J Chron Dis Vol. 34, pp. 463 to 468, 1981
Printed in Great Britain.

0021-9681/81/090463-06$02.00/0
Pergamon Press Ltd

ON SUMMARY ESTIMATORS OF RELATIVE RISK

ROBERT E. TARONE

National Cancer Institute, Landow Building, Room 5C25 Bethesda, MD 20205, U.S.A.

(*Received in revised form* 21 *April* 1980)

Abstract—Two summary relative risk estimators, which are analogues of the Mantel–Haenszel summary odds ratio, are derived for use in prospective studies with stratified data. One of the proposed summary relative risks is shown to be closely related to the maximum likelihood estimator of a common risk ratio, assuming a Poisson distribution for the number of cases in each stratum. This estimator is compared to a recently proposed index of mortality, the Relative Risk Index [1].

INTRODUCTION

THE PROBLEM of obtaining a summary estimate of relative risk from stratified incidence or mortality data arises often in epidemiological studies. For data obtained in a retrospective study, Mantel and Haenszel have proposed a summary odds ratio which provides an estimate of relative risk if the incidence or mortality rates being compared are small [2]. A summary relative risk estimator which is an analogue of the Mantel–Haenszel summary odds ratio is derived for use in prospective studies. This estimator is compared in several examples to the Relative Risk Index proposed recently by Lilienfeld and Pyne [1].

DERIVATION OF SUMMARY RELATIVE RISK ESTIMATOR

Consider a population from which incidence or mortality data have been collected for some disease. Suppose that the data for males and females have been reported for m age strata indexed by $i = 1, 2, \ldots m$. The data from stratum i can be summarized in a 2×2 table such as that shown in Table 1. For stratum i, let the probability of a case in males be denoted by p_{1i} and the probability of a case in females be denoted by p_{2i}, and let $q_{ji} = 1 - p_{ji}$ for $j = 1, 2$. The odds ratio for stratum i is given by $\Psi_i = (p_{1i}q_{2i})/(q_{1i}p_{2i})$ and the relative risk is given by $\phi_i = p_{1i}/p_{2i}$. Let $\hat{p}_{ji} = x_{ji}/n_{ji}$ be the observed rate in stratum i for sex category j. Then the maximum likelihood estimators for Ψ_i and ϕ_i are given by $\hat{\Psi}_i = (\hat{p}_{1i}\hat{q}_{2i})/(\hat{q}_{1i}\hat{p}_{2i})$ and $\hat{\phi}_i = \hat{p}_{1i}/\hat{p}_{2i}$. The problem is to obtain a summary estimator of relative risk by combining the age specific risk ratios, $\hat{\phi}_i$.

TABLE 1. SUMMARY OF DISEASE DATA FROM ONE AGE STRATUM

	Male	Female	Total
Cases	x_{1i}	x_{2i}	$x_{\cdot i}$
Non cases	$n_{1i} - x_{1i}$	$n_{2i} - x_{2i}$	$n_{\cdot i} - x_{\cdot i}$
Total	n_{1i}	n_{2i}	$n_{\cdot i}$

It will be useful to consider the problem of combining odds ratios before proceeding to a discussion of summary relative risk estimates. One method of combining several odds

ratios is to form a weighted average, with each odds ratio weighted by the reciprocal of its asymptotic variance [3], that is by

$$\omega_i = \Psi_i^{-2}\left[\frac{1}{n_{1i}p_{1i}q_{1i}} + \frac{1}{n_{2i}p_{2i}q_{2i}}\right]^{-1}$$

If $\Psi_i = 1$, so that $p_{1i} = p_{2i} = p_i$, these weights reduce to

$$w_i = \frac{n_{1i}n_{2i}}{n_{\cdot i}} p_i q_i.$$

The Mantel–Haenszel estimate of summary odds ratio is obtained if w_i is estimated by $\hat{w}_i = n_{1i}n_{2i}\hat{p}_{2i}\hat{q}_{1i}/n_{\cdot i}$ [4]. While it would be possible to estimate p_iq_i in w_i by $\hat{p}_i\hat{q}_i$, where $\hat{p}_i = x_{\cdot i}/n_{\cdot i}$ and $\hat{q}_i = 1 - \hat{p}_i$, the Mantel–Haenszel weights have the advantage of guaranteeing that the summary odds ratio will be finite even if some of the individual odds ratios $\hat{\Psi}_i$ are infinite because of one or more zero entries in the 2×2 tables.

The above heuristic derivation of the Mantel–Haenszel estimator utilized the assumption that all of the individual odds ratios were equal to 1. It can be shown that the Mantel–Haenszel estimator is efficient only if $\Psi_i = 1$ for all strata [4]. The Mantel–Haenszel estimator usually performs quite well, however, when the individual odds ratios are not equal to 1. One explanation for this fact is that the Mantel–Haenszel estimator can be obtained as the first iteration in the calculation of the maximum likelihood estimator of a common odds ratio (Clayton cited by Armitage [5]) using the method of scoring [6]. The likelihood function and the information matrix required for the calculation of the maximum likelihood estimator are given by Gart [3] and the initial estimates which lead to the Mantel–Haenszel estimator on the first iteration are $\Psi^0 = 1$ and $p_{1i}^0 = p_{2i}^0 = \frac{1}{2} - (\frac{1}{4} - \hat{p}_{2i}\hat{q}_{1i})^{1/2}$. The initial values of p_{1i}^0 and p_{2i}^0 will not always be well defined for proportions near $\frac{1}{2}$, however they will be well defined for most epidemiological data, for which the probabilities are usually substantially less than $\frac{1}{2}$.

A summary estimator of relative risk can be derived by analogy to the above heuristic derivation of the Mantel–Haenszel summary odds ratio. The asymptotic variance of the age-specific relative risk $\hat{\phi}_i$ is equal to

$$v_i = \phi_i^2\left(\frac{q_{1i}}{n_{1i}p_{1i}} + \frac{q_{2i}}{n_{2i}p_{2i}}\right).$$

Thus if $\phi_i = 1$, so that $p_{1i} = p_{2i} = p_i$, we can weight each individual relative risk by the reciprocal of its asymptotic variance, that is by

$$W_i = \frac{n_{1i}n_{2i}}{n_{\cdot i}}\frac{p_i}{q_i}.$$

If we follow the analogy to the derivation of the Mantel–Haenszel summary odds ratio and estimate p_i/q_i in W_i by $\hat{p}_{2i}/\hat{q}_{1i}$, the resulting summary risk ratio calculated for males relative to females would not be equal to the reciprocal of the summary risk ratio calculated for females relative to males. If we, however, estimate p_i/q_i in W_i by $\hat{p}_{2i}/\hat{q}_i$, we obtain the combined estimator

$$\hat{\Phi}_1 = \frac{\Sigma W_i\hat{\phi}_i}{\Sigma W_i} = \left(\Sigma\frac{n_{2i}}{n_{\cdot i}\hat{q}_i}x_{1i}\right)\Big/\left(\Sigma\frac{n_{1i}}{n_{\cdot i}\hat{q}_i}x_{2i}\right).$$

Like the Mantel–Haenszel summary odds ratio, this estimator will be finite even if some of the x_{2i} are equal to zero, and the summary risk ratio for males relative to females will be the reciprocal of the risk ratio for females relative to males. The asymptotic variance of $\hat{\Phi}_1$ is equal to

$$V_1 = \Sigma W_i^2\phi_i^2\left(\frac{q_{1i}}{n_{1i}p_{1i}} + \frac{q_{2i}}{n_{2i}p_{2i}}\right)\Big/(\Sigma W_i)^2$$

Note that the variance estimator obtained by substituting the estimates $\hat{p}_{1i}$ and $\hat{p}_{2i}$ into the expression for V_1 will be infinite if $x_{2i} = 0$ for some i. Thus inference based on the estimator $\hat{\Phi}_1$ may prove difficult in studies with very low incidence or mortality rates.

If the probabilities p_{1i} and p_{2i} are small we find that

$$\hat{\Phi}_1 \cong \left(\Sigma \frac{n_{2i}}{n_{\cdot i}} x_{1i}\right) \Big/ \left(\Sigma \frac{n_{1i}}{n_{\cdot i}} x_{2i}\right) = \hat{\Phi}_2.$$

This latter form of the summary relative risk estimator, which is an analogue of the Mantel–Haenszel estimator for odds ratios, has been proposed by Rothman and Boice [7]. They state without documentation that this estimator provides a good approximation to the maximum likelihood estimator. In fact, $\hat{\Phi}_2$ can be obtained as the first iteration in the calculation of the maximum likelihood estimator for the common risk ratio assuming a Poisson distribution. The likelihood function can be obtained by assuming that x_{2i} is distributed as a Poisson variable with mean $n_{2i}p_{2i}$ and x_{1i} is distributed as a Poisson variable with mean $n_{1i}\phi p_{2i}$. The initial estimates which lead to $\hat{\Phi}_2$ as the first iteration in the calculation of the maximum likelihood estimator of ϕ are $\phi^0 = 1$ and $p_{1i}^0 = p_{2i}^0 = \hat{p}_{2i}$. The asymptotic variance of $\hat{\Phi}_2$ can be estimated by

$$\hat{V}_2 = [\Sigma x_{1i}(n_{2i}/n_{\cdot i})^2 + \hat{\Phi}_2^2 \Sigma x_{2i}(n_{1i}/n_{\cdot i})^2]/(\Sigma x_{2i} n_{1i}/n_{\cdot i})^2.$$

This variance estimator, like $\hat{\Phi}_2$, will be well defined provided at least one x_{2i} is not equal to zero. In addition, the summary risk ratio of males relative to females will be equal to the reciprocal of the summary risk ratio of females relative to males when using the estimator $\hat{\Phi}_2$.

A noniterative test for heterogeneity of the stratum-specific relative risks can be obtained using the summary relative risk estimator $\hat{\Phi}_2$. Letting $\bar{p}_{1i} = x_{\cdot i}\hat{\Phi}_2/(n_{1i}\hat{\Phi}_2 + n_{2i})$ and $\bar{p}_{2i} = x_{\cdot i}/(n_{1i}\hat{\Phi}_2 + n_{2i})$, define the statistic

$$X_F^2 = \sum_{j=1}^{2} \sum_{i=1}^{m} (x_{ji} - n_{ji}\bar{p}_{ji})^2/(n_{ji}\bar{p}_{ji}).$$

If all of the stratum-specific relative risks are equal, then the statistic X_F^2 will be distributed asymptotically as a chi-squared random variable with $m - 1$ degrees of freedom.

EXAMPLES AND DISCUSSION

It is informative to examine the weighted combination obtained by estimating p_i in W_i by $\hat{p}_i = x_{\cdot i}/n_{\cdot i}$. The weight for $\hat{\phi}_i$ would then be

$$\hat{W}_i = \frac{n_{1i}n_{2i}}{n_{\cdot i}} \frac{\hat{p}_i}{\hat{q}_i}.$$

If $n_{1i} = n_{2i} = n_{\cdot i}/2$ and q_i is approximately equal to 1, it follows that

$$\hat{W}_i \cong \frac{n_{\cdot i}}{4} \hat{p}_i = \frac{x_{\cdot i}}{4}.$$

That is, each relative risk would be weighted by the total number of cases observed in the appropriate age stratum. In arguing against a related weighting scheme, Lilienfeld and Pyne state that "the number of deaths (expected or observed) has little relevance to the epidemiologist, whose principle concern is with death rates and populations." [1] Lilienfeld and Pyne choose to weight each age-specific relative risk by $n_{1i}n_{2i}/n_{\cdot i}$. If $n_{1i} = n_{2i} = n_{\cdot i}/2$ these weights reduce to $n_{\cdot i}/4$. That is, Lilienfeld and Pyne would weight each relative risk by the total population size in the appropriate age stratum.

The major argument given by Lilienfeld and Pyne to justify their choice of the weights $n_{1i}n_{2i}/n_{\cdot i}$ is based on their assertion that these weights were given by Cochran [8] and by Mantel and Haenszel [2] for estimating different summary parameters associated with

the combination of 2 × 2 tables. Although Lilienfeld and Pyne state that "Cochran estimated differences in risk", the fact is that Cochran derived an overall significance test and not a summary estimator. In the appendix of his excellent paper [8], Cochran stated that if differences are constant on the logit scale the test statistic obtained using the weights $n_{1i}n_{2i}/n_{\cdot i}$ "should be close to optimum in power." Radhakrishna showed that these weights *are optimal* for constant differences on the logit scale [9]. Since constant differences on the logit scale are equivalent to the assumption of constant odds ratios, the optimality of the weights $n_{1i}n_{2i}/n_{\cdot i}$ given by Cochran follows from a consideration of odds ratios. Furthermore, the weights used by Mantel and Haenszel for combining odds ratios were $n_{1i}n_{2i}\hat{p}_{2i}\hat{q}_{1i}/n_{\cdot i}$, not simply $n_{1i}n_{2i}/n_{\cdot i}$. Thus there is no reason to expect the Lilienfeld and Pyne weights to be appropriate for estimating any overall parameter based upon the combination of 2 × 2 tables.

A few examples will illustrate important differences between the estimator obtained using the Lilienfeld and Pyne weights, which depend only on the population structure, and the estimator $\hat{\Phi}_2$, which utilizes weights which depend upon the number of events in each stratum as well as the population structure. Consider the contrived example represented by the data in Table 2. The data for males and females in Table 2 represent a population which has been divided into two age strata. The cases could represent incidence or mortality associated with some disease. Gart has presented techniques for the analysis of ratios of rates using maximum likelihood methods based on the Poisson distribution [10]. Using Gart's methods, we find the overall maximum likelihood estimate of the risk ratio for males relative to females to be 1.01. A test of the validity of the assumption of constant risk ratios in both strata is not significant ($p = 0.32$), and a test of the null hypothesis that the underlying population relative risk is equal to 1 is not significant ($p > 0.8$). Thus, the observed rates are consistent with the hypothesis of equal risk for males and females in both strata.

TABLE 2. NUMBER OF CASES AND POPULATION SIZES FOR TWO AGE STRATA

	Males		Females		
Age group	Cases	Population size	Cases	Population size	Relative risk
1	3	2,000,000	1	2,000,000	3.0
2	400	2,000,000	400	2,000,000	1.0

In populations for which the number of males at risk in each stratum is equal to the number of females at risk, we have seen that the Lilienfeld and Pyne estimate of summary relative risk is obtained by weighting each age-specific relative risk by the corresponding stratum population size. Thus, the Lilienfeld and Pyne estimate for the data in Table 2 gives a value of 2.0, which is almost twice as large as the maximum likelihood estimate of the underlying risk ratio. Because the strata are of equal size, the relative risk of 3.0 from stratum 1 is given the same weight in the Lilienfeld and Pyne estimate as the relative risk of 1.0 from stratum 2, in spite of the fact that the observed rates are consistent with an underlying relative risk of 1.0 in stratum 1 as well as in stratum 2. The use of the estimator $\hat{\Phi}_2$ gives a summary relative risk estimate of 1.00, a value virtually identical to the maximum likelihood estimate.

A slight variation of the above example will demonstrate that if the strata have different population sizes, then the difference between the estimates can be even more striking.

TABLE 3. NUMBER OF CASES AND POPULATION SIZES FOR TWO AGE STRATA

	Males		Females		
Age group	Cases	Population size	Cases	Population size	Relative risk
1	3	2,000,000	1	2,000,000	3
2	400	100,000	400	100,000	1

TABLE 4. NUMBER OF CASES AND POPULATION SIZES FOR TWO AGE STRATA

	Males		Females		
Age group	Cases	Population size	Cases	Population size	Relative risk
1	3	2,000,000	1	2,000,000	3.0
2	200	2,000,000	400	2,000,000	0.5

Table 3 is identical to Table 2 with the exception that stratum 2 now has a population size of 200,000. The methods of Gart [10] again show that the data are consistent with an equal risk for males and females in both strata. The Lilienfeld and Pyne estimate of 2.90 for this data is almost three times as large as the maximum likelihood estimate of 1.01. The estimator $\hat{\Phi}_2$ again gives a summary relative risk estimate of 1.00.

Lilienfeld and Pyne apply their method to data on mortality due to diabetes, for which some of the age-specific relative risks are significantly less than one and others are significantly greater than one. No single relative risk adequately summarizes such heterogeneous data; however, a third example will illustrate how the different weighting systems can affect the estimates of relative risk for nonhomogeneous data. Table 4 is identical to Table 2 with the exception that only 200 cases are observed for males in age group 2. The Lilienfeld and Pyne estimate of summary risk ratio for males relative to females for the data in Table 4 is 1.75. This summary relative risk is greater than 1, in spite of the fact that the male population had approximately 1/2 the number of cases observed in the female population of the same size. The estimator $\hat{\Phi}_2$ gives a summary relative risk estimate of 0.51. For data such as those in Table 4 and in the diabetes example discussed by Lilienfeld and Pyne, it is recommended that no attempt be made to summarize the stratum-specific risk ratios with a single summary relative risk estimate.

Consider now a comparison of non-melanoma skin cancer incidence rates for women in Dallas–Fort Worth and in Minneapolis–St. Paul. The numbers of cases and the sizes of the populations at risk are presented in Table 5, stratified by age [11].

TABLE 5. NUMBER OF CASES OF NON-MELANOMA SKIN CANCER AND POPULATION SIZE FOR WOMEN IN DALLAS–FORT WORTH AND MINNEAPOLIS–ST. PAUL

	Dallas–Fort Worth		Minneapolis–St. Paul		
Age	Cases	Population size	Cases	Population size	Relative risk
15–24	4	181,343	1	172,675	3.81
25–34	38	146,207	16	123,065	2.00
35–44	119	121,374	30	96,216	3.14
45–54	221	111,353	71	92,051	2.57
55–64	259	83,004	102	72,159	2.21
65–74	310	55,932	130	54,722	2.33
75–84	226	29,007	133	32,195	1.89
85+	65	7,538	40	8,328	1.80

These data have been analyzed by Gart using maximum likelihood methods based on the Poisson model [12]. A test of the hypothesis of no differences in incidence rates between the two cities using the methods of Armitage [13] indicates a highly significant elevated risk for women in Dallas–Fort Worth ($p < 10^{-4}$). Using the methods of Gart [10], the maximum likelihood estimate of summary relative risk is found to be 2.24. The associated test of the assumption of constant risk ratios across age strata gives a chi-squared statistic of 8.05 with 7 degrees of freedom, indicating no significant heterogeneity of age-specific risk ratios.

Using the methods described above for small p_i, we find that $\hat{\Phi}_2 = 2.24$ with an estimated standard error of 0.12. Testing for heterogeneity of stratum risks, we find

$X_F^2 = 8.22$, indicating an adequate fit of the homogeneous risk model. Thus the results of the noniterative methods based on $\hat{\Phi}_2$ agree closely with the results of the iterative maximum likelihood methods. By way of comparison, the Lilienfeld and Pyne summary estimate of relative risk is 2.77. One reason for the higher value of the Lilienfeld and Pyne estimator is that it gives much greater relative weight than $\hat{\Phi}_2$ or the maximum likelihood estimator do to the risk ratio of 3.81 in the first age stratum.

In conclusion, the estimators $\hat{\Phi}_1$ and $\hat{\Phi}_2$ derived in this paper provide simple summary estimators of relative risk for data collected in a prospective study. If all stratum-specific rates are small, the estimator $\hat{\Phi}_2$ will closely approximate the maximum likelihood estimator of a common risk ratio. The variance estimator for $\hat{\Phi}_2$ will provide well defined estimates except in the degenerate situation in which no cases are observed in one of the two populations being compared. In addition, we have seen that the recently proposed Relative Risk Index can give an estimate which is substantially different from the maximum likelihood estimate, even in cases for which the data are consistent with a constant risk ratio across strata.

Acknowledgement—The author wishes to thank Virginia Grubar for preparing the manuscript.

REFERENCES

1. Lilienfeld DE, Pyne DA: On indices of mortality: deficiencies, validity, and alternatives. **J. Chron. Dis.** 32: 463–468, 1979
2. Mantel N, Haenszel W: Statistical aspects of the analysis of data from retrospective studies of disease. **J. Natl. Can. Inst.** 22: 719–748, 1959
3. Gart JJ: On the combination of relative risks. **Biometrics** 18: 601–610, 1962
4. Hauck WW: The large sample variance of the Mantel–Haenszel estimator of a common odds ratio. **Biometrics** 35: 817–819, 1979
5. Armitage P: The use of the cross-ratio in aetiological surveys. In: **Perspectives in Probability and Statistics.** Gani J (Ed.). New York: Academic Press, 1975, pp. 349–355.
6. Rao CR: **Linear Statistical Inference and Its Applications.** New York: Wiley, 1973
7. Rothman KJ, Boice JD: **Epidemiologic Analysis with a Programmable Calculator.** Washington, DC: NIH Publication No. 79–1649, 1979
8. Cochran WG: Some methods for strengthening the common χ^2 tests. **Biometrics** 10: 417–451, 1954
9. Radhakrishna S: Combination of results from several 2 × 2 contingency tables. **Biometrics** 21: 86–98, 1965
10. Gart JJ: The analysis of ratios and cross-product ratios of Poisson variates with applications to incidence rates. **Commun. Statist.-Theor. Meth.** A7(10): 917–937, 1978
11. Scotto J, Kopf AW, Urbach F: Non-melanoma skin cancer among Caucasians in four areas of the United States. **Cancer** 34: 1333–1338, 1974
12. Gart JJ: Statistical analyses of relative risk. **Environ. Hlth Perspect.** 32: 157–167, 1979
13. Armitage P: The chi-square test for heterogeneity of proportions, after adjustment for stratification. **J. R. Statist. Soc. B** 28: 150–163, 1966