Does a dose–response relationship reduce sensitivity to hidden bias?

PAUL R. ROSENBAUM

*Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia,
PA 19104-6340, USA*
rosenbaum@stat.wharton.upenn.edu

SUMMARY

It is often said that an important consideration in judging whether an association between treatment and response is causal is the presence or absence of a dose–response relationship, that is, larger ostensible treatment effects when doses of treatment are larger. This criterion is widely discussed in textbooks and is often mentioned in empirical papers. At the same time, it is well known through both important examples and elementary theory that a treatment may cause dramatic effects with no dose–response relationship, and hidden biases may produce a dose–response relationship when the treatment is without effect. What does a dose–response relationship say about causality? It is observed here that a dose–response relationship may or may not reduce sensitivity to hidden bias, and whether it has or has not can be determined by a suitable analysis using the data at hand. Moreover, a study without a dose–response relationship may or may not be less sensitive to hidden bias than another study with such a relationship, and this, too, can be determined from the data at hand. An example concerning cytogenetic damage among professional painters is used to illustrate.

Keywords: Causal inference; Confounding; Dose-control design; Dose-response relationship; Hill's criteria for causality; Observational study; Sensitivity analysis.

1. DOSE–RESPONSE AND HIDDEN BIAS

In his President's Address to the Royal Society of Medicine, Hill (1965) asked:

> Our observations reveal an association between two variables, perfectly clear-cut and beyond what we would care to attribute to the play of chance. What aspects of that association should we especially consider before deciding that the most likely interpretation of it is causation?

Hill then discussed nine considerations or criteria to aid in judging whether an association is causal, one of which was the presence or absence of a 'biological gradient or dose–response', such as the observation that heavy smokers have higher rates of lung cancer than do light smokers. Hill's criteria have been highly influential, particularly in epidemiology, where they are discussed, with varying degrees of enthusiasm, in most texts on epidemiologic methods. Moreover, the criteria are commonly discussed in empirical articles.

At the same time, there are, in the literature, frequent expressions of puzzlement about some of Hill's criteria. For example, Rothman (1986, p18) writes:

> Some causal associations, however, show no apparent trend of effect with dose; an example is the association between DES and adenocarcinoma of the vagina . . . Associations that do show a dose–response trend are not necessarily causal; confounding can result in such a trend between a noncausal risk factor and disease if the confounding factor itself demonstrates a biologic gradient in its relation with disease.

In an interesting reinterpretation of dose–response, Weiss (1981, p.488) echoes one of Rothman's points:

> . . . one or more confounding factors can be related closely enough to both exposure and disease to give rise to [a dose response relationship] in the absence of cause and effect.

Of what value, then, is a dose–response relationship?

In common practice, a dose–response relationship is claimed if the null hypothesis of no treatment effect is rejected in a test against the alternative of a dose–response relationship, where the test explicitly or implicitly assumes there are no hidden biases due to unobserved covariates. However, as Hill suggested, the interesting aspect of a dose–response relationship is the evidence it provides, or fails to provide, about whether the treatment caused its ostensible effects or whether they are produced by a hidden bias. To test for a dose–response relationship assuming hidden biases are absent, as is commonly done, is to assume away the question one hoped to investigate.

A sensitivity analysis can clarify what a dose–response relationship says about hidden biases. Unlike a test that assumes hidden biases are absent, a sensitivity analysis assumes that hidden biases may be present. A dose–response relationship may strengthen the evidence of causality—more precisely, it may reduce the degree of sensitivity to hidden bias—or it may not, and whether it does or does not is visible upon analysis in the data at hand. In particular, I will present an example in which a dose–response relationship reduces sensitivity to hidden biases, mention a second example with parallel structure where there is no reduction in sensitivity to bias, and then observe that both these examples are far more sensitive to hidden bias than the studies of DES and adenocarcinoma of the vagina, to which Rothman appropriately refers, where there is no dose–response relationship. To reach these conclusions, the method of analysis must assume that hidden biases may be present.

2. Sensitivity to hidden bias

A sensitivity analysis determines the magnitude of hidden bias that would need to be present to alter the conclusions of an observational study. The first sensitivity analysis was conducted by Cornfield *et al.* (1959), who determined that very large hidden biases would need to be present to alter the conclusion that heavy smoking causes lung cancer; see also Greenhouse (1982) and Gastwirth *et al.* (1998a). A sensitivity analysis replaces the correct but uninformative qualitative observation, 'association does not imply causation', by a quantitative statement, specific to the results of a particular study, saying that to explain away the association actually observed, biases of such and such a magnitude would have to be present. Whether hidden biases of a particular magnitude are present or not is a matter of scientific speculation—perhaps informed speculation, perhaps not, but in either case, speculation referring to something beyond the data immediately at hand—but the degree of sensitivity to hidden bias is a fact of the matter, something visible in the data at hand.

A fairly general method of sensitivity analysis, similar in spirit to the method of Cornfield *et al.* (1959), is built in the following way. The method will be described for matched treated/control pairs using signed rank statistics (Rosenbaum, 1987, 1991a), but it is applicable also to case-control studies, unmatched groups, matching with multiple controls, stratified groups, using rank sum statistics, Fisher's exact test, McNemar's test, the Mantel–Haenszel test, aligned rank tests, Gehan's test, and many others;

see Rosenbaum (2002, Section 4). There are I matched pairs, $i = 1, \ldots, I$, with two subjects in each pair, $j = 1, 2$, where the treated subject is identified by $Z_{ij} = 1$ and the control by $Z_{ij} = 0$, so $1 = Z_{i1} + Z_{i2}$ for each i. In pair i, the treatment is given to the treated subject at dose $d_i \geqslant 0$, whereas the untreated control receives a zero or negligible dose. Let r_{ij} be the observed response of the jth subject in pair i.

In a paired randomized experiment, Z_{i1} would be determined by the flip of a fair coin, $\Pr(Z_{i1} = 1) = \frac{1}{2}$, $Z_{i2} = 1 - Z_{i1}$, with flips for distinct pairs being mutually independent, and randomization would create the only probability distributions required for inference—it would form the 'reasoned basis for inference' in Fisher's (1935) sense. Consider, for example, testing the null hypothesis of no treatment effect, which asserts that the response r_{ij} of each subject is not changed by providing or withholding treatment. Then the null distribution of Wilcoxon's signed rank statistic may be derived from random assignment of treatments within pairs, without further assumptions; see Lehmann (1998, Section 3). In contrast, in an observational study, subjects may have differing, unknown probabilities of receiving treatment, so that, for example, the distribution of the signed rank statistic is unknown. Typically, in an observational study, subjects are matched for observed covariates, such as age, in an effort to control biases due to the absence of randomization, but this strategy may not control for unobserved covariates not used in matching. The sensitivity analysis model asserts that two subjects matched for observed covariates may differ in their chances of receiving the treatment by at most a factor of $\Gamma \geqslant 1$,

$$\frac{1}{\Gamma} \leqslant \frac{\Pr(Z_{i1} = 1)}{1 - \Pr(Z_{i1} = 1)} = \frac{\Pr(Z_{i1} = 1)}{\Pr(Z_{i2} = 1)} \leqslant \Gamma \text{ for } i = 1, \ldots, I, \tag{1}$$

with independent treatment assignments in distinct pairs. For $\Gamma = 1$, (1) implies $\Pr(Z_{i1} = 1) = \frac{1}{2}$, yielding conventional randomization inferences. For $\Gamma > 1$, the distribution of treatment assignments is unknown but bounded, and this yields a range of distributions for, say, the signed rank statistic, resulting in a range of inferences, say a range of significance levels. The sensitivity analysis computes this range for several values of Γ, thereby displaying the sensitivity of the inference, that is, the magnitude of hidden bias that would be required to alter the qualitative conclusions of the study. In case-control studies of rare diseases with binary outcomes and exposures, (1) is very similar to the risk ratio inequality of Cornfield *et al.* (1959) together with allowance for sampling variability (Rosenbaum, 1991b, p. 93; 1995, p. 1429), so (1) may be thought of as a generalization of that inequality to cover other situations. Alternatively, in terms of propensity scores, (1) may be derived assuming a logit model for treatment assignment in the population in terms of observed covariates and a bounded unobserved covariate with coefficient $\log(\Gamma)$, with subsequent matching to control the observed covariates (Rosenbaum, 2002, Section 4.2.2).

Consider two test statistics, one that ignores doses—namely, Wilcoxon's signed rank statistic—the other statistic being similar in structure but that uses doses—namely, the univariate version of the coherent signed rank statistic (Rosenbaum, 1997, 2002, Section 9). Let q_i be the rank of the absolute difference $|r_{i1} - r_{i2}|$ in pair i, ranking from 1 to I, with average ranks used for ties between pairs. For the 'signs', write $s_{i1} = 1$ if $r_{i1} > r_{i2}$, $s_{i1} = 0$ otherwise, and write $s_{i2} = 1$ if $r_{i2} > r_{i1}$, $s_{i2} = 0$ otherwise, so that $s_{i+} = s_{i1} + s_{i2} = 0$ if there is a tie within a pair, $r_{i1} = r_{i2}$. Then Wilcoxon's signed rank statistic is the sum of the ranks for pairs in which the treated subject had the higher response, $T = \sum_{i=1}^{I} q_i \sum_{j=1}^{2} Z_{ij} s_{ij}$. The signed rank test with doses is $D = \sum_{i=1}^{I} d_i q_i \sum_{j=1}^{2} Z_{ij} s_{ij}$. The signed rank statistic T weights pairs based on the magnitude of the differences in responses, whereas D uses that magnitude together with the magnitude of the dose. When doses are irrelevant or very poorly measured, T is more useful, but when doses are strongly predictive of differences in responses, D is more useful. See Rosenbaum (1997) for development of properties of D.

Let $\overline{\overline{T}}$ be a random variable formed as the sum of I independent parts, where part i takes the value q_i with probability $\Gamma/(1 + \Gamma)$ and the value 0 otherwise. Define $\overline{T}$ in the same way, but let part i take the value q_i with probability $1/(1 + \Gamma)$ and the value 0 otherwise. Now, under (1), the distribution of the

signed rank statistic T is unknown, but it is possible to show (Rosenbaum, 1987, 2002, Section 4) that

$$\Pr\left(\overline{\overline{T}} \geqslant k\right) \geqslant \Pr\left(T \geqslant k\right) \geqslant \Pr\left(\overline{T} \geqslant k\right) \text{ for every } k, \qquad (2)$$

so for each fixed $\Gamma \geqslant 1$, the unknown distribution of T is sharply bounded by two known distributions. From (2), sharp bounds on significance levels are obtained directly, and from these, by inversion, one obtains bounds on confidence intervals and point estimates for treatment effects (Rosenbaum, 1993, 1995, 2002, Sections 4–5). When $\Gamma = 1$, the upper and lower bounds in (2) are equal to each other and equal to the usual null distribution for Wilcoxon's signed rank test. In parallel, let $\overline{\overline{D}}$ be a random variable formed as the sum of I independent parts, where part i takes the value $d_i q_i$ with probability $\Gamma/(1 + \Gamma)$ and 0 otherwise, and define $\overline{D}$ similarly but with $1/(1 + \Gamma)$ in place of $\Gamma/(1 + \Gamma)$. Then the known distributions of $\overline{\overline{D}}$ and $\overline{D}$ sharply bound the unknown distribution of the dose-signed rank statistic D. For small i, the bounding distributions may be determined exactly, and for large I, with well behaved doses d_i, the central limit theorem yields an approximation. For example, $\overline{\overline{T}}$ has expectation $E\left(\overline{\overline{T}}\right) = \frac{\Gamma}{1+\Gamma} \sum q_i s_{i+}$ and variance $var\left(\overline{\overline{T}}\right) = \frac{\Gamma}{(1+\Gamma)^2} \sum q_i^2 s_{i+}$, both of which reduce to the usual null expectation and variance of the signed rank statistic when $\Gamma = 1$ and there are no ties. Similarly, $E\left(\overline{\overline{D}}\right) = \frac{\Gamma}{1+\Gamma} \sum d_i q_i s_{i+}$ and $var\left(\overline{\overline{D}}\right) = \frac{\Gamma}{(1+\Gamma)^2} \sum (d_i q_i)^2 s_{i+}$. Then as $I \to \infty$, one approximates the upper bound $\Pr\left(\overline{\overline{D}} \geqslant k\right)$ by

$$1 - \Phi\left\{\frac{k - E\left(\overline{\overline{D}}\right)}{\sqrt{var\left(\overline{\overline{D}}\right)}}\right\},$$

where $\Phi\left(\cdot\right)$ is the standard Normal cumulative distribution, with parallel approximations for the other bounds.

See Normand *et al.* (2001), Aakvik (2001), Davanzo *et al.* (2001) and Li *et al.* (2001) for four recent applications of this approach to sensitivity analysis. There are many other forms of sensitivity analysis; see, for instance, Bross (1967), Schlesselmann (1978), Rosenbaum and Rubin (1983), Rosenbaum (1986), Manski (1990), Balke and Pearl (1997), Copas and Li (1997), Gastwirth *et al.* (1998b), Lin *et al.* (1998), Robins *et al.* (1999), and Copas and Eguchi (2001). In particular, Gastwirth *et al.* (1998b, henceforth GKR) use two sensitivity parameters, Γ and Δ, where Γ describes, as here, the strength of the relationship between an unobserved covariate and treatment assignment, and Δ describes the strength of the relationship between the response r_{ij} and the unobserved covariate. The statistics T and D are both of the type covered in that paper (GKR, expression 1, p. 908), so that two-dimensional sensitivity analysis may alternatively be applied to appraise whether a dose response relationship reduces sensitivity to hidden bias. In the very simplest situations, such as McNemar's test for paired binary responses, a simultaneous sensitivity analysis with two parameters, Γ and Δ, is identical to a sensitivity analysis based on (1) alone, but with a different value of Γ, so in these cases, the one and two parameter sensitivity analyses might be thought of as using different units of measurement to measure the same thing (e.g. inches versus meters) rather than as measuring different things (e.g. meters versus kilograms).

Does a dose–response relationship reduce sensitivity to hidden bias?

3. EXAMPLE: CYTOGENETIC DAMAGE AMONG PAINTERS

Professional painters are regularly exposed to varied substances, including lead, in paint and paint thinners, and these exposures might cause genetic damage. Pinto *et al.* (2000) compared male public

Table 1. *Micronuclei/1000 for 22 Painters and Age-Matched Controls, All Male, with Years of Exposure for Painters, and Their Base 2 Logs*

Age painter	Age control	Years	$\log_2$	Micronuclei painter	Micronuclei control
18	18	1.6	0.68	0.32	0.00
20	18	2.0	1.00	0.00	0.65
20	20	1.6	0.68	0.00	0.00
22	22	4.7	2.23	0.66	0.00
22	22	4.0	2.00	0.99	0.33
23	23	5.0	2.32	0.00	0.00
23	23	3.0	1.58	0.95	0.00
26	27	2.0	1.00	0.00	0.00
29	29	1.8	0.85	0.32	0.32
30	30	3.0	1.58	0.00	0.62
31	31	3.5	1.81	0.33	1.96
35	36	8.0	3.00	0.65	0.00
40	39	4.6	2.20	0.33	0.00
40	39	1.6	0.68	0.00	0.00
41	40	25.0	4.64	1.99	1.33
42	42	5.0	2.32	1.63	0.66
47	47	4.0	2.00	1.54	0.00
48	49	8.0	3.00	1.64	0.64
52	51	3.6	1.85	1.99	0.00
60	58	8.6	3.10	4.84	0.00
60	63	40.0	5.32	2.89	0.32
62	64	11.0	3.46	5.35	1.30

building painters in Merida, Yucatan, Mexico to male clerks, matched for age. The painters worked without masks or gloves.

Table 1 presents 22 matched pairs from Pinto *et al.* (2000). The pairs of men are matched for age. The outcome is a measure of cytogenetic damage, namely micronuclei in oral epithelial cells (MN/1000). The measure of dose is the number of years of exposure for the painters. Obviously, years of exposure and age are related—a 60 year old painter can have 40 years of exposure, but an 18 year old painter cannot—hence the importance of matching on age. Because of the extreme skewness, the $\log_2$ of years of exposure is used as the dose. (It is easy to view log's with base 2, because magnitudes refer to doublings: for painters #8 and #12, $\log_2(8) - \log_2(2) = 3 - 1 = 2$ says that painter #8's exposure must be doubled twice to yield painter #12's exposure. In Pinto *et al.* (2000), there are 24 painters and 23 controls 'frequency matched' for age, although one painter was missing years of exposure and another was missing the micronuclei measure, leaving 22 painters. In Table 1, the 22 painters are individually pair matched for age with 22 controls.)

Notice the pattern in Table 1. For many individuals, there are no micronuclei, and low frequencies are also common among both painters and controls. Eight painters and three controls have MN/1000 of 1.3 or more, and all of these men are over 30 years old, with the highest values among painters with long exposure.

The research design in Table 1 is a dose-control design, in which exposed subjects have doses, and there are controls with negligible doses. It is a common design in occupational and environmental health, although the quality and relevance of the dose information can vary markedly from study to study.

Table 2 presents two sensitivity analyses for the painter data, one ignoring the doses using the signed

Table 2. *Sensitivity analysis for significance levels,*
with and without doses

Γ	Ignoring doses		Using doses	
1	max	0.0032	max	0.0025
	min	0.0032	min	0.0025
2	max	0.064	max	0.038
	min	0.0000096	min	0.000014
2.2	max	0.085	max	0.049
	min	0.0000031	min	0.0000052
2.5	max	0.12	max	0.067
	min	0.00000056	min	0.0000016

rank statistic, T, the other incorporating doses using the dose-signed rank statistic, D. The analysis presents upper and lower bounds on the one-sided significance levels derived from the two tests, for several possible magnitudes of hidden bias, Γ. If there is no hidden bias—that is, if $\Gamma = 1$ as in a randomized experiment—then the upper and lower bounds are equal to each other, and they equal the usual significance level from the randomization distribution. If there were no hidden bias, the effect of exposure would be highly significant whether doses are used or ignored, although the significance level is slightly smaller with doses. If $\Gamma = 2$, then matched subjects may not have equal chances of exposure because they differ in unobserved ways; rather, one subject might be twice as likely as the other to be exposed. In this case, the analysis without doses yields both significance levels slightly above and substantially below the conventional one-sided 0.05 level, so, in this specific sense, it is sensitive to a bias of magnitude $\Gamma = 2$, but the analysis with doses yields only levels below 0.05, so it is insensitive to a bias of magnitude $\Gamma = 2$. For $\Gamma = 2.2$, there is a similar pattern, the maximum significance level being 0.085 without doses and 0.049 with doses.

In this one study, a dose–response relationship has, to a moderate extent, reduced sensitivity to hidden bias. In other words, somewhat larger hidden biases would be required to explain the dose–response relationship than to explain the exposed-control relationship ignoring doses. The qualitative impression that dose is associated with response has been replaced by a quantitative finding: in this particular study, the dose–response relationship has reduced sensitivity to hidden bias, but the magnitude of the reduction, though not trivial, is far from dramatic.

A stronger association between dose and response in Table 1 would have produced a greater reduction in sensitivity to hidden bias. Specifically, if the 22 treated-minus-control differences in responses had been ordered in the same way as the doses, then the upper bound on the significance level from D would equal 0.05 at $\Gamma = 2.8$. In contrast, ignoring doses, the upper bound on the significance level using the signed rank statistic is 0.16 for $\Gamma = 2.8$. A bias of magnitude $\Gamma = 2.8$ could explain the difference in responses between exposed and control subjects, but it could not also explain such a strong relationship between dose and response. A dose–response relationship has the potential to substantially reduce sensitivity to hidden bias, but an analysis is needed to determine whether and to what extent a reduction in sensitivity has occurred. The Appendix demonstrates that this is generally true and not an idiosyncrasy of these data: when doses are replaced by their ranks, at most half of the $I!$ arrangements of the doses lead to a reduction in sensitivity when compared to ignoring the doses, and this is true even if the difference between treated subjects and controls is highly significant and insensitive to bias for all $I!$ rearrangements of the doses.

Does the absence of a dose–response relationship indicate a high degree of sensitivity to hidden bias? It may not. Recall that Rothman (1986) gave the association between DES and vaginal cancer as an example of an association, believed today to be causal, for which there is no dose–response relationship. That

association is much less sensitive to hidden bias than the relationship described in Table 2. Specifically, the study by Herbst *et al.* (1971) of DES and vaginal cancer becomes sensitive to hidden bias at about $\Gamma = 7$, so only enormous hidden biases could explain the association; see Rosenbaum (2002, Section 4) for details. In other words, the dose–response relationship in Table 1 somewhat reduces sensitivity to hidden bias in the study of genetic damage among painters, but even the strengthened evidence is much weaker than the evidence in a study where a dose–response relationship was absent. In short, a dose–response relationship in one study may reduce sensitivity to hidden bias in that study, but another study without such a relationship may, nonetheless, be far less sensitive to hidden bias.

Does a dose–response relationship always reduce sensitivity to hidden bias? It does not. In the example in Rosenbaum (1997), there is a dose–response relationship: extremely high doses are often found but not always found with extremely high responses, but the pattern is sufficiently erratic that the dose–response relationship does not reduce sensitivity to hidden bias. That is, in this case, hidden biases that would explain the exposed-control relationship ignoring doses would also be sufficient to also explain the dose–response relationship—there is no reduction in sensitivity to hidden bias.

4. SUMMARY

A dose–response relationship is relevant to thinking about hidden biases, but the mere presence or absence of such a relationship provides little guidance. As shown in examples, a dose–response relationship may reduce sensitivity to hidden bias, or it may not, and if it does the reduction may be either substantial or slight; in any particular study, this is easily determined by sensitivity analysis. A study may lack a dose–response relationship yet be highly insensitive to hidden bias, and this too is easily determined.

ACKNOWLEDGMENT

APPENDIX: THE BEHAVIOR OF D AND T AS A FUNCTION OF THE DOSE PATTERN

It was argued that a dose–response relationship, even if statistically significant and perhaps insensitive to hidden bias, may or may not reduce sensitivity when compared to an analysis that ignores doses, so that one needs to check in each study whether a reduction in sensitivity has been obtained. In other words, even when the test based on D is statistically significant and perhaps insensitive to bias, it is not automatically less sensitive than the test based on T: whether or not there is reduced sensitivity depends on the strength of the association between dose and response. This appendix briefly restates the point formally, thereby demonstrating that this observation is not an idiosyncrasy of particular data sets. In a sense that is formalized below, *at most* half of the $I!$ arrangements of the I doses yields a reduction in sensitivity to hidden bias when compared to ignoring the doses, even when all $I!$ arrangements yield highly significant results that are insensitive to bias.

Viewed as a function of the doses, $\mathbf{d} = (d_1, \ldots, d_I)^T$, the standardized deviate, say $\widetilde{D}$, for upper bound on the the dose-signed rank statistic, $\overline{\overline{D}}$, is

$$\widetilde{D} = h(\mathbf{d}) = \frac{\sum \left(B_i - \frac{\Gamma}{1+\Gamma} \right) d_i \, q_i \, s_{i+}}{\sqrt{\frac{\Gamma}{(1+\Gamma)^2} \sum (d_i \, q_i)^2 \, s_{i+}}}, \tag{3}$$

where $B_i = \sum_{j=1}^{2} Z_{ij} \, s_{ij}$. The standardized deviate, say $\widetilde{T}$, for the upper bound on the usual signed rank statistic, $\overline{\overline{T}}$, is obtained from this formula when the doses are constant, $d_1 = \ldots = d_I$.

One way to understand how $\widetilde{D}$ and $\widetilde{T}$ are related for fixed data is to understand how $h(\mathbf{d})$ behaves as a function of $\mathbf{d}$. The doses were assumed to be nonnegative; assume now that at least one dose is strictly positive, $0 < \max_i d_i$, in at least one untied pair, $s_{i+} = 1$. This minor assumption ensures that the denominator in (3) is not zero. Now, $h(\mathbf{d})$ is the ratio of an affine function of $\mathbf{d}$ and a positive convex function of $\mathbf{d}$, so $h(\mathbf{d})$ is concave (Avriel *et al.*, 1988, p. 213). Hence, $h\{\lambda \mathbf{d} + (1-\lambda)\mathbf{d}^*\} \geqslant \lambda h(\mathbf{d}) + (1-\lambda) h(\mathbf{d}^*)$ for $0 \leqslant \lambda \leqslant 1$.

To discuss a simple, clear example, suppose the doses are replaced by their untied ranks when used in (3). (Incidentally, if tied ranks were used for doses instead of $\log_2$'s, the upper bound on the significance level using D for $\Gamma = 2.5$ is reduced from 0.067 in Table 2 to 0.057.) For any untied, ranked doses, say $\mathbf{d}$, there is a reversed version, $\mathbf{d}^*$, in which $d_i^* = I + 1 - d_i$, so with $I = 22$ as in Table 2, rank 22 is reversed to rank 1, rank 21 is reversed to rank 2, and so on. Moreover, $\lambda \mathbf{d} + (1-\lambda)\mathbf{d}^*$ is constant for $\lambda = \frac{1}{2}$, so $\widetilde{T} = h\left\{\frac{1}{2}(\mathbf{d} + \mathbf{d}^*)\right\}$. From the concavity of $h(\cdot)$, it follows that $\widetilde{T} = h\left\{\frac{1}{2}(\mathbf{d} + \mathbf{d}^*)\right\} \geqslant \frac{1}{2}\{h(\mathbf{d}) + h(\mathbf{d}^*)\}$, that is, the deviate for the signed rank statistic $\widetilde{T}$ is at least as large as the average of the two deviates, $h(\mathbf{d})$ and $h(\mathbf{d}^*)$, for the dose-signed rank statistic with, respectively, doses $\mathbf{d}$ and $\mathbf{d}^*$. This means that $\widetilde{T}$ is at least as large as the smaller of $h(\mathbf{d})$ and $h(\mathbf{d}^*)$, and if $\widetilde{T}$ is strictly smaller than one of these dose-signed rank statistics, then it is strictly larger than the other. If exposed subjects usually have higher responses than controls, then most B_i equal 1, and both $h(\mathbf{d})$ and $h(\mathbf{d}^*)$ may be large, statistically significant and insensitive to bias—that is, one may have both $h(\mathbf{d}) \geqslant 1.65$ and $h(\mathbf{d}^*) \geqslant 1.65$ for quite large Γ—and yet $\widetilde{T}$ is at least as large as $\min\{h(\mathbf{d}), h(\mathbf{d}^*)\}$, so the signed rank statistic ignoring doses is at least as insensitive to hidden bias as one of these two. Now the $I!$ rearrangements of the dose ranks partition into $(I!)/2$ pairs, $\mathbf{d}$ and $\mathbf{d}^*$, where $\mathbf{d}^*$ reverses the ranks in $\mathbf{d}$. In each of these pairs, $\widetilde{T}$ is at least as large as the smaller $\widetilde{D}$. It follows that for at least half of the rearrangements of the dose ranks, the upper bound on the deviate $\widetilde{T}$ for the signed rank statistic that ignores the doses is at least as large as the upper bound on the deviate $\widetilde{D}$ for the dose-signed rank statistic. Moreover $\widetilde{T} \geqslant \widetilde{D} = h(\mathbf{d})$ for at least $(I!)/2$ of the $I!$ dose rearrangements even if $h(\mathbf{d}) \geqslant 1.65$ for quite large Γ and for all of the $I!$ dose rearrangements. This is one way of formalizing the claim, stated in the body of the paper, that a statistically significant and insensitive dose–response relationship may or may not lead to a reduction in sensitivity to hidden bias; indeed, a reduction takes place for at most half of the $I!$ dose arrangements.

REFERENCES

AAKVIK, A. (2001). Bounding a matching estimator: the case of a Norwegian training program. *Oxford Bull. Econ. Statist.* **63**, 115–143.

AVRIEL, M., DIEWERT, W. E., SCHAIBLE, S. AND ZANG, I. (1988). *Generalized Concavity*. New York: Plenum.

BALKE, A. AND PEARL, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association* **92**, 1172–1176.

BROSS, I. D. J. (1967). Pertinency of an extraneous variable. *Journal of Chronic Diseases* **20**, 487–495.

COPAS, J. B. AND LI, H. G. (1997). Inference for non-random samples (with discussion). *Journal of the Royal Statistical Society,* Series B **59**, 55–96.

COPAS, J. AND EGUCHI, S. (2001). Local sensitivity approximations for selectivity bias. *Journal of the Royal Statistical Society,* Series B **63**, 871–896.

CORNFIELD, J., HAENSZEL, W., HAMMOND, E. *et al.* (1959). Smoking and lung cancer: Recent evidence and a discussion of some questions. *Journal of the National Cancer Institute* **22**, 173–203.

DAVANZO, P., THOMAS, M. A., YUE, K., OSHIRO, T., BELIN, T., STROBER, M. AND McCRACKEN, J. (2001). Decreased anterior cingulate myo-inositol/creatine spectoscopy resonance with lithium treatment in children with bipolar disorder. *Neuropsychopharmacology* **24**, 359–369.

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

GASTWIRTH, J. L., KRIEGER, A. M. AND ROSENBAUM, P. R. (1998a). Cornfield's inequality. Armitage, P. and Colton, T. (eds), *Encyclopedia of Biostatistics*. New York: Wiley, pp. 952–955.

GASTWIRTH, J. L., KRIEGER, A. M. AND ROSENBAUM, P. R. (1998b). 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,* Series B **62**, 545–555.

GREENHOUSE, S. (1982). Jerome Cornfield's contributions to epidemiology. *Biometrics* Supplement, 33–45.

HERBST, A., ULFELDER, H. AND POSKANZER, D. (1971). Adenocarcinoma of the vagina: Association of maternal stilbestrol therapy with tumor appearance in young women. *New England Journal of Medicine* **284**, 878–881.

HILL, A. B. (1965). The environment and disease: Association or causation? *Proc. Roy. Soc. Med.* **58**, 295–300.

LEHMANN, E. L. (1998). *Nonparametrics: Statistical Methods Based on Ranks*, Revised first edition. Upper Saddle River, NJ: Prentice-Hall.

LI, Y. P., PROPERT, K. J. AND ROSENBAUM, P. R. (2001). Balanced risk set matching. *Journal of the American Statistical Association* **96**, 870–882.

LIN, D. Y., PSATY, B. M. AND KRONMAL, R. A. (1998). Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics* **54**, 948–963.

MANSKI, C. (1990). Nonparametric bounds on treatment effects. *American Economic Review* 319–323.

NORMAND, S. T., LANDRUM, M. B., GUADAGNOLI, E. *et al.* (2001). Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: A matched analysis using propensity scores. *Journal of Clinical Epidemiology* **54**, 387–398.

PINTO, D., CEBALLOS, J. M., GARCIA, G. *et al.* (2000). Increased cytogenetic damage in outdoor painters. *Mutation Research* **467**, 105–111.

ROBINS, J. M., ROTNITZKEY, A. AND SCHARFSTEIN, D. (1999). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In Halloran, E. and Berry, D. (eds), *Statistical Models in Epidemiology*, New York: Springer, pp. 1–94.

ROSENBAUM, P. R. (1986). Dropping out of high school in the United States: An observational study. *J. Ed. Statist.* **11**, 207–224.

ROSENBAUM, P. R. (1987). Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika* **74**, 13–26.

ROSENBAUM, P. R. (1991a). Discussing hidden bias in observational studies. *Annals of Internal Medicine* **115**, 901–905.

ROSENBAUM, P. R. (1991b). Sensitivity analysis for case-control studies. *Biometrics* **47**, 87–100.

ROSENBAUM, P. R. (1993). Hodges–Lehmann point estimates of treatment effect in observational studies. *Journal of the American Statistical Association* **88**, 1250–1253.

ROSENBAUM, P. R. (1995). Quantiles in nonrandom samples and observational studies. *Journal of the American Statistical Association* **90**, 1424–1431.

ROSENBAUM, P. R. (1997). Signed rank statistics for coherent predictions. *Biometrics* **53**, 556–566.

ROSENBAUM, P. R. (2002). *Observational Studies*, Second Edition. New York, NY: Springer.

ROSENBAUM, P. R. AND RUBIN, D. B. (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society,* Series B **45**, 212–218.

ROTHMAN, K. J. (1986). *Modern Epidemiology*. Boston, MA: Little, Brown.

SCHLESSELMANN, J. J. (1978). Assessing the effects of confounding variables. *American Journal of Epidemiology* **108**, 3–8.

WEISS, N. (1981). Inferring causal relationships: Elaboration of the criterion of 'dose–response'. *American Journal of Epidemiology* **113**, 487–490.