ALTERNATIVES TO ROTHMAN'S TEST FOR SYNERGISM (OR ANTAGONISM) IN COHORT STUDIES

by

Michael D. Hogan[1], Lawrence L. Kupper[2],
Bernard M. Most[2], and Joseph K. Haseman[1]

[1]Biometry Branch
National Institute of Environmental Health Sciences

[2]Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1139

August 1977

ALTERNATIVES TO ROTHMAN'S TEST FOR SYNERGISM (OR ANTAGONISM) IN COHORT STUDIES *

by

Michael D. Hogan[1], Lawrence L. Kupper[2],
Bernard M. Most[2], and Joseph K. Haseman[1]

---

[1]Biometry Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, 27709

[2]Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, North Carolina, 27514.

*Dr Kupper's work on this project was supported by NIH Research Career Development Award (No. 1-K04-ES00003) from the National Institute of Environmental Health Sciences.

*Dr. Most's work on this project was supported by NIH Research Fellowship Award #1-F22-ES01766-02.

ABSTRACT

Rothman has explored in some detail the issue of assessing the potential presence of synergism (or antagonism) in data generated from either a cohort or a case-control study. Arguing that the "natural" scale for quantifying the joint effects of two or more factors acting in combination is the probability scale, he has proposed a procedure based on a ratio-type index for evaluating two-factor interaction in the presence of non-zero background effects. This paper reviews the rationale underlying Rothman's approach for a cohort study and presents a simpler and more appropriate test procedure (utilizing a linear contrast of the observed risks) for the additive approximation to his basic probabilistic model of "no interaction." A likelihood ratio test based on his original model is also proposed, as well as a closed form approximation to it. Finally, the assessment of interaction in cohort studies involving exposure factors measured at more than two levels is addressed.

1. INTRODUCTION

Rothman ([5], [6]) has expressed concern that the evaluation of synergism/antagonism or its statistical counterpart, interaction, can be a rather subjective process because of its dependence on the scale of measurement employed in the analysis. He has also asserted that for the binary (e.g., present or absent) types of variables typically encountered in epidemiologic studies of mortality or morbidity, the "natural" scale for quantifying the joint effects of two or more factors acting in combination is the probability scale. Operating within such a probability-based framework, Rothman has proposed a procedure which involves the use of an index for assessing two-factor synergism (or antagonism, which hereafter will be assumed to be encompassed in the single term synergism) in the presence of various background phenomena which may also be contributing to disease risk.

This paper examines in detail the underlying basis of Rothman's approach and presents an alternative test procedure for his modified model of "no synergism" in cohort studies involving two factors, each at two levels. A likelihood ratio test of his original union of independent events model is proposed, and a closed form approximation to this test is also considered. In addition, some methodological objections are raised concerning Rothman's suggested procedure for analyzing data with factors measured at more than two levels. The scope of this paper is restricted to data generated in a cohort framework since the union of independent events probability model originally proposed by Rothman does not have a direct analog in the case-control situation.

2. DISCUSSION OF ROTHMAN'S APPROACH

The rationale underlying Rothman's methodology can be described as follows: Let P_A (or P_B) denote the probability that the disease in question will occur if exposure factor A (or B) acts in isolation (i.e., without background

factors present), and let P_C be the probability that the disease is produced by background factors (C) acting alone. Now, suppose that A and C are both present but act independently of one another in such a way that the expected number developing the disease because of exposure to A depends on C only in the sense that the number at risk to A would be those not already affected by C, and vice versa. Thus, if N subjects are exposed simultaneously to both A and C, then, under the type of independence described above, NP_C subjects would be expected to develop the disease because of C. Of the remaining $(N-NP_C)$ subjects at risk, the expected number incurring disease due to A would then be $(N-NP_C)P_A$; and the total expected number of subjects developing the disease would be given by

$$NP_C + (N-NP_C)P_A,$$

or

$$N(P_A+P_C-P_AP_C) .$$

(This same expression would be obtained, of course, if the exposure sequence were reversed.) The expression $(P_A+P_C-P_AP_C)$ is immediately recognized as the probability of the union of two events which are independent in the "usual" probabilistic sense.

Analogously, the probability of disease occurring given the simultaneous exposure to two factors A and B in the presence of C (with independence assumptions as described above) would be

$$P_A+P_B+P_C-P_AP_B-P_AP_C-P_BP_C+P_AP_BP_C, \qquad (1)$$

which is exactly the probability structure expected under Rothman's definition of "no synergism."

Rothman notes that neither P_A nor P_B is directly observable in a cohort study, since background factors are invariably present. Thus, it is necessary to reparametrize the "no interaction" model (1) in terms of disease probabilities that are directly estimable from cohort data. In particular, under the assumption that the background C acts independently of factors A and B (an assumption that may be difficult to verify in actual practice), (1) can be re-written as

$$R_{00} + (R_{10}-R_{00}) + (R_{01}-R_{00}) - \frac{(R_{10}-R_{00})(R_{01}-R_{00})}{(1-R_{00})} , \qquad (2)$$

where

$$R_{00} = P_C,$$
$$R_{10} = P_A+P_C-P_AP_C,$$
$$R_{01} = P_B+P_C-P_BP_C .$$

The probabilities (or risks) R_{00}, R_{10} and R_{01} can be estimated directly from an appropriately designed cohort study, but not from case-control data. (The probability delineated in equation (2) could be divided by R_{00}, thereby converting it to a statement about relative risk. It would then involve some quantities that are ordinarily estimable using case-control data, but would still depend on R_{00}.)

Now, if R_{11} represents the actual probability of disease occurring due to the simultaneous exposure to factors A and B in the presence of C (a rate which is also estimable from cohort data), then a reasonable measure of the degree of synergism would necessarily involve a comparison of R_{11}

with (2), or similarly, a comparison of the deviations of each of these quantities from the background rate R_{00}. In this regard, Rothman has suggested the use of a ratio-type parameter of the form

$$S = \frac{(R_{11}-R_{00})}{(R_{10}-R_{00}) + (R_{01}-R_{00}) - \frac{(R_{10}-R_{00})(R_{01}-R_{00})}{(1-R_{00})}} .$$

Synergism, antagonism, or no interaction between factors A and B would be indicated, respectively, by S being greater than, less than, or equal to 1. However, Rothman does not actually work with the index S. Instead, he notes that in most epidemiologic investigations the effects of the individual exposure factors relative to background are sufficiently small so that a product term like $(R_{10}-R_{00})(R_{01}-R_{00})/(1-R_{00})$ can be neglected for all practical purposes. As a result, he drops this particular term in the denominator of S and proceeds to consider only the modified index

$$S' = \frac{(R_{11}-R_{00})}{(R_{10}-R_{00}) + (R_{01}-R_{00})} .$$

To make inferences about S', Rothman assumes that

$$\ln\hat{S}' = \ln(\hat{R}_{11}-\hat{R}_{00}) - \ln(\hat{R}_{10}+\hat{R}_{01}-2\hat{R}_{00}),$$

(where $\hat{R}_{ij}$ is an appropriate estimator of R_{ij}) has an approximate normal distribution for large samples with an estimated standard error $SE(\ln \hat{S}')$ based on a first-order Taylor series approximation. Evaluation of the potential presence of synergism in the data set is then determined by whether or not the $100(1-\alpha)\%$ confidence interval

$$\exp\left[\ell n\hat{S}' \pm Z_{1-\frac{\alpha}{2}} SE(\ell n\hat{S}')\right]$$

includes the value 1.

3. ALTERNATIVE TEST PROCEDURES

For the modified index S', "no interaction" is equivalent to "S' = 1", or

$$(R_{11}-R_{00}) = (R_{10}-R_{00}) + (R_{01}-R_{00}),$$

which, on further simplification, yields

$$R_{11}-R_{10}-R_{01}+R_{00} = 0.$$

But,

$$T' = (R_{11}-R_{10}-R_{01}+R_{00})$$

is nothing more than the usual analysis of variance-type interaction contrast for the situation involving two factors (say, A and B), each at two levels (e.g., A_0 & A_1 and B_0 & B_1). Thus, a simple, direct approach for testing H_0: S'=1 is to consider instead the standard statistical hypothesis H_0: T'=0 of no interaction (i.e., additivity) for the following two-way table of risks:

	B_0	B_1
A_0	R_{00}	R_{01}
A_1	R_{10}	R_{11}

The test that is commonly employed in this instance (e.g., see [7], pp. 495-496) is based on the statistic

$$Z' = \frac{\hat{T}' - 0}{\sqrt{\hat{V}(\hat{T}')}}, \qquad (3)$$

where

$$\hat{T}' = \hat{R}_{11} - \hat{R}_{10} - \hat{R}_{01} + \hat{R}_{00} ,$$

$$\hat{V}(\hat{T}') = \sum_{i=0}^{1} \sum_{j=0}^{1} \frac{\hat{R}_{ij}(1-\hat{R}_{ij})}{N_{ij}} ,$$

N_{ij} is the number of subjects at risk in the (i,j)th cell, and $\hat{R}_{ij}$ is assumed to be a binomial (N_{ij}, R_{ij}) random variable. Under the null hypothesis, the test statistic Z' has an approximate standard normal distribution given reasonably-sized $|N_{ij}|$.

Although a ratio-type estimator such as $\hat{S}'$ has a certain amount of intuitive appeal because of its parallelism to traditional epidemiologic indices like relative risk, there are a number of problems associated with its actual use. In the first place, there are occasions when the observed value of $\ell n\hat{S}'$ will not be defined. For example, if $\frac{1}{2}(\hat{R}_{10}+\hat{R}_{01}) < \hat{R}_{00}$, then the denominator of $\hat{S}'$ is negative, so that $\ell n\hat{S}'$ does not exist for any $\hat{R}_{11} > \hat{R}_{00}$. And, if $\frac{1}{2}(\hat{R}_{10}+\hat{R}_{01}) = \hat{R}_{00}$, then $\ell n\hat{S}'$ is undefined. Yet, if the underlying individual effect probabilities R_{10} and R_{01} are small, these types of observed outcomes are frequently encountered. [E.g., if $R_{00}=0.001$, $R_{10}=R_{01}=0.002$, and $N_{ij}=500$ for all (i,j), then $Pr(\hat{R}_{10}+\hat{R}_{01} \leq 2\hat{R}_{00}) = 0.373$.] Secondly, even if R_{10} and R_{01} are large enough to ensure that the likelihood of observing an outcome for which $\ell n\hat{S}'$ is undefined is essentially zero, the test based on $\ell n\hat{S}'$ is still apt to be unduly conservative. [E.g., if $R_{00}=0.01$, $R_{10}=R_{01}=0.20$, and $N_{ij}=50$

for all (i, j), then $Pr(\hat{R}_{10}+\hat{R}_{01} \leq 2\hat{R}_{00})$ is only 3×10^{-6}. However, the true significance level (determined by actual enumeration) associated with a nominal 5% level test is 0.015, as compared to a corresponding value of 0.055 for the test procedure based on the statistic $\hat{T}'$.] This conservatism arises in part because of Rothman's implicit assumption ([6], p. 507) that the $|\hat{R}_{ij}|$ are distributed as Poisson random variables, which tends to inflate $SE(\ell n\hat{S}')$, particularly when the observed $|\hat{R}_{ij}|$ are large. In contrast, the test statistic given by equation (3) is defined for all possible values of the observed risks and leads to a procedure which appears to operate at approximately the proper significance level for a reasonably wide range of trial cases. Furthermore, $\hat{T}'$ provides a direct measure of the degree of synergism present in the observed data set, and is greater than, equal to or less than zero according as $\hat{S}'$ (when defined) is greater than, equal to, or less than one. Thus, the test based on $\hat{T}'$ would seem to be the appropriate one to use when considering the modified hypothesis H_0: $S'=1$.

Now, there may well be occasions when the probabilities R_{10} and R_{01} are large enough to preclude the use of an additive model as an approximation to the probabilistic conceptualization of no synergism as defined by equations (1) and (2). When this situation occurs, a test of Rothman's original hypothesis of no interaction (namely, that R_{11} has the structure given by (2)) can be developed using a likelihood ratio approach. Given the earlier assumption that we are dealing with four independent binomial distributions corresponding to the four exposure categories, it follows that the likelihood function for the data is

$$L = \prod_{i=0}^{1}\prod_{j=0}^{1}\binom{N_{ij}}{n_{ij}}R_{ij}^{n_{ij}}(1-R_{ij})^{N_{ij}-n_{ij}},$$

where n_{ij} denotes the number of subjects who develop disease out of the N_{ij} at risk in the (i,j)-th category. Based on this likelihood function, a likelihood ratio statistic ($\hat{\lambda}$, say) can be calculated as

$$\hat{\lambda} = \hat{L}_0/\hat{L},$$

where $\hat{L}_0$ is the value of L when maximized subject to the constraint (2) and $\hat{L}$ is the unconstrained maximum value of L based on the observed data. It is well known that $-2\ell n\hat{\lambda}$ has approximately a χ^2_1 distribution for large $|N_{ij}|$ under the "no interaction" null hypothesis of equation (2). Since explicit expressions for the estimates of the $|R_{ij}|$ under the constrained maximization procedure cannot be obtained, it is necessary to employ an iterative function maximization computer algorithm to compute the value of $\hat{L}_0$. For example, a direct search program such as MAXLIK can be used (see [3] for a description).

To obtain some information concerning the properties of the proposed likelihood ratio test procedure, simulation studies were conducted. Table 1 contrasts the observed (i.e., empirical) significance levels determined by simulation with corresponding nominal levels of 0.10 and 0.05. For purposes of comparison, the "true" significance level associated with the test based on $\ell n\hat{S}'$ is also included. (In any given instance, this value is determined by summing the probabilities of occurrence of all outcomes that lead to a rejection of the null hypothesis. Since outcomes for which the test statistic $\ell n\hat{S}'$ is undefined have been ignored in the summation process, the probability of observing such outcomes is given in Table 1 in conjunction with the "true" significance level.)

TABLE 1. Comparison of Empirical and Nominal Significance Levels for the Likelihood Ratio Test and the Test Based on $\ell n\hat{S}'$

						Significance Levels			
	Number of	Population Rates					Empirical		
N[1.]	Replicates[2.]	R_{00}	R_{10}	R_{01}	R_{11}[3.]	Nominal	L.R.T.	$\ell n\hat{S}'$	
20	600	.01	.10	.10	.1818	.10,.05	.119,.063	.033,.005	(.060)[4.]
50	400						.115,.058	.052,.021	(.004)
20	200	.01	.20	.20	.3535	.10,.05	.095,.030	.045,.014	(.003)
50	200						.100,.065	.061,.025	(<.001)

1. The N_{ij} were taken equal to a common N to simplify calculations
2. For the L.R.T. simulations
3. Values of R_{11} were determined using the "no interaction" model (2)
4. Probability of undefined outcomes

The findings displayed in Table 1 suggest that the likelihood ratio procedure is operating at the correct significance level (at least for the limited number of cases that were considered), and that the use of the $\ell n\hat{S}'$ test leads to an unduly conservative procedure. Because of this conservatism, it did not seem meaningful to compare the powers of the two tests under discussion. [However, simulation studies did suggest that the likelihood ratio test appears to attain a reasonable level of power (i.e., .70-.80) when $N \geq 50$ and R_{11} is about twice its null value under (2)].

As an alternative to the likelihood ratio test procedure discussed above, one can define a specific measure for quantifying synergism and can then develop a closed-form test based on this measure. A logical measure to consider would be the <u>difference</u> between the probability of disease given joint exposure to both factors A and B and the null value of this probability under the model defined by equations (1) and (2). If equation (2) is re-written as

$$1 - \frac{(1-R_{10})(1-R_{01})}{(1-R_{00})} = R_{11}^{(o)} \text{ , say,} \qquad (4)$$

then the proposed measure would be

$$T = R_{11} - R_{11}^{(o)} \text{ .} \qquad (5)$$

Following the approach of Rothman in the development of his asymptotic test based on $\ell n\hat{S}'$, we can obtain a <u>biased</u> estimate of $R_{11}^{(o)}$ by replacing the population rates in equation (4) by their unconstrained maximum likelihood estimates; i.e., we can set

$$\hat{R}_{11}^{(o)} = 1 - \frac{(1-\hat{R}_{10})(1-\hat{R}_{01})}{(1-\hat{R}_{00})} \text{ ,}$$

where the $|\hat{R}_{ij}|$ are the observed sample proportions or risks. Then, a test of the null hypothesis H_o: $T=0$ could be based on the statistic

$$Z = \frac{\hat{T} - 0}{\sqrt{\hat{V}(\hat{T})}}, \tag{6}$$

where

$$\hat{T} = \hat{R}_{11} - \hat{R}_{11}^{(o)} ,$$

and

$$\hat{V}(\hat{T}) \doteq \frac{\hat{R}_{11}(1-\hat{R}_{11})}{N_{11}} + \frac{(1-\hat{R}_{10})(1-\hat{R}_{01})}{(1-\hat{R}_{00})^2} \left\{ \frac{\hat{R}_{00}(1-\hat{R}_{10})(1-\hat{R}_{01})}{(1-\hat{R}_{00})\, N_{00}} + \frac{\hat{R}_{10}(1-\hat{R}_{01})}{N_{10}} + \frac{\hat{R}_{01}(1-\hat{R}_{10})}{N_{01}} \right\} .$$

The performance of the statistic Z as an approximate unit normal random variable was examined for a series of different cases by comparing its actual significance level with the corresponding nominal significance level. The results of this examination are summarized in Table 2. The corresponding results for a test based on $\hat{T}'$ are also included to provide some insight into the reasonableness of the additive approximation to the underlying probabilistic model (2).

TABLE 2. Comparison of Actual and Nominal Significance Levels for Tests Based on $\hat{T}$ and $\hat{T}'$

	Population Rates				Significance Levels		
						Actual	
N	R_{00}	R_{10}	R_{01}	R_{11}[1]	Nominal	$\hat{T}$	$\hat{T}'$
10	.01	.10	.10	.1818	.10,.05	.119,.086	.120,.065
20						.115,.061	.113,.061
50						.109,.056	.107,.054
100						.103,.052	.105,.053
10	.01	.20	.20	.3535	.10,.05	.134,.086	.139,.073
20						.113,.062	.121,.065
50						.108,.055	.122,.065
100						.103,.052	.139,.076

1. Values of R_{11} were determined using the "no interaction" model (2).

The results displayed in Table 2 suggest that the test based on $\hat{T}$ seems to be operating at about the same level as the likelihood ratio procedure (see the corresponding entries in Table 1), and that the actual significance level, which appears to be consistently inflated, is converging to the nominal level with increasing N. As would be expected, when the departure of the additive approximation from the underlying probabilistic model becomes more drastic, an increase in N does not lower the actual (inflated) significance level associated with the $\hat{T}'$-based procedure.

4. INTERACTION WITH FACTORS AT MORE THAN TWO LEVELS

When at least one of the exposure factors in a cohort study is measured at more than two levels (e.g., is no longer regarded just as being either present or absent), the probabilistic interpretation of synergism based on the union of independent events concept (see Section 1) is no longer obvious. Given that an additive model in terms of observable risks (the model implicitly assumed under Rothman's modified null hypothesis) is a reasonable representation of no synergism when multiple levels are involved, Rothman's procedure is not the approach that typically would be employed to test for synergism.

Under his approach, all possible two-by-two sub-tables that can be formed from the original (RXC) table with R_{00} as one of the cell entries are generated, and the level of "synergism" for the (i, j)-th subtable is assessed using the ratio parameter

$$S'_{ij} = \frac{(R_{ij}-R_{00})}{(R_{i0}-R_{00}) + (R_{0j}-R_{00})} ; \quad \begin{matrix} i = 1,2,\ldots,R-1, \\ j = 1,2,\ldots,C-1 . \end{matrix}$$

The final conclusion regarding the presence or absence of interaction in the overall data set is based on whether or not a confidence interval for a weighted average of the $|S'_{ij}|$, say

$$\bar{S}' = \sum_{i=1}^{R-1}\sum_{j=1}^{R-1} W_{ij}S'_{ij} \Big/ \sum_{i=1}^{R-1}\sum_{j=1}^{C-1} W_{ij} \quad ,$$

contains the value one. According to Rothman, if the $\left|\hat{S}'_{ij}\right|$ are not equal, then standardization is required, with the $\left|W_{ij}\right|$ being based on some suitable external standard. On the other hand, if the $\left|\hat{S}'_{ij}\right|$ do not vary, then a pooled effect measure is desired; and he recommends that W_{ij} be chosen inversely proportional to the estimated variance of $\ell n\hat{S}'_{ij}$.

There are a variety of concerns that can be raised about this proposed extension of Rothman's approach. As indicated in Section 3, test procedures based on the ratio-type estimators $\left|\hat{S}'_{ij}\right|$ may not always be defined. In addition, Rothman suggests a specific form for the weighting factors $\left|W_{ij}\right|$ only for the special case when all of the $\left|\hat{S}'_{ij}\right|$ are equal. Yet, it is far from obvious that this is the situation usually encountered in actual practice. If the $\left|\hat{S}'_{ij}\right|$ cannot be assumed to be equal (which is an issue that needs to be resolved analytically by some appropriate statistical test), then several different problems can arise. For instance, there may be occasions when the individual $\left|\hat{S}'_{ij}\right|$ exhibit marked departures (in both directions) from one and yet the weighted $\hat{\bar{S}}'$ "cancels out" these effects and gives a misleading indication of no interaction. Moreover, it is not immediately apparent that an appropriate, external standard can always be defined. (For example, if the exposure factors are cigarette smoking and alcohol consumption, then it is not clear how an external population for standardization would be selected or utilized in a meaningful way.) Finally, even if the $\left|\hat{S}'_{ij}\right|$ can be regarded as uniform, it is not obvious what physical interpretation is to be assigned to the weighted index $\hat{\bar{S}}'$.

All of these problems can be avoided by employing any of the standard procedures for analyzing complex contingency tables that have appeared in the statistical literature. For example, the Grizzle-Starmer-Koch [2] linear models approach to the analysis of categorical data is a very general procedure that permits one to analyze the type of data under discussion in an analysis of variance framework (e.g., see [4]). Under this approach, important individual components of the overall (R-1)(C-1) degrees of freedom interaction effect can be examined, and potentially confounding variables can be included in the model and evaluated in a logical manner.

Bishop, Feinberg and Holland [1] (in a comprehensive text that reviews this entire field) point out that if several of the observed cell entries are zero, then some problems may be encountered with the estimation procedure employed by Grizzle et. al. To circumvent this problem, Bishop and her co-authors have proposed a method for dealing with zero cell counts. They also support the use of a maximum likelihood approach as an alternative to the procedure advocated by Grizzle et. al.

5. CONCLUDING REMARKS

Most of the emphasis of this paper has been placed on a probabilistic model for synergism derived from a union of independent events concept. Yet, while there may well be a variety of situations that can be assumed to be adequately represented or at least approximated by the probabilistic conceptualization of synergism under discussion, it is not the purpose of this paper to advocate the indiscriminate use of this model. It is the prerogative as well as the responsibility of the individual investigator to draw upon his or her past experience and knowledge of the biological mechanisms underlying a given data set in order to select the most appropriate model for assessing the potential presence of synergism.

REFERENCES

(1) Bishop, Y. M. M., Feinberg, S. E., and Holland, P. W. (1975). Discrete Multivariate Analysis: Theory and Practice, The M.I.T. press.

(2) Grizzle, J.E., Starmer, C. F., and Koch, G. G. (1969). Analysis of categorical data by linear models. Biometrics, 25, (3), 489-504.

(3) Kaplan, E. B., and Elston, R. C. (1972). A subroutine package for maximum likelihood estimation (MAXLIK). U.N.C. Inst. of Stat. Mimeo Series No. 823.

(4) Koch, G. G., Johnson, W. D., and Tolley, H. D. (1972). A linear models approach to the analysis of survival and extent of disease in multi-dimensional contingency tables. Jour. of Amer. Stat. Assoc., 67, (340), 783-796.

(5) Rothman, K. J. (1974). Synergy and antagonism in cause-effect relationships. Amer. Jour. of Epid., 99, (6), 385-388.

(6) -------(1976). The estimation of synergy or antagonism. Amer. Jour. of Epid., 103, (5), 506-511.

(7) Snedecor, G. W., and Cochran, W. G. (1967). Statistical Methods, Sixth Edition, Iowa State University Press.