Estimation of a Common Effect Parameter from Sparse Follow-Up Data

Sander Greenland; James M. Robins

Biometrics, Vol. 41, No. 1 (Mar., 1985), 55-68.

Stable URL:
http://links.jstor.org/sici?sici=0006-341X%28198503%2941%3A1%3C55%3AEOACEP%3E2.0.CO%3B2-I

Biometrics is currently published by International Biometric Society.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/ibs.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org/
Mon Mar 28 14:58:26 2005

BIOMETRICS 41, 55–68
March 1985

Estimation of a Common Effect Parameter from Sparse Follow-Up Data

Sander Greenland

Division of Epidemiology, UCLA School of Public Health,
Los Angeles, California 90024, U.S.A.

James M. Robins

Occupational Health Program, Harvard School of Public Health,
Boston, Massachusetts 02115, U.S.A.

SUMMARY

Breslow (1981, *Biometrika* **68,** 73–84) has shown that the Mantel–Haenszel odds ratio is a consistent estimator of a common odds ratio in sparse stratifications. For cohort studies, however, estimation of a common risk ratio or risk difference can be of greater interest. Under a binomial sparse-data model, the Mantel–Haenszel risk ratio and risk difference estimators are consistent in sparse stratifications, while the maximum likelihood and weighted least squares estimators are biased. Under Poisson sparse-data models, the Mantel–Haenszel and maximum likelihood rate ratio estimators have equal asymptotic variances under the null hypothesis and are consistent, while the weighted least squares estimators are again biased; similarly, of the common rate difference estimators the weighted least squares estimators are biased, while the estimator employing "Mantel–Haenszel" weights is consistent in sparse data. Variance estimators that are consistent in both sparse data and large strata can be derived for all the Mantel–Haenszel estimators.

1. Introduction

It is common in epidemiology to encounter cross-classifications that are "sparse," in that a large number of entries in the classification are small or 0. Breslow (1981) developed a large-sample theory to study odds-ratio estimation in sparse data, and demonstrated the consistency of the Mantel–Haenszel odds ratio within that theory. Nevertheless, while the odds ratio is the parameter of interest in case–control studies, it has come to enjoy this status primarily because it approximates the risk ratio under most case–control study circumstances. In typical cohort studies this approximation may no longer hold, and the parameter of interest becomes the risk ratio itself. We examine the performance of estimators of a common risk ratio under the sparse-data theory developed by Breslow. We also present parallel results for estimators of a common rate ratio and estimators of a common risk difference.

Published follow-up data rarely appear in the sparse-stratification format that predominates in case–control research. We believe this appearance is deceptive, however, and that common practice has concealed the actual extent of sparse-data problems in follow-up studies. It has been our experience that the lack of widely disseminated methods for dealing with sparse follow-up data has led to the employment of ad hoc methods for coalescing small strata into large strata before analysis and publication. Probably the most common approach has been to limit the number of stratifying variables to two or three and collapse

Key words: Binary data; Cohort studies; Epidemiologic methods; Mantel–Haenszel estimate; Risk difference; Risk ratio; Weighted least squares.

over categories of polytomous variables so that "large-stratum" methods can be employed. This approach is unsatisfactory, if only for the reason that the number of important confounders often exceeds the number that can be controlled while preserving large strata; as a result, there may be considerable residual bias (confounding) in the estimates obtained from the reduced stratification. An example of this situation is provided by an analysis of the effect of maternal marijuana use on meconium passage by neonates (Greenland et al., 1982): Data were available on numerous potential confounders, yet it was impossible to stratify on more than two variables at a time and still avoid zero cells. Another ad hoc approach that has been applied to sparse follow-up data is the "multivariate confounder score" method of Miettinen (1976), which again results in the data being summarized in relatively few large strata. The study by Neutra et al. (1978) of the effect of fetal monitoring on neonatal death provides an example: Here, data on eighteen potential confounders were used to create six strata. The confounder-score approach is also rather unsatisfactory; in particular, it is known to yield invalid hypothesis tests (Pike, Anderson, and Day, 1979). In each of the cited studies logistic regression was also employed. Unfortunately, unconditional logistic estimates are biased if the cross-classification of the regressors with the outcome is sparse (Pike, Hill, and Smith, 1980). And while conditional logistic regression is unbiased in sparse data, it is based on a common odds-ratio model and cannot provide risk ratio or risk difference estimates.

Consider a series of K pairs of independent binomial observations (x_k, y_k) with denominators (n_k, m_k) and "success" probabilities (p_{1k}, p_{0k}) for $k = 1, \ldots, K$, as would arise in a fixed cohort study of an exposure and disease; x_k and y_k are the numbers of persons contracting the disease out of n_k exposed and m_k unexposed persons. Such data typically arise from nonexperimental studies involving complete follow-up of individuals over very short risk periods; examples are common among perinatal studies. Let us suppose we have a situation in which the risk ratio $\phi = p_{1k}/p_{0k}$ is the preferred measure of association and remains constant across k. Unlike the odds-ratio situation considered by Breslow (1981), in making inferences about ϕ it is necessary to employ the unconditional distribution of (x_k, y_k),

$$\Pr(x, y \mid n_k, m_k) = \binom{n_k}{x}(\phi p_{0k})^x(1 - \phi p_{0k})^{n_k - x}\binom{m_k}{y} p_{0k}^y q_{0k}^{m_k - y}, \tag{1}$$

which involves the K nuisance parameters $p_{0k} = 1 - q_{0k}$. For large values of n_k and m_k this presents no difficulty in principle, as we could estimate the set of parameters ϕ, p_{0k} $(k = 1, \ldots, K)$ by maximizing the likelihood based on (1) (subject to the constraints that the ϕp_{0k} and p_{0k} lie between 0 and 1); this requires iterative solution of a set of $K + 1$ equations, given by Rothman and Boice (1979). We will denote the estimator so obtained by $\hat{\phi}_U$. Alternatively, one could employ the weighted least squares (WLS) approach of Grizzle, Starmer, and Koch (1969), which yields a closed-form estimator $\hat{\phi}_W$ of ϕ defined by

$$\log \hat{\phi}_W = \left(\sum_k W_k \log \hat{\phi}_k\right) \Big/ \sum_k W_k, \tag{2}$$

where $W_k = (1/x_k - 1/n_k + 1/y_k - 1/m_k)^{-1}$ and $\hat{\phi}_k = x_k m_k/(y_k n_k)$. A constant, usually $\frac{1}{2}$, is commonly added to each cell to avoid division by 0. The quantity $\hat{\phi}_W$ also derives from inverse-variance weighted averaging of the log risk ratios. Rothman and Boice (1979) replace W_k with the inverse of the asymptotic variance of log $\hat{\phi}_k$ at $\phi = 1$, $W_{0k} = n_k m_k t_k/[N_k(N_k - t_k)]$, to obtain a "null-weighted" least squares estimator $\hat{\phi}_{W0}$. Here, $t_k = x_k + y_k$ and $N_k = m_k + n_k$.

Consideration of inverse asymptotic null variance weighting schemes led Tarone (1981) to propose, as an analog to the famous Mantel–Haenszel odds ratio, the estimator

$$\hat{\phi}_{\mathrm{T}} = \frac{\sum_k y_k m_k/(N_k - t_k)}{\sum_k y_k n_k/(N_k - t_k)}. \tag{3}$$

A constant may be added to the term $N_k - t_k$ to avoid division by 0. Tarone also considered a slightly different "Mantel–Haenszel" risk ratio estimator

$$\hat{\phi}_{\mathrm{MH}} = \frac{\sum_k x_k m_k/N_k}{\sum_k y_k n_k/N_k}, \tag{4}$$

which was also independently proposed by Nurminen (1981) and Kleinbaum, Kupper, and Morgenstern (1982).

Under the assumption that all the n_k and m_k tend to infinity (the "large-stratum" assumption) it is not difficult to show that all of the above estimators are consistent asymptotically normal. Our concern here, however, is with the situation in which this assumption is not justified. In this case we turn to the "sparse-data" large-sample theory introduced by Breslow (1981). Under this theory we will show that only the last estimator, $\hat{\phi}_{\mathrm{MH}}$, is in general consistent for ϕ.

2. Large-Sample Results

The limiting model employed here is similar to that used by Breslow. We assume there is a finite number of possible denominator configurations and K_i tables of the type with denominators (n_i, m_i). The cells of the jth table of the ith type are denoted x_{ij}, y_{ij}, $n_i - x_{ij}$, and $m_i - y_{ij}$ for $j = 1, \ldots, K_i$. As the sample size grows, the ratio K_i/K is assumed to approach a limit π_i which depends on the distribution of exposure across strata. Omission of the subscripts for x and y will indicate we are considering them as random variables. In many instances, for each i the nuisance parameters p_{0ij} can be assumed to have been randomly sampled from a probability distribution $F_i(z)$. Although none of our results depend on this assumption (Robins, Breslow, and Greenland, 1985), employing it yields the following expression for the unconditional distribution of (x, y) in a table of type i:

$$\Pr(x, y \mid n_i, m_i) = \int_0^{1/\phi} \binom{n_i}{x} (\phi z)^x (1 - \phi z)^{n_i - x} \binom{m_i}{y} z^y (1 - z)^{m_i - y}\, dF_i. \tag{5}$$

The terms x_k, y_k and $N_k - t_k$ have positive probability of being 0, so $\hat{\phi}_{\mathrm{W}}$ and $\hat{\phi}_{\mathrm{W0}}$ must be modified by adding a nonzero constant c to each cell in order for their sparse-data asymptotic means to exist. Given this modification, their respective asymptotic means ϕ_{W} and ϕ_{W0} may be computed from the unconditional distribution (5) using arguments that completely parallel those given by Breslow (1981) for the logit estimator of the odds ratio. Numerical calculations described below show that, with $c = \frac{1}{2}$ or less, ϕ_{W} and ϕ_{W0} may depart markedly from ϕ, even with moderately large strata.

The term $N_k - t_k$ in (3) has positive probability of being 0, so $\hat{\phi}_{\mathrm{T}}$ must be modified by adding a nonzero constant b to the terms $N_k - t_k$ in order for the asymptotic mean of $\hat{\phi}_{\mathrm{T}}$ to exist. Given this modification, and again employing (5) along with arguments parallel to Breslow (1981) to compute ϕ_{T}, we show below that, taking $b = 1$ or less, the asymptotic mean ϕ_{T} of $\hat{\phi}_{\mathrm{T}}$ may be far from ϕ when the strata are small.

Let ϕ_U be the value to which the unconditional maximum likelihood estimator $\hat{\phi}_U$ converges. By direct computation of ϕ_U in special cases, we have found that ϕ_U does not appear to have a simple or even continuous relationship to the underlying parameters; this is due in large part to its dependence on the nuisance parameter distributions. Furthermore, in certain cases $\hat{\phi}_U$ diverges. Because of the analytic complexity of the general case, we present systematic results only for the case $n = m = 1$. Analytic results for this case are given in Appendix 1. Numerical results given below show that, for values of ϕ in the neighborhood of unity, the bias in $\hat{\phi}_U$ can be severe.

Using again the unconditional distribution (5) and arguments that completely parallel those given by Breslow for the Mantel–Haenszel odds ratio, we find that in sparse data $\hat{\phi}_{MH}$ is asymptotically normal with mean ϕ and asymptotic variance

$$\frac{\sum_i \pi_i \mathrm{var}_i(R - \phi S)}{K[\sum_i \pi_i \mathrm{E}_i(S)]^2} = \mathrm{var}^{A}(\hat{\phi}_{MH}), \tag{6}$$

where var_i and E_i are computed under the distribution (5), $R = xm_i/N_i$ and $S = yn_i/N_i$.

3. Numerical Evaluations

Table 1 gives selected numerical evaluations of the asymptotic means of $\hat{\phi}_W$ and $\hat{\phi}_{W0}$, assuming $c = \frac{1}{2}$ is added to each cell, but tables with $t_k = 0$ are thrown out (the latter modification improved both estimators). Also given is the asymptotic mean of $\hat{\phi}_T$, defined by adding $b = 1$ to the terms $N_k - t_k$ in (3). The values of n, m, and p_0 were assumed to be constant across strata. Under these conditions stratification is unnecessary, as the crude (unstratified) estimator would always equal $\hat{\phi}_{MH}$. These assumptions are made for computational simplicity; the results do, however, unequivocally demonstrate the bias inherent in the estimators in sparse data. A large number of other values for p_0 and ϕ were also used; the four combinations presented in the tables exemplify the general results.

The bias in $\hat{\phi}_W$ can be severe, but is never a simple function of any of the parameters except sample size. The expectation of $\hat{\phi}_W$ appears to be very sensitive to the exposure ratio, n/m. Other values for the constant c and other schemes for handling cell zeros (such as adding a constant only when the cell is zero) were tried, but none improved the performance of the estimators. Note that the bias in $\hat{\phi}_W$ can remain large even for large values of n and m. The performance of $\hat{\phi}_{W0}$ is usually somewhat better than that of $\hat{\phi}_W$ though not consistently so; the estimator $\hat{\phi}_{W0}$ still exhibits large bias and sensitivity to the n/m ratio.

With $b = 1$, $\hat{\phi}_T$ exhibited very predictable behavior: Its bias was always toward the null value ($\phi = 1$) and dropped off as a direct function of the total stratum size $N = n + m$. For values of N of 10 or greater, $\hat{\phi}_T$ always showed less than 10% bias; on the other hand, its bias could be large when N was small. Using smaller values of b or adding a constant to each cell only worsened the bias of $\hat{\phi}_T$.

A further set of evaluations was done to examine the dependence of $\hat{\phi}_W$ and $\hat{\phi}_{W0}$ on the cell expectations. For brevity these are not presented here. In general, the bias in $\hat{\phi}_W$ and $\hat{\phi}_{W0}$ decreased with increasing smallest cell expectation, but in some cases the bias in $\hat{\phi}_W$ exceeded 10% even with a smallest cell expectation of 6.

Table 2 shows the limiting value ϕ_U of the unconditional maximum likelihood estimator $\hat{\phi}_U$ as a function of ϕ and a constant p_0 for the case $n = m = 1$. Generally, the relative bias of $\hat{\phi}_U$ is greatest for values of ϕ near 1, and for small values of p_0. Note, however, the rather pathological shift in the bias at the points in Table 2 at which $|\phi - 1|/\phi = p_0$. Examination of the likelihood for cases of n, $m \geqslant 2$ indicates that the bias, although smaller with larger n and m, will be subject to similar discontinuities.

Table 1
Asymptotic means of the weighted least squares and null-weighted risk ratio estimators as a function of the sample sizes n, m (E = *smallest cell expectation*)

n	m	$\phi = 2,\ p_0 = .1$				$\phi = 2,\ p_0 = .4$			
		E	ϕ_W	ϕ_{WO}	ϕ_T	E	ϕ_W	ϕ_{WO}	ϕ_T
1	1	0.1	1.36	1.41	1.83	0.2	1.23	1.34	1.56
1	2	0.2	1.49	1.45	1.90	0.2	1.32	1.30	1.63
2	2	0.2	1.40	1.50	1.92	0.4	1.32	1.49	1.64
1	4	0.2	1.78	1.59	1.95	0.2	1.49	1.31	1.73
2	4	0.4	1.66	1.64	1.96	0.4	1.50	1.50	1.75
4	4	0.4	1.44	1.62	1.97	0.8	1.48	1.67	1.77
1	8	0.2	2.30	1.88	1.97	0.2	1.65	1.37	1.85
2	8	0.4	2.08	1.81	1.98	0.4	1.69	1.54	1.85
4	8	0.8	1.79	1.78	1.98	0.8	1.68	1.69	1.86
8	8	0.8	1.50	1.78	1.98	1.6	1.66	1.82	1.88
1	16	0.2	3.08	2.27	1.99	0.2	1.76	1.42	1.92
2	16	0.4	2.67	2.03	1.99	0.4	1.83	1.58	1.92
4	16	0.8	2.25	1.90	1.99	0.8	1.84	1.72	1.92
8	16	1.6	1.88	1.88	1.99	1.6	1.82	1.83	1.93
16	16	1.6	1.61	1.92	1.99	3.2	1.80	1.90	1.94
n	m	$\phi = 8,\ p_0 = .05$				$\phi = 8,\ p_0 = .1$			
		E	ϕ_W	ϕ_{WO}	ϕ_T	E	ϕ_W	ϕ_{WO}	ϕ_T
1	1	0.05	2.13	2.27	6.00	0.1	1.96	2.19	4.89
1	2	0.1	2.67	2.83	6.76	0.2	2.47	2.74	5.80
2	2	0.1	2.37	2.75	6.96	0.2	2.42	3.04	5.90
1	4	0.2	3.42	3.55	7.33	0.2	3.17	3.49	6.70
2	4	0.2	3.18	3.66	7.40	0.4	3.17	3.96	6.74
4	4	0.2	2.76	3.55	7.51	0.4	3.17	4.39	6.82
1	8	0.4	4.41	4.33	7.65	0.2	4.04	4.21	7.29
2	8	0.4	4.23	4.69	7.67	0.4	4.10	4.88	7.31
4	8	0.4	3.79	4.80	7.71	0.8	4.10	5.48	7.33
8	8	0.4	3.41	4.84	7.77	0.8	4.12	5.96	7.38
1	16	0.4	5.80	5.12	7.82	0.2	5.05	4.78	7.63
2	16	0.8	5.54	5.56	7.83	0.4	5.14	5.51	7.63
4	16	0.8	5.02	5.93	7.84	0.8	5.16	6.20	7.64
8	16	0.8	4.56	6.20	7.86	1.6	5.17	6.76	7.66
16	16	0.8	4.29	6.38	7.89	1.6	5.19	7.16	7.69

Table 2
Limiting values ϕ_U *of the unconditional maximum likelihood risk ratio estimator when* $n = m = 1$ (U = *Undefined*, I = *Impossible combination of* ϕ *and* p_0)

ϕ	p_0					
	0.0[a]	0.1	0.25	0.33	0.8	1.0[a]
0.20	0.17	0.17	0.17	0.18	0.19	0.20
0.50	0.33	0.34	0.36	0.38	0.45	0.50
0.75	0.43	0.45	0.48	U	1.00	1.00
1.33	2.33	2.20	U	1.00	I	I
2.00	3.00	2.80	2.50	2.33	I	I
5.00	6.00	5.50	I	I	I	I

[a] Limiting values as p_0 approaches 0 or 1.

4. Rate Ratio Estimation Under a Poisson Model

Under some circumstances, the data of interest are modeled as a series of K pairs of independent Poisson observations (x_k, y_k) with fixed "person-time" denominators (n_k, m_k)

and means given by $r_{1k}n_k$ and $r_{0k}m_k$, where r_{1k} and r_{0k} are the disease rates. Such data arise in vital statistics and in follow-up studies of dynamic populations. Consider first the situation in which the rate ratio $\omega = r_{1k}/r_{0k}$ remains constant across k. Similar to the odds-ratio case, inference can proceed using the distribution of x_k conditional on the total number of cases $t_k = x_k + y_k$:

$$\Pr(x \mid n_k, m_k, t_k) = \binom{t_k}{x}\left(\frac{\omega n_k}{\omega n_k + m_k}\right)^x\left(\frac{m_k}{\omega n_k + m_k}\right)^y. \tag{7}$$

The conditional maximum likelihood estimator $\hat{\omega}_C$ is found by iterative solution of the likelihood based on (7) (Rothman and Boice, 1979). If, instead, an unconditional likelihood is formed using the original Poisson distributions with nuisance parameters r_{0k}, $k = 1, \ldots, K$, the resulting unconditional maximum likelihood estimator will be identical to $\hat{\omega}_C$ (Andersen, 1970).

One closed-form estimator of ω is the weighted least squares estimator $\hat{\omega}_W$, defined by

$$\log \hat{\omega}_W = \left(\sum_k W_k \log \hat{\omega}_k\right) \Big/ \sum_k W_k, \tag{8}$$

where $W_k = (1/x_k + 1/y_k)^{-1}$ and $\hat{\omega}_k = x_k m_k/(y_k n_k)$. A constant c may be added to x_k and y_k to avoid division by 0. Rothman and Boice (1979) replace W_k with the inverse of the variance of log $\hat{\omega}_k$ at $\omega = 1$, $W_{0k} = t_k n_k m_k/N_k^2$, to obtain a "null-weighted" least squares estimator $\hat{\omega}_{W0}$. Another closed-form estimator of ω is the Mantel–Haenszel rate ratio proposed by Rothman and Boice (1979):

$$\hat{\omega}_{MH} = \frac{\sum_k x_k m_k/N_k}{\sum_k y_k n_k/N_k}. \tag{9}$$

Under the assumption that the n_k and m_k tend to infinity, all the above estimators of ω are consistent asymptotically normal. Two "sparse-data" limiting models suggest themselves for this situation. One could assume a joint distribution π for n, m, and r_0, so that (n_k, m_k, r_{0k}) is one of a sequence of K independent identically distributed random vectors with strictly positive components and finite covariance matrix. Alternatively, one could assume a joint distribution π for n, m, and t, the marginal total of events. The results given below hold under either model, as well as for fixed sequences of n, m, and r_0 or t.

In order for the asymptotic means ω_W and ω_{W0} of $\hat{\omega}_W$ and $\hat{\omega}_{W0}$ to exist, a constant c must be added to x_k and y_k. Under either sparse-data model it is not difficult to show, in a manner paralleling the odds ratio and risk ratio results, that neither ω_W nor ω_{W0} will in general equal ω.

Using arguments that completely parallel those given by Breslow for the odds ratio, we find that conditioning on the margin t, $\hat{\omega}_{MH}$ is consistent asymptotically normal with large-sample variance

$$\frac{E_\pi[\text{var}(R - \omega S \mid n, m, t)]}{E_\pi^2[E(S \mid n, m, t)]} = \text{var}^A(\hat{\omega}_{MH}),$$

where the inner variance and expectation are over the binomial distribution (7). This formula also applies with r_0 substituted for t under the other "sparse-data" model. From the results of Andersen (1970) we have that $\hat{\omega}_C$ (and hence $\hat{\omega}_U$) is also consistent asymptot-

Table 3
Asymptotic means of the weighted least squares and null-weighted rate ratio estimators as a function of the cell expectations $\phi r_0 n$ and $r_0 m$

$\phi r_0 n$	$r_0 m$	$\omega = 2.0$		$\omega = 10.0$		$\omega = 0.5$	
		ω_W	ω_{WO}	ω_W	ω_{WO}	ω_W	ω_{WO}
1	1	2.00	2.00	10.00	10.00	0.50	0.50
1	2	2.56	2.16	12.80	10.82	0.64	0.54
2	2	2.00	2.00	10.00	10.00	0.50	0.50
1	4	3.07	2.25	15.34	11.23	0.77	0.56
2	4	2.39	2.03	11.97	10.14	0.60	0.51
4	4	2.00	2.00	10.00	10.00	0.50	0.50
1	8	3.39	2.30	16.97	11.48	0.85	0.57
2	8	2.66	2.04	13.31	10.19	0.67	0.51
4	8	2.23	2.00	11.16	9.99	0.56	0.50
8	8	2.00	2.00	10.00	10.00	0.50	0.50

ically normal, with large-sample variance

$$\frac{\omega^2}{K\mathrm{E}_\pi[\mathrm{var}(x \mid n, m, t)]} = \mathrm{var}^{\mathrm{A}}(\hat{\omega}_{\mathrm{C}}),$$

where $\mathrm{var}(x \mid n, m, t) = tpq$ is the binomial variance from (7), $p = \omega n/(\omega n + m)$, and $q = 1 - p$. As noted by Breslow (1984), if $\omega = 1$,

$$\mathrm{var}^{\mathrm{A}}(\hat{\omega}_{\mathrm{MH}}) = \mathrm{var}^{\mathrm{A}}(\hat{\omega}_{\mathrm{C}})$$

again, completely analogous to odds-ratio results (Breslow, 1981). Finally, it is not difficult to show that if the exposure ratio n_i/m_i is constant, $\hat{\omega}_{\mathrm{C}} = \hat{\omega}_{\mathrm{MH}}$ and thus the two will (trivially) have equal variances in this case.

Table 3 displays the asymptotic means of $\hat{\omega}_{\mathrm{W}}$ and $\hat{\omega}_{\mathrm{WO}}$, assuming $c = \frac{1}{2}$ is added to both x_k and y_k unless $t_k = 0$, in which case the table is thrown out. The quantities r_0, n, and m are assumed constant across strata. Under this assumption $\hat{\omega}_{\mathrm{C}} = \hat{\omega}_{\mathrm{MH}}$, and the values of all the tabulated quantities depend on the baseline rate r_0 and the denominators n and m only through the parameter ω and the cell expectations $r_0 n$ and $r_0 m$. Consequently, the tabulated quantities are presented as functions of ω and the cell expectations.

It can be seen that the bias in $\hat{\omega}_{\mathrm{W}}$ and $\hat{\omega}_{\mathrm{WO}}$ is directly proportional to the ratio of the cell expectations and inversely proportional to their total, with no bias present if the expectations are equal. It appears that $\hat{\omega}_{\mathrm{W}}$ has several times the bias of $\hat{\omega}_{\mathrm{WO}}$, and while the bias in $\hat{\omega}_{\mathrm{W}}$ can be severe, the bias in $\hat{\omega}_{\mathrm{WO}}$ exceeds 10% only for small and imbalanced cell expectations.

5. Risk and Rate Difference Estimation

Results paralleling those for the risk ratio can be found for the problem of estimation of a common risk difference. Specifically, suppose now we have a situation in which the risk difference $\delta = p_{1k} - p_{0k}$ is the preferred measure of association and remains constant across k. Again, inference must proceed using the unconditional distribution of (x_k, y_k):

$$\Pr(x, y \mid n_k, m_k) = \binom{n_k}{x}(p_{0k} + \delta)^x(1 - p_{0k} - \delta)^{n_k - x}\binom{m_k}{y}p_{0k}^y q_{0k}^{m_k - y}. \tag{10}$$

Because the algebraic details closely parallel those for the risk ratio results, we will only review our findings regarding estimation of δ from sparse data.

Table 4
Asymptotic means of the weighted least squares and null-weighted risk difference estimators as a function of the sample sizes n, m

n	m	$\delta = .1,\ p_0 = .1$		$\delta = .6,\ p_0 = .1$		$\delta = .1,\ p_0 = .4$	
		δ_W	δ_{WO}	δ_W	δ_{WO}	δ_W	δ_{WO}
1	1	.179	.174	.411	.398	.062	.056
1	2	.217	.177	.476	.455	.046	.036
2	2	.152	.137	.472	.422	.080	.065
1	4	.220	.193	.505	.492	.047	.040
2	4	.159	.157	.526	.468	.078	.062
4	4	.113	.104	.537	.470	.095	.077
1	8	.221	.213	.504	.501	.050	.047
2	8	.150	.170	.546	.493	.080	.064
4	8	.104	.124	.575	.501	.099	.078
8	8	.091	.086	.581	.526	.099	.088
1	16	.230	.229	.494	.492	.050	.049
2	16	.151	.178	.550	.503	.082	.066
4	16	.096	.134	.593	.521	.104	.079
8	16	.082	.102	.601	.545	.103	.088
16	16	.088	.084	.594	.561	.100	.094

Let δ_U be the value to which the unconditional maximum likelihood estimator $\hat{\delta}_U$ based on (10) converges. Similar to the risk ratio case, the relationship between δ_U and the underlying parameters appears complex, and in certain cases $\hat{\delta}_U$ diverges.

The weighted least squares estimator is

$$\hat{\delta}_W = \sum_k W_k \hat{\delta}_k \Big/ \sum_k W_k, \tag{11}$$

where $W_k = [x_k(n_k - x_k)/n_k + y_k(m_k - y_k)/m_k]^{-1}$ and $\hat{\delta}_k = x_k/n_k - y_k/m_k$; a constant may be added to each cell to avoid division by 0. Rothman and Boice (1979) replace W_k with $W_{0k} = n_k m_k N_k/[t_k(N_k - t_k)]$ to derive a null-weighted estimator $\hat{\delta}_{WO}$. Table 4 presents some of the results of our numerical studies of $\hat{\delta}_W$ and $\hat{\delta}_{WO}$, assuming a constant $c = \frac{1}{2}$ is added to each cell, but tables with $t_k = 0$ thrown out (the latter modification improved both slightly). The quantities p_0, n, and m are assumed constant across strata. Other methods of dealing with zeros were tried (including adding $c = \frac{1}{2}$ only in the calculation of the weights), but none produced better performance. For both estimators, the bias can be severe, and both are very sensitive to the n/m ratio.

In situations in which the data follow the Poisson model of §4, δ becomes the rate difference $r_{1k} - r_{0k}$, and inference must proceed using the unconditional distribution. The weighted least squares estimators have the same form as in (11), but with $W_k = [x/n_k^2 + y/m_k^2]^{-1}$ for $\hat{\delta}_W$ and $W_{0k} = n_k m_k/t_k$ for $\hat{\delta}_{WO}$ (Rothman and Boice, 1979). A constant may be added to x and y to avoid division by 0; again, both estimators become biased in sparse data because of the necessity of ad hoc methods for handling zeros. We have been unable to derive results concerning the consistency of the unconditional maximum likelihood estimator of δ in the sparse-data Poisson (rate difference) case.

The Mantel–Haenszel risk and rate ratios can be written as ratios of standardized rates with standard weights $n_k m_k/N_k$. Thus, a sparse-data risk or rate difference can be found by taking the difference of the same standardized rates, leading to

$$\hat{\delta}_{MH} = \sum_k (x_k m_k/N_k - y_k n_k/N_k) \Big/ \left(\sum_k n_k m_k/N_k\right) \tag{12}$$

(Greenland, 1982). It is straightforward to show that $\hat{\delta}_{\mathrm{MH}}$ is consistent in both sparse data and large strata.

6. Variance Estimation

Robins et al. (1985) have proposed an estimator for the variance of the Mantel–Haenszel odds ratio based on the unconditional distribution that consistently estimates the asymptotic variance under both the large-stratum and sparse-data limiting models. Their approach extends in a straightforward manner to the other Mantel–Haenszel estimators, yielding the following variance estimator for the log risk ratio:

$$\widehat{\mathrm{var}}^{\mathrm{A}}(\log \hat{\phi}_{\mathrm{MH}}) = \frac{\sum_k D_k}{\left(\sum_k R_k\right)\left(\sum_k S_k\right)}, \tag{13}$$

where

$$D_k = (m_k n_k t_k - x_k y_k N_k)/N_k^2.$$

Under the unconditional distribution, the variance estimator for the log of the rate ratio $\hat{\omega}_{\mathrm{MH}}$ is identical to (13) except that $D_k = n_k m_k t_k/N_k^2$. Another estimator of this variance derived by Tarone (1981) for large strata is also consistent in sparse data. Both our estimator and Tarone's differ from the estimator based on the conditional distribution derived by Breslow (1984), which is

$$\widehat{\mathrm{var}}^{\mathrm{A}}(\log \hat{\omega}_{\mathrm{MH}}) = \frac{\sum_k n_k m_k t_k/N_k^2}{\hat{\omega}_{\mathrm{MH}}\left[\sum_k \frac{n_k m_k t_k}{N_k(\hat{\omega}_{\mathrm{MH}} n_k + m_k)}\right]^2}.$$

As Breslow (1984) notes, the latter depends on x_k and y_k only through $\hat{\omega}_{\mathrm{MH}}$ and thus might be expected to perform better than other estimators; nevertheless, the analogous odds-ratio variance estimators exhibit negligible differences in performance (Robins et al., 1985).

For the risk difference, the formula analogous to (13) is

$$\widehat{\mathrm{var}}^{\mathrm{A}}(\hat{\delta}_{\mathrm{MH}}) = \sum_k L_k \Bigg/ \left(\sum_k n_k m_k/N_k\right)^2 \tag{14}$$

where

$$L_k = [x_k(n_k - x_k)m_k^3 + y_k(m_k - y_k)n_k^3]/(n_k m_k N_k^2)$$

For the rate difference (Poisson) case, L_k becomes $(x_k m_k^2 + y_k n_k^2)/N_k^2$.

An alternative method of estimating the variance of $\hat{\phi}_{\mathrm{MH}}$ would be to parallel Breslow and Liang (1982) and construct a general variance estimate as a weighted average of sparse-data and large-stratum variance estimates. In the odds-ratio case, however, this type of estimator does not appear to perform as well as the unconditional estimator, and thus we have not pursued this approach further.

All the variance formulas presented here were derived under the assumption of homogeneity across strata of the effect parameter. Guilbaud (1983) has shown that in the odds-ratio case the large-stratum variance formula derived under homogeneity (Hauck, 1979) is inappropriate when heterogeneity is present. We would expect that in a similar fashion the variance formulas presented here do not apply when heterogeneity is present. However, if important heterogeneity is present, the Mantel–Haenszel estimator will not estimate a meaningful parameter, and thus will also be inappropriate (Greenland, 1982).

7. Conditions Under Which Stratification Is Unnecessary

It is well known that if in follow-up data the exposure ratios n_k/m_k are constant across strata, the Mantel–Haenszel estimate of ϕ, ω, or δ will equal the crude estimate obtained from the single crude table, i.e., based on $\sum_k x_k$, $\sum_k y_k$, $\sum_k n_k$, and $\sum_k m_k$ (Kleinbaum et al., 1982). In such a situation one often sees the stratification discarded in favor of the simpler, crude analysis, including a variance estimate derived from the crude table. For the rate ratio and rate difference, it is easily verified that the crude variance estimator is consistent for the true asymptotic variance. But, for the risk ratio and risk difference, if the nuisance parameters (p_{0k}) are not constant across strata, then under either limiting model the crude variance estimator converges to a value larger than the true asymptotic variance (a proof of this is given in Appendix 2). Thus, if the nuisance parameters are not constant, then conditional on the fixed stratum margins (n_k, m_k), the use of the crude data alone will result in overly conservative confidence intervals even if the crude point estimator is conditionally unbiased. On the other hand, if the nuisance parameters are known to be constant, stratification will be unnecessary, as the crude point estimator will be unbiased (regardless of variation in the exposure ratios) and the crude variance estimator will yield valid confidence intervals.

8. Efficiency Considerations

As discussed in §4, the Mantel–Haenszel rate ratio $\hat{\omega}_{\mathrm{MH}}$ is efficient on the null hypothesis $\omega = 1$ and also whenever the ratio n/m is constant; in fact, in the latter case it is identical to the conditional maximum likelihood estimate. These properties can easily be shown to hold in the large-stratum case, as well as the sparse-data case considered here. Thus, we would expect that $\hat{\omega}_{\mathrm{MH}}$ would have good efficiency in general, much like the Mantel–Haenszel odds ratio (Hauck, 1979; Breslow, 1981).

Consider next estimation of a common risk ratio ϕ. For large strata, the maximum likelihood estimator $\hat{\phi}_{\mathrm{U}}$ and the ordinary weighted least squares estimator $\hat{\phi}_{\mathrm{W}}$ are fully efficient, with asymptotic variance

$$\phi^2 \Big/ \sum_k w_k = \sigma_{\mathrm{R}}^2,$$

where

$$w_k = [1/(\phi p_{0k} n_k) + 1/(p_{0k} m_k) - 1/n_k - 1/m_k]^{-1}.$$

The large-stratum variances of $\hat{\phi}_{\mathrm{MH}}$ and $\hat{\phi}_{\mathrm{T}}$ are

$$\phi^2 \sum_k s_k^2 w_k^{-1} \Big/ \left(\sum_k s_k\right)^2 = \sigma_{\mathrm{MH}}^2$$

and

$$\phi^2 \sum_k u_k^2 w_k^{-1} \Big/ \left(\sum_k u_k\right)^2 = \sigma_{\mathrm{T}}^2,$$

respectively, where $s_k = p_{0k} m_k n_k / N_k$ and $u_k = p_{0k} m_k n_k / (N_k - \phi p_{0k} n_k - p_{0k} m_k)$ (Tarone, 1981). When $\phi = 1$, $\sigma_{\mathrm{R}}^2 = \sigma_{\mathrm{T}}^2$, i.e., $\hat{\phi}_{\mathrm{T}}$ is fully efficient under the null hypothesis, as is the null-weighted estimator $\hat{\phi}_{\mathrm{W0}}$. In contrast, $\hat{\phi}_{\mathrm{MH}}$ is not (Tarone, Gart, and Hauck, 1983); for example, taking $K = 2$, $p_{01} = .2$, $p_{02} = .8$, $\phi = 1$, and $n = m$ constant yields a large-stratum variance ratio of $\sigma_{\mathrm{R}}^2/\sigma_{\mathrm{MH}}^2 = .74$, i.e., $\hat{\phi}_{\mathrm{MH}}$ is only 74% efficient in this case, even though $\phi = 1$ and n and m are constant and equal.

Finally, consider estimation of a common risk or rate difference. For large strata, the maximum likelihood and the ordinary weighted least squares estimators are fully efficient; under the null hypothesis the null-weighted estimators are also efficient. But again, the estimators employing the "Mantel–Haenszel" weights are inefficient, and as in the risk ratio case they may have less than 80% efficiency in certain cases. These results are to be expected, since in neither case is $\hat{\delta}_{\mathrm{MH}}$ an inverse-variance weighted estimator.

9. Discussion

As we have shown, $\hat{\phi}_{\mathrm{MH}}$ and $\hat{\delta}_{\mathrm{MH}}$ are the only currently available risk ratio and risk difference estimators that are consistent in sparse data. Nevertheless, their potentially considerable inefficiency in large strata suggests that their use be limited to sparse data, and that more efficient sparse-data estimators should be sought. In contrast, we note again that the inefficiency of the Mantel–Haenszel estimators of a common odds ratio and a common rate ratio appears small.

The theoretical and empirical results presented here provide an interesting parallel between odds ratio estimation (Breslow, 1981), risk and rate ratio estimation, and risk and rate difference estimation in sparse data. For the fixed cohort case, weighted least squares estimators appear generally unsatisfactory due to their large and erratic bias; the unconditional maximum likelihood estimators are also biased. For the Poisson case, the weighted least squares estimators again tend to be unsatisfactory.

These conclusions undoubtedly extend to situations involving modeling of risks or rates as a function of several covariates. In particular, our results indicate that weighted least squares methods will yield severely biased estimates of model parameters in sparse cross-classifications; for binomial outcomes, unconditional maximum likelihood will also be biased in sparse data. In light of this, a conservative approach to sparse fixed cohort data would be to perform a separate Mantel–Haenszel analysis for each covariate, stratifying on the others. Unfortunately, this approach is inefficient, in that interactions can be dealt with only by data-splitting; for estimation purposes continuous variables must be treated categorically, and the Mantel–Haenszel estimators are themselves inefficient. Alternatively, conditional modeling of the odds ratio may be employed (Thomas, 1981), but this generally imposes functional restrictions on the interactions between covariates and stratifying factors that exclude constant risk ratio or risk difference assumptions. In our opinion, a fully satisfactory approach to the analysis of sparse multivariate data has yet to be developed.

We have not attempted to address the issue of the appropriateness or utility of the common risk ratio or risk difference assumptions. Nevertheless, a particularly noteworthy objection to the common risk ratio assumption is the severe range limitation imposed on the parameter of interest by the nuisance parameters: for example, a common risk ratio cannot exceed 4 if the nuisance parameters p_{0k} can range as high as .25. (Analogous comments apply for the risk difference.) If, on the other hand, the nuisance parameters are always so small that no important limit is imposed on the common risk ratio, the odds ratio would serve as a practical substitute for the risk ratio, and by using the odds ratio all inference could proceed conditionally. Unfortunately, no such alternative to the risk difference is available. With these points in mind, we suggest that common risk ratio or risk difference estimators or models be used cautiously.

Acknowledgements

We are grateful to Kenneth Rothman and Nathan Mantel for their helpful comments, and to Tamba McIntire-Morgan and Virginia Hansen for preparation of this manuscript.

Résumé

Breslow (1981, *Biometrika* **68,** 73–84) a montré que le rapport des probabilités de Mantel–Haenszel est un estimateur consistent du rapport des probabilités communes dans des stratifications peu denses. Pour les études de cohortes, pourtant, l'estimation du rapport ou de la différence des risques communs peut avoir un plus grand intérêt. Sous un modèle binomial de données peu denses, les estimateurs de Mantel–Haenszel, du rapport et de la différence des risques, sont consistents dans des stratifications peu denses, tandis que les estimateurs du maximum de vraisemblance et des moindres carrés pondérés sont biaisés. Sous les modèles de Poisson des données peu denses, les estimateurs de Mantel–Haenszel et du maximum de vraisemblance du rapport des risques ont des variances asymptotiques égales sous l'hypothèse nulle et sont consistents, tandis que les estimateurs des moindres carrés pondérés sont encore biaisés; de façon semblable, pour les estimateurs de la différence des risques communs, les estimateurs des moindres carrés pondérés sont biaisés, tandis que l'estimateur qui utilise les pondérations "Mantel–Haenszel" est consistent pour les données peu denses. Des estimateurs de la variance qui sont consistents aussi bien pour les données peu denses que pour les larges strates peuvent être trouvés pour tous les estimateurs de Mantel–Haenszel.

References

Andersen, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistical Society, Series B* **32,** 283–301.

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

Breslow, N. E. (1984). Elementary methods of cohort analysis. *International Journal of Epidemiology* **13,** 112–115.

Breslow, N. E. and Liang, K. Y. (1982). The variance of the Mantel–Haenszel estimator. *Biometrics* **38,** 943–952.

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

Greenland, S., Staisch, K. J., Brown, N., and Gross, S. J. (1982). Effects of marijuana on human pregnancy, labor, and delivery. *Neurobehavioral Toxicology and Teratology* **4,** 447–450.

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

Guilbaud, O. (1983). On the large-sample distribution of the Mantel–Haenszel odds-ratio estimator. *Biometrics* **39,** 523–525.

Hauck, W. W. (1979). The large sample variance of the Mantel–Haenszel estimator of a common odds ratio. *Biometrics* **35,** 817–819.

Kleinbaum, D. G., Kupper, L. L., and Morgenstern, H. (1982). *Epidemiologic Research: Principles and Quantitative Methods.* Belmont, California: Lifetime Learning Publications.

Miettinen, O. S. (1976). Stratification by a multivariate confounder score. *American Journal of Epidemiology* **104,** 609–620.

Neutra, R. R., Fienberg, S. E., Greenland, S., and Friedman, E. A. (1978). Effect of fetal monitoring on neonatal death rates. *New England Journal of Medicine* **299,** 324–326.

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

Pike, M. C., Anderson, J., and Day, N. (1979). Some insights into Miettinen's multivariate confounder score approach to case–control study analysis. *Journal of Epidemiology and Community Health* **33,** 104–106.

Pike, M. C., Hill, A. P., and Smith, P. G. (1980). Bias and efficiency in logistic analyses of stratified case–control studies. *International Journal of Epidemiology* **9,** 89–95.

Robins, J. M., Breslow, N. E., and Greenland, S. (1985). Estimators of the Mantel–Haenszel variance consistent in both sparse-data and large-strata limiting models. *Biometrics* (to appear).

Rothman, K. J. and Boice, J. D. (1979). *Epidemiologic Analysis with a Programmable Calculator.* NIH Publ. 79-1649. Washington, D. C.: U.S. Government Printing Office.

Tarone, R. E. (1981). On summary estimators of relative risk. *Journal of Chronic Diseases* **34,** 463–468.

Tarone, R. E., Gart, J. J., and Hauck, W. W. (1983). On the asymptotic inefficiency of certain noniterative estimators of a common relative risk or odds ratio. *Biometrika* **70,** 519–522.

Thomas, D. C. (1981). General relative-risk models for survival time and matched case–control analysis. *Biometrics* **37,** 673–686.

Received January 1984; *revised November* 1984.

Appendix 1

Large-Sample Limiting Value of $\hat{\phi}_U$ when $n = m = 1$

For the case $n = m = 1$ there are only four possible tables, according to whether (x, y) equals (1, 1), (1, 0), (0, 1), or (0, 0). Let π_1, π_2, π_3, and π_4 be the respective probabilities of observing each table type. If $F(z)$ is the nuisance parameter distribution, we have that

$$\pi_1 = \mathrm{E}[(\phi z)(z)] = \phi \mathrm{E}(z^2),$$

$$\pi_2 = \mathrm{E}[\phi z(1 - z)] = \phi[\mathrm{E}(z) - \mathrm{E}(z^2)],$$

$$\pi_3 = \mathrm{E}[(1 - \phi z)(z)] = \mathrm{E}(z) - \phi \mathrm{E}(z^2),$$

$$\pi_4 = \mathrm{E}[(1 - \phi z)(1 - z)] = 1 - (\phi + 1)\mathrm{E}(z) + \phi \mathrm{E}(z^2) = 1 - \pi_1 - \pi_2 - \pi_3.$$

Let $K_1, \ldots, K_4$ be the number of tables of the four types. Note that the nuisance parameter estimates will be identical for two tables of the same type, and let $\hat{p}_{01}, \ldots, \hat{p}_{04}$ be the maximum likelihood estimates of the nuisance parameters for each table type. By substituting directly into the likelihood, we find $\hat{p}_{01} = \min(1, 1/\hat{\phi}_U)$, $\hat{p}_{02} = \min(\frac{1}{2}, 1/\hat{\phi}_U)$, $\hat{p}_{03} = \min[1, 1/(2\hat{\phi}_U)]$, and $\hat{p}_{04} = 0$. Substituting these results directly into the log-likelihood in the regions $\hat{\phi}_U < \frac{1}{2}$, $\frac{1}{2} < \hat{\phi}_U < 1$, $1 < \hat{\phi}_U < 2$, and $2 < \hat{\phi}_U$, and maximizing with respect to $\hat{\phi}_U$, we obtain

(i) $\hat{\phi}_U = (K_1 + K_2 + K_3)/(K_1 + K_3)$ if $K_1 < K_2 - K_3$;

(ii) $\hat{\phi}_U = 1$ if $K_1 > |K_2 - K_3|$;

(iii) $\hat{\phi}_U = (K_1 + K_2)/(K_1 + K_2 + K_3)$ if $K_1 < K_3 - K_2$;

(iv) If $K_1 = |K_2 - K_3|$, a unique value for $\hat{\phi}_U$ does not exist.

Noting that K_i/K converges to π_i in probability, we obtain the following results by replacing K_i with π_i and $\hat{\phi}_U$ with ϕ_U in (i)–(iv) above, and then reexpressing the π_i in terms of $\mathrm{E}(z)$, $\mathrm{E}(z^2)$, and ϕ. Let $R = \mathrm{E}(z^2)/\mathrm{E}(z)$ and $\theta = (\phi - 1)/\phi$. Then

(i) If $R < \theta$, then $\phi_U = \phi + 1 - \phi R$;

(ii) If $R > |\theta|$, then $\phi_U = 1$;

(iii) If $R < -\theta$, then $\phi_U = \phi/(\phi + 1 - \phi R)$;

(iv) If $R = |\theta|$, then ϕ_U is undefined.

Note that $0 \le R \le \sup(z) \le \min(1, 1/\phi)$. From this and (i)–(iii) we find that if $R \ne |\theta|$, $\phi > 1$ implies $\phi < \phi_U \le \phi + 1$, $\phi < 1$ implies $\phi/(\phi + 1) < \phi_U < \phi$, and $\phi = 1$ implies $\phi_U = 1$. Thus, if $R \ne |\theta|$, the bias in $\hat{\phi}_U$ does not exceed 1, and is 0 if $\phi = 1$.

Appendix 2

Proof that the Crude Risk Ratio Variance Estimator Is Inconsistent if the Nuisance Parameters Are Not Constant

Consider the sparse-data model of §2. The quantities y_{ij}, K_i, π_i, $F_i(z)$, n_i, m_i, N_i, K, ϕ, and $\hat{\phi}_{MH}$ are as defined in §§1 and 2. Given $n_i = \mu N_i$ for all i, the crude variance estimator of the log of the crude risk ratio, $\log \hat{\phi}_C$, is defined as

$$\hat{V}_C(\log \hat{\phi}_C) \equiv \frac{1 - \hat{\phi}_C \hat{\mathrm{E}}(z)}{K\mu \hat{N}_+ \hat{\phi}_C \hat{\mathrm{E}}(z)} + \frac{1 - \hat{\mathrm{E}}(z)}{K(1 - \mu)\hat{N}_+ \hat{\mathrm{E}}(z)},$$

where $\hat{N}_+ \equiv (\sum_i N_i K_i)/K$, $\hat{\mathrm{E}}(z) \equiv \sum_{ij} y_{ij}/\sum_i K_i m_i$, and $\hat{\phi}_C = \hat{\phi}_{MH}$. It is straightforward to show that $K\hat{V}_C(\log \hat{\phi}_C)$ converges in probability to KV_C, where

$$KV_C = \frac{1 - \phi \mathrm{E}(z)}{\mu N_+ \phi \mathrm{E}(z)} + \frac{1 - \mathrm{E}(z)}{(1 - \mu)N_+ \mathrm{E}(z)},$$

$$\mathrm{E}(z) \equiv \sum_i \pi_i m_i \mathrm{E}_i(z) \Big/ \sum_i \pi_i m_i, \quad \text{and} \quad N_+ \equiv \sum_i \pi_i N_i.$$

Theorem 1. Under the sparse-data model of §2, if $n_i = \mu N_i$ for all i, then $\text{var}^{A}(\log \hat{\phi}_{MH}) \le V_C$, with equality if and only if $E_i(z)$ does not depend on i and $\text{var}_i(z) = 0$ for all i.

The proof proceeds by the following lemmas.

Lemma 1. Let $a_i = \pi_i n_i m_i E_i(z)/N_i$, $A = \sum_i a_i$, and

$$H_i = \frac{1 - \phi E_i(z)}{K\pi_i n_i \phi E_i(z)} + \frac{1 - E_i(z)}{K\pi_i m_i E_i(z)}.$$

Then

$$\text{var}^{A}(\log \hat{\phi}_{MH}) = \frac{K \sum_i a_i^2 H_i - \sum_i a_i \text{var}_i(z)/E_i(z)}{KA^2}$$

Proof. This follows after substantial algebra from equation (6) of §2.

Lemma 2. Let C be a constant, $U = \sum_i a_i^2 H_i/A^2$, and suppose $n_i = \mu N_i$ for all i. Then subject to the linear constraint $E(z) = C$, U attains a maximum of V_C when $E_i(z) = E(z)$ for all i.

Proof. From $n_i = \mu N_i$ we obtain $m_i = (1 - \mu)N_i$,

$$E(z) = \frac{\sum_i \pi_i N_i E_i(z)}{\sum_i \pi_i N_i},$$

and $a_i = \mu(1 - \mu)\pi_i N_i E_i(z)$, so that, if $E(z)$ is constrained to equal a constant C, A will be the constant $\mu(1 - \mu) \sum_i \pi_i N_i E_i(z)$. Consequently,

$$\sum a_i^2 H = \mu^2(1 - \mu)^2 \sum_i [\pi_i N_i E_i(z)]^2 H_i = \left(\frac{1 - \mu}{\phi} + \mu\right)A - \mu(1 - \mu)F = UA^2$$

where $F = \sum_i \pi_i N_i E_i^2(z)$. Thus, to maximize U with respect to the $E_i(z)$, we need only minimize F. Using the method of Lagrange multipliers we find that, subject to $E(z) = C$, F is minimized when $E_i(z) = C = E(z)$. Direct substitution then yields V_C as the constrained maximum of U.

Theorem 1 now follows by noting that U bounds $\text{var}^{A}(\log \hat{\phi}_{MH})$ from above, with equality if and only if $\text{var}_i(z) = 0$ for all i.

We note that Theorem 1 can easily be extended to the large-stratum case: One parallels the above argument by taking i to index strata (rather than denominator configurations), K = number of strata, $\pi_i = 1/K$, and $E_i(z) = p_{0i}$; $\text{var}_i(z)$ is necessarily 0 in this case.