Ann. Inst. Statist. Math.
Vol. 44, No. 3, 529–535 (1992)

ON INCONSISTENCY OF THE COMMON RATE DIFFERENCE ESTIMATORS FROM SPARSE FOLLOW-UP DATA

TOSIYA SATO*

*Department of Epidemiology, School of Health Sciences,
University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan*

(Received January 4, 1991; revised October 2, 1991)

Abstract. It has been shown that the unconditional maximum likelihood estimator of the common odds ratio, risk ratio and risk difference parameters are inconsistent in sparse statification. Under a Poisson sparse-data model, the maximum likelihood estimator for the rate difference, which is the difference of the disease incidence rates among the exposed and the unexposed, is also shown to be biased. The sparse-data asymptotic bias of the maximum likelihood estimator is evaluated numerically and compared with that of the weighted least squares estimators.

Key words and phrases: Bias, epidemiologic methods, follow-up studies, Poisson observations, sparse-data.

1. Introduction

In many epidemiologic studies, we often encounter stratified tables that are 'sparse', in which a large number of cell frequencies are small or zero. Breslow (1981) showed the asymptotic bias of the unconditional maximum likelihood estimator of the common odds ratio in his sparse-data large sample theory. Greenland and Robins (1985) studied the estimation of effect parameters from sparse follow-up data and showed the asymptotic bias of the maximum likelihood estimator of the common risk ratio and difference which are ratio and difference between two binomial proportions. They also studied the estimation of ratio and difference between two Poisson rates, the rate ratio and difference, in sparse stratification. While the maximum likelihood estimator of the common rate ratio is consistent in sparse-data, they failed to show any results concerning the consistency of the maximum likelihood estimator of the common rate difference.

From the general theory of the maximum likelihood estimator (Andersen (1970)), it can be expected that the maximum likelihood estimator of the common rate difference is inconsistent in sparse-data, we know little about the sparse-data

* Now at The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan.

performance of it. Through numerical evaluation, this note shows the bias of the maximum likelihood estimator of the common rate difference under a Poisson sparse-data model. The asymptotic bias in the weighted least squares estimators is also calculated and compared to that of the maximum likelihood estimator.

2. Numerical evaluations

Consider a series of K pairs of independent Poisson observations (x_k, y_k) with means $(n_k r_{1k}, m_k r_{0k})$ for $k = 1, \ldots, K$. In follow-up studies of dynamic populations, x_k and y_k denote the number of persons contracting the disease under study out of n_k exposed and m_k unexposed fixed person-time at risk, and r_{1k} and r_{0k} are the incidence rates of the exposed and the unexposed in the k-th stratum. We assume that the rate difference $\delta = r_{1k} - r_{0k}$ is the preferred effect measure and remains constant across strata. The maximum likelihood estimator $\hat{\delta}_{\mathrm{ML}}$ for δ is obtained based on the unconditional distribution of (x_k, y_k):

$$\mathrm{pr}(x, y \mid n_k, m_k) = \frac{[n_k(\delta + r_{0k})]^x}{x!} e^{-n_k(\delta + r_{0k})} \frac{(m_k r_{0k})^y}{y!} e^{-m_k r_{0k}}, \tag{2.1}$$

with the K nuisance parameters r_{0k}. The maximum likelihood estimate is only defined as the iterative solution of a set of $K+1$ equations (Rothman and Boice (1982)).

Three closed-form rate difference estimators are available as an alternative to the maximum likelihood estimator. The weighted least squares approach (Grizzle *et al.* (1969)) yields an estimator $\hat{\delta}_{\mathrm{W}}$ of δ defined by

$$\hat{\delta}_{\mathrm{W}} = \sum_k W_k \hat{\delta}_k \Big/ \sum_k W_k,$$

where $W_k = (x_k/n_k^2 + y_k/m_k^2)^{-1}$ and $\hat{\delta}_k = x_k/n_k - y_k/m_k$. Rothman and Boice (1982) replace W_k with $W_{0k} = n_k m_k / t_k$ which is the inverse of the asymptotic variance of $\hat{\delta}_k$ at $\delta = 0$, to derive a null-weighted least squares estimator $\hat{\delta}_{\mathrm{W}_0}$. Here, $t_k = x_k + y_k$. A constant, usually $1/2$, is added to x_k and y_k, or tables with $t_k = 0$ are thrown out to avoid division by zero. Analogous to the odds ratio and other effect measures, Greenland (1982) replace W_k with the standard weights $S_k = n_k m_k / N_k$ to obtain the Mantel-Haenszel estimator $\hat{\delta}_{\mathrm{MH}}$, where $N_k = n_k + m_k$.

Under large-stratum assumption that all the n_k and m_k tend to infinity, all the above estimators are consistent asymptotically normal. On the other hand, under the sparse-data limiting model introduced by Breslow (1981), the weighted least squares estimators are biased, while the Mantel-Haenszel estimator is still consistent asymptotically normal (Greenland and Robins (1985)). The limiting model considered here is identical with that used by Greenland and Robins. We could assume that (n_k, m_k, r_{0k}) is one of a sequence of K independent identically distributed random vectors with positive components and finite covariance matrix, and K tends to infinity.

Table 1. Asymptotic means of the maximum likelihood, weighted least squares and null-weighted rate difference estimators (per 1000 person-time).

r_0	n[2]	m[2]	$\delta = 0$[1]	$\delta = 5$			$\delta = 10$			$\delta = 20$		
			δ_W	δ_{ML}	δ_W	δ_{W_0}	δ_{ML}	δ_W	δ_{W_0}	δ_{ML}	δ_W	δ_{W_0}
10	25	100	6.80	1.64	10.29	5.27	10.99	13.76	10.01	20.91	20.73	18.44
10	50	100	0.23	4.88	4.53	4.61	10.60	8.64	8.67	20.89	16.56	16.03
10	100	100	0.00	5.50	4.44	4.01	10.65	8.53	7.70	20.71	16.61	15.17
10	50	200	1.92	5.70	4.84	3.91	10.59	7.94	7.70	20.34	14.67	15.17
10	100	200	−0.46	5.56	3.29	3.79	10.58	7.20	7.58	20.39	15.53	15.52
10	200	200	0.00	5.44	4.18	3.88	10.48	8.48	8.02	20.42	17.55	17.02
10	100	400	−0.51	5.19	2.74	3.94	10.14	6.39	8.02	20.07	14.66	16.56
10	200	400	−0.68	5.17	3.46	4.14	10.14	7.90	8.51	20.08	17.29	17.70
10	400	400	0.00	5.11	4.50	4.42	10.13	9.17	9.07	20.10	18.75	18.65
25	20	80	4.79	5.06	7.66	3.97	12.04	10.59	7.86	21.77	16.68	15.49
25	40	80	−1.16	5.76	2.55	3.82	11.33	6.32	7.60	21.42	14.04	15.15
25	80	80	0.00	5.92	4.15	3.82	10.99	8.33	7.71	21.15	16.85	15.83
25	40	160	−1.28	5.55	1.85	3.90	10.54	5.15	7.86	20.40	12.22	15.93
25	80	160	−1.70	5.32	2.33	4.06	10.39	6.51	8.23	20.39	15.23	16.85
25	160	160	0.00	5.16	4.44	4.33	10.25	8.96	8.79	20.31	18.22	17.98
25	80	320	−3.05	5.06	0.94	4.45	10.07	5.14	8.95	20.07	14.02	18.06
25	160	320	−1.28	5.04	3.27	4.57	10.06	7.92	9.19	20.06	17.41	18.57
25	320	320	0.00	5.02	4.72	4.70	10.04	9.48	9.45	20.05	19.11	19.06

[1] When $\delta = 0$, the maximum likelihood and null-weighted estimators are not biased.

[2] Person-time denominators.

In numerical evaluation of the asymptotic means for $\hat{\delta}_{\mathrm{ML}}$, $\hat{\delta}_{\mathrm{W}}$ and $\hat{\delta}_{\mathrm{W}_0}$, the values of n, m and r_0 were assumed to remain constant across strata. These assumptions are made for computational simplicity, so that the asymptotic means may be defined by only four parameters, n, m, δ and r_0. Under these conditions, $\hat{\delta}_{\mathrm{W}}$ converges to a real value δ_{W} which is given by

$$\delta_{\mathrm{W}} = \mathrm{E}(W\hat{\delta})/\mathrm{E}(W),$$

where E denotes the exact expectation under the unconditional distribution (2.1). Similarly, the asymptotic mean δ_{W_0} of $\hat{\delta}_{\mathrm{W}_0}$ is defined. It is necessary that a constant is added to x and y and/or tables with $t = 0$ are thrown out in order for the sparse-data asymptotic means for $\hat{\delta}_{\mathrm{W}}$ and $\hat{\delta}_{\mathrm{W}_0}$ to exist. The asymptotic mean of the maximum likelihood estimator, δ_{ML}, is determined as the solution of the limiting equation

$$\mathrm{E}\left(\frac{x}{\delta_{\mathrm{ML}} + \tilde{r}_0(\delta_{\mathrm{ML}})}\right) = n,$$

where $\tilde{r}_0(\delta)$ is the maximum likelihood estimator of r_0 under a specific δ which is given by

$$\tilde{r}_0(\delta) = [t - N\delta + \sqrt{(t - N\delta)^2 + 4yN\delta}]/(2N).$$

Table 1 gives selected numerical evaluations for four values of δ and two of r_0. For the asymptotic means of $\hat{\delta}_{\mathrm{W}}$ and $\hat{\delta}_{\mathrm{W}_0}$, treatments (1) adding 1/2 to x and y, (2) throwing out tables with $t = 0$ and (3) both (1) and (2) were tried. Results of (3) for δ_{W} and (2) for δ_{W_0}, which are less biased than the other two modifications, are shown in Table 1.

Unless $n = m$, the weighted least squares estimator $\hat{\delta}_{\mathrm{W}}$ is biased even when $\delta = 0$ and its bias tends to be negative except when the smallest cell expectation is smaller than 1. When $\delta > 0$, the bias in $\hat{\delta}_{\mathrm{ML}}$ is generally positive and stable regardless of the values of δ, while the bias in $\hat{\delta}_{\mathrm{W}}$ and $\hat{\delta}_{\mathrm{W}_0}$ is generally conservative and tends to be larger with increase in δ. The asymptotic mean of $\hat{\delta}_{\mathrm{W}}$ is very sensitive to the exposure ratio, n/m. The performance of $\hat{\delta}_{\mathrm{W}_0}$ is usually better than that of $\hat{\delta}_{\mathrm{W}}$.

3. Example

In this section I will illustrate aforementioned common rate difference estimators by using a study of arsenic exposure and respiratory cancer deaths in Montana smelter workers (presented in Breslow and Day (1987), Table 3.14). Table 2 summarizes the data stratified into 13 age and calendar-period categories. Table 3 presents point estimates and 95% confidence intervals for δ. The approximate confidence intervals associated with four estimates are given as follows: for 'Maximum likelihood', the score-based interval was calculated, which is given by the two appropriate solutions to

$$\frac{[\sum_k x_k/(\delta + \tilde{r}_{0k}) - \sum_k n_k]^2}{\sum_k[(\delta + \tilde{r}_{0k})/n_k + \tilde{r}_{0k}/m_k]^{-1}} = 3.84.$$

Table 2. The Montana smelter workers study of arsenic exposure and respiratory cancer (from Breslow and Day (1987), Table 3.14).

		Calendar period							
	Exposure	1938–1949		1950–1959		1960–1969		1970–1977	
Age	level	Deaths	Person-years	Deaths	Person-years	Deaths	Person-years	Deaths	Person-years
40–49	High	0	337.29	0	121.00				
	Low	2	3075.27	0	936.75				
			0.11[1]		0.13				
50–59	High	4	626.72	3	349.53	1	142.33		
	Low	2	2849.76	3	2195.59	3	747.77		
			0.22		0.16		0.19		
60–69	High	9	672.09	7	441.10	3	244.82	1	100.64
	Low	2	2085.43	7	1675.91	10	1501.73	1	440.21
			0.32		0.26		0.16		0.23
70–79	High	1	277.25	2	268.27	1	197.20	2	92.75
	Low	3	833.61	6	973.32	6	1027.12	6	674.44
			0.33		0.28		0.19		0.14

[1] Ratio of person-years between high and low exposure levels.

Table 3. Estimates of the common rate difference and 95% confidence intervals, per 1000 person-years.

	Full tables results		Collapsed tables results	
	δ	95% C.I.	δ	95% C.I.
Maximum likelihood	5.94	(3.26, 9.43)	5.65	(2.99, 9.26)
Mantel-Haenszel	5.90	(3.20, 9.37)	5.82	(3.13, 9.28)
Weighted least squares	3.41	(1.16, 5.66)	5.60	(3.55, 7.66)
Null-weighted least squares	2.92	(1.53, 4.32)	4.93	(3.69, 6.17)

For 'Mantel-Haenszel', the Fieller-like interval (Sato (1990)) which is the two solutions to the quadratic equation

$$\frac{(\sum_k S_k\hat{\delta}_k - \delta\sum_k S_k)^2}{\delta\sum_k P_k + \sum_k Q_k} = 3.84,$$

where $P_k = n_k m_k(m_k - n_k)/N_k^2$ and $Q_k = n_k m_k/N_k^2$. And for 'Weighted least squares' and 'Null-weighted least squares',

$$\hat{\delta}_{\mathrm{W}} \pm 1.96\sqrt{1/\sum_k W_k}$$

(change W to W_0 for the null-weighted least squares method). In the calculation of weighted least squares estimators, the same treatments mentioned in the previous section have done.

The left-half of Table 3 presents the $K = 13$ tables analysis given in Table 2. While the maximum likelihood and the Mantel-Haenszel methods gave close point estimates and confidence intervals, the weighted least squares methods gave smaller point estimates. A referee suggested a modification of the weighted least squares methods that one could collapse together all tables with approximately the same exposure ratio (n_k/m_k) and that calculate $\hat{\delta}_{\mathrm{W}}$ and $\hat{\delta}_{\mathrm{W}_0}$. Seeing exposure ratios in Table 2, I tentatively collapsed tables (1) age (40–49, 50–59) and periods (1938–1949, 1950–1959), (2) age (60–69, 70–79) and periods (1938–1949, 1950–1959) and (3) periods (1960–1969, 1970–1977). The right-half of Table 3 gives the results from the collapsed $K = 3$ data. This modification gives an inconsistent estimate of δ, except when all exposure ratios are exactly the same value in collapsed tables. However, the maximum likelihood and the Mantel-Haenszel estimates changed only a little, and the weighted least squares estimates improved dramatically. While the suggested modification works well for the point estimate, it appears to give narrow confidence intervals in both weighted least squares methods.

4. Discussion

Similar to the results of the other effect measures (Breslow (1981), Greenland and Robins (1985)), the weighted least squares estimators for the common rate difference are biased severely under the Poisson sparse-data model. The maximum likelihood estimator is also biased in sparse-data, however $\hat{\delta}_{\mathrm{ML}}$ performed better than $\hat{\delta}_{\mathrm{W}}$ or $\hat{\delta}_{\mathrm{W}_0}$. The bias in $\hat{\delta}_{\mathrm{ML}}$ is usually less than 0.5 per 1000 person-time denominator under the range of selected parameter values.

The Mantel-Haenszel estimator is the only currently available rate difference estimator which is consistent in sparse-data (Greenland and Robins (1985)) and easily calculable confidence interval method is available. However, $\hat{\delta}_{\mathrm{MH}}$ is sometimes very inefficient in large-strata even when $\delta = 0$. Hence Greenland and Robins suggest that its use might be limited to sparse-data. In large-strata, though the calculation is somewhat complicated, the maximum likelihood estimator associated with the score-based confidence interval may be used.

We need further study to compare mean squared errors between the Mantel-Haenszel and the maximum likelihood estimators, and performances between the Fieller-like and the score-based confidence interval methods under both sparse-data and large-strata settings.

References

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators, *J. Roy. Statist. Soc. Ser. B*, **32**, 283–301.

Breslow, N. E. (1981). Odds ratio estimators when the data are sparse, *Biometrika*, **68**, 73–84.

Breslow, N. E. and Day, N. E. (1987). *Statistical Methods in Cancer Research, Vol. 2—The Design and Analysis of Cohort Studies*, Oxford University Press, New York.

Greenland, S. (1982). Interpretation and estimation of summary ratios under heterogeneity, *Statistics in Medicine*, **1**, 217–227.

Greenland, S. and Robins, J. M. (1985). Estimation of a common effect parameter from sparse follow-up data, *Biometrics*, **41**, 55–68.

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

Rothman, K. J. and Boice, J. D. (1982). *Epidemiologic Analysis with a Programmable Calculator*, Epidemiology Resources Inc., Boston.

Sato, T. (1990). Confidence intervals for effect parameters common in cancer epidemiology, *Environmental Health Perspectives*, **87**, 95–101.