senior in high school (sesp57), denominational control of high school – Catholic, other private, public (hsdm57), percentile rank of grades in high school (hsrankq), whether or not teacher consider graduate outstanding (tchevl), IQ score (gwiiq_bm) and main religious preference of family in 1957 (relfml). Table 2 shows the standardized biases and p -values for two sample t-tests comparing the measured confounders between the veterans and nonveterans (only the two largest categories of religious denomination of family are shown). The variables are ordered by absolute standardized biases.

There are significant differences between veterans and nonveterans on their high school performance and their IQs with veterans doing worse in high school on average and having lower IQs on average.

Table 2. Standardized biases and p-values for two sample t-tests comparing veterans to nonveterans

Variable	Standardized Bias	p-value for t-test
High School Rank	-0.33	< 0.01
Teacher Evaluation	-0.14	< 0.01
IQ Score	-0.11	< 0.01
Family Income	-0.06	0.17
Catholic	0.05	0.19
Home Town Population	-0.04	0.27
Father's Education (years)	-0.04	0.29
Catholic School	-0.01	0.81
Mother's Education (years)	0.00	0.91
Lutheran Religion	0.00	0.96

3.3 Matching

We use full matching to control for differences on the measured confounders between veterans and nonveterans. There was missing data on some of the measured confounders. We include missing data indicators (edfa57qmis, edmo57qmis, sesp57mis and hsrnkqmis) as additional covariates for these confounders with missing data. By including missing data indicators as variables in the matching, we balance both the observed data and missingness among the treatment and control groups as suggested by Rosenbaum and Rubin (1984). We estimated the propensity score for being a veteran by logistic regression of veteran status on the measured confounders and the missing data indicators. For units with missing variables, we substitute the mean value of the variable for the units for which it is nonmissing. Because we are also including the missing data indicators in the logistic regression, this mean substitution does not affect the propensity scores. We do not include observations with missing values of a covariate when assessing balance on the covariate. We use propensity caliper matching by constructing a distance using the Mahalanobis distance on the ranks of the variables (the measured confounders and the missing data indicators) and then adding a large penalty if the difference in logit propensity score between two units is greater than 0.2 standard deviations of the logit propensity score. We carried out the full matching using Hansen's (2004) optmatch package. The third column of Table 3 compares the standardized biases of the logit propensity score for no matching, a full matching and several full matchings with restrictions on the number of units in each stratum. The second column of Table 3 reports a measure of the effective sample size in matched pairs for each matching (the sum of the harmonic means of the treated and control units in each stratum).

For a good balance between the effective sample size in matched pairs and control of bias, we will use the full matching with a maximum of 4 units in each stratum henceforth. The structure of this matching is that it contains 62 strata

Table 3. Effective sample size in terms of matched pairs of various matchings and standardized biases for various matchings

Type of Matching	Sample Size	Standardized Bias
No matching	-	0.30
Matched pairs	1273	0.22
Full matching, maximum of 3 units in stratum	1229	0.08
Full matching, maximum of 4 units in stratum	1201	0.06
Unrestricted full matching	1181	0.06

Table 4. Standardized biases before matching and after matching

Variable	Before Matching	After Matching
High School Rank	-0.33	-0.02
Teacher Evaluation	-0.14	-0.01
IQ Score	-0.11	0.01
Family Income	-0.06	0.00
Roman Catholic	0.05	-0.01
Home Town Population	-0.04	0.00
Father's Educations (years)	-0.04	0.00
Father's Education Missing	0.03	0.02
Mother's Education Missing	0.03	0.02
Family Income Missing	0.02	0.02
HS Rank Missing	-0.02	0.00
Catholic	-0.01	0.02
Mother's Education (years)	0.00	0.01
Lutheran	0.00	0.00

with 3 veterans and 1 nonveteran, 68 strata with 2 veterans and 1 nonveteran, 793 matched pairs, 77 strata with 1 veteran and 2 nonveterans and 81 strata with 1 veteran and 3 nonveterans. The matching approximately balances all of the covariates. Table 4 reports the standardized biases before and after matching. The numerator of the standardized bias after matching is the average, over the treated units, of the difference between a treated unit's covariate value and the average of the control units' covariate values in the treated unit's stratum. The denominator of the standardized bias after matching is the same as the denominator of the standardized bias before matching, the square root of the average of the treated units' and control units' standard deviations for the covariate. The covariates are listed in order of their absolute standardized biases before matching. After matching, the absolute standardized bias is less than or equal to 0.02 for all of the covariates.

The matched data set is available at <http://www-stat.wharton.upenn.edu/~dsmall/>. The entire Wisconsin Longitudinal Study is available at <http://www.ssc.wisc.edu/wlsresearch/>.

3.4 Sensitivity analysis

Using the full matching described in the above section, an aligned rank test of the null hypothesis that military service does not cause an increase in pack years smoked gives a p-value of 0.00004. This test is valid if treatment assignment is randomized within sets. To investigate the impact of an unmeasured covariate, we carry out a simultaneous sensitivity analysis that examines the impact of specified associations between the unmeasured covariate and becoming a veteran and the unmeasured covariate and the outcome. We use the following approach to make the sensitivity parameter Δ interpretable. We replace the data with the aligned ranks of the data divided by half the maximum of these ranks and then let \mathbf{q} be the identity function. In other words, we take each observation minus the mean in its stratum, rank these residuals, let \mathbf{r}_C be the ranked residuals divided by half of the maximum of the ranks and let $\mathbf{q}(\mathbf{r}_C) = \mathbf{r}_C$. To interpret the sensitivity parameter Δ , consider two observations in the same stratum, one treated and one control, for which one has aligned rank equal to the first quartile of the aligned ranks and the other has aligned rank equal to the third quartile of the aligned ranks, and for which one has unobserved covariate u_{ij} equal to 1 and the other has unobserved covariate u_{ij} equal to 0. Then the odds that the observation with the higher unobserved covariate also has the higher aligned rank (and hence the higher response) is at least $\exp(-\delta) = 1/\Delta$ and at most $\exp(\delta) = \Delta$.

The maximum p-values for various combinations of the sensitivity parameters Γ and Δ are shown in Table 5. If either the treatment is unrelated to the unobserved covariate, $\Gamma = 1$, or the response is unrelated to the unobserved covariate, $\Delta = 1$, then the significance level of .00004 from the randomized experiment remains valid. An unobserved covariate that increased the odds of serving in the military by 10% could not explain the large test statistic no matter how strongly that covariate is related to pack years ($\Gamma = 1.1, \Delta = \infty$); however, a covariate that increased the odds of serving in the military by 20% could explain it ($\Gamma = 1.2, \Delta = \infty$). On the other hand, an unobserved covariate that increased both the odds of serving in the military and the odds of a higher pack years level by 20% could not explain the large test statistic ($\Gamma = 1.2, \Delta = 1.2$). An unobserved covariate that increased the odds of serving in the military by 50% and the odds of a higher pack years level by 150% could explain the large test statistic

Table 5. Sensitivity analysis: maximum p-values

Γ	$\Delta = 1$	$\Delta = 1.1$	$\Delta = 1.2$	$\Delta = 1.5$	$\Delta = 2.5$	$\Delta = \infty$
1	.00004	.00004	.00004	.00004	.00004	.00004
1.1	.00004	.00005	.00007	.00013	.00046	.00363
1.2	.00004	.00007	.00011	.00039	.00329	.06464
1.5	.00004	.00013	.00040	.00428	.11298	.92932
2.5	.00004	.00054	.00418	.14282	.98001	1
∞	.00004	.01397	.27703	.99979	1	1

($\Gamma = 1.5, \Delta = 2.5$), as could an unobserved covariate that increased the odds of serving in the military by 150% and the odds of a higher pack years level by 50% ($\Gamma = 2.5, \Delta = 1.5$).

4. BINARY VERSUS BOUNDED u : RESULTS FOR $n_i = 3$ AND A CONJECTURE

In previous work, the unobserved covariate u was not binary, $u \in \{0, 1\}$, but rather bounded, $u \in [0, 1]$, and an explicit argument showed that the extreme u was either 0 or 1. In contrast, in the current paper, we assume u is binary, rather than assuming $u \in [0, 1]$ and deducing that the bounds are provided by binary u . The current section shows this to be the case when there are three units in a matched set: the extreme $u \in [0, 1]$ is always 0 or 1. We are not, so far, able to show this in general, but we conjecture it is true in general. For somewhat related discussion, see Wang and Krieger (2006).

In this section we consider mathematical issues relating to the dual sensitivity analysis considered in other sections. For ease of exposition, we consider three individuals per stratum. One of the three individuals receives the treatment and the other two the control. The values for the unobserved covariate for the three individuals are 0, x , and 1. The three responses in this stratum are denoted by $R_1 < R_2 < R_3$. We can assume that $R_1 = 0$, without loss of generality.

The unobserved covariate affects both the probabilities that the three individuals receive the three responses and simultaneously affects which of the three individuals receives the treatment. Specifically, the probability that individual i receives the treatment is:

$$\begin{aligned} Q_1(x; \Gamma) &= 1/(1 + \Gamma^x + \Gamma) \\ Q_2(x; \Gamma) &= \Gamma^x/(1 + \Gamma^x + \Gamma) \\ Q_3(x; \Gamma) &= \Gamma/(1 + \Gamma^x + \Gamma), \end{aligned}$$

where $\Gamma = e^\gamma$ for some $\gamma > 0$. There are six permutations for the probabilities that the individuals receive each of the three responses. Consider P_{i_1, i_2, i_3} which denotes the probability that individuals 1, 2, and 3 receive responses R_{i_1} , R_{i_2} and R_{i_3} respectively where (i_1, i_2, i_3) represents a permutation of the integers 1, 2, and 3. Let $\Delta_i = e^{\delta R_i}$, where $\delta > 0$. Then these six probabilities are:

$$\begin{aligned} P_{123} &= (\Delta_2^x + \Delta_3)/p_T \\ P_{132} &= (\Delta_3^x + \Delta_2)/p_T \\ P_{213} &= \Delta_3/p_T \\ P_{231} &= \Delta_3^x/p_T \\ P_{312} &= \Delta_2/p_T \\ P_{321} &= \Delta_2^x/p_T \end{aligned}$$

where $p_T = P_{123} + P_{132} + P_{213} + P_{231} + P_{312} + P_{132}$.

We want to determine the behavior of test statistics as a function of the value of x . To this end, we first consider the

probability that the treated person receives the largest response. The probability that each of the individuals receives the largest response is,

$$\begin{aligned} P_1(x; \Delta_2, \Delta_3) &= P_{312} + P_{321} \\ P_2(x; \Delta_2, \Delta_3) &= P_{132} + P_{231} \\ P_3(x; \Delta_2, \Delta_3) &= P_{123} + P_{213}. \end{aligned}$$

Since the argument fixes the values of Δ_2 and Δ_3 , the specification of these parameters is omitted in the exposition that follows (i.e., we denote $P_i(x; \Delta_2, \Delta_3)$ as $P_i(x)$).

To maximize the probability that the treated individual receives the treatment is equivalent to maximizing

$$g(x; \Gamma) = P_1(x)Q_1(x; \Gamma) + P_2(x)Q_2(x; \Gamma) + P_3(x)Q_3(x; \Gamma).$$

Theorem 1. *The function $g(x; \Gamma)$ is maximized when x is either 0 or 1.*

Proof. This turns out to be a nontrivial calculus problem since $g(x; \Gamma)$ is not convex. We begin with a few preliminaries:

- i) Since $Q_1(x; 1) = Q_2(x; 1) = Q_3(x; 1) = \frac{1}{3}$ and $P_1(x) + P_2(x) + P_3(x) = 1$, it follows that $g(x; 1) = \frac{1}{3}$ for all x .
- ii) Let

$$\beta_i(x; \Gamma) = \frac{Q_i(x; \Gamma) - Q_i(0; \Gamma)}{Q_i(1; \Gamma) - Q_i(0; \Gamma)}.$$

It turns out from the definition of $Q_i(x; \Gamma)$ (notably, $Q_3(x; \Gamma)/Q_1(x; \Gamma) = \Gamma$ and $Q_1(x; \Gamma) + Q_2(x; \Gamma) + Q_3(x; \Gamma) = 1$) that $\beta_1(x; \Gamma) = \beta_2(x; \Gamma) = \beta_3(x; \Gamma) \equiv \beta(x; \Gamma)$. Hence,

$$Q_i(x; \Gamma) = \beta(x; \Gamma)Q_i(1; \Gamma) + (1 - \beta(x; \Gamma))Q_i(0; \Gamma).$$

It suffices to show that

$$v(x; \Gamma) = \beta(x; \Gamma)g(1; \Gamma) + (1 - \beta(x; \Gamma))g(0; \Gamma) - g(x; \Gamma) \geq 0.$$

But

$$\begin{aligned} g(x; \Gamma) &= \beta(x; \Gamma)\{P_1(x)Q_1(1; \Gamma) + P_2(x)Q_2(1; \Gamma) \\ &\quad + P_3(x)Q_3(1; \Gamma)\} + (1 - \beta(x; \Gamma)) \\ &\quad \times \{P_1(x)Q_1(0; \Gamma) + P_2(x)Q_2(0; \Gamma) \\ &\quad + P_3(x)Q_3(0; \Gamma)\}. \end{aligned}$$

Hence it is sufficient to show that

$$\text{i) } (P_1(1) - P_1(x))Q_1(1; \Gamma) + (P_2(1) - P_2(x))Q_2(1; \Gamma) \\ + (P_3(1) - P_3(x))Q_3(1; \Gamma) \geq 0$$

and

$$\text{ii) } (P_1(0) - P_1(x))Q_1(0; \Gamma) + (P_2(0) - P_2(x))Q_2(0; \Gamma) \\ + (P_3(0) - P_3(x))Q_3(0; \Gamma) \geq 0.$$

We first consider i). Since $Q_2(1; \Gamma) = Q_3(1; \Gamma) = \Gamma Q_1(1; \Gamma)$, the left-hand side can be written as

$$\begin{aligned} Q_1(1; \Gamma)\{P_1(1) + \Gamma(1 - P_1(1)) - P_1(x) - \Gamma(1 - P_1(x))\} \\ = Q_1(1; \Gamma)(\Gamma - 1)(P_1(x) - P_1(1)) \geq 0. \end{aligned}$$

The last inequality follows because $P_1(x)$ is a decreasing function of x as is shown in the Appendix.

Similarly for ii) $\Gamma Q_1(0; \Gamma) = \Gamma Q_2(0; \Gamma) = Q_3(0; \Gamma)$, the left-hand side can be written as

$$\begin{aligned} Q_1(0; \Gamma)\{(1 - P_3(0)) + \Gamma P_3(0) - (1 - P_3(x)) - \Gamma P_3(x)\} \\ = Q_1(0; \Gamma)(\Gamma - 1)(P_3(0) - P_3(x)) \geq 0. \end{aligned}$$

The last inequality follows because $P_3(x)$ is a decreasing function of x as is shown in the Appendix. \square

A similar argument will show that the probability that the treated subject receives either R_2 or R_3 occurs when $x = 0$ or $x = 1$. Even if the P-value were optimized by a binary covariate, the issue of finding whether $x = 0$ or $x = 1$ for each individual grows exponentially with the number of strata. The following example illustrates the complication.

Example. Consider the formulation above with $R_1 = 0$, $R_2 = 2$ and $R_3 = 2.2$. Also let $\gamma = \delta = 2$ so that $\Gamma = e^2$, $\Delta_2 = e^4$ and $\Delta_3 = e^{4.4}$. Then

- i) if $x = 0$, the probabilities the treated person receives R_1 , R_2 and R_3 are respectively .1115, .3776, .5109.
- ii) if $x = 1$, the probabilities the treated person receives R_1 , R_2 and R_3 are respectively .0754, .4611, .4635.

This setup first shows that the value of x that maximizes the P-value might be 0 or 1 (depending on the critical value). In this example, if we want to maximize the probability that the treated person receives a response exceeding 2, then $x = 0$ provides the maximum. On the other hand if we want to maximize the probability that the treated person receives a response that is at least 2, then $x = 1$ provides the maximum.

More importantly, this example shows that the problem is not separable when there is more than one strata. Consider two strata each with the same values of R_i as above. Consider the sensitivity analysis that requires maximizing $P(\text{Sum of the responses} \geq 4.2)$. If both individuals have $x = 0$, this probability is $.5109^2 + 2(.5109)(.3776) = .6469$. If both individuals have $x = 1$, this probability is $.4635^2 + 2(.4635)(.4611) = .6423$. But if one individual has a value of $x = 0$ and the other $x = 1$, the probability is $.5109(.4635) + (.5109)(.4611) + (.4635)(.3776) = .6474$. Since having dispersed x values yields a larger optimum, this implies that if we set $x = 0$ in the first stratum, then we need to assign $x = 1$ in the second and vice versa, hence the values are linked.

5. SUMMARY

We have considered a simultaneous, two parameter sensitivity analysis for general forms of matching, including matching with multiple controls and full matching. The method extends ideas in Gastwirth, Krieger and Rosenbaum (1998) for matched pairs. Results were illustrated in a study of the effects of military service on smoking habits.

6. APPENDIX: DETAILS OF A PROOF

In this appendix we show that $P_1(x)$ and $P_3(x)$ are decreasing functions of x . Let

$$D(x) = \Delta_2 + \Delta_2^x + \Delta_3^x(1 + \Delta_2) + \Delta_3(1 + \Delta_2^x)$$

Then $P_1(x) = \frac{\Delta_2 + \Delta_2^x}{D(x)}$. Hence

$$\begin{aligned} D^2(x)P_1'(x) &= [\Delta_2 + \Delta_2^x + \Delta_3^x(1 + \Delta_2) + \Delta_3(1 + \Delta_2^x)] \\ &\quad \times \Delta_2^x \log(\Delta_2) - (\Delta_2 + \Delta_2^x)[\Delta_2^x \log(\Delta_2) \\ &\quad + \Delta_3^x \log(\Delta_3)(1 + \Delta_2) + \Delta_3 \Delta_2^x \log(\Delta_2)] \\ &\leq (\Delta_2 \Delta_3)^x (1 + \Delta_2)(\log(\Delta_2) - \log(\Delta_3)) \\ &\quad + \Delta_3 \Delta_2^x \log(\Delta_2)(1 - \Delta_2) \leq 0. \end{aligned}$$

Similarly, $P_3(x) = \frac{\Delta_3(1 + \Delta_2^x)}{D(x)}$. Hence

$$\begin{aligned} D^2(x)P_3'(x)/\Delta_3 &= [\Delta_2 + \Delta_2^x + \Delta_3^x(1 + \Delta_2) + \Delta_3(1 + \Delta_2^x)] \\ &\quad \times \Delta_2^x \log(\Delta_2) - (1 + \Delta_2^x)[\Delta_2^x \log(\Delta_2) \\ &\quad + \Delta_3^x \log(\Delta_3)(1 + \Delta_2) + \Delta_3 \Delta_2^x \log(\Delta_2)] \\ &\leq (\Delta_2 \Delta_3)^x (1 + \Delta_2)(\log(\Delta_2) - \log(\Delta_3)) \\ &\quad + \Delta_2 \Delta_2^x \log(\Delta_2) - (1 + \Delta_2) \Delta_3^x \log(\Delta_3) \\ &\leq 0. \end{aligned}$$

Received 30 September 2008

REFERENCES

Bray, R. M., Marsden, M. E. and Peterson, M. R. (1991). Standardized comparisons of the use of alcohol, drugs, and cigarettes among military personnel and civilians, *American Journal of Public Health*, **81**, 865–869.

Cochran, W. G. (1965). The planning of observational studies of human populations (with Discussion), *Journal of the Royal Statistical Society*, **A**, **128**, 134–155.

Cornfield, J., Haenszel, W., Hammond, E., Lilienfeld, A., Shimkin, M., and Wynder, E. (1959). Smoking and lung cancer, *Journal of the National Cancer Institute*, **22**, 173–203.

Dawid, A. P. (1979). Conditional independence in statistical theory, *Journal of the Royal Statistical Society* **B**, **41**, 1–31. [MR0535541](#)

Diprete, T. A. and Gangl, M. (2004). Assessing bias in the estimation of causal effects, *Sociological Methodology*, **34**, 271–310.

Fisher, R. A. (1935). *The Design of Experiments*. Edinburgh: Oliver & Boyd.

Feigelman, W. (1994). Cigarette smoking among former military service personnel: a neglected social issue, *Preventive Medicine*, **23**, 235–241.

Gastwirth, J. L. (1992). Methods for assessing the sensitivity of statistical comparisons used in Title VII cases to omitted variables, *Jurimetrics*, **33**, 19–34.

Gastwirth, J. L., Krieger, A. M. and Rosenbaum, P. R. (1998). Dual and simultaneous sensitivity analysis for matched pairs, *Biometrika*, **85**, 907–920.

Gastwirth, J. L., Krieger, A. M. and Rosenbaum, P. R. (2000). Asymptotic separability in sensitivity analysis, *Journal of the Royal Statistical Society* **B**, **62**, 545–555. [MR1772414](#)

Hansen, B. B. (2004). Full matching in an observational study of coaching for the SAT, *Journal of the American Statistical Association*, **99**, 609–618. [MR2086387](#)

Hansen, B. B. and Klopfer, S. O. (2006). Optimal full matching and related designs via network flows, *Journal of Computational and Graphical Statistics*, **15**, 609–627. [MR2280151](#)

Hansen, B. B. (2007). Optmatch: flexible, optimal matching for observational studies, *R News*, **7**, 18–24.

Haviland, A., Nagin, D. S., Rosenbaum, P. R. (2007). Combining propensity score matching and group-based trajectory analysis in an observational study, *Psychological Methods*, **12**, 247–267.

Haviland, A., Nagin, D. S., Rosenbaum, P. R. and Tremblay, R. (2008). Combining group-based trajectory modeling and propensity score matching for causal inferences in nonexperimental longitudinal data, *Developmental Psychology*, **44**, 422–436.

Hurwitz, E. S., Barrett, M. J., Bregman, D., et al. (1985). Public Health Service study on Reye's syndrome and medications, *New England Journal of Medicine*, **313**, 849–57.

Imbens, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation, *American Economic Review*, **93**, 126–132.

Klevens, R. M., Giovino, G. A., Peddicord, J. P., Nelson, D. E., Mowery, P. and Grummer-Strawn, L. (1995). The association between veteran status and cigarette-smoking behaviors, *American Journal of Preventive Medicine*, **11**, 245–250.

Kroutil, L. A., Bray, R. M. and Marsden, M. E. (1994). Cigarette smoking in the U.S. military: findings from the 1992 worldwide survey, *Preventive Medicine*, **23**, 521–528.

McKinney, W. P., McIntire, D. D., Carmody, T. J. and Joseph, A. (1997). Comparing the smoking behavior of veterans and nonveterans, *Public Health Reports*, **112**, 212–217.

Ming, K. and Rosenbaum, P. R. (2000). Substantial gains in bias reduction from matching with a variable number of controls, *Biometrics*, **56**, 118–124.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. Reprinted in *Statistical Science* 1990, **5**, 463–480. [MR1092985](#)

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*, Vienna: R Foundation, <http://www.R-project.org>.

Rosenbaum, P. R. (1987). Sensitivity analysis for certain permutation inferences in matched observational studies, *Biometrika*, **74**, 13–26. [MR0885915](#)

Rosenbaum, P. R. (1991). A characterization of optimal designs for observational studies, *Journal of the Royal Statistical Society*, **B**, **53**, 597–610. [MR1125717](#)

Rosenbaum, P. R. (2002). Attributing effects to treatment in matched observational studies, *Journal of the American Statistical Association*, **97**, 183–192. [MR1963391](#)

Rosenbaum, P. R. (2003). Exact confidence intervals for nonconstant effects by inverting the signed rank test, *American Statistician*, **57**, 132–138. [MR1969770](#)

Rosenbaum, P. and Rubin, D. (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome, *Journal of the Royal Statistical Society* **B**, **45**, 212–218.

Rosenbaum, P. R. and Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score, *Journal of the American Statistical Association*, **79**, 516–524.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies, *Journal of Educational Psychology*, **66**, 688–701.

- Singer, B., Ryff, C. D., Carr, D., Magee, W. J. (1998). Linking life histories and mental health, *Sociological Methodology*, **28**, 1–51.
- Smith, H. (1997). Matching with multiple controls to estimate treatment effects in observational studies, *Sociological Methodology*, **27**, 325–353.
- Wang, L. and Krieger, A. M. (2006). Causal conclusions are most sensitive to unobserved binary covariates, *Statistics in Medicine*, **25**, 2257–2271. [MR2240099](#)
- Warren, J. R., Hauser, R. M., Sheridan, J. T. (2002). Occupational stratification across the life course: Evidence from the Wisconsin Longitudinal Study, *American Sociological Review*, **67**, 432–455.

Dylan Small
400 Huntsman Hall
3730 Walnut St.
Philadelphia, PA 19104, USA
E-mail address: dsmall@wharton.upenn.edu

Joseph L. Gastwirth

Abba M. Krieger

Paul R. Rosenbaum
Department of Statistics
The Wharton School
University of Pennsylvania
473 Jon M. Huntsman Hall
3730 Walnut Street
Philadelphia, PA 19104-6340, USA
E-mail address:
rosenbaum@stat.wharton.upenn.edu