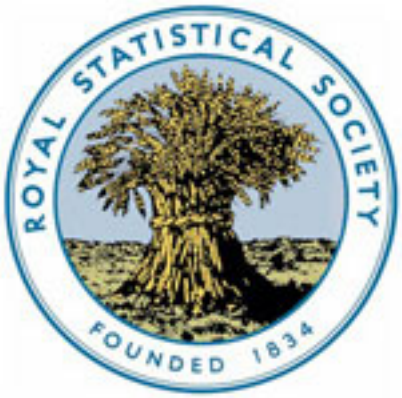

Appl. Statist. (1985)
34, *No.* 1, *pp.* 42–48

Small Sample Properties of some Estimators of a Common Hazard Ratio

By ALEXANDER M. WALKER

International Agency for Research on Cancer, France

[Received October 1983. Revised April 1984]

SUMMARY

The simulated small-sample behaviours of several estimators of a common hazard ratio are compared. In most circumstances, a modified version of the empirical logit (Haldane 1955; Anscombe 1956) seems to be the analytic technique of choice. The performance of the Standardized Mortality Ratio (*SMR*) is indistinguishable from that of the *ML* estimator when the denominator experience is much larger than that of the numerator experience. When many cells have small expectations, then the empirical logit becomes biased, and a variant of the Mantel–Haenszel estimator (Mantel and Haenszel, 1959) is the most widely reliable non-iterative estimator.

Keywords: Hazard ratio; Small samples

1. Introduction

Accumulated times at risk (person-years of experience) in cohort studies are commonly assigned to strata within which a reasonable approximation is to assume a uniform hazard (or incidence) function. The stratifying variables may be fixed (such as gender) or time dependent (such as age); they determine the baseline stratum-specific hazards and thus define the context within which the effect of any further characteristics (such as an industrial exposure) can be examined. Often the ratio of hazard rates in the presence *vs.* absence of the additional characteristic under study is taken to be fixed, conditionally on stratifying variables, and the analytic goal is to estimate the common hazard ratio over strata. Maximum likelihood estimates of common hazard ratios can be readily obtained with mathematical modelling programmes that take into account the Poisson structure of the likelihood function (Clayton, 1982; Berry, 1983) but these may be unavailable, or cumbersome when the object is simply to analyse stratified data from simple tables. A review of current journals of epidemiology and industrial medicine, where cohort studies appear most often, suggests that non-iterative estimates are used far more commonly in practice than are maximum likelihood procedures, and that there are many studies in which at least some cells contain few or no observed events. The purpose of this note then is to compare the theoretical and observed variances of several simple, consistent estimators of the common hazard ratio, with the intent of identifying situations in which such estimators are most useful, and of signalling some data configurations in which they might provide misleading results.

2. Estimators

For each stratum $i = 1, 2, \ldots, k$ let there be a_i events in a total observation time C_i (including both censored and non-censored individuals) and b_i events in a total observation time D_i. The two observed incidence or hazard rates are $I_{1i} = a_i/C_i$ and $I_{2i} = b_i/D_i$. Let λ_{1i} and λ_{2i} be the

Present address: Unit of Analytical Epidemiology, International Agency for Research on Cancer, 150 cours Albert Thomas, 69372 Lyon Cedex 08, France.

 0035–9254/85/34042 $2.00

corresponding parameters with $\lambda_{1i}/\lambda_{2i} = \psi$ for all i, and let $\phi = \ln(\psi)$. The stratum-specific estimates of ψ are $\hat{\psi}_i = I_{1i}/I_{2i}$. Although C_i and D_i are random variables, the a_i and b_i can be analysed as if they were independently Poisson with parameters $\alpha_i = \lambda_{1i}C_i$ and $\beta_i = \lambda_{2i}D_i$ for two reasons: the first two moments of the distributions of I_{1i} and I_{2i} are identical to those that would be obtained if indeed C_i and D_i were fixed, and the kernel of the likelihood for the count and person-time observations is identical to that for the count alone with fixed person-time (Breslow, 1977, 1978, Clayton, 1982; Berry, 1983).

The most studied non-iterative estimator which could be applied to such data is the empirical logit, coupled with an offset to account for varying levels of C_i and D_i over strata. The empirical logit for stratum i, modified so as to avoid undefined values (Haldane, 1955; Anscombe, 1956; Cox, 1970), is

$$\hat{\phi}_i = \ln\left[(a_i + 0.5)/(b_i + 0.5)\right] - \ln(C_i/D_i),$$

whose variance (Gart and Zweifel, 1967) is estimated by

$$\hat{V}_{\phi_i} = (n_i + 1)(n_i + 2)/\left[n_i(a_i + 1)(b_i + 1)\right],$$

where $n_i = a_i + b_i$.

An estimator of the common hazard ratio across strata is then

$$\hat{\psi}_{LGT} = \exp(\hat{\phi}_{LGT}) = \exp\left[(\Sigma \hat{V}_{\phi_i}^{-1}\hat{\phi}_i)/\Sigma \hat{V}_{\phi_i}^{-1}\right],$$

with summations (here and below) taken over $i = 1$ to k. An estimate of the variance of the empirical logit is

$$\hat{V}_{\phi_{LGT}} = (\Sigma \hat{V}_{\phi_i}^{-1})^{-1}. \tag{1}$$

Use of the empirical logit is infrequent in contemporary epidemiologic analysis.

Commonly used estimators, and two new ones, can be derived as information-weighted averages of the stratum specific hazard ratios. Nurminen (1981) has shown that a common ratio estimate obtained as a weighted average of stratum specific estimates is asymptotically efficient, provided that the weights are asymptotically proportional to the inverse of the variance of the stratum specific estimates. The variance of the weighted average approaches the Cramér-Rao lower bound, given by the inverse of the sum of stratum-specific Fisher informations. Estimators derived by this device have the attractive property of yielding $\hat{\psi} = \hat{\psi}_i$ when the $\hat{\psi}_i$ are identical across strata.

By a first order Taylor series expansion, it can be shown that the variance of the stratum-specific estimate $\hat{\psi}_i$ is approximately equal to

$$\hat{V}_{\psi_i} = \psi^2(\alpha_i^{-1} + \beta_i^{-1}). \tag{2}$$

After substitution of observed counts for the parameters in this expression, an average of the $\hat{\psi}_i$, weighted according to the inverse of the variances, simplifies to

$$\hat{\psi}_{IV} = \frac{\Sigma a_i^2 D_i/(C_i n_i)}{\Sigma a_i b_i/n_i}.$$

The variance of the natural logarithm of $\hat{\psi}_{IV}$, found again by a first order Taylor series expansion, is

$$\hat{V}_{\phi_{IV}} = \frac{\Sigma a_i^3(a_i + 4b_i)D_i^2/(n_i^3 C_i^2)}{\left[\Sigma a_i^2 D_i/(C_i n_i)\right]^2} + \frac{\Sigma a_i b_i (a_i^3 + b_i^3)/n_i^4}{(\Sigma a_i b_i/n_i)^2} + \frac{\Sigma 2a_i^2 b_i(a_i - 2b_i)D_i/(n_i^3 C_i)}{(\Sigma a_i^2 D_i/C_i n_i)(\Sigma a_i b_i/n_i)}. \tag{3}$$

The numeric value of this cumbersome expression is in most instances very close to $\left[\Sigma(a_i^{-1} + b_i^{-1})^{-1}\right]^{-1}$.

The stratum-specific variance of equation (2) can be reparameterized to

$$\hat{V}_{\psi_i} = \psi^2(\psi^{-1}D_i + C_i)/(\beta_i C_i). \tag{4}$$

When (4) is used as a basis for deriving weights, the leading term (ψ^2) cancels out. Substituting b_i for β_i and 1 for the ψ within the parentheses, one obtains a simple estimate of ψ first proposed by Rothman and Boice (1979) (RB hereafter), as an analogue of the Mantel-Haenszel (Mantel and Haenszel, 1959) estimate of the common odds ratio over a series of 2 x 2 tables:

$$\hat{\psi}_{\mathrm{RB}} = \frac{\Sigma a_i D_i / T_i}{\Sigma b_i C_i / T_i},$$

where $T_i = C_i + D_i$. The variance of the natural logarithm of the expression is approximately

$$\hat{V}_{\phi_{\mathrm{RB}}} = \frac{\Sigma a_i (D_i/T_i)^2}{(\Sigma a_i D_i/T_i)^2} + \frac{\Sigma b_i (C_i/T_i)^2}{(\Sigma b_i C_i/T_i)^2}. \tag{5}$$

When the D_i are all much larger than the C_i, then expression (4) becomes

$$\hat{V}_{\psi_i} = \psi D_i (\beta_i C_i)^{-1}. \tag{6}$$

Substituting, as above, b_i for β_i and weighting the ψ_i according to the inverse of (6), one obtains the standardized mortality ratio

$$\hat{\psi}_{SMR} = \frac{\Sigma a_i}{\Sigma b_i / C_i D_i},$$

the variance of the logarithm of which is

$$\hat{V}_{\phi_{SMR}} = (\Sigma a_i)^{-1} + \frac{\Sigma b_i (C_i/D_i)^2}{(\Sigma b_i C_i/D_i)^2}. \tag{7}$$

Any consistent estimate $\tilde{\psi}$ of ψ substituted into equation (4) will lead with inverse-variance weighting to an asymptotically efficient "two-step" estimator,

$$\hat{\psi}_{TS} = \frac{\Sigma a_i D_i / (\tilde{\psi}^{-1} D_i + C_i)}{\Sigma b_i C_i / (\tilde{\psi}^{-1} D_i + C_i)}. \tag{8}$$

Generally $\tilde{\psi}$ is taken as $\hat{\psi}_{\mathrm{RB}}$. The efficiency of $\hat{\psi}_{TS}$ will be superior to that of $\hat{\psi}_{\mathrm{RB}}$ as a consequence of using a more accurate estimate of ψ in the weighting function.

Clayton (1982) has pointed out that recursive solution of equation (8) leads to the maximum likelihood estimate, $\hat{\psi}_{ML}$. It follows immediately that when C_i/D_i is a constant, H, across strata, then the pooled estimate

$$\hat{\psi}_p = (\Sigma a_i)(H \Sigma b_i)^{-1}$$

is the maximum-likelihood estimate of ψ. The pooled estimate is also, for constant H, equal to $\hat{\psi}_{\mathrm{RB}}$, $\hat{\psi}_{TS}$, and $\hat{\psi}_{SMR}$. Because it does not allow discrimination among the various estimators, constant H will be considered a degenerate case and avoided in the comparisons of the following section.

3. Simulated Behaviour

In order to test the performance of the measures presented in Section 2, a series of simulations have been carried out. Each analysis consisted of 10 000 random replications of a two-stratum, two-sample study comparing the incidence rates shown in Table 1. The complete series consisted of 10 distinct distributions of "large" and "small" expected numbers over the four cells, large for this purpose being taken as 30 and small as 3, except where adjustment was necessary to avoid a constant C_i/D_i across strata. For each study, a set of expected counts was chosen for a_1, b_1, a_2, and b_2; person times at risk (C_1, D_1, C_2, D_2) were chosen so as to produce the expected incidence rates of Table 1. Each random replication was generated using the expected counts as the parameters of four separate Poisson distributions. All uninformative strata ($a_i + b_i = 0$) were re-

TABLE 1

Incidence rates underlying the randomizations in Table 2, expressed in events per 1000 person-years

Stratum	λ_{1i}	λ_{2i}	ψ_i
1	0.3	0.1	3.0
2	1.2	0.4	3.0

$\psi = 3$ $\quad \ln(\psi) = 1.099$

randomized. Replications for which $\Sigma a_i b_i = 0$ were replaced with new ones; replacement was necessary by this criterion for from 0 to 1 per cent of all samples, depending on the study, and is therefore unlikely to have affected the simulations in any important way. The random counts were combined with the fixed times at risk and analysed by each of the techniques of Section 2.

Table 2 presents the randomization parameters chosen, the Cramér-Rao lower bound to the variance of $\hat{\phi}$ (i.e. $\ln(\hat{\psi})$), calculated from the parameters, and (for each technique) the mean value of $\hat{\phi}$, the expected variance (as derived from substituting parameter values into equations (1), (3), (5) and (7)), the observed variance of $\hat{\phi}$, the observed coefficient of skewness (the third moment about the mean divided by 1.5th power of the variance), and the mean square error (observed variance plus the square of the observed bias). Of the four non-recursive estimators on the log scale, $\hat{\phi}_{LGT}$ appears to be the least biased over the range of circumstances examined; it has furthermore small observed variances, and uniformly the lowest mean square error. Of some interest is that the variance of $\hat{\phi}_{LGT}$ is frequently less than the asymptotic estimate provided by the Cramér-Rao lower bound, most notably when small cells predominate. $\hat{\phi}_{\mathrm{RB}}$ is only slightly more biased on the average, and performs better than $\hat{\phi}_{LGT}$ in some circumstances. Over the range of simulations, the direction and magnitude of the observed bias in $\hat{\phi}_{\mathrm{RB}}$ correlates very closely with its coefficient of skewness, suggesting that the variations in the average observed estimates are closely tied to efficacy of the logarithmic transformation in rendering the distribution of $\hat{\phi}_{\mathrm{RB}}$ symmetrical. $\hat{\phi}_{IV}$ has a positive bias which becomes non-negligible as the small cells predominate. $\hat{\phi}_{SMR}$ has largest variance in most circumstances, and by and large a more skewed sampling distribution than the other log estimates, although in the important circumstances of large b_i and D_i, its performance is, as expected, indistinguishable from the *MLE*.

The recursive estimators, the $\hat{\phi}_{TS}$ and the *MLE*, are closely similar with observed variances close to the lower bound, and mean values which tend slightly closer to the log parameter than $\hat{\phi}_{\mathrm{RB}}$. As with $\hat{\phi}_{\mathrm{RB}}$, deviations in the mean estimate appear to be principally a function of skewness in the log distributions.

Departures from log symmetry follow a regular pattern for all the estimators. Within the range of examples studied, small counts in a_i coupled with large corresponding b_i result in distributions skewed negatively on the log scale, and large a_i together with small corresponding b_i give distributions skewed positively. In mixed data sets the effect of a positive skewness seems to predominate, and in data sets with a_i and b_i similar for all i, the observed sampling distributions are very nearly symmetric.

When counts from cells with small expected values play an important role in the estimates, the distribution of the latter can become not only skew but also polymodal, particularly in the longer tail, as the discreteness of the small cell distributions becomes dominant. As an example, in the most commonly encountered estimation problem from Table 2 (all a_i small, all b_i large), two secondary peaks containing over 3.0 per cent of the probability mass for $\hat{\phi}_{\mathrm{RB}}$, $\hat{\phi}_{SMR}$, $\hat{\phi}_{TS}$, and $\hat{\phi}_{ML}$ lie beyond the estimate corresponding to the null hypothesis, despite the fact that $\hat{\psi} = 1$ is $\ln(3)/(0.181)^{\frac{1}{2}} = 2.58$ standard errors below the parameter.

When randomizations were extended to larger numbers of strata and proportionately smaller counts in the small cells, the relative performance of the various estimators remained the same, with the exception of the empirical logit. Although optimal when the expected counts in the small

TABLE 2

Performance of estimators of a common hazard ratio in 10 000 random repetitions

Randomization parameters				*Lower bound*	*Method*	*Observed mean*	*Expected variance*	*Observed variance*	*Coefficient of skewness*	*Mean square error*
α_1	β_1	α_2	β_2							
4	3	3	3	0.311	IV	1.300	0.311	0.413	0.219	1.300
					RB	1.112	0.312	0.408	0.018	0.408
					SMR	1.114	0.313	0.410	0.030	0.410
					LGT	1.082	0.273	0.258	-0.021	0.258
					TS	1.111		0.406	0.006	0.406
					MLE	1.111		0.406	0.006	0.406
30	3	3	3	0.237	IV	1.275	0.237	0.302	0.626	1.275
					RB	1.199	0.249	0.335	0.963	0.345
					SMR	1.266	0.309	0.543	1.577	0.571
					LGT	1.037	0.194	0.196	0.422	0.199
					TS	1.158		0.277	0.676	0.280
					MLE	1.170		0.282	0.668	0.287
3	30	3	3	0.237	IV	1.225	0.237	0.262	-0.235	1.225
					RB	1.042	0.243	0.295	-0.542	0.299
					SMR	1.053	0.258	0.312	-0.440	0.314
					LGT	1.161	0.194	0.193	-0.387	0.197
					TS	1.020		0.274	-0.634	0.280
					MLE	1.028		0.278	-0.637	0.283
40	30	3	3	0.054	IV	1.131	0.054	0.057	0.135	1.131
					RB	1.103	0.054	0.057	0.133	0.057
					SMR	1.103	0.054	0.057	0.131	0.057
					LGT	1.099	0.052	0.053	0.121	0.053
					TS	1.103		0.057	0.134	0.057
					MLE	1.103		0.057	0.133	0.057
40	3	30	3	0.181	IV	1.199	0.181	0.235	0.847	1.199
					RB	1.184	0.181	0.235	0.847	0.242
					SMR	1.186	0.184	0.239	0.859	0.247
					LGT	0.972	0.143	0.155	0.531	0.171
					TS	1.184		0.235	0.847	0.242
					MLE	1.184		0.235	0.847	0.242

TABLE 2 (*continued*)

Randomization parameters				*Lower bound*	*Method*	*Observed mean*	*Expected variance*	*Observed variance*	*Coefficient of skewness*	*Mean square error*
α_1	β_1	α_2	β_2							
30	3	3	30	0 183	IV	1.212	0.183	0.189	0.008	1.212
					RB	1.169	0.214	0.255	0.427	0.260
					SMR	1.270	0.306	0.528	1.297	0.557
					LGT	1.098	0 145	0.156	-0.039	0.156
					TS	1.062		0.167	0.035	0.168
					MLE	1.098		0.174	-0.038	0.174
3	40	3	30	0 181	IV	1.179	0.181	0.202	-0 524	1.179
					RB	1.008	0 181	0.234	-0.843	0.242
					SMR	1 008	0.181	0.234	-0.843	0.242
					LGT	1 221	0.143	0.154	-0.526	0.169
					TS	1 008		0.234	-0.844	0.242
					MLE	1.008		0.234	-0.844	0.242
3	30	30	30	0.056	IV	1 138	0.056	0.061	-0.047	1.138
					RB	1.093	0.057	0.059	-0.074	0.060
					SMR	1.094	0.058	0.060	-0.062	0.060
					LGT	1.129	0.054	0.056	-0.078	0.057
					TS	1.091		0.059	-0.089	0.059
					MLE	1.092		0.059	-0.087	0.059
30	3	30	30	0.056	IV	1 113	0.056	0.056	0.100	1.113
					RB	1.109	0.059	0.061	0.130	0.061
					SMR	1.141	0.108	0.115	0.261	0.117
					LGT	1.066	0.054	0.054	0.089	0.056
					TS	1.102		0.057	0.097	0.057
					MLE	1.104		0.057	0.096	0.057
40	30	30	30	0.031	IV	1.117	0.031	0.032	0.047	1.117
					RB	1.102	0.031	0.032	0.038	0.032
					SMR	1.102	0.031	0.032	0.041	0.032
					LGT	1.098	0.031	0.030	0.042	0.030
					TS	1.102		0.032	0.041	0.032
					MLE	1.102		0.032	0.038	0.032

cells were on the order of three, and the number of strata limited to two, the empirical logit became almost always more biased on the log scale than $\hat{\phi}_{\mathrm{RB}}$, $\hat{\phi}_{SMR}$, or $\hat{\phi}_{TS}$ with 10 strata and expected counts of 0.6 in the smallest cells.

4. Acknowledgements

Drs R. Brookmeyer, N. Day, J. Kaldor and J. Wahrendorf provided advice for this report. Dr J. R. Anderson suggested the two-step estimator.

References

Anscombe, F. J. (1956) On estimating binomial response relations. *Biometrika*, **43**, 461–464.

Berry, G. (1983) The analysis of mortality by the subject-years method. *Biometrics*, **39**, 173–184.

Breslow, N. (1977) Some statistical models useful in the study of occupational mortality. In *Environmental Health: Quantitative Methods*. (A. Whittemore ed.), pp. 88–103. Philadelphia: Society for Industrial and Applied Mathematics.

———(1978) The proportional hazards model: Applications to epidemiology. *Commun. in Statist.*, A7, 315–332.

Clayton, D. G. (1982) The analysis of prospective studies of disease aetiology. *Commun. in Statist.*, **A11**, 2129–2155.

Cox, D. R. (1970) *Analysis of Binary Data*, Section 3.2. London: Chapman and Hall.

Gart, J. J. and Zweifel, J. R. (1967) On the bias of various estimators of the logit and its variance with application to quantal bioassay. *Biometrika*, **54**, 181–187.

Haldane, J. B. S. (1955) The estimation and significance of the logarithm of a ratio of frequencies. *Ann. Human Genetics*, **20**, 309–311.

Mantel, N. and Haenszel, W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease. *J. Nat. Cancer Inst.*, **22**, 719–748.

Nurminen, M. (1981) Asymptotic efficiency of general non-iterative estimators of common relative risk. *Biometrika*, **68**, 525–530.

Rothman, K. J. and Boice, J. D. (1979) *Epidemiologic Analysis with a Programmable Calculator*. Washington D.C. NIH Publication 79-1649.

Appendix

An editor, in addition to offering many helpful suggestions which have been incorporated into the text, has pointed out that the estimators $\hat{\psi}_{\mathrm{RB}}$, $\hat{\psi}_{SMR}$, and $\hat{\psi}_{TS}$ can all be derived from likelihood functions. Let $H_i = C_i/D_i$. Then conditional on n_i, $a_i \sim$ Binomial $(n_i, H_i(1 + H_i)^{-1})$. Setting the first derivative of the conditional likelihood equal to zero, one obtains

$$\Sigma(a_i - b_i\hat{\psi}H_i)\,(1 + \hat{\psi}H_i)^{-1} = 0$$

which suggests the recursive relation

$$\hat{\psi} = \Sigma a_i\,(1 + \psi^* H_i)^{-1}\ [\Sigma b_i H_i (1 + \psi^* H_i)^{-1}]^{-1}.$$

Taking $\psi^* = 0$ gives the *SMR*, $\psi^* = 1$ gives the Rothman-Boice estimator, $\psi^* = \psi_{\mathrm{RB}}$ gives the two step estimator.