

Exact, Nonparametric Inference When Doses Are Measured With Random Errors

Paul R. ROSENBAUM

Studies that estimate the effects of exposure to a possibly harmful agent often compare exposed subjects who received varied doses with matched controls who received zero dose. If the doses are measured with error, then one may wish to use the fallible doses to estimate a linear relationship between the unobserved true dose and the observed response. If one is willing to assume that the dose errors for exposed subjects are symmetrically distributed about 0—that is, the dose errors are pure errors and not, say, systematic underreporting of exposure—then the presence of zero-dose controls is all that is needed to obtain exact, distribution-free confidence intervals and tests, and consistent point estimates. The method is simpler for matched pairs than for matched sets with two or more matched subjects, and it is illustrated using two studies, one of each kind. With matched pairs, as in this first example, the method uses Wilcoxon's signed rank test as the basis for inference. When there are several zero-dose controls matched to each exposed subject, the familiar null distribution of the signed rank statistic is no longer applicable because of dependence within matched sets, so the appropriate exact distribution and large-sample approximation are developed.

KEY WORDS: Dose-control design; Dose-response; Errors in variables; Full matching; Multiple controls; Observational studies; Wilcoxon's signed rank test.

1. THE DOSE-CONTROL DESIGN

1.1 Introduction: Motivation and Outline

In the dose-control design, the effects of exposure to a possibly harmful agent are estimated by comparing exposed subjects who received varied doses to matched controls who received zero dose. The doses are often measured with error, and one may wish to use the fallible doses to estimate a linear relationship between the unobserved true dose and the observed response. This is a common but special case of the traditional problem of fitting a line when both variables are subject to error. Even if one were willing to assume that the dose errors are normally distributed with expectation 0, the solution to the traditional problem is usually thought to require some additional information, such as duplicate measures of the doses with independent errors, true doses for a subsample of individuals, knowledge of the theoretical reliability of the dose measurements, or an instrumental variable. Madansky (1959), Kendall and Stuart (1973, sec. 29), Fuller (1987, sec. 1), and Cheng and Van Ness (1999) have provided detailed discussion of these techniques. It turns out, however, that if one is willing to assume that the dose errors for exposed subjects are symmetrically distributed about 0—that is, the dose errors are pure errors and not, say, systematic underreporting of exposure—then the presence of zero-dose controls is all that is needed to obtain exact, distribution-free confidence intervals and tests and consistent point estimates.

The method is simpler for matched pairs, discussed in Section 2, than for matched sets with two or more matched subjects, discussed in Section 3, and the method is illustrated using two studies, one of each kind. The first study discussed in Section 1.2 concerns the dose-response relationship between the average number of cigarettes smoked per day and the level of a tobacco-specific carcinogen, NNK, in the cervical mucus of female smokers and nonsmokers. With matched pairs, as in this first example, the method uses Wilcoxon's signed rank test as the basis for inference; see Sections 2.1 and 2.2. When there are several zero-dose controls matched to each exposed subject, the

familiar null distribution of the signed rank statistic is no longer applicable because of dependence within matched sets, so the appropriate exact distribution and large-sample approximation are developed in Section 3.2. Other matched designs (e.g., with one control matched to several exposed subjects) permit similar analyses, as is briefly discussed in Sections 3.3 and 3.4. The method for matched sets is illustrated using a study of the relationship between radiation exposure and cytogenetic damage in Section 3.5.

The methods proposed herein are contrasted with other methods in Sections 2.4 and 3.6. In particular, in Section 2.4 a simple, special case is considered, with many additional assumptions involving normal distributions and certain types of independence, thereby obtaining a procedure based on the *t*-test rather than the signed rank test. Questions of efficiency, robustness, and sample size are then discussed. With multiple controls, the traditional test is not the signed rank test discussed in Section 3.2, but rather the aligned rank test of Hodges and Lehmann (1962). Unfortunately, the aligned rank test is not applicable with dose errors, and Section 3.6 explains why. Several important limitations of the method are discussed in Section 4.

1.2 Example: A Matched Dose-Control Study

Concerned that smoking may contribute to cervical cancer, Prokopczyk, Cox, Hoffman, and Waggoner (1997) asked whether a potent carcinogen specific to tobacco, namely nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) was found in higher concentrations in the cervical mucus of female smokers than in female nonsmokers. Table 1 describes nine pairs of women matched for age, one smoker and one nonsmoker, recording NNK levels (in ng/g), along with self-reported average frequency of smoking for the smokers, recorded as average number cigarettes per day. Presumably, the self-reported number of cigarettes per day for smokers is, at best, a fallible measure of true average daily cigarette consumption.

Paul R. Rosenbaum is Robert G. Putzel Professor, Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104 (E-mail: rosenbaum@stat.wharton.upenn.edu). This work was supported by grant SES-0345113 from the Methodology, Measurement, and Statistics Program and the Statistics and Probability Program of the National Science Foundation.

Table 1. NNK in Cervical Mucus (ng/g) of Nine Smokers and Nine Age-Matched Nonsmoking Controls

Pair	Smoker cigarettes/day	Smoker NNK	Nonsmoker NNK
1	7.5	25.9	9.7
2	10.0	84.9	6.7
3	4.0	115.0	19.1
4	5.0	19.7	6.0
5	10.0	42.7	18.3
6	5.0	30.8	4.1
7	20.0	64.2	10.6
8	10.0	58.0	30.8
9	20.0	43.0	22.7

SOURCE: Prokopczyk et al. (1997).

1.3 Dose Errors in the Dose-Control Design

There are I pairs, $i = 1, \dots, I$, of two subjects, $j = 1, 2$, receiving true doses $x_{ij} > 0$ and yielding responses

$$Y_{ij} = \alpha_i + \beta x_{ij} + \varepsilon_{ij}, \quad (1)$$

where β , α_i , and x_{ij} are fixed quantities and the errors ε_{ij} are continuous, mutually independent and identically distributed (iid), that is,

$$\varepsilon_{ij} \stackrel{\text{iid}}{\sim} F \in \mathcal{C}, \quad (2)$$

where \mathcal{C} is the set of continuous cumulative distribution functions. In the dose-control design, the control $j = 2$ is known to have received dose $x_{i2} = 0$, but the dose x_{i1} for the exposed subject, $j = 1$, is not observed, and instead $X_{i1} > 0$ is observed,

$$X_{i1} = x_{i1} + \eta_{i1}, \quad (3)$$

where the dose errors η_{i1} are mutually independent and independent of the ε 's, identically distributed from a continuous distribution symmetric about 0, that is,

$$\eta_{i1} \stackrel{\text{iid}}{\sim} H \in \mathcal{S}, \quad (4)$$

where \mathcal{S} is the set of continuous distributions symmetric about 0.

In terms of Section 1.2, the model (1) says the typical response may vary with α_i from one pair i to another i' , because subjects in different pairs have different ages, and that NNK levels rise proportionally with the true average number of cigarettes smoked per day, x_{ij} . Also, (3) says that the reported dose levels X_{i1} for smokers are fallible, that they deviate from the true dose levels, x_{i1} , by an error of measurement, η_{i1} , that is symmetric about 0. Although NNK is specific to tobacco, one can be exposed to tobacco while smoking $x_{ij} = 0$ cigarettes through, for instance, passive smoking (i.e., inhaling the smoke from a cigarette that someone else is smoking), so (1) has $Y_{ij} = \alpha_i + \varepsilon_{ij}$ when $x_{ij} = 0$, not $Y_{ij} = 0$.

For exposed subjects, $j = 1$, the model, (1) and (3), is a matched nonparametric version of the traditional problem of errors in variables (see, e.g., Wald 1940; Neyman and Scott 1951; Madansky 1959; Kendall and Stuart 1973, sec. 29; Fuller 1987, sec. 1) and it is well known that conventional ways of regressing Y_{i1} on X_{i1} do not generally provide consistent inferences about β , even if all the nuisance parameters, the α 's are equal, $\alpha_1 = \dots = \alpha_I$; see Section 2.4 for some specifics. In essence,

the problem is that even when the null hypothesis $H_0: \beta = \beta_0$ is true for $\beta_0 \neq 0$, the residuals,

$$\begin{aligned} Y_{i1} - \beta_0 X_{i1} &= (\alpha_i + \beta x_{i1} + \varepsilon_{i1}) - \beta_0(x_{i1} + \eta_{i1}) \\ &= \alpha_i + \varepsilon_{i1} - \beta_0 \eta_{i1} \end{aligned} \quad (5)$$

are not independent of the predictors, $X_{i1} = x_{i1} + \eta_{i1}$, because both involve η_{i1} . Regression estimators, such as least squares or Theil's nonparametric slope estimator (Hollander and Wolfe 1999, sec. 9.3), are typically inconsistent when x_{i1} is measured with error, because the standard measures try to find a value θ that makes the residuals $Y_{i1} - \theta X_{i1}$ independent of the predictor X_{i1} , but if θ does this, then $\theta \neq \beta$.

Various solutions to inference about β have been proposed. Some of these involve replicate measures, X_{ij} and \tilde{X}_{ij} with the same true x_{ij} and independent errors, $X_{ij} = x_{ij} + \eta_{ij}$, $\tilde{X}_{ij} = x_{ij} + \tilde{\eta}_{ij}$, whereas others involve measuring the true x_{ij} for a subsample of units, and still others involve finding instrumental variables (see Madansky 1959; Kendall and Stuart 1973, sec. 29; Fuller 1987, sec. 1, for surveys of these and other approaches). As it turns out, however, the presence of zero-dose controls will provide everything needed for exact, nonparametric inference about β under the model (1) and (3), without replicate doses or correct doses for a subsample. The method has a slight relationship to an idea of Wald (1940), but without the issues raised by Neyman and Scott (1951), and moreover yields exact inferences.

2. INFERENCE IN THE DOSE-CONTROL DESIGN

2.1 Hypothesis Tests

To test the null hypothesis $H_0: \beta = \beta_0$, calculate

$$\begin{aligned} Z_{i,\beta_0} &= (Y_{i1} - \beta_0 X_{i1}) - Y_{i2} \\ &= (\alpha_i + \beta x_{i1} + \varepsilon_{i1}) - \beta_0(x_{i1} + \eta_{i1}) - (\alpha_i + \varepsilon_{i2}) \\ &= (\beta - \beta_0)x_{i1} + (\varepsilon_{i1} - \varepsilon_{i2} - \beta_0 \eta_{i1}) \end{aligned} \quad (6)$$

for $i = 1, \dots, I$. Now the $\varepsilon_{i1} - \varepsilon_{i2} - \beta_0 \eta_{i1}$ are continuous, mutually independent, and symmetric about 0. If $H_0: \beta = \beta_0$ is true, then the Z_{i,β_0} themselves are continuous, mutually independent, and symmetric about 0, so a distribution-free, exact test of this null hypothesis is obtained by applying Wilcoxon's signed rank test to the Z_{i,β_0} (see Hettmansperger 1984, sec. 2, for discussion of the signed rank test).

It is easy to see that this test is consistent, as $I \rightarrow \infty$, providing that the doses x_{i1} do not get smaller and smaller as the sample size increases. Recall that Wilcoxon's signed rank statistic equals the number of positive Walsh averages, $(Z_{i,\beta_0} + Z_{k,\beta_0})/2$, for $i \leq k$, and, under the null hypothesis, the chance that a Walsh average is positive is $\frac{1}{2}$ (see, e.g., Hettmansperger 1984, sec. 2). If the null hypothesis is false with $\beta > \beta_0$, then each Z_{i,β_0} is symmetric about a positive quantity, namely $(\beta - \beta_0)x_{i1}$, so that $\Pr(Z_{i,\beta_0} + Z_{k,\beta_0} > 0) > \frac{1}{2}$ for every $i \leq k$. If as $I \rightarrow \infty$, there is a positive constant $\kappa > 0$ such that at least a fraction $\lambda > 0$ of the exposed subjects have $x_{i1} > \kappa > 0$, then the proportion of positive Walsh averages will tend to a quantity strictly greater than $\frac{1}{2}$ and its variance will drop to 0, and the probability that the signed rank test will reject H_0 will tend to 1. A parallel argument works for $\beta < \beta_0$.

Remark 1. The distributional assumptions of Section 1.3 are more than is needed; they are sufficient but not necessary. Specifically, the argument of this section requires only that the I quantities $\varepsilon_{i1} - \varepsilon_{i2} - \beta\eta_{i1}$ be continuous, mutually independent, and symmetric about 0, and this can happen in other ways. For instance, in Section 1.3 the $2I$ errors in (1), namely ε_{i1} and ε_{i2} , $i = 1, \dots, I$, were assumed to be iid, which implies that $\varepsilon_{i1} - \varepsilon_{i2}$ is symmetric about 0. If, instead, the $2I$ errors in (1) were mutually independent and continuous, and if the errors ε_{i1} for exposed subjects came from one distribution $F_E \in \mathcal{S}$ symmetric about 0 while the errors ε_{i2} for control subjects came from a different distribution $F_C \in \mathcal{S}$ symmetric about 0, then $\varepsilon_{i1} - \varepsilon_{i2}$ would still be symmetric about 0, and this would also suffice. For example, this would permit the exposed subjects to have responses Y_{i1} exhibiting greater dispersion than the responses of controls Y_{i2} , quite apart from the varied doses x_{i1} . More generally, under certain models, ε_{i1} and ε_{i2} can be both asymmetric and unequal in dispersion. For instance, suppose that $\varepsilon_{ij} = \bar{\varepsilon}_{ij} + \tilde{\varepsilon}_{ij}$, where all of the $\bar{\varepsilon}_{ij}$ and $\tilde{\varepsilon}_{ij}$ are independent and

$$\begin{aligned} \bar{\varepsilon}_{ij} &\stackrel{\text{iid}}{\sim} F \in \mathcal{C}, & j = 1, 2; \\ \tilde{\varepsilon}_{i1} &\stackrel{\text{iid}}{\sim} F_E \in \mathcal{S}; & \text{and} & \tilde{\varepsilon}_{i2} \stackrel{\text{iid}}{\sim} F_C \in \mathcal{S}; \end{aligned}$$

then ε_{i1} and ε_{i2} may have distributions that are asymmetric with different shapes and dispersions, yet $\varepsilon_{i1} - \varepsilon_{i2}$ and $\varepsilon_{i1} - \varepsilon_{i2} - \beta\eta_{i1}$ are symmetric about 0.

2.2 Confidence Intervals and Point Estimates

An exact, distribution-free, $100(1 - \alpha)\%$ confidence set for β is obtained by inverting the test in Section 2.1 (e.g., Lehmann 1959, sec. 3.5); that is, the confidence set is the set of hypotheses $H_0: \beta = \beta_0$ not rejected at level α . Because $X_{i1} > 0$, the difference Z_{i,β_0} declines steadily as β_0 increases, which implies that the confidence set is always an interval. Because the test is consistent, the length of the interval tends to 0 in probability as $I \rightarrow \infty$.

Inverting the test also yields a consistent point estimate by the method of Hodges and Lehmann (1962). Specifically, the estimator is approximately the value, $\hat{\beta}$, such that the signed rank statistic computed from $Z_{i,\hat{\beta}}$ equals the null expectation of the signed rank statistic, namely $I(I + 1)/4$. More precisely, as a function of β_0 , the signed rank statistic computed from Z_{i,β_0} is a decreasing step function. As a result, the estimator $\hat{\beta}$ is either the unique point where this step function passes $I(I + 1)/4$ or the midpoint of the interval where it equals $I(I + 1)/4$.

2.3 Example: Dose-Response for NNK

In Section 1.2, β is ng/g of NNK per cigarette per day. The confidence interval for β in this example is computed as follows. With a sample of size $I = 9$, under the null hypothesis, Wilcoxon’s signed rank statistic is 40 or more with probability .020 and is 5 or less with probability .020. When computed from the Z_{i,β_0} in (6) with $\beta_0 = 1.8475$, the signed rank statistic is 40, so $H_0: \beta = 1.8475$ is rejected at level .020; but when computed with $\beta_0 = 1.8475001$, the signed rank statistic is 39, so that this hypothesis is not rejected at level .020. Similarly, when computed from the Z_{i,β_0} in (6) with $\beta_0 = 8.79286$, the signed rank statistic is 5, so $H_0: \beta = 8.79286$ is rejected at level .020;

but when computed with $\beta_0 = 8.79285$, the signed rank statistic is 6, so that this hypothesis is not rejected at level .020. It follows that a $96\% = 100(1 - 2 \times .020)\%$ confidence interval for β is $[1.847, 8.793]$.

To compute $\hat{\beta}$, calculate the null expectation of the signed rank statistic, $I(I + 1)/4 = 9(9 + 1)/4 = 22.5$. If the signed rank statistic is computed from Z_{i,β_0} then with $\beta_0 = 3.211$ the value is 23, whereas with $\beta_0 = 3.212$ it is 22, so $\hat{\beta} = 3.21$.

2.4 A Normal Model: Efficiency, Sample Size, and Robustness

This section discusses a very special case of the model in Section 1.3 as described in Remark 1, in which all quantities are sampled from normal distributions, including the errors, ε_{ij} and η_{i1} , the doses, x_i , and the pair effects, α_i . This model is useful in certain theoretical comparisons; it is not offered for use as a practical model. That is, the normal model and related models in this section make stronger assumptions than are needed for validity of the method, so these assumptions should not be made in practice, but the assumptions are useful in studying relative efficiency, power, robustness, and other theoretical properties of the method. How well does the method perform in simple situations?

Specifically, in the current section only, suppose that the unobservable quantities are normal, $\varepsilon_{i1} \stackrel{\text{iid}}{\sim} N(0, \sigma_{\varepsilon_1}^2)$, $\varepsilon_{i2} \stackrel{\text{iid}}{\sim} N(0, \sigma_{\varepsilon_2}^2)$, $\eta_{i1} \stackrel{\text{iid}}{\sim} N(0, \sigma_{\eta}^2)$, $x_{i1} \stackrel{\text{iid}}{\sim} N(\mu_x, \sigma_x^2)$, and $\alpha_i \stackrel{\text{iid}}{\sim} N(\mu_\alpha, \sigma_\alpha^2)$, where all of these quantities are mutually independent and the observable data, (Y_{i1}, Y_{i2}, X_{i1}) , are generated from (1) and (3) as before. Because the doses x_i are always positive in the dose-control design, this normal model can be a reasonable approximation only when $\mu_x > 0$ and μ_x/σ_x is not small, perhaps 3 or more. The covariance of (Y_{i1}, Y_{i2}) is then σ_α^2 .

With these added assumptions, conventional tests of the independence of Y_{i1} and X_{i1} —that is, tests of $H_0: \beta = 0$ —have the correct level in the presence of measurement error η_{i1} , because when that null hypothesis is true, $Y_{i1} = \alpha_i + \varepsilon_{i1}$ is independent of $X_{i1} = x_{i1} + \eta_{i1}$, but the power of these tests is affected (see Lagakos 1988). This conclusion requires the independence of the pair effects α_i and the doses x_{i1} —the independence of age and smoking in Table 1—but this independence assumption was not needed in Section 2.1. Tests of independence of Y_{i1} and X_{i1} cannot be inverted to construct a confidence interval for β ; see the discussion of (5) in Section 1.3.

As is well known, even with all of these additional assumptions, β cannot be estimated without the data, Y_{i2} , from the zero-dose controls, or else without some other form of information not given in the formulation of the problem. For instance, in the familiar manner, if $\beta > 0$, then the regression of Y_{i1} on X_{i1} has slope $\beta\sigma_\alpha^2/(\sigma_x^2 + \sigma_\eta^2) < \beta$ whenever there is measurement error, $\sigma_\eta^2 > 0$. Moreover, the distribution of the bivariate observations, (Y_{i1}, X_{i1}) , $i = 1, \dots, I$, without the responses, Y_{i2} , from the zero-dose controls is not identified; there is, for example, no maximum likelihood estimate (see also Cheng and Van Ness 1999, sec. 1.2.1). These are negative results indicating that even under the best circumstances and even with the strongest assumptions, the data, Y_{i2} , on zero-dose controls are needed for identification. Only these negative results use the assumptions that $\alpha_i \stackrel{\text{iid}}{\sim} N(\mu_\alpha, \sigma_\alpha^2)$ and that α_i is independent of x_{i1} ;

they are not needed in Section 2.2 and later in this section. In actual practice, one would not want to assume that α_i is independent of x_{i1} . For instance, in Table 1, α_i reflects the matching on age and x_{i1} is the intensity of smoking, and there is no reason to believe that age and smoking intensity are unrelated; moreover, there is no need to believe this when using the method in Section 2.2. In fact, matching on age is intended to prevent a difference in age from being mistaken for an effect of smoking. In actual practice, it is best to view the α_i 's as nuisance parameters.

Under the normal model in the dose-control design, exact inference about β may be based on a t -test in place of Wilcoxon's signed rank test in Section 2.1. Consider testing the hypothesis $H_0: \beta = \beta_0$ by computing (6), where $E(Z_{i,\beta_0}) = (\beta - \beta_0)\mu_x$, $\text{var}(Z_{i,\beta_0}) = (\beta - \beta_0)^2\sigma_x^2 + \sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2 + \beta_0^2\sigma_\eta^2 = \nu$, say, and

$$Z_{i,\beta_0} \stackrel{\text{iid}}{\sim} N\{(\beta - \beta_0)\mu_x, \nu\}, \quad (7)$$

and applying the one-sample t -test with $I - 1$ degrees of freedom to the Z_{i,β_0} , testing that they have expectation 0. For all $I \geq 2$, if $H_0: \beta = \beta_0$ is true, then this test will have the correct level. With specified variances, the power of this t -test against $H_A: \beta = \beta_A$ can be computed from the noncentral t -distribution.

Consider the relative efficiency of this t -test and the nonparametric procedure in Section 2.1 when the normal model is true. For a given null and alternative hypothesis, (β_0, β_A) , for given values of the other parameters, and for a given level φ and power ψ , there is a necessary sample size, $I_{\beta_0, \beta_A, \varphi, \psi}^{(t)}$, for this t -test under (7), and, analogously, a necessary sample size, $I_{\beta_0, \beta_A, \varphi, \psi}^{(W)}$, for Wilcoxon's signed rank test, and $\rho_{\beta_0, \beta_A, \varphi, \psi} = I_{\beta_0, \beta_A, \varphi, \psi}^{(W)} / I_{\beta_0, \beta_A, \varphi, \psi}^{(t)}$ is the exact relative efficiency. In words, the exact relative efficiency, $\rho_{\beta_0, \beta_A, \varphi, \psi}$, for the dose-control design is the same as the familiar exact relative efficiency of the t -test and Wilcoxon test for location shift in a normal distribution. With normal data, the relative efficiency of the Wilcoxon signed rank test compared with the t -test is quite high even in quite small samples (Hettmansperger 1984, table 2.4, p. 75; Lehmann 1998, table 4.4, p. 174).

The normal model (7) may also be used in sample size calculations when planning a study. In particular, Noether (1987, sec. 2.2) gave the sample size required for Wilcoxon's signed rank test in terms of a quantity, $p' = \Pr(Z_{i,\beta_0} + Z_{j,\beta_0} > 0)$ for $i \neq j$, which under (7) and $H_A: \beta = \beta_A > \beta_0$ is $p' = \Phi\{\sqrt{2/\nu}(\beta_A - \beta_0)\mu_x\}$, where $\Phi(\cdot)$ is the standard normal cumulative distribution (e.g., Lehmann 1998, p. 167). Using this p' in Noether's formula gives the required sample size, $I_{\beta_0, \beta_A, \varphi, \psi}^{(W)}$.

If $H_0: \beta = \beta_0$ is true under (7), then $0 = E(Y_{i1} - \beta_0 X_{i1} - Y_{i2})$. Using this as an estimating equation yields the analog of the Hodges-Lehmann estimate for the t -statistic, namely $\tilde{\beta} = (\bar{Y}_1 - \bar{Y}_2) / \bar{X}_1$, where $\bar{Y}_j = (1/I) \sum_{i=1}^I Y_{ij}$, $j = 1, 2$, and $\bar{X}_1 = (1/I) \sum_{i=1}^I X_{i1}$. Recall that the controls, Y_{i2} , received zero dose, so $\tilde{\beta}$ is very similar to Wald's (1940) slope estimate. Under the normal model, $\tilde{\beta}$ is the ratio of two normal random variables, so $\tilde{\beta}$ does not have a normal distribution.

In the example in Table 1, the t -test procedure yields a 95% confidence interval for β of [1.49, 8.60], which is just slightly longer than the nonparametric interval in Section 2.2,

namely [1.85, 8.79], and is just slightly closer to 0. Of course, the nonparametric interval requires no assumption about the distribution of x_{i1} and has weaker assumptions about the error distributions. The point estimate, $\hat{\beta} = 3.89$, is somewhat further from 0 than the Hodges-Lehmann estimate of $\hat{\beta} = 3.21$. The point estimates are not at the center of their confidence intervals, in part because $\hat{\beta}$ does not have a normal distribution.

What happens if the distributions are not normal? In this paragraph only, suppose that ε_{i1} , ε_{i2} , η_{i1} , and x_{i1} are mutually independent, with ε_{i1} , ε_{i2} , and η_{i1} each symmetric about 0 and x_{i1} symmetric about $\mu_x > 0$, but perhaps these four distributions are not normal. When testing the hypothesis $H_0: \beta = \beta_0$, compute $Z_{i,\beta_0} = (\beta - \beta_0)x_{i1} + \varepsilon_{i1} - \varepsilon_{i2} - \beta_0\eta_{i1}$ as before, and the Z_{i,β_0} 's are iid and symmetric about $(\beta - \beta_0)\mu_x$. So, once again, the nonnull properties of the dose-control design may be deduced from familiar facts about inference for the center of symmetry of a symmetric distribution. For instance, if $(\beta - \beta_0)(x_{i1} - \mu_x) + \varepsilon_{i1} - \varepsilon_{i2} - \beta_0\eta_{i1}$ were double exponential, then the foregoing t -test procedure would be about 2/3 as efficient as the nonparametric procedure in Section 2.1 (see Hettmansperger 1984, p. 74). With long-tailed distributions, the t -test will often be inefficient but conservative (see Benjamini 1983).

3. DOSE-CONTROL DESIGNS WITH MATCHED SETS

3.1 One Exposed Subject Matched to Several Controls

In Section 1.3, each exposed subject, $j = 1$, was matched to a single control. For many hazardous exposures, exposed subjects are difficult to obtain but unexposed subjects are plentiful, and power can be increased by matching each exposed subject to several unexposed controls. The dose-control design with multiple controls has I matched sets, $i = 1, \dots, I$, with one exposed subject, $j = 1$, and $n_i - 1 \geq 1$ unexposed control subjects, $j = 2, \dots, n_i$. Write $N = \sum n_i$ for the total number of subjects. The exposed subject received dose $x_{i1} > 0$, but only a fallible measure of it, $X_{i1} > 0$, is recorded, where (3) and (4) hold. The controls received dose $x_{ij} = 0$ for $j = 2, \dots, n_i$. Dose and response are related by (1), where the N errors ε_{ij} are sampled as in (2) and are independent of the I dose errors η_{i1} . As in Remark 1, alternative assumptions are possible for the ε_{ij} .

A natural tactic is to proceed as before in Section 2, calculating (6) for each of the $n_i - 1$ controls in matched set i and applying the signed rank test to the resulting $M = N - I = \sum (n_i - 1)$ differences. The obvious difficulty with this tactic is that in Section 2 the I differences (6) were independent, a fact that was important in deriving the usual null distribution of the signed rank statistic, whereas with multiple controls the $n_i - 1$ differences in matched set i are dependent on one another, because they are all computed using the same exposed subject. In Section 3.2 the distribution is obtained for the signed rank statistic when matching with multiple controls.

The approach taken in Section 3.2 is similar in spirit to that of Sen and Puri (1967) for multivariate one-sample rank tests, but the two approaches differ in many details, because multivariate observations differ in certain key respects from univariate observations with multiple controls. Details follow. Specifically, the method of Sen and Puri is intended for use with I independent observations on a p -dimensional multivariate outcome.

Because of this, Sen and Puri computed absolute ranks separately for each of the p outcomes, then combined them using a quadratic form (their expression 2.8) that resembles Hotelling's T^2 statistic; that statistic is asymptotically chi-squared with p degrees of freedom under the null hypothesis. In contrast, in the current context, (a) the I matched sets are independent of one another, but observations in set i are related through α_i ; (b) the outcome is unidimensional; and (c) the dimension of the differences within set i is not a constant p , but rather varies with i . Specifically, with one exposed subject and $n_i - 1$ controls in set i , there are $n_i - 1$ correlated matched-pair differences between the exposed subject and the $n_i - 1$ controls. Because of this, the conventional Wilcoxon signed rank statistic is used, albeit with an unconventional exact null distribution reflecting the dependence within matched sets, and there is a limiting normal distribution under the null hypothesis, but not the conventional normal distribution for the signed rank statistic that arises with I independent pairs of two subjects. What our method and the method of Sen and Puri (1967) have in common is that in both cases a certain form of multivariate symmetry about 0 is present, and this leads to a uniform distribution over a certain set of signed reflections of the data, specifically the set \mathcal{W}_{β_0} defined in (9) in Section 3.2. Other multivariate extensions of the signed rank statistic exist (see Hettmansperger, Nyblom, and Oja 1992). Öhrvik (1998) proposed an extension of the signed rank statistic as a quadratic form in $g - 1$ degrees of freedom for use in crossover studies using Williams' time-balanced Latin squares. For p -dimensional observations, Hettmansperger, Möttönen, and Oja (1997) developed an affinely invariant multivariate signed rank statistic that is asymptotically chi squared on p degrees of freedom under the null hypothesis.

3.2 The Distribution of the Signed Rank Statistic With Multiple Controls

To test the null hypothesis $H_0: \beta = \beta_0$, calculate the M differences, $i = 1, \dots, I, j = 2, \dots, n_i$,

$$W_{ij,\beta_0} = (Y_{i1} - \beta_0 X_{i1}) - Y_{ij} = (\beta - \beta_0)x_{i1} + (\varepsilon_{i1} - \varepsilon_{ij} - \beta_0\eta_{i1}), \tag{8}$$

and form the I vectors $\mathbf{w}_{i,\beta_0} = (W_{i2,\beta_0}, \dots, W_{i,n_i,\beta_0})$ of dimension $n_i - 1, i = 1, \dots, I$. For $i \neq i'$, the vectors \mathbf{w}_{i,β_0} and \mathbf{w}_{i',β_0} are independent, but the $n_i - 1$ coordinates of \mathbf{w}_{i,β_0} are dependent because they involve the same ε_{i1} and η_{i1} . Write $\mathbf{W}_{\beta_0} = (\mathbf{w}_{1,\beta_0}, \dots, \mathbf{w}_{I,\beta_0})$ for the M -dimensional vector of \mathbf{w}_{i,β_0} 's.

If the null hypothesis $H_0: \beta = \beta_0$ is true, then \mathbf{w}_{i,β_0} exhibits a form of multivariate symmetry about 0 discussed by Sen and Puri (1967); specifically, \mathbf{w}_{i,β_0} and $-\mathbf{w}_{i,\beta_0}$ have the same distribution. In parallel with Sen and Puri (1967, sec. 2), let \mathcal{W}_{β_0} be the set containing the 2^I vectors

$$\{(-1)^{a_1}\mathbf{w}_{1,\beta_0}, \dots, (-1)^{a_I}\mathbf{w}_{I,\beta_0}\}, \quad \text{where } a_i = 0 \text{ or } a_i = 1. \tag{9}$$

It follows that under the null hypothesis $H_0: \beta = \beta_0$, the conditional distribution of \mathbf{W}_{β_0} given $\mathbf{W}_{\beta_0} \in \mathcal{W}_{\beta_0}$ takes the same probability, 2^{-I} , for all elements of \mathcal{W}_{β_0} .

Because the distributions are continuous, ties of all kinds do not occur. The extension to ties is discussed separately at the end of this section. Rank the $M = N - I$ absolute differences

$|W_{ij,\beta_0}|$ from 1 to M , writing r_{ij,β_0} for the rank. Also, write $s_{ij,\beta_0} = 1$ if $W_{ij,\beta_0} > 0$ or $s_{ij,\beta_0} = 0$ if $W_{ij,\beta_0} < 0$. If $W_{ij,\beta_0} > 0$, then $s_{ij,\beta_0}r_{ij,\beta_0}$ is the rank of $|W_{ij,\beta_0}|$, and otherwise is 0. In parallel, if $W_{ij,\beta_0} < 0$, then $(1 - s_{ij,\beta_0})r_{ij,\beta_0}$ is the rank of $|W_{ij,\beta_0}|$, and otherwise is 0. Now write

$$\mathbf{q}_i = (s_{i2,\beta_0}r_{i2,\beta_0}, \dots, s_{i,n_i,\beta_0}r_{i,n_i,\beta_0})$$

and

$$\tilde{\mathbf{q}}_i = \{(1 - s_{i2,\beta_0})r_{i2,\beta_0}, \dots, (1 - s_{i,n_i,\beta_0})r_{i,n_i,\beta_0}\},$$

with q_{ij} and \tilde{q}_{ij} signifying the j th coordinates, $j = 2, \dots, n_i$, where $q_{ij} + \tilde{q}_{ij} = r_{ij,\beta_0}$. Also, write $q_{i+} = \sum_{j=2}^{n_i} q_{ij}$, $\tilde{q}_{i+} = \sum_{j=2}^{n_i} \tilde{q}_{ij}$, and $r_{i+} = \sum_{j=2}^{n_i} r_{ij,\beta_0}$. The signed rank statistic computed from the M differences (8) is the sum of the ranks of the positive differences, namely $T_{\beta_0} = \sum_{i=1}^I \sum_{j=2}^{n_i} s_{ij,\beta_0}r_{ij,\beta_0} = \sum_{i=1}^I \sum_{j=2}^{n_i} q_{ij} = \sum_{i=1}^I q_{i+}$.

Assume throughout this paragraph that the null hypothesis $H_0: \beta = \beta_0$ is true for the purpose of testing it. Given $\mathbf{W}_{\beta_0} \in \mathcal{W}_{\beta_0}$, matched set i yields either \mathbf{q}_i or $\tilde{\mathbf{q}}_i$ each with probability $\frac{1}{2}$, with independent selections in distinct matched sets. The conditional null distribution of the signed rank statistic T_{β_0} , given $\mathbf{W}_{\beta_0} \in \mathcal{W}_{\beta_0}$, is a known distribution that does not depend on the unknown distributions of the ε_{ij} 's and η_{i1} ; specifically, it is the distribution of the sum of I independent random variables taking values q_{i+} or \tilde{q}_{i+} each with probability $\frac{1}{2}$. For moderate I , the exact null conditional distribution of T_{β_0} may be obtained by the algorithm of Pagano and Trichtler (1983, sec. 3.3), whereas as $I \rightarrow \infty$, a normal approximation may be used with expectation $\tau = \frac{1}{2} \sum_{i=1}^I (q_{i+} + \tilde{q}_{i+}) = \frac{1}{2} \sum_{i=1}^I r_{i+} = M(M + 1)/4$ and variance $\omega^2 = \frac{1}{4} \sum_{i=1}^I (q_{i+} - \tilde{q}_{i+})^2$, comparing $(T - \tau)/\omega$ with the standard normal distribution.

Ties are handled as follows. If there are ties among the $|W_{ij,\beta_0}|$, then use average ranks in computing r_{ij,β_0} . If some $W_{ij,\beta_0} = 0$, then set $s_{ij,\beta_0} = \frac{1}{2}$, with the consequence that such a W_{ij,β_0} contributes a constant to T_{β_0} . With these alterations in the definitions of r_{ij,β_0} and s_{ij,β_0} , the exact null distribution of T_{β_0} and the moments for the large-sample normal approximation are computed exactly as described earlier.

As in Section 3, a confidence interval for β is obtained by inverting the test of the hypothesis $H_0: \beta = \beta_0$, and a point estimate $\hat{\beta}$ is the approximate solution to $T_{\hat{\beta}} = M(M + 1)/4$.

3.3 Several Exposed Subjects Matched to One Control

In Section 3.1 a single exposed subject with a fallible positive dose was matched to one or more controls with dose zero. If, instead, exposed subjects are plentiful and zero-dose controls are rare, then each control might be matched to several exposed subjects. One might hope to apply the same methods, even though there is an asymmetry between the fallible doses for exposed subjects and the known zero doses for controls. This section briefly confirms that the method of Section 3.2 is applicable nonetheless.

Suppose that the first subject in matched set i is the control with $x_{i1} = 0$. The remaining $n_i - 1$ subjects in set i are exposed with positive true doses x_{i2}, \dots, x_{i,n_i} , which are fallibly measured as X_{i2}, \dots, X_{i,n_i} , generated by $X_{ij} = x_{ij} + \eta_{ij}, j = 2, \dots, n_i$, with $\eta_{ij} \stackrel{iid}{\sim} H \in \mathcal{S}$ in parallel with (3) and (4). As before, true

dose and response are related by (1) and (2), where the response errors ε_{ij} are independent of the dose errors η_{ij} .

Compute the M differences, $i = 1, \dots, I, j = 2, \dots, n_i$,

$$V_{ij,\beta_0} = (Y_{ij} - \beta_0 X_{ij}) - Y_{i1} = (\beta - \beta_0)x_{ij} + (\varepsilon_{ij} - \varepsilon_{i1} - \beta_0\eta_{ij}), \quad (10)$$

and write $\mathbf{v}_{i,\beta_0} = (V_{i2,\beta_0}, \dots, V_{i,n_i,\beta_0})$. As with the \mathbf{w}_{i,β_0} in Section 3.2 the coordinates of \mathbf{v}_{i,β_0} are dependent, because ε_{i1} is present in all of them, but \mathbf{v}_{i,β_0} and \mathbf{v}_{k,β_0} are independent for $i \neq k$. Moreover, if the null hypothesis $H_0: \beta = \beta_0$ is true, then \mathbf{v}_{i,β_0} and $-\mathbf{v}_{i,\beta_0}$ have the same distribution. In other words, \mathbf{w}_{i,β_0} has a single η_{i1} and \mathbf{v}_{k,β_0} has $n_i - 1$ different η_{ij} , so (8) and (10) are not quite parallel, but the properties used in Section 3.2 to obtain inferences with multiple controls continue to hold in the current section with multiple exposed subjects. That is, one may draw inferences about β by applying the extended signed rank statistic in Section 3.2 to the V_{ij,β_0} in (10).

3.4 Full Matching

In full matching, a matched set may consist of either a single exposed subject with one or more controls or a single control with one or more exposed subjects. When stratifying on continuous covariates, the optimal stratification—the one that makes covariates as similar as possible within strata—is always a full matching, and both theory and simulations show that full matching is much better than pair matching (Rosenbaum 1991; Gu and Rosenbaum 1993).

With full matching, one computes \mathbf{w}_{i,β_0} in a matched set i with several controls or \mathbf{v}_{k,β_0} in a matched set k with several exposed subjects, yielding a collection of I vectors composed of some \mathbf{w}_{i,β_0} 's and some \mathbf{v}_{k,β_0} 's. In parallel with the reasoning in Section 3.3, the extended signed rank statistic of Section 3.2 may be applied to these I vectors to draw inferences about β .

3.5 Example: Radiation and Cytogenetic Damage

In Taiwan in 1983 and 1984, some 180 residential buildings were constructed using radioactive ^{60}Co -contaminated steel rods. Chang et al. (1999) examined 16 residents of these buildings exposed to radiation from the rods and 7 unexposed controls who lived elsewhere. They looked at several measures of cytogenetic damage, one measure being the number of centromere-positive signals per 1,000 binucleated cells, or $C + /\text{BN}$. Table 2 records $C + /\text{BN}$ and their estimate of the radiation dose (in mSv) received from the rods for the 16 residents.

In Table 2 the seven available controls are matched to the exposed subjects on the basis of age and gender. The matching is exact for gender and is as close as possible on age subject to the restrictions that all seven controls are used and no control is matched to more than four exposed subjects (see Ming and Rosenbaum 2000).

Table 3 illustrates the computations using T_{β_0} , in this case testing $H_0: \beta = 0$. The four exposed subjects in matched set $i = 1$ are each compared with the control in matched set $i = 1$, so that, for example, the difference for the first exposed subject is $V_{ij,0} = 12.0 - 0 \times 14 - 8.3 = 3.7$. The q_{ij} are the ranks of $|V_{ij,0}|$ for $V_{ij,0} > 0$, and the \tilde{q}_{ij} are the ranks for $V_{ij,0} < 0$. The total of the positive ranks in set $i = 1$ is $q_{i+} = 19$, and the

Table 2. Radiation and Cytogenetic Damage for 16 Exposed Subjects (E) and 7 Unexposed Controls (C)

Matched set	Subject	$C + /\text{MN}$	Radiation dose
1	E1	12.0	14
	E5	6.8	209
	E7	10.0	219
	E13	12.5	766
2	C3	8.3	0
	E2	19.0	30
3	C1	3.2	0
	E3	31.8	75
4	E6	8.4	209
	E14	24.0	1,166
	C2	8.8	0
	E4	3.7	181
5	E15	20.0	1,265
	C7	5.1	0
	E8	22.7	544
6	E9	11.3	567
	E16	18.7	1,493
	C4	9.5	0
	E10	8.5	572
7	C6	11.9	0
	E11	36.0	638
	E12	33.3	660
	C5	14.0	0

SOURCE: Chang et al. (1999).

total of the negative ranks is $\tilde{q}_{i+} = 3$. The signed rank statistic, $T_{\beta_0} = T_0$, is the sum of the ranks of the positive differences, $\sum_{i,j} q_{ij} = \sum_i q_{i+} = 124$. If the null hypothesis were true, then the V_{ij,β_0} would be symmetric about 0, and, given \mathcal{W}_{β_0} , the two possible rank totals for set i , namely q_{i+} and \tilde{q}_{i+} , would each contribute to T_{β_0} with probability $\frac{1}{2}$, independently for different sets $i \neq i'$; for instance, either 19 or 3 from the first matched

Table 3. Computing the Signed Rank Test With Several Exposed Subjects Matched to One Control

Matched set i	$C + /\text{MN}$	Subject	$V_{ij,0}$	q_{ij}	\tilde{q}_{ij}	q_{i+}	\tilde{q}_{i+}
1	12.0	E1	3.7	7	0	19	3
	6.8	E5	-1.5	0	3		
	10.0	E7	1.7	4	0		
	12.5	E13	4.2	8	0		
	8.3	C3	•				
2	19.0	E2	15.8	13	0	13	0
	3.2	C1	•				
3	31.8	E3	23.0	16	0	28	1
	8.4	E6	-4	0	1		
	24.0	E14	15.2	12	0		
	8.8	C2	•				
4	3.7	E4	-1.4	0	2	11	2
	20.0	E15	14.9	11	0		
	5.1	C7	•				
5	22.7	E8	13.2	10	0	24	0
	11.3	E9	1.8	5	0		
	18.7	E16	9.2	9	0		
	9.5	C4	•				
6	8.5	E10	-3.4	0	6	0	6
	11.9	C6	•				
7	36.0	E11	22.0	15	0	29	0
	33.3	E12	19.3	14	0		
	14.0	C5	•				
T_0				124		124	

set $i = 1$. Now \mathcal{W}_{β_0} contains $2^7 = 128$ sign changes, only one of which—switching just $i = 6$ —produces a larger value of the signed rank statistic, so the one-sided exact p value is $2/128 = .0156$. The null expectation of T_0 is $\frac{1}{2} \sum (q_{i+} + \tilde{q}_{i+}) = M(M+1)/4$ or $16(16+1)/4 = 68$, and the null variance is $\frac{1}{4} \sum (q_{i+} - \tilde{q}_{i+})^2 = \frac{1}{4} \{(19-3)^2 + \dots + (29-0)^2\} = 672$, yielding a standardized deviate of $(124 - 68)/\sqrt{672} = 2.1602$ and approximate p value of $.0154 = 1 - \Phi(2.1602)$ from the standard normal cumulative distribution, $\Phi(\cdot)$.

The 95% confidence interval is the set of hypotheses $H_0: \beta = \beta_0$ rejected by neither of two one-sided .025 level tests. For instance, $H_0: \beta = .00281$ yields exact p value $4/128 = .03125$, whereas $H_0: \beta = .002805$ yields exact p value $3/128 = .0234375$, so the lower endpoint of the confidence interval is approximately .00281. The upper endpoint is obtained in the same way, yielding the exact 95% confidence interval [.00281, .03571]. Now T_{β_0} is $68 = M(M+1)/4$ for $.01455 \leq \beta_0 \leq .01464$, so $\hat{\beta} = (.01455 + .01464)/2 = .0146$.

3.6 Aligned Ranks

When matching with multiple controls, the standard non-parametric test is not the signed rank test of Section 3.2, but rather the aligned rank test of Hodges and Lehmann (1962). With matched pairs, the aligned rank test is very similar to Wilcoxon's signed rank test. In the aligned rank test, the mean response in matched set i is subtracted from all responses in matched set i ; then all of these aligned responses in all of the matched sets are ranked together from 1 to N , and the test statistic is the sum of the ranks for treated subjects. Unfortunately, the standard theory of the aligned rank test does not apply when there are doses measured with errors. This section, which may be skipped, explains why. Briefly, the aligned rank statistic is based on permutation invariance within matched sets, which does not hold with dose errors, whereas the signed rank test with multiple controls in Section 3.2 uses invariance under sign reflections. In matched pairs, these two types of invariance are essentially the same, but this is no longer true with multiple controls.

Suppose that the null hypothesis, $H_0: \beta = \beta_0$, is true for the purpose of testing it. In the situation of Section 3.2, the first subject, $j = 1$, in matched set i was exposed at dose x_{i1} which was inaccurately recorded as $X_{i1} = x_{i1} + \eta_{i1}$, whereas subjects $j = 2, \dots, n_i$ received dose $x_{ij} = 0$. Define the aligned responses $A_{i1, \beta_0} = Y_{i1} - \beta_0 X_{i1} - \bar{Y}_{i, \beta_0}$ and $A_{ij, \beta_0} = Y_{ij} - \bar{Y}_{i, \beta_0}$, $j = 2, \dots, n_i$, where $\bar{Y}_{i, \beta_0} = (1/n_i) \{Y_{i1} - \beta_0 X_{i1} + \sum_{j=2}^{n_i} Y_{ij}\}$. The aligned rank statistic ranks the A_{ij, β_0} from 1 to N and sums the ranks of the I exposed subjects. If the doses were measured without error, so that $\eta_{i1} = 0$ and $X_{i1} = x_{i1}$, then, under the null hypothesis, the aligned responses, $A_{ij, \beta_0} = \varepsilon_{ij} - (\frac{1}{n_i}) \sum_{k=1}^{n_i} \varepsilon_{ik} = \varepsilon_{ij} - \bar{\varepsilon}_{i+}$, say, would be exchangeable within each matched set, because the ε_{ij} are iid, and this fact yields the null distribution of the aligned rank statistic. In contrast, with dose errors, even when $H_0: \beta = \beta_0$ is true, the aligned responses are $A_{i1, \beta} = \varepsilon_{i1} - (1 - \frac{1}{n_i}) \beta \eta_{i1} - \bar{\varepsilon}_{i+}$ and $A_{ij, \beta} = \varepsilon_{ij} + \frac{\beta \eta_{i1}}{n_i} - \bar{\varepsilon}_{i+}$, $j = 2, \dots, n_i$, which are not exchangeable, so the usual null distribution for the aligned rank statistic is not available. For instance, if the dispersion of $\beta \eta_{i1}$ was extremely large compared with the dispersion of ε_{ij} , then the aligned response for the exposed subject in matched set i , namely $A_{i1, \beta}$, would too often

be the largest or smallest aligned response in set i , receiving either the largest or smallest rank in this set, and this would not be correctly reflected by a variance of the rank for $A_{i1, \beta}$ computed assuming exchangeability.

4. SUMMARY, LIMITATIONS, AND DISCUSSION

The dose-control design is a simple solution to the traditional problem of estimating a linear dose-response relationship when both the responses and the doses are subject to error. It does this with the aid of controls known to have received zero dose.

The traditional errors-in-variables model given by (1)–(4) may be more realistic than a model that asserts the doses are measured without error, but it remains a very simple model, perhaps an oversimplification in some cases. For instance, the model does not permit systematic underreporting of smoking in Section 1.2.

Hill (1965) suggested that a dose-response relationship is important in nonrandomized observational studies, such as those described in Sections 1.2 and 3.5, because it helps in judging whether the agent under study caused it ostensible effects or whether there is some unobserved bias in the comparison. Specifically, in observational studies, Hill suggested that there is stronger evidence that the treatment is the actual cause of its ostensible effects if larger doses are associated with larger effects, and he believed that investigators should make this check routinely. Hill's suggestion, which has been highly influential in epidemiology, is correct in certain formal senses (Rosenbaum 2003, 2004). In contrast, the analysis in Section 2 escapes bias from dose errors by paying little attention to whether exposed subjects with high doses also have high responses, attributing a weak dose-response relationship among exposed subjects to dose errors. In the dose-control design, large dose errors do not prevent estimation of β , but they do limit the value of one check that is commonly applied to observational data.

[Received August 2003. Revised September 2004.]

REFERENCES

- Benjamini, Y. (1983), "Is the t -Test Conservative When the Parent Distribution Is Long-Tailed?" *Journal of the American Statistical Association*, 78, 645–654.
- Chang, W. P., Hsieh, W. A., Chen, D., Lin, Y., Hwang, J., Hwang, J., Tsai, M., and Hwang, B. (1999), "Change in Centromeric and Acentromeric Micronucleus Frequencies in Human Populations After Chronic Radiation Exposure," *Mutagenesis*, 14, 427–432.
- Cheng, C.-L., and Van Ness, J. W. (1999), *Statistical Regression With Measurement Error*, New York: Oxford University Press.
- Fuller, W. A. (1987), *Measurement Error Models*, New York: Wiley.
- Gu, X. S., and Rosenbaum, P. R. (1993), "Comparison of Multivariate Matching Methods: Structures, Distances and Algorithms," *Journal of Computational and Graphical Statistics*, 2, 405–420.
- Hettmansperger, T. P. (1984), *Statistical Inference Based on Ranks*, New York: Wiley.
- Hettmansperger, T. P., Möttönen, J., and Oja, H. (1997), "Affine-Invariant Multivariate One-Sample Signed-Rank Tests," *Journal of the American Statistical Association*, 92, 1591–1600.
- Hettmansperger, T. P., Nyblom, J., and Oja, H. (1992), "On Multivariate Notions of Sign and Rank," in *L₁-Statistical Analysis and Related Methods*, ed. Y. Dodge, New York: North-Holland, pp. 267–278.
- Hill, A. B. (1965), "The Environment and Disease: Association or Causation?" *Proceedings of the Royal Society of Medicine*, 58, 295–300.
- Hodges, J. L., and Lehmann, E. L. (1962), "Rank Methods for Combination of Independent Experiments in the Analysis of Variance," *The Annals of Mathematical Statistics*, 33, 482–497.
- (1963), "Estimates of Location Based on Ranks," *The Annals of Mathematical Statistics*, 34, 598–611.

- Hollander, M., and Wolfe, D. A. (1999), *Nonparametric Statistical Methods*, New York: Wiley.
- Kendall, M. G., and Stuart, A. (1973), *The Advanced Theory of Statistics, Vol. 2, Inference and Relationship* (3rd ed.), New York: Hafner.
- Lagakos, S. W. (1988), "Effects of Mismodelling and Mismeasuring Explanatory Variables on Tests of Their Association With a Response Variable," *Statistics in Medicine*, 7, 257–274.
- Lehmann, E. L. (1959), *Testing Statistical Hypotheses*, New York: Wiley.
- (1998), *Nonparametrics*, Englewood Cliffs, NJ: Prentice-Hall.
- Madansky, A. (1959), "The Fitting of Straight Lines When Both Variables Are Subject to Error," *Journal of the American Statistical Association*, 54, 173–205.
- 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., and Scott, E. L. (1951), "On Certain Methods of Estimating the Linear Structural Relation," *The Annals of Mathematical Statistics*, 22, 352–361.
- Noether, G. E. (1987), "Sample Size Determination for Some Common Nonparametric Tests," *Journal of the American Statistical Association*, 82, 645–647.
- Öhrvik, J. (1998), "Nonparametric Methods in Crossover Studies," *Biometrical Journal*, 40, 771–789.
- Pagano, M., and Tritchler, D. (1983), "Obtaining Permutation Distributions in Polynomial Time," *Journal of the American Statistical Association*, 78, 435–440.
- Prokopczyk, B., Cox, J. E., Hoffman, D., and Waggoner, S. E. (1997), "Identification of Tobacco-Specific Carcinogen in the Cervical Mucus of Smokers and Nonsmokers," *Journal of the National Cancer Institute*, 89, 868–873.
- Rosenbaum, P. R. (1991), "A Characterization of Optimal Designs for Observational Studies," *Journal of the Royal Statistical Society, Ser. B*, 53, 597–610.
- (2003), "Does a Dose-Response Relationship Reduce Sensitivity to Hidden Bias?" *Biostatistics*, 4, 1–10.
- (2004), "Design Sensitivity in Observational Studies," *Biometrika*, 91, 153–164.
- Sen, P. K., and Puri, M. (1967), "On Rank Order Tests for Location in the Multivariate One-Sample Problem," *The Annals of Mathematical Statistics*, 38, 1216–1228.
- Wald, A. (1940), "The Fitting of Straight Lines if Both Variables Are Subject to Error," *The Annals of Mathematical Statistics*, 11, 284–300.