The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: www.tandfonline.com/journals/utas20

On Judging the Significance of Differences by Examining the Overlap Between Confidence Intervals

Nathaniel Schenker & Jane F Gentleman

To cite this article: Nathaniel Schenker & Jane F Gentleman (2001) On Judging the Significance of Differences by Examining the Overlap Between Confidence Intervals, The American Statistician, 55:3, 182-186, DOI: 10.1198/000313001317097960

To link to this article: https://doi.org/10.1198/000313001317097960

Published online: 01 Jan 2012.

Submit your article to this journal

Article views: 3045

View related articles

Citing articles: 155 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=utas20

On Judging the Significance of Differences by Examining the Overlap Between Confidence Intervals

Nathaniel SCHENKER and Jane F. GENTLEMAN

To judge whether the difference between two point estimates is statistically significant, data analysts often examine the overlap between the two associated confidence intervals. We compare this technique to the standard method of testing significance under the common assumptions of consistency, asymptotic normality, and asymptotic independence of the estimates. Rejection of the null hypothesis by the method of examining overlap implies rejection by the standard method, whereas failure to reject by the method of examining overlap does not imply failure to reject by the standard method. As a consequence, the method of examining overlap is more conservative (i.e., rejects the null hypothesis less often) than the standard method when the null hypothesis is true, and it mistakenly fails to reject the null hypothesis more frequently than does the standard method when the null hypothesis is false. Although the method of examining overlap is simple and especially convenient when lists or graphs of confidence intervals have been presented, we conclude that it should not be used for formal significance testing unless the data analyst is aware of its deficiencies and unless the information needed to carry out a more appropriate procedure is unavailable.

KEY WORDS: Efficiency; Inference; Power; Test of significance; Two-sample problem; Type I error.

1. INTRODUCTION

To judge whether the difference between two point estimates is statistically significant, data analysts sometimes examine the overlap between the two associated confidence intervals. If there is no overlap, the difference is judged significant, and if there is overlap, the difference is not judged significant. This method is simple to use because it is easy to compare boundary values from two confidence intervals to see whether the intervals overlap. Moreover, it seems natural to inspect graphed confidence intervals for overlap. Over a number of years, however, we have tried to discourage the practice of examining overlap to assess significance among authors of articles that we have reviewed, because the procedure can lead to mistaken conclusions.

A nonexhaustive search of abstracts and articles in the health sciences found more than 60 articles in which the method of examining overlap was used or recommended to demonstrate significance and/or lack of significance, either formally or informally. Recent examples were found in 22 different sources: *American Heart Journal*; *American Journal of Epidemiology*; *Annals of the Rheumatic Diseases*; *British Medical Journal*; *Cancer Epidemiology Biomarkers & Prevention*; *Cancer Research*; *CDC Surveillance Summaries*; *Clinical and Experimental Allergy*; *Clinical Cancer Research*; *Clinical Neuropharmacology*; *Emerging Infectious Diseases*; *Health Education Research*; *International Journal of Epidemiology*; *Journal of Clinical Epidemiology*; *Journal of Medical Entomology*; *Journal of Medical Genetics*; *Journal of the National Cancer Institute*; *Journal of Nuclear Medicine*; *The Clinical Journal of Pain*; *The Cochrane Library*; *The Journal of Nutrition*; and *The Lancet*. Respective examples of articles are: Abergel et al. (1998); Sont et al. (2001); Gotzsche (2000); Appleby and Bell (2000); Millikan et al. (2000); Martinez et al. (1999); Kann et al. (2000); Ezeamuzie et al. (1999); Tersmette et al. (2001); Inzelberg et al. (2000); Chalmers and Salmon (2000); Mcbride (2000); Wong et al. (1999); Mancuso, Peterson, and Charlson (2001); Debboun et al. (2000); Slavotinek et al. (1999); Djordjevic, Stellman, and Zang (2000); Eising and von der Ohe (1998); Herr et al. (1998); Collins et al. (2001); Vatassery, Smith, and Quach (1998); and Mehta, Eikelboom, and Yusuf (2000).

Discussion of properties of the method of examining overlap in the literature is limited and contradictory. For example, an exchange in the *Journal of the American Academy of Dermatology* (Bigby and Gadenne 1996; Rahlfs 1997; Bigby and Gadenne 1997) offered contradictory advice concerning the method. A follow-up letter in that journal by Cole and Blair (1999) provided a more accurate assessment of the method along with some Monte Carlo results for a few situations.

This article considers the common task of comparing estimates that are assumed to be consistent, asymptotically normal, and asymptotically independent. It is shown that rejection of the null hypothesis by the method of examining overlap implies rejection by the standard method, whereas failure to reject by the method of examining overlap does not imply failure to reject by the standard method. Thus, the method of examining overlap is more conservative (i.e., rejects the null hypothesis less often) than the standard method when the null hypothesis is true, and it mistakenly fails to reject the null hypothesis more frequently than does the standard method when the null hypothesis is false. We expand on the work of Cole and Blair (1999) by providing more theoretical insight into the properties of the methods and by developing asymptotic results on efficiency, probability of Type I error, and power for a broad range of situations. Our goal is to provide an accessible discussion for statistical practitioners in subject-matter areas and for statistical consultants. Throughout the article, we provide simple examples and figures that can be used in consulting situations to illustrate the results.

Nathaniel Schenker is Senior Scientist for Research and Methodology, National Center for Health Statistics, Hyattsville, MD 20782 (E-mail: nschenker@cdc.gov). Jane F. Gentleman is Director, Division of Health Interview Statistics, National Center for Health Statistics, Hyattsville, MD 20782 (E-mail: jgentleman@cdc.gov). The authors thank Jennifer Madans of the National Center for Health Statistics for her encouragement.

 In the Public Domain

Section 2 describes our general setup and then derives the main result underlying the conservatism and low power of the method of examining overlap relative to the standard method. Section 3 presents asymptotic numerical comparisons of the methods. Section 4 briey discusses another simple but sometimes misleading method of examining overlap that we have seen used. Section 5 discusses extensions and other aspects of our results.

2. SETUP AND MAIN RESULT

2.1 Definitions and Assumptions

Suppose that we wish to make inference about the difference between two population quantities, Q_1 and Q_2, which are estimated by $\hat{Q}_1$ and $\hat{Q}_2$, and suppose that the associated standard errors, SE_1 and SE_2, are estimated by $\widehat{\mathrm{SE}}_1$ and $\widehat{\mathrm{SE}}_2$.

Our development assumes that we have consistent estimates and that $\hat{Q}_1$ and $\hat{Q}_2$ are asymptotically normal and independent. More formally, we assume that $\hat{Q}_1 \xrightarrow{P} Q_1$, $\hat{Q}_2 \xrightarrow{P} Q_2$, $\widehat{\mathrm{SE}}_1/\mathrm{SE}_1 \xrightarrow{P} 1$, $\widehat{\mathrm{SE}}_2/\mathrm{SE}_2 \xrightarrow{P} 1$, $(\hat{Q}_1 - Q_1)/\mathrm{SE}_1 \xrightarrow{D} Z_1$, and $(\hat{Q}_2 - Q_2)/\mathrm{SE}_2 \xrightarrow{D} Z_2$ as the sample sizes increase to ∞, where Z_1 and Z_2 are independent standard normal random variables.

It follows from standard theorems on convergence (see, e.g., Cramér 1946, pp. 254–255) that $(\hat{Q}_1 - Q_1)/\widehat{\mathrm{SE}}_1$, $(\hat{Q}_2 - Q_2)/\widehat{\mathrm{SE}}_2$, and $[(\hat{Q}_1 - \hat{Q}_2) - (Q_1 - Q_2)]/\sqrt{\widehat{\mathrm{SE}}_1^2 + \widehat{\mathrm{SE}}_2^2}$ have limiting standard normal distributions. Based on these three approximate pivotal quantities (Brownlee 1965, sec. 2.16), nominal 95% confidence intervals for Q_1, Q_2, and $Q_1 - Q_2$ are given by

$$\hat{Q}_1 \pm 1.96\widehat{\mathrm{SE}}_1, \tag{1}$$

$$\hat{Q}_2 \pm 1.96\widehat{\mathrm{SE}}_2, \tag{2}$$

and

$$(\hat{Q}_1 - \hat{Q}_2) \pm 1.96\sqrt{\widehat{\mathrm{SE}}_1^2 + \widehat{\mathrm{SE}}_2^2}, \tag{3}$$

respectively.

By the duality between $100(1-\alpha)\%$ confidence intervals and significance tests at level α (Bickel and Doksum 1977, p. 178), a test at nominal level .05 of the null hypothesis that $Q_1 - Q_2 = 0$ can be carried out by examining whether the nominal 95% interval (3) contains 0. The null hypothesis is rejected if and only if the interval does not contain 0. We will refer to this as the "standard" method. In contast, the "overlap" method conducts a test at nominal level .05 by rejecting the null hypothesis if and only if the nominal 95% intervals (1) and (2) do not overlap.

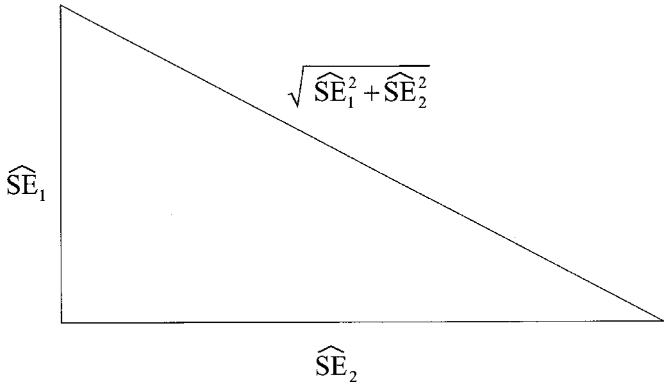

Figure 1. Right triangle depicting the relationship between the estimated standard errors of $\hat{Q}_1$, $\hat{Q}_2$, and $\hat{Q}_1 - \hat{Q}_2$.

2.2 Example: Comparing Proportions

Suppose we wish to compare two large populations with respect to the proportions of people with a given attribute. Independent simple random samples, each of size 200, have been drawn, and 112 people in the first sample have the attribute, whereas 88 people in the second sample have the attribute.

The estimated proportions and their estimated standard errors are $\hat{Q}_1 = 112/200 = .56$, $\hat{Q}_2 = 88/200 = .44$, $\widehat{\mathrm{SE}}_1 = \sqrt{(.56)(.44)/200} = .0351$, and $\widehat{\mathrm{SE}}_2 = \sqrt{(.44)(.56)/200} = .0351$, respectively. (Finite population corrections have been ignored since the populations are large.) Thus, intervals (1) and (2) for the population proportions Q_1 and Q_2 are $.56 \pm (1.96)(.0351) = [.49, .63]$ and $.44 \pm (1.96)(.0351) = [.37, .51]$, respectively. The overlap method leads us to declare that the proportions *are not* significantly different.

In contrast, since interval (3) for the difference $Q_1 - Q_2$, $(.56 - .44) \pm 1.96\sqrt{(.0351)^2 + (.0351)^2} = [.02, .22]$, does not contain 0, the standard method would lead us to declare that the proportions *are* significantly different.

2.3 Main Result

The preceding example illustrates that the overlap method can fail to reject the null hypothesis when the standard method would reject it. In general, rejection of the null hypothesis by the overlap method implies rejection by the standard method, but not vice versa. Thus, the overlap method is more conservative and less powerful than the standard method.

To derive this result, we begin by noting that intervals (1) and (2) overlap if and only if the interval

$$(\hat{Q}_1 - \hat{Q}_2) \pm 1.96(\widehat{\mathrm{SE}}_1 + \widehat{\mathrm{SE}}_2) \tag{4}$$

contains 0. Thus, the difference between the standard method and the overlap method is only in the width of the underlying intervals (3) and (4). The ratio of the width of interval (4) to the width of interval (3) is

$$\frac{\widehat{\mathrm{SE}}_1 + \widehat{\mathrm{SE}}_2}{\sqrt{\widehat{\mathrm{SE}}_1^2 + \widehat{\mathrm{SE}}_2^2}}. \tag{5}$$

Figure 1 displays a right triangle depicting the relationship between the estimated standard errors of $\hat{Q}_1$, $\hat{Q}_2$, and $\hat{Q}_1 - \hat{Q}_2$. Expression (5) is equal to the ratio of the sum of the lengths of the two sides adjacent to the right angle to the length of the hypotenuse. It is clear from the figure that the ratio is greater than 1, which demonstrates our result.

It is also straightforward to show that ratio (5) achieves its maximum value, $\sqrt{2}$, when the two adjacent sides of the triangle in Figure 1 are of equal length, and that it approaches its minimum value, 1, as one adjacent side increases in length rel-

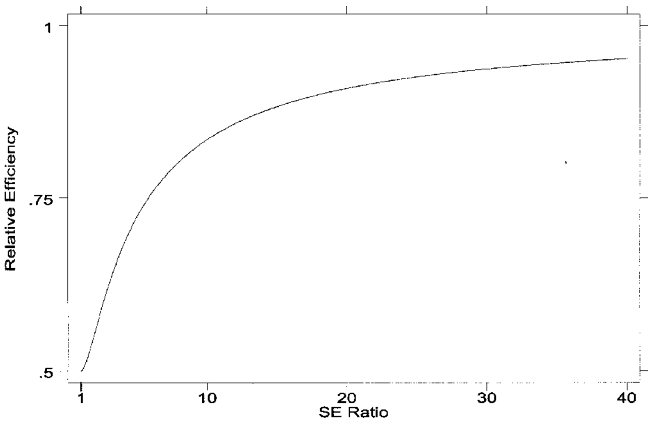

Figure 2. Asymptotic relative efficiency of the overlap method to the standard method.

ative to the other. Thus, the overlap method is expected to be more deficient relative to the standard method when SE_1 and SE_2 are nearly equal, and less deficient when one standard error is large relative to the other.

3. ASYMPTOTIC COMPARISONS

To illustrate the deficiencies of the overlap method due to the length of its underlying interval relative to the standard method, we provide asymptotic numerical comparisons (as the sample sizes increase to ∞) concerning efficiency, probability of Type I error, and power. Our results all follow from standard theorems on convergence (see, e.g., Cramér 1946, pp. 254–255). Derivations are available from the authors upon request. We investigate performance under various limiting values of the ratio of one standard error to the other. Our results when the limiting value of $\mathrm{SE}_2/\mathrm{SE}_1$ is k are the same as those when the limiting value of $\mathrm{SE}_1/\mathrm{SE}_2$ is k, so we simply refer to the limiting value as the "SE ratio" and consider only ratios that are greater than or equal to 1.

3.1 Efficiency

The asymptotic relative efficiency (ARE) of the overlap method to the standard method may be defined as the limit in probability of the ratio of the squared width of interval (3) to the squared width of interval (4), that is, the limit in probability of the squared inverse of expression (5). For an SE ratio of k, the ARE is given by $(1+k^2)/(1+k)^2$. The ARE is displayed in Figure 2 for a range of values of the SE ratio. As anticipated at the end of Section 2.3, the minimum ARE, 1/2, occurs when the SE ratio is 1. As the SE ratio increases, the ARE approaches 1.

Suppose, as in the example of Section 2.2, that the standard errors of $\hat{Q}_1$ and $\hat{Q}_2$ are of the form $\sigma_1/\sqrt{n_1}$ and $\sigma_2/\sqrt{n_2}$, that is, inversely proportional to the square roots of the sample sizes underlying $\hat{Q}_1$ and $\hat{Q}_2$. Then the ARE can be interpreted as the asymptotic ratio of the sample sizes that the standard method and the overlap method would need to produce underlying intervals of the same width and to achieve the same power. For example, when the SE ratio is 1, the standard method needs asymptotically only half of the sample sizes that the overlap method needs.

3.2 Probability of Type I Error and Power

Since the limiting power of either the standard method or the overlap method is 1 for any fixed, nonzero hypothesized alternative value of $Q_1 - Q_2$, we consider a sequence of alternative values such that $(Q_1 - Q_2)/\sqrt{\mathrm{SE}_1^2 + \mathrm{SE}_2^2}$ approaches a finite value, which we call the "standardized difference," as the sample sizes increase. Thus, we effectively specify alternative values of $Q_1 - Q_2$ that tend to 0, and we force the powers of the two methods to converge to values that are less than 1 and not necessarily equal. This allows a more meaningful comparison of the two methods. {See Cox and Hinkley (1974, pp. 317–318), for similar uses of sequences of alternative hypotheses.}

For an SE ratio of k and a standardized difference of d, when tests of significance are conducted at nominal level .05, the asymptotic powers of the standard method and the overlap method are given by

$$\Phi(-1.96 + d) + \Phi(-1.96 - d), \tag{6}$$

and

$$\Phi\left(\frac{-1.96(1+k)}{\sqrt{1+k^2}} + d\right) + \Phi\left(\frac{-1.96(1+k)}{\sqrt{1+k^2}} - d\right), \tag{7}$$

respectively, where Φ denotes the standard normal cumulative distribution function.

The asymptotic probability of Type I error is just the asymptotic power when the standardized difference is 0 (i.e., when the null hypothesis is true), so expressions (6) and (7) can be used to calculate the asymptotic probability of Type I error by setting $d = 0$. Figure 3 displays asymptotic probabilities of Type I error for the standard method and overlap method for a range of values of the SE ratio. The probability for the standard method is equal to the nominal level of .05 regardless of the value of the SE ratio, whereas the probability for the overlap method is smaller than .05 and does not even reach .03 until the SE ratio nears the value 9.

Figure 4 displays the asymptotic powers of the standard method and the overlap method as functions of the standardized difference, for SE ratios of 1 and 8. For each method and SE ratio, the asymptotic power for a standardized difference of

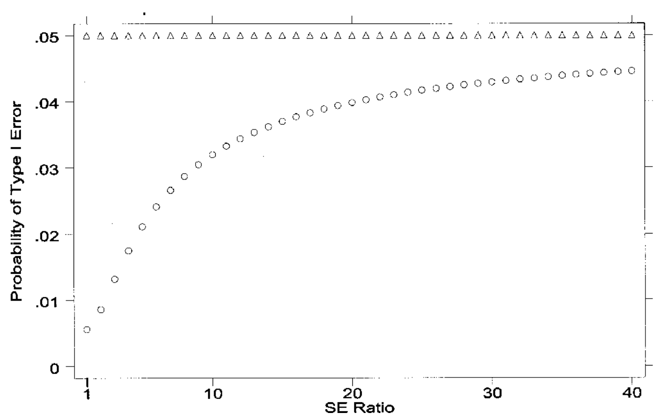

Figure 3. Asymptotic probabilities of Type I error for the overlap method and the standard method when the nominal level of tests is .05. ○ = overlap method; △ = standard method.

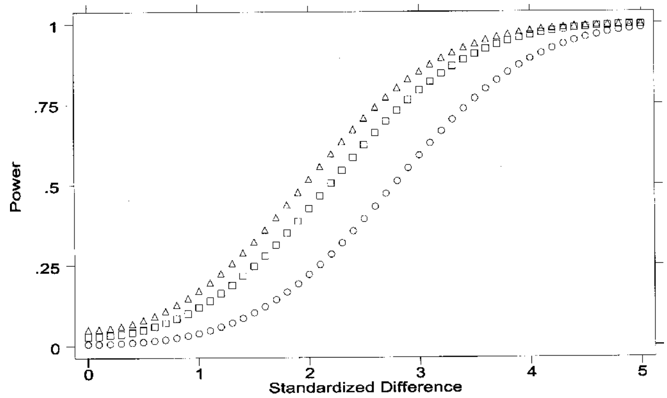

Figure 4. Asymptotic powers of the overlap method and the standard method when the nominal level of tests is .05. ○ = overlap method (SE ratio = 1); □ = overlap method (SE ratio = 8); △ = standard method.

d is equal to that for a standardized difference of $-d$, so we only show values for standardized differences that are greater than or equal to 0. Moreover, for a given value of the standardized difference, the asymptotic power of the standard method does not depend on the SE ratio, which is why there is only one graph for the standard method.

For the SE ratio of 1, large differences in asymptotic power occur for intermediate values of the standardized difference. For example, for a standardized difference of 2.8, which corresponds to the difference between Q_1 and Q_2 being about the same size as the width of interval (1) for Q_1 or interval (2) for Q_2, the power of the standard method is .80, whereas the power of the overlap method is .51. The power of the overlap method is much closer to that of the standard method when the SE ratio is 8. For example, for a standardized difference of 2.8, the power of the overlap method is .73.

4. EXAMINING WHETHER EITHER INTERVAL CONTAINS THE OTHER POINT ESTIMATE

Although it is not the focus of this article, we briey mention another method that we have seen used, which declares a difference to be significant if and only if interval (1) does not contain the value $\hat{Q}_2$ and interval (2) does not contain the value $\hat{Q}_1$. {Another version declares significance if one interval does not contain the other point estimate; see, e.g., Cromwell et al. (1996).} Under the assumptions of Section 2.1, it can be shown that this method tends to be anti-conservative; that is, it rejects the null hypothesis more often than it should when the null hypothesis is true.

As a simple example, consider the setup of Section 2.2, but suppose that the first and second samples have, respectively, 108 and 92 people with the attribute. Then the point estimates are $\hat{Q}_1 = .54$ and $\hat{Q}_2 = .46$, and the confidence intervals for Q_1 and Q_2 are $[.47, .61]$ and $[.39, .53]$, respectively. The method of examining whether either interval contains the other point estimate leads one to declare that the two proportions are significantly different. The standard analysis, however, declares the difference to be nonsignificant since the interval for $Q_1 - Q_2$, $[-.02, .18]$, contains 0.

5. DISCUSSION

5.1 Relationship to Study Design

If a study has been designed specifically for the comparison of two population quantities, then it is likely that the two associated standard errors will not be very different and thus that the overlap method will have substantial deficiencies. For example, consider a stratified random sample with a major goal being the comparison of two stratum means for a variable. If we ignore finite population corrections, the variance of the difference between the stratum estimates is of the form $\sigma_1^2/n_1 + \sigma_2^2/n_2$, where σ_1^2 and σ_2^2 are the stratum variances and n_1 and n_2 are the sample sizes. The variance of the difference is minimized for a given total sample size by setting $n_1/n_2 = \sigma_1/\sigma_2$. Thus, provided that σ_1 and σ_2 do not differ drastically, the individual standard errors, $\mathrm{SE}_1 = \sigma_1/\sqrt{n_1}$ and $\mathrm{SE}_2 = \sigma_2/\sqrt{n_2}$ will not have very different relative sizes. Moreover, the effects of any differences between σ_1 and σ_2 on the individual standard errors will be mitigated somewhat by the fact that more of the sample is allocated to the stratum with the higher variance. Conversely, when a study has not been designed for a specific comparison, as might be the case with a general-purpose survey, the relative sizes of the standard errors have more of a chance of being very different, in which case the overlap method could be less deficient.

5.2 Effects of Correlation

Another consideration is the effect of correlation between the point estimates $\hat{Q}_1$ and $\hat{Q}_2$, which is not an issue when the estimates are independent, as we have assumed to be true asymptotically up to now. In the presence of a non-zero correlation, say ρ, the standard error of $\hat{Q}_1 - \hat{Q}_2$ is $\sqrt{\mathrm{SE}_1^2 + \mathrm{SE}_2^2 - 2\rho \mathrm{SE}_1 \mathrm{SE}_2}$. Thus, if $\rho > 0$, then the overlap method tends to be even more conservative and less powerful, relative to a method that uses an appropriate estimate of standard error, than has been indicated by our results. If $-1 < \rho < 0$, then the overlap method still tends to be deficient, but less so than has been indicated by our results. The degree of deficiency decreases as the correlation becomes more negative. In the unlikely case of a correlation equal to -1, the standard error of $\hat{Q}_1 - \hat{Q}_2$ reduces to $\mathrm{SE}_1 + \mathrm{SE}_2$, and the interval (4) underlying the overlap method corresponds to an interval formed using an appropriate estimate of standard error. It follows that, in the case of perfect negative correlation, the overlap method will tend to be appropriate.

Examples of situations in which positive correlations can occur are blocked experiments, longitudinal studies, and analyses of data in which there is overlap in the sets of elements used to compute the two estimates being compared. Complex survey designs can result in either positive or negative correlations, even when there is no overlap in the sets of elements used for the two estimates. For example, in a comparison of males with females, a correlation could result from the two groups having households in common.

5.3 Conclusions

The overlap method is simple, and it is convenient when lists or graphs of confidence intervals are presented. It can be useful as a quick and relatively rough method for exploratory data analysis. It should not be regarded as an optimal method for significance testing, however, given its conservatism and low power relative to the standard method in the common situation that we have considered. Thus, the overlap method should not be used for formal significance testing unless the data analyst is aware of its deficiencies and unless the information needed to carry out a more appropriate procedure is unavailable.

{Received June 2000. Revised April 2001.}

REFERENCES

Abergel, E., Linhart, A., Chatellier, G., Gariepy, J., Ducardonnet, A., Diebold, B., and Menard, J. (1998), "Vascular and Cardiac Remodeling in World Class Professional Cyclists," *American Heart Journal*, 136, 818–823.

Appleby, J., and Bell, A. (2000), "Reporting NHS Performance: How Did the Media Perform?" *British Medical Journal*, 321, 248.

Bickel, P. J., and Doksum, K. A. (1977), *Mathematical Statistics: Basic Ideas and Selected Topics*, San Francisco: Holden-Day.

Bigby, M., and Gadenne, A. (1996), "Understanding and Evaluating Clinical Trials," *Journal of the American Academy of Dermatology*, 34, 555–590.

——— (1997), Reply to comments on "Understanding and Evaluating Clinical Trials," *Journal of the American Academy of Dermatology*, 37, 804–805.

Brownlee, K. A. (1965), *Statistical Theory and Methodology in Science and Engineering*, New York: Wiley.

Chalmers, R. M., and Salmon, R. L. (2000), "Primary Care Surveillance for Acute Bloody Diarrhea, Wales," *Emerging Infectious Diseases*, 6, 412–414.

Cole, S. R., and Blair, R. C. (1999), "Overlapping Confidence Intervals," *Journal of the American Academy of Dermatology*, 41, 1051–1052.

Collins, S. L., Edwards, J. E., Moore, R. A., and McQuay, H. J. (2001), "Single Dose Dextropropoxyphene, Alone and with Paracetamol (Acetaminophen), for Postoperative Pain" (Cochrane Review), in *The Cochrane Library*, 1, 2001, Oxford: Update Software.

Cox, D. R., and Hinkley, D. V. (1974), *Theoretical Statistics*, London: Chapman and Hall.

Cramér, H. (1946), *Mathematical Methods of Statistics*, Princeton: Princeton University Press.

Cromwell, G. L., Davis, G. W., Morrow, W. E., Primo, R. A., Rozeboom, D. W., Sims, M. D., Stanisiewski, E. P., and Ho, C. H. (1996), "Efficacy of the Antimicrobial Compound U-82,127 as a Growth Promoter for Growing-Finishing Pigs," *Journal of Animal Science*, 74, 1284–1287.

Debboun, M., Strickman, D., Solberg, V. B., Wilkerson, R. C., McPherson, K. R., Golenda, C., Keep, L., Wirtz, R. A., Burge, R., and Klein, T. A. (2000), "Field Evaluation of Deet and a Piperidine Repellent Against *Aedes Communis* (*Diptera: Culicidae*) and *Simulium Venustum* (*Diptera: Simuliidae*) in the Adirondack Mountains of New York," *Journal of Medical Entomology*, 37, 919–923.

Djordjevic, M. V., Stellman, S. D., and Zang, E. (2000), "Doses of Nicotine and Lung Carcinogens Delivered to Cigarette Smokers," *Journal of the National Cancer Institute*, 92, 106–111.

Eising, E. G., and von der Ohe, M. R. (1998), "Differentiation of Prolonged Colonic Transit Using Scintigraphy with Indium-111-Labeled Polystyrene Pellets," *Journal of Nuclear Medicine*, 39, 1062–1066.

Ezeamuzie, C. I., Al-Ali, S. F., Al-Dowaisan, A., Khan, M., Hijazi, Z., and Thomson, M. S. (1999), "Reference Values of Total Serum IgE and Their Significance in the Diagnosis of Allergy Among the Young Adult Kuwaiti Population," *Clinical and Experimental Allergy*, 29, 375–381.

Gotzsche, P. C. (2000), "Do Patients with Osteoarthritis Get the Clinical Research They Need?" *Annals of the Rheumatic Diseases*, 59, 407–408.

Herr, K. A., Mobily, P. R., Kohout, F. J., and Wagenaar, D. (1998), "Evaluation of the Faces Pain Scale for Use with the Elderly," *The Clinical Journal of Pain*, 14, 29–38.

Inzelberg, R., Carasso, R. L., Schechtman, E., and Nispeanu, P. (2000), "A Comparison of Dopamine Agonists and Catechol-O-Methyltransferase Inhibitors in Parkinson's Disease," *Clinical Neuropharmacology*, 23, 262–266.

Kann, L., Kinchen, S. A., Williams, B. I., Ross, J. G., Lowry, R., Grunbaum, J. A., and Kolbe, L. J. (2000), "Youth Risk Behavior Survellance—United States, 1999," *CDC Surveillance Summaries*, MMWR 2000, 49(SS-5).

Mancuso, C. A., Peterson, M. G. E., and Charlson, M. E. (2001), "Comparing Discriminative Validity Between a Disease-Specific and a General Health Scale in Patients with Moderate Asthma," *Journal of Clinical Epidemiology*, 54, 263–274.

Martinez, M. E., Maltzman, T., Marshall, J. R., Einspahr, J., Reid, M. E., Sampliner, R., Ahnen, D. J., Hamilton, S. R., and Alberts, D. S. (1999), "Risk Factors for Ki-Ras Protooncogene Mutation in Sporadic Colorectal Adenomas," *Cancer Research*, 59, 5181–5185.

Mcbride, N. (2000), "The Western Australian School Health Project: Comparing the Effects of Intervention Intensity on Organizational Support for School Health Promotion," *Health Education Research*, 15, 59–72.

Mehta, S. R., Eikelboom, J. W., and Yusuf, S. (2000), Reply to comments on "Intracranial Haemorrhage with Bolus Thrombolytic Agents," *The Lancet*, 356, 1850.

Millikan, R., Pittman, G., Tse, C.-K., Savits, D. A., Newman, B., and Bell, D. (2000), "Glutathione *S*-Transferases M1, T1, and P1 and Breast Cancer," *Cancer Epidemiology Biomarkers & Prevention*, 9, 567–573.

Rahlfs, V.W. (1997), Comment on "Understanding and Evaluating Clinical Trials," *Journal of the American Academy of Dermatology*, 37, 803–804.

Slavotinek, A., Rosenberg, M., Knight, S., Gaunt, L., Fergusson, W., Killoran, C., Clayton-Smith, J., Kingston, H., Campbell, R. H., Flint, J., Donnai, D., and Biesecker, L. (1999), "Screening for Submicroscopic Chromosome Rearrangements in Children with Idiopathic Mental Retardation Using Microsatellite Markers for the Chromosome Telomeres," *Journal of Medical Genetics*, 36, 405–411.

Sont, W. N., Zielinski, J. M., Ashmore, J. P., Jiang, H., Krewski, D., Fair, M. E., Band, P. R., and Létourneau, E. G. (2001), "First Analysis of Cancer Incidence and Occupational Radiation Exposure Based on the National Dose Registry of Canada," *American Journal of Epidemiology*, 153, 309–318.

Tersmette, A. C., Petersen, G. M., Offerhaus, G. J. A., Falatko, F. C., Brune, K. A., Goggins, M., Rozenblum, E., Wilentz, R. E., Yeo, C. J., Cameron, J. L., Kern, S. E., and Hruban, R. H. (2001), "Increased Risk of Incident Pancreatic Cancer Among First-Degree Relatives of Patients with Familial Pancreatic Cancer," *Clinical Cancer Research*, 7, 738–744.

Vatassery, G. T., Smith, W. E., and Quach, H. T. (1998), "α-Tocopherol in Rat Brain Subcellular Fractions is Oxidized Rapidly During Incubations with Low Concentrations of Peroxynitrite," *The Journal of Nutrition*, 128, 152–157.

Wong, C. M., Hu, Z. G., Lam, T. H., Hedley, A. J., and Peters, J. (1999), "Effects of Ambient Air Pollution and Environmental Tobacco Smoke on Respiratory Health of Non-Smoking Women in Hong Kong," *International Journal of Epidemiology*, 28, 859–864.