10

Power of the Rank Test for Multi-Strata Case-Control Studies with Ordinal Exposure Variables

Grzegorz A. Rempala and Stephen W. Looney

1. Introduction

In epidemiological studies of rare diseases (i.e., rare types of cancer) researchers often face major difficulty in obtaining enough cases of the disease to make valid comparisons using odds-ratio estimators. Moreover, they may wish to adjust for the influence of certain extraneous factors so that the effect of the variables of interest can be more clearly visible. This is especially so in case-control studies when it is known that the effects of the risk factor are confounded with such variables as age, sex, and individual physical characteristics of the subjects. These confounding variables often make it difficult (or even impossible) to directly compare the exposed and unexposed groups. Typically, to evaluate the effect of the risk factor in these situations within the odds-ratio framework, methods based on data stratification and within-stratum dichotomization are used. The latter is usually accomplished by classifying cases and controls within each stratum as either exposed or unexposed to the risk factor under investigation. Whereas the stratification is often unavoidable, it may not be practical to dichotomize exposure. Instead, one might wish to consider multiple levels of the exposure variable, based on some appropriate ordinal or even continuous scale (cf., e.g., Greenberg and Tamburro ***[1]***).

The statistical problem of testing for the exposure effect in such settings has been considered by several authors and a variety of approaches have been discussed. In particular, a test based on the rank of the exposure level of each case within a group of individually matched controls was proposed ***(2)*** and a study of its large sample properties under the null hypothesis followed ***(3)***. The analysis in **ref. *3*** indicated that the rank test method is asymptotically efficient

From: *Methods in Molecular Biology, vol. 184: Biostatistical Methods*
Edited by: S. W. Looney © Humana Press Inc., Totowa, NJ

when compared to the best parametric tests under a logistic shift model and under most circumstances significantly outperforms tests based on a dichotomized exposure variable, even for fairly small sample sizes. The successful application of the rank test in several case-control studies when the exposure variable was ordinal rather then dichotomized ***(1,4,5)*** has proved it to be a valuable alternative to more complicated methods, such as multistrata conditional logistic regression. It appears that the rank approach is especially suitable in case-control studies where the exposure variable is poorly characterized but the rank of the exposure of each case among its matching controls is relatively easily established (cf. **Subheading 6.**).

The purpose of this chapter is to examine some properties of the rank-based method within the multistrata case-control study framework. In particular, we are especially concerned here with methods of approximating the power of the appropriate tests for small and moderate sample sizes. This issue is particularly important in the analysis of retrospective/prospective studies of rare diseases (or common diseases of low frequency) when the number of cases is limited. Herein we provide a simple bootstrap algorithm for calculating the approximate power of the appropriate test statistics under translation alternatives. Further, we compare the obtained results with the formulas for exact power of the proposed rank tests under the logistic shift model. The latter is obtained by considering the distributions of appropriate stratum-specific exceedance statistics under the logistic translation alternative. The bootstrap algorithm considered here is somewhat similar to the one for the two-sample Wilcoxon statistic presented by Collings and Hamilton ***(6)***, but unlike their algorithm it appears to be consistent.

2. Testing Against a Shift Alternative with Multiple Strata

Suppose that in our retrospective study we have total of n cases of disease under investigation. Here and elsewhere we assume that n is a fixed and, usually, a small number. For instance, in a typical study of rare diseases one would have $n \leq 10$. Corresponding to the ith case $i = 1,...,n$ are n_i controls, usually matched for known or suspected sources of unwanted variation. The case together with its controls form a stratum, and throughout this chapter we assume that there is only one case per stratum and no cases or controls belong to more then one stratum. These assumptions are made mostly for convenience, and the method presented here can be extended to accommodate more complicated study designs (*see*, e.g., Cuzick ***[3]***). For each stratum i ($i = 1,...,n$) let R_i be the rank of the exposure of the case among all $(n_i + 1)$ individuals in this stratum. There have been at least two tests based on the sum of the R_i's proposed in the literature ***(2,3)***. One is based on the statistic

$$W_1 = \sum_{i=1}^{N} R_i \tag{2.1}$$

and is simply a combination of stratum-specific Wilcoxon two-sample statistics (with respective sample sizes 1 and n_i). The second one is a weighted version of W_1,

$$W_2 = \sum_{i=1}^{N} \frac{R_i}{n_i + 2} \tag{2.2}$$

By considering the marginal likelihood of ranks for all strata combined, it can be argued that the test based on W_2 is locally most powerful for testing against a logistic shift alternative (cf., e.g., Randles and Wolfe [*8*, chap. 9]).

Let $F_i(x)$ be the distribution function of the control exposure in the ith stratum for $i = 1,...,n$. If we assume that the corresponding case exposure distribution is $F_i(x - \Delta)$, where Δ is the across-strata treatment effect independent of i, then either Eq. (2.1) or Eq. (2.2) could be used to test

$$H_0 : \Delta = 0 \text{ vs } H_1 : \Delta > 0 \tag{2.3}$$

rejecting H_0 for large values of the statistic. Whereas for small Δ (local alternative) the test based on W_2 would perhaps be preferred over W_1, especially if we have reason to assume that F_i's are logistic cumulative distribution functions (cdf's), the simpler form of W_1 makes it a reasonable competitor for nonlocal alternatives. Obviously, tests based on W_1 and W_2 will coincide when all strata are of equal size.

As in the one-stratum Wilcoxon rank-sum statistic, both W_1 and W_2 are distribution-free under H_0 but depend upon particular forms of the F_i's under H_1. Thus, without some additional assumptions about the forms of the F_i's, there is no closed-form expression for the power of the tests based on **Eq. (2.1)** or **Eq. (2.2)**. Often, even if such assumptions about the stratum exposure distribution are made, the small number of individuals within strata makes their empirical validation (e.g., by means of a goodness-of-fit test) virtually impossible. This is typically the case in the studies of rare diseases with which we are concerned here.

It seems to be of interest, therefore, to introduce a procedure that would allow us to obtain a reasonable approximation of the power of tests based on **Eq. (2.1)** or **Eq. (2.2)** for testing **Eq. (2.3)** without any reference to the particular form of the stratum-specific exposure distributions. This can be accomplished by implementing a bootstrap method similar to that used for approximating the power of the one-stratum Wilcoxon rank-sum test. However, there is one important difference: in our case the bootstrap algorithm will have to perform in a multistrata setting with only one case exposure value per stratum.

3. Bootstrap Algorithm for Estimating Power

Suppose that the amount of the across-strata shift (Δ) between cases and controls is known to be equal either to 0 or some positive constant Δ and let us denote the power of an α-level test ($0 < \alpha < 1$) based on either W_1 or W_2 against a simple alternative $\Delta = d > 0$ by $\Pi(d,\alpha)$. As indicated earlier, $\Pi(\cdot)$ depends also on the F_i's — the exposure distributions for each strata — but for the sake of simplicity we will not reflect this fact in our notation. It should also be noted here that owing to the discrete nature of the test statistics, α may take only finitely many values and thus when performing exact tests we may achieve only finitely many ("natural") α levels.

To approximate $\Pi(d, \alpha)$ we first must obtain estimates of the exposure distributions F_i for $i = 1,...,n$. A number of different approaches are possible here, depending, for instance, on whether we measure exposure on the continuous or ordinal scale. Because in our study of angiosarcoma (cf. **Subheading 6.**) we may take the F_i's to be continuous, we consider the following version of a stratum-specific empirical cdf.

For the i-th stratum let $z_{(1)},... z_{(n_i + 1)}$ be the ordered values of observed exposure levels for all n_i controls belonging to strata i, combined with the translated i-th case exposure level (i.e., the i-th case exposure value minus the quantity d).We assume for convenience that there are no ties among the z's — should that not be the case the procedure described here applies with minor modifications, as long as there are at least two distinct z's. Define $z_{(0)} = 2z_{(1)} - z_{(2)}$ and $z_{(n_i+2)} = 2z_{(n_i + 1)} - z_{(n_i)}$.

Let $\hat{F}_i$ denote the continuous cdf obtained by assigning probability $1/(n_i + 2)$ uniformly over each interval $(z_{(k)}, z_{(k+1)})$ for $k = 0,...,n_i + 1$. Given the F_i's we may estimate $\Pi(d, \alpha)$ as follows:

1. For each stratum i, draw a computer generated sample of size $n_i + 1$ from $\hat{F}_i$. Add the quantity d to the first observation. This will simulate the shift in the exposure distributions between the case and the controls.
2. Using the first observation as the case value, the statistic W_i ($i = 1, 2$) is calculated from the sample obtained in **step 1**, yielding $W_i^{(0)}$. Let τ be the critical value of the test determined by the condition $\Pi(0, \alpha) = \alpha$. If $W_i^{(0)} \geq \tau$, a success is recorded; otherwise a failure is recorded.
3. Repeat **steps 1** and **2** B times. The bootstrap approximation $\hat{\Pi}(d, \alpha)$ to $\Pi(d, \alpha)$ is given by the binomial proportion of successes among B repetitions. (In all the cases discussed in this chapter, $B \geq 2000$.)

As described in **Subheading 4.**, the simulation study appears to indicate that the above algorithm provides a reasonable estimator of $\Pi(d, \alpha)$ for $n \geq 6$ when the total number of subjects (cases and controls) in all strata combined is at least 36. Asymptotically, the algorithm is consistent, that is, under the sequence of hypotheses $\Delta_N = d_N \to 0$ such that $\Pi(d_N, \alpha) \to \text{const} \geq a$, the differ-

ence between $\Pi(d_N, \alpha)$ and its bootstrap approximation converges in probability to 0 as the number of cases and controls increases to infinity. This may seem somewhat unusual at first glance, as, in general, the distribution of uncentered *U*-statistics cannot be approximated consistently by their bootstrapped versions. However, let us note that in the preceding algorithm we do not use a simple bootstrap replica of W_1 (or W_2), but rather its counterpart based on the correctly shifted exposures of the stratum specific case and matching controls.

If the number of cases and controls is not too small (*see* **Subheading 4.**) the bootstrap CLT provides the following alternative approximation of power. After completing **step 1** as in the preceding, in **step 2** we simply calculate the value of W_i, say W_i^0, and then **steps 1** and **2** are repeated B times to obtain the usual approximations to the mean and variance of the bootstrapped version of W_i

$$\mathrm{E}^*_B(W_i) = \sum_{b=1}^{B} W_i^0(b)/B$$

$$\mathrm{Var}^*_B(W_i) = \frac{1}{B-1} \sum_{b=1}^{B} \left[W_i^0(b) - \mathrm{E}^*_B(W_i) \right]^2 \tag{3.1}$$

The resulting approximate power formula is then given by

$$\Pi(d,\alpha) \approx 1 - \Phi\left(\frac{\tau - \mathrm{E}^*_B(W_i)}{\sqrt{\mathrm{Var}^*_B(W_i)}} \right) \tag{3.2}$$

where $\Phi(\cdot)$ stands for the standard normal cdf. The sketch of the mathematical argument supporting this claim can be found in Rempala et al. *(7)*. In this chapter, we are mostly interested in the performance of the bootstrap approximation for a small number of cases (typically about 10) and are less concerned with its large-sample properties. The results of the computer-simulated study of its accuracy for small to moderate sample sizes under the logistic shift model is presented in the next section.

4. Comparison with the Exact Power Under A Logistic Shift Model

Let X_i^0 be the exposure level of the case in stratum i and let X_i^j, $j = 1,\ldots,n_i$ be the levels of the corresponding controls. The X_i^j's for $j = 1,\ldots,n_i$ are therefore distributed according to F_i. To be able to compare the results of the bootstrap power approximation with the true power we have to derive the exact power formula for testing **Eq. (2.3)** using **Eq. (2.1)** or **Eq. (2.2)** and hence impose at this point some additional assumptions on the exposure distributions F_i for $i = 1,\ldots,n$. In what follows we assume the validity of the so called "shift model," that is, we suppose that for each individual exposure level X_i^j we have

$$X_i^j = \Delta\delta_{0i} + \gamma_i + Z_j \tag{4.1}$$

where δ_{0i} is Kronecker's function, γ_i's are some (unknown) positive constants, and Z_j's are independent, identically distributed random variables with any distribution. For the purpose of the computer simulation described in this subheading we have taken the Z_j's to be logistic $L(0,1)$ random variables. With this particular choice of the Z_j's, we refer to **Eq. (4.1)** throughout the paper as "the logistic shift model" (cf. also Cuzick *[3]*).

The calculation of the exact power of the location shift test (**Eq. [2.3]**) based on **(2.1)** or **(2.2)** under **Eq. (4.1)** with the Z_j's being logistic random variables can be accomplished by considering the distributions of stratum-specific exceedance statistics (cf. Katzenbeisser *[9,10]*) and then combining the results by multiple convolution. Plots of the distributions of the statistic **Eq. (2.1)** and **Eq. (2.2)** under the logistic shift model with eight cases and number of controls per stratum as in the angiosarcoma study (*see* **Subheading 6.**) for the shift values $\Delta = 0, 1, 2, 3$ are presented in **Figs. 1** and **2**. The general formula for the distribution of **Eqs. (2.1)** and **(2.2)** in terms of exceedance statistics under the logistic shift model is provided in Rempala et al. *(7)*. As can be seen from the plots the normal approximation is fairly accurate even for large values of Δ. In fact, the normal approximation works reasonably well for Δ between 0 and 3 for statistics W_1 and W_2 (for W_1 we need to apply a continuity correction) as long as the number of strata is at least 6, the number of controls per stratum is at least 3, and the total number of controls is at least 36.

Having obtained the exact distribution of statistics (**Eqs. [2.1]** and **[2.2]**), for logistic shifts, we may calculate the true power of tests based on W_1 or W_2, which in turn allows us to assess the accuracy of our bootstrap approximation algorithm. Such a comparison with eight cases and numbers of controls coinciding with the numbers from the angiosarcoma study described in **Subheading 6.** is given in **Fig. 3**. As can be seen from the plot, for the logistic shift model, the overall performance of our bootstrap approximation appears to be quite satisfactory even for a relatively small number of cases ($n = 8$).

5. Estimation of the Across-Strata Treatment Effect

So far we have assumed that the amount of the across-strata shift between the exposure distributions of cases and controls (Δ) is known. In practical situations this is obviously rarely the case and we need a way to estimate Δ. One of the standard nonparametric approaches would be to take

$$\hat{\Delta} = \frac{1}{n} \sum_{i=1}^{n} \left(X_i^{\,0} - \overline{X_i} \right) \tag{5.1}$$

but we do not propose to do so because the above estimator of Δ is nonrobust against outliers and for small sample sizes may be quite misleading. More robust estimators, say $\hat{\Delta}_1$ and $\hat{\Delta}_2$, may be obtained by using the Hodges–

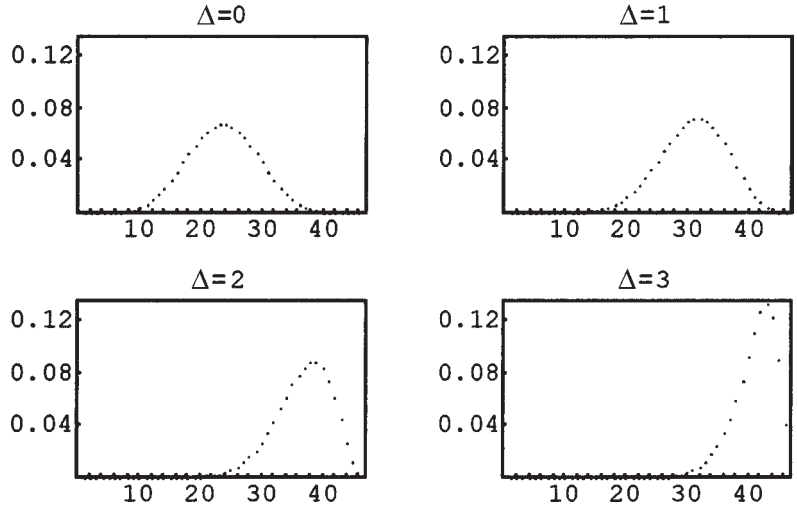

Fig. 1. Relative frequencies for statistic W_1, under different location shift (Δ) values, with eight cases and the number of controls per case equal to 2, 3, 3, 4, 7, 8, 8, 10, respectively. The Central Limit Effect is clearly visible even for moderate $\Delta \leq 2$.

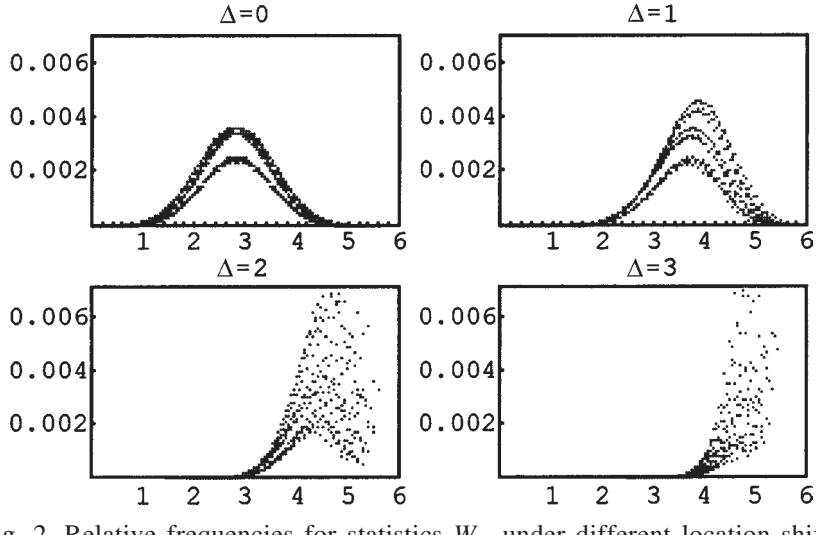

Fig. 2. Relative frequencies for statistics W_2, under different location shift (Δ) values, with eight cases and the number of controls per case equal to 2, 3, 3, 4, 7, 8, 8, 10, respectively. The Central Limit Effect is clearly visible even for moderate $\Delta \leq 2$.

Lehmann method ***(11)*** based on the statistics **Eqs. (2.1)** and **(2.2)**, respectively. The use of the Hodges–Lehmann technique is appropriate here because, under the null hypothesis, W_1 and W_2 are symmetric about

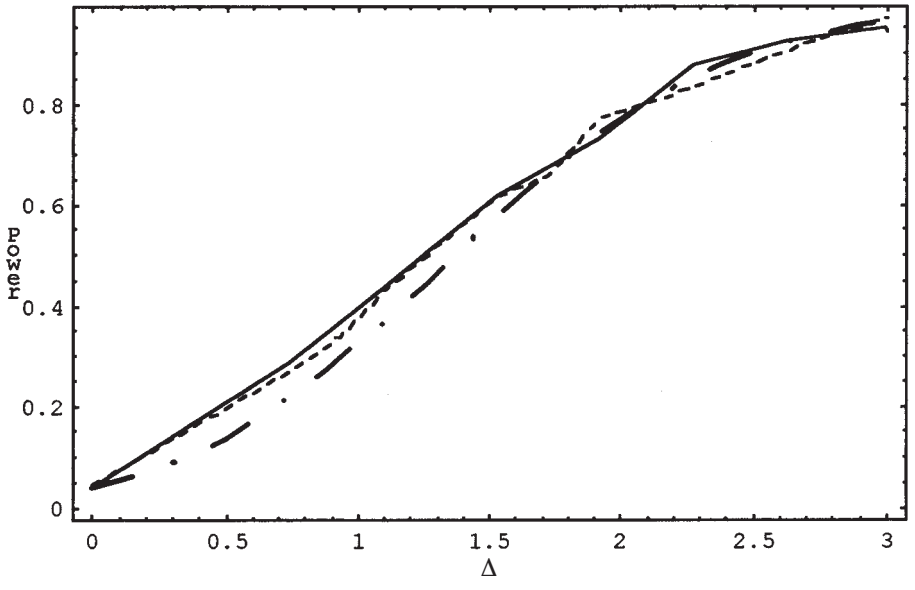

Fig. 3. True power under logistic shift *(chain-dot line)* and its bootstrap approximations using binomial proportion *(solid line)* and bootstrap CLT *(dashed line)* for test statistic W_1 with eight cases and number of controls per case equal to 2, 3, 3, 4, 7, 8, 8, 10, respectively.

$n + \sum_{i=1}^{n} n_i / 2$ and $n + \sum_{i=1}^{n} n_i / (2n_i + 4)$, respectively. To find the explicit forms for the estimators, let us define $D_{ij} = X_i^0 - X_i^j$ for $i = 1,...,n$ and $j = 1,...,n_i$ to be the within-stratum exposure level differences between the case and matched controls. Then, using the Hodges–Lehmann argument (cf., e.g., Randles and Wolfe [*8*, p. 208]), it is easy to show that

$$\hat{\Delta}_1 = med\{D_{ij}\} \tag{5.2}$$

the median of the D_{ij}'s. Similar reasoning applies to $\hat{\Delta}_2$ although now we have to consider a "weighted version" of the D_{ij}'s, due to the weights present in **Eq. (2.2)**. Namely, let $n^* = \Pi_{i=1}^{n} (n_i + 2)$ and let us consider the extended list of D_{ij}'s in which each D_{ij} is repeated exactly $n^*/(n_i + 2)$ times. Then

$$\hat{\Delta}_1 = med\{\text{extended list of } D_{ij}\} \tag{5.3}$$

Obviously, **Eqs. (5.2)** and **(5.3)** coincide when all n_i are equal. If the exposure distributions F_i's are symmetric, then $\hat{\Delta}_1$ and $\hat{\Delta}_2$ are unbiased for Δ. Otherwise, under most circumstances they are median unbiased (or almost so). In our setting, **Eqs. (5.2)** and **(5.3)** appear to be more appropriate estimators of the treatment effect than **Eq. (5.1)**. The estimators of the standard errors of

Eqs. (5.2) and **(5.3)** may again be obtained using the bootstrap method in a fashion similar to **Eq. (3.1)**. Under most circumstances, the standard errors for $\hat{\Delta}$,$\hat{\Delta}_1$, $\hat{\Delta}_2$ and will be of order $O(n^{-1})$ and hence, for small n, the variability of all three estimators could be quite high (cf. **Subheading 6.**). In such cases, in addition to directly estimating Δ one might also wish to estimate $\theta_i = P(X_i^0 > X_i^j)$, the probability that the i-th case exposure exceeds that of its control. This can be easily done, since the quantity $(R_i - 1)/n_i$ is always an unbiased estimator of θ_i. Under the shift model (**Eq. [4.1]**) this parameter, say θ, is the same in all strata and may be estimated unbiasedly by the statistics (cf. also Cuzick ***[3]***):

$$\hat{\theta}_1 = \frac{\sum_{i=1}^{N} (R_i - 1)}{\sum_{i=1}^{N} n_i} = \frac{W_1 - n}{\sum_{i=1}^{N} n_i}$$

and

$$\hat{\theta}_2 = \frac{\sum_{i=1}^{N} (R_i - 1)/(n_i + 2)}{\sum_{i=1}^{N} n_i/(n_i + 2)} = \frac{W_2 - \sum_{i=1}^{N} 1/(n_i + 2)}{\sum_{i=1}^{N} n_i/(n_i + 2)}$$

Because $\hat{\theta}_1$ and $\hat{\theta}_2$ are both linear combinations of stratum specific Wilcoxon–statistics their standard errors will typically be much smaller than that of **Eqs. (5.1)–(5.3)**. Under **Eq. (4.1)** the parameter θ can sometimes be expressed as an explicit function of Δ. For instance, if Δ is close to zero the delta method shows that

$$\Delta \approx \frac{\theta - 1/2}{\int f^2(x)\, dx}$$

provided that $\int f^2(x)\, dx < \infty$, where $f(x)$ is the density function of F_1. Under the logistic shift model, for instance, this yields $\Delta \approx 6\theta - 3$.

6. A Study of Angiosarcoma Occurrences Among Chemical Industry Workers

In this subheading, we illustrate the method with an example that has in fact motivated our study of multistrata rank-based methods. The data presented were collected over the past 20 yr by researchers from the Division of Occupational Toxicology at the University of Louisville School of Medicine as part of an effort to examine the relationship between occupational exposure to suspected carcinogens and the development of disease. In 1974, in response to the discovery of cases of hepatic angiosarcoma among its employees, the B. F. Goodrich Louisville Chemical Plant jointly with the University of Louisville School of Medicine, developed an exposure monitoring system utilizing rank ordering of exposures for highly suspected chemicals. The exposure index combined two components: work history and a

job exposure category based on a 7-point scale that rated the monthly exposure of each employee on any particular job from 0 (absent from the environment) to 6 (very frequent intimate skin contact or high inhalation). The exposure levels for different jobs performed during the month by a given employee were then weighted by their duration to give the total monthly exposure index for that employee. This index was accumulated across months of employment to give the cumulative exposure rank months (CERM). Obviously, CERM is, by its design, a very imprecise measure of exposure but was thought to be the best available, because only a minimal amount of information concerning historical exposures could be found in the company records. Thus, owing to the nature of CERM the standard analysis based on the conditional likelihood method and logistic regression or the hypergeometic distribution (as offered, for instance, by StatXact) could yield misleading results. The rank method based only on the relative magnitude of exposures seems to be more appropriate.

In **Tables 1** and **2**, we present the total stratum-standardized CERM for the exposure to the chemical vinyl chloride, along with corresponding rank-statistics, for the eight cases of angiosarcoma identified among B. F. Goodrich Chemical Plant employees between the time of the first case (diagnosed in January, 1974) and January, 1998, together with the total stratum-standardized CERM for the controls matched by sex, age, and length of employment. Standardization of CERM within any given strata is obtained by dividing the raw CERM values by the stratum-specific CERM standard deviation for the controls. Of course, the underlying assumption here is that, within-strata, exposures of the case and its corresponding controls are measured on the same scale.

It is easily seen from the values of the statistics reported in **Table 2** that the data show significant association of the exposure to vinyl chloride (as measured in CERM) with the development of hepatic angiosarcoma. When testing **Eq. (2.3)** by means of **Eq. (2.1)** and **Eq. (2.2)**, we find the normal approximations of the P-values to be 0.01262 and 0.01252, respectively. The achieved power of the tests is estimated to be about 0.60 for W_1 and about 0.65 for W_2 at the 5% significance level.

7. Summary

The rank method of analyzing multistrata case-control studies was considered. The consistent bootstrap algorithm for approximating the power of the rank-based test was presented. The computer simulation study presented in this chapter indicates that, under a logistic shift alternative, the bootstrap algorithm provides a reasonably good approximation of the true power, even for relatively small sample sizes, provided that the across-strata shift (treatment effect) is known. For an unknown shift, a consistent and robust estimator of the Hodges–Lehmann type was proposed.

Table 1
Values of the Total Stratum-Standardized CERM for Vinyl Chloride of the Eight Angiosarcoma Cases and Their Matched Controls Not Developing the Disease Among the Employees of B. F. Goodrich Chemical Plant (Matching Done by Sex, Age, and the Length of Employment)

Case value	Number of controls (n_i)	Control values	Case rank (R_i)
3.91	8	2.14, 2.66, 2.76, 2.99, 4.39, 1.13, 1.89, 4.41	7
4.05	7	2.41, 2.39, 4.04, 0.98, 1.69, 3.27, 2.07	8
2.76	8	2.99, 1.82, 1.8, 3.63, 2.00, 1.18, 2.33, 2.85	6
4.21	10	2.4, 1.56, 3.27, 1.93, 2.3, 1.09, 2.53, 4.56, 3.59, 2.96	10
5.78	3	5.25, 4.5, 3.36	4
1.87	2	1.18, 0.82	3
2.82	4	1.24, 3.57, 1.63, 2.63	4
4.85	3	5.16, 3.13, 5.23	2

Table 2
The Values of Test Statistics W_1 and W_2, Along with the Corresponding Estimates of the Shift (Δ), Mean Percentile Shift (θ), and Achieved Power

	W_i	P-Value	$\hat{\Delta}$ (s.e.)	$\hat{\Delta}_i$ (s.e.)	$\hat{\theta}_i$ (s.e.)	$\hat{\prod}(\hat{\Delta}_i, 0.05)$
$i = 1$	43.00	0.01262	1.03 (0.20)	1.13 (0.15)	0.8 (0.056)	0.60
$i = 2$	5.53	0.01252	1.03 (0.20)	1.18 (0.10)	0.79 (0.05)	0.65

References

1. Greenberg, R. A. and Tamburro, C. H. (1981) Exposure indices for epidemiological surveillance of carcinogenic agents in an industrial chemical environment. *J. Occup. Med.* **23,** 353–358.
2. van Elteren, P. (1960) On the combination of independent two-sample tests of Wilcoxon. *Bull. Int. Statist. Instit.* **37,** 351–361.
3. Cuzick, J. (1985) A method for analyzing case-control studies with ordinal exposure variables. *Biometrics* **41,** 609–621.

4. Cuzick, J., Bulstrode, J. C., Stratton, I.,Thomas, B. S., Bulbrook, R. D., and Hayward, J. L. (1983) A prospective study of urinary androgen levels and ovarian cancer. *Int. J. Cancer* **32,** 13–19.
5. Kwa, H. G., Cleton, F., Wang, D. Y., Bulbrook, R. D., Bulstrode, J. C., Hayward, J. L., et al. (1981) A prospective study of plasma prolactin levels and subsequent risk of breast cancer. *Int. J. Cancer* **28,** 673–676.
6. Collings, B. J. and Hamilton, M. A. (1988) Estimating the power of the two-sample Wilcoxon test for location shift. *Biometrics* **44**, 847–860.
7. Rempala, G., Looney, S., Tamburro, C., and Fortwengler, P (1998) *Power of the rank test for multi-strata case-control studies*. University of Louisville Department of Mathematics Technical Report no. 1/98.
8. Randles, R. H. and Wolfe, D. A. (1979) *Introduction to the Theory of Nonparametric Statistics.* Krieger, Malabar, FL.
9. Katzenbeisser, W. (1985) The distribution of two-sample location exceedance statistic under Lehmann alternatives. *Statistische Hefte* **26,** 131–138.
10. Katzenbeisser, W. (1989) The exact power of two-sample location tests based on exceedance statistics against shift alternatives. *Statistics* **20,** 47–54.
11. Hodges, J. L. and Lehmann, E. L. (1963) Estimates of location based on rank tests. *Ann. Math. Statist.* **34,** 598–611.