BIOMETRICS 42, 311–323
June 1986

Estimators of the Mantel–Haenszel Variance Consistent in Both Sparse Data and Large-Strata Limiting Models

James Robins

Occupational Health Program, Harvard School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115, U.S.A.

Norman Breslow

Department of Biostatistics, SC-32, School of Public Health and Community Medicine, University of Washington, Seattle, Washington 98195, U.S.A.

and

Sander Greenland

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

SUMMARY

This paper proposes a new estimator of the variance of the Mantel–Haenszel estimator of common odds ratio that is easily computed and consistent in both sparse data and large-strata limiting models. Monte Carlo experiments compare its performance to that of previously proposed variance estimators.

1. Introduction

The Mantel–Haenszel (1959) estimator of the common odds ratio in a series of 2×2 tables is widely used by practicing statisticians and epidemiologists. In spite of this popularity, until recently it lacked a robust variance estimator. Hauck (1979) proposed a variance based on a "large-stratum" limiting model (Model I) wherein the number of tables remained fixed but individual cell sizes increased without bound. Breslow (1981), using a "sparse-data" limiting model (Model II) in which the number of tables increased but the cell sizes remained bounded, proposed one variance based on the conditional distribution of the observations in each table given fixed marginal totals and a second empirical variance that was more easily calculated. Since this latter variance is inconsistent under Model I, Breslow and Liang (1982) later suggested a weighted average of the Hauck and empirical variances that has a theoretical justification under either limiting model.

This paper proposes a Mantel–Haenszel variance based on the unconditional distribution of the data that is both easily computed and consistent under each limiting model. Monte Carlo experiments compare its performance with that of the previously proposed estimators. The discussion attempts to clarify the relationship between the two limiting models.

2. Sparse Data

Consider a series of K 2×2 tables with the first row of each table formed by pairs of independent binomial observations (X_k, Y_k) with denominators (n_k, m_k), success probabil-

This work was funded in part by National Institute of Environmental Health Sciences Grants 5 P30 ES00002 and 5 R23 ES03405, and U.S. Public Health Service Grant 1-RO1-CA40644 from the National Cancer Institute.

Key words: Case–control studies; Cox proportional hazards model; Mantel–Haenszel estimator; Odds ratio.

ities (p_{1k}, p_{0k}), $k \in \{1, \ldots, K\}$, and common odds ratio ψ. We assume that as $K \to \infty$ only a finite number of different denominator configurations (n_k, m_k) occur (Model II). This would be the appropriate limiting model for a matched or finely stratified case-control study in which n_k, m_k denote the number of cases and controls and X_k, Y_k the number exposed in each set. Let $N_k = n_k + m_k$, $t_k = X_k + Y_k$, $R_k = X_k(m_k - Y_k)/N_k$, $S_k = Y_k(n_k - X_k)/N_k$, and $N_+ = \sum_k N_k$. The Mantel–Haenszel estimator is $\tilde{\psi}_{\mathrm{MH}} = \sum_k R_k / \sum_k S_k$. Under mild restrictions on the sequences of nuisance parameters $\{p_{0k}\}$, $\tilde{\psi}_{\mathrm{MH}}$ will be asymptotically normal with asymptotic mean ψ. (Since R_k, S_k are bounded random variables, a sufficient condition is that ψ be finite and nonzero and furthermore that the p_{0k} do not converge toward 0 or 1 as $K \to \infty$.) Let $\mathrm{var}^{\mathrm{A}}$ and E^{A} represent asymptotic variances and expectations while var and E represent their exact counterparts conditional on p_{0k}. Writing $\tilde{\psi}_{\mathrm{MH}} - \psi = \sum (R_k - \psi S_k)/\sum S_k$ and noting that $\mathrm{E}(R_k - \psi S_k) = 0$, we can conclude from the central limit theorem and Slutsky's theorem that

$$\lim_{K\to\infty} K \,\mathrm{var}^{\mathrm{A}}(\tilde{\psi}_{\mathrm{MH}}) = \frac{\lim_{K\to\infty} \sum_{k=1}^{K} \mathrm{var}(R_k - \psi S_k)/K}{[\lim_{K\to\infty} \sum_{k=1}^{K} \mathrm{E}(S_k)/K]^2}. \tag{1}$$

Note that if the p_{0k} were independent and identically distributed with common distribution function F, then the asymptotic normality of $\tilde{\psi}_{\mathrm{MH}}$ would hold both conditional on p_{0k} and unconditionally. Furthermore, since $\mathrm{E}(R_k - \psi S_k \mid p_{0k}) = 0$, by the weak law of large numbers the above variance expression would hold in either case. Thus, our results cover both random and fixed sequences $\{p_{0k}\}$.

Breslow (1981) proposed both a "conditional" estimator of the Mantel–Haenszel variance defined by

$$K\, \widetilde{\mathrm{var}}_{\mathrm{C}}(\tilde{\psi}_{\mathrm{MH}}) = \frac{\sum_k \mathrm{var}(R_k - \psi S_k \mid t_k; \psi)/K}{[\sum_k \mathrm{E}(S_k \mid t_k; \psi)/K]^2},$$

where the estimator $\tilde{\psi}_{\mathrm{MH}}$ is substituted for the unknown ψ in actual computation, and an empirical estimator given by

$$K\, \widetilde{\mathrm{var}}_{\mathrm{E}}(\tilde{\psi}_{\mathrm{MH}}) = \frac{\sum_k (R_k - \tilde{\psi}_{\mathrm{MH}} S_k)^2/K}{(\sum_k S_k/K)^2}.$$

This paper proposes two new estimators of the Mantel–Haenszel variance defined by

$$K\, \widetilde{\mathrm{var}}_{\mathrm{U}}(\tilde{\psi}_{\mathrm{MH}}) = \frac{\sum_k \tilde{\phi}_{\mathrm{U}_k}(\tilde{\psi}_{\mathrm{MH}})/K}{(\sum_k S_k/K)^2},$$

where

$$\tilde{\phi}_{\mathrm{U}}(\psi) = \frac{mn}{N^2}\left\{\frac{\psi(n - X)Y}{n}\left[1 + (\psi - 1)\frac{Y}{m}\right] + \frac{X(m - Y)}{m}\left[\psi - (\psi - 1)\frac{X}{n}\right]\right\}$$

and we drop the subscript k when referring to a single 2×2 table; and

$$K\, \widetilde{\mathrm{var}}_{\mathrm{US}}(\tilde{\psi}_{\mathrm{MH}}) = K\left[\frac{\sum_k P_k R_k}{2R_+^2} + \frac{\sum_k (P_k S_k + Q_k R_k)}{2R_+ S_+} + \frac{\sum_k Q_k S_k}{2S_+^2}\right](\tilde{\psi}_{\mathrm{MH}})^2,$$

where $P_k = (X_k + m_k - Y_k)/N_k$, $Q_k = (Y_k + n_k - X_k)/N_k$, $R_+ = \sum_k R_k$, and $S_+ = \sum_k S_k$.

The corresponding estimators of $K\,\mathrm{var}^{\mathrm{A}}(\ln \tilde{\psi}_{\mathrm{MH}})$ are

$$KV_i \equiv K\, \widetilde{\mathrm{var}}_i(\tilde{\psi}_{\mathrm{MH}})/(\tilde{\psi}_{\mathrm{MH}})^2, \quad i \in \{\mathrm{U}, \mathrm{US}, \mathrm{C}, \mathrm{E}\}.$$

In contrast to V_{C}, V_{US} and V_{U} are easily computed. In addition, we will show that, like V_{C} and in contrast with V_{E}, V_{U} and V_{US} are consistent under both large-strata and sparse data limiting models. V_U is not invariant under the interchange of rows in each 2×2 table. V_{US} is the arithmetic average of V_U and V_U recomputed upon interchange of rows and serves as a symmetrized version of V_U. V_{US} is also symmetric under interchange of rows with columns, i.e., $n_k \leftrightarrow t_k$, $m_k \leftrightarrow n_k + m_k - t_k$, $X_k \leftrightarrow X_k$, $Y_k \leftrightarrow n_k - X_k$.

We show all four estimators of $K\ \mathrm{var}^{\mathrm{A}}(\tilde{\psi}_{\mathrm{MH}})$ are consistent under limiting Model II. An estimator will be consistent if it can be written as the ratio of consistent estimators of the numerator and the denominator of (1). If ψ is known and we can find an unbiased estimator $\tilde{\phi}(\psi)$ of $\mathrm{var}(R - \psi S)$, then $\sum_k \tilde{\phi}_k(\psi)/K$ will be consistent for the numerator by the weak law of large numbers applied to sums of nonidentically distributed, bounded, and unbiased estimators. Furthermore, since $\tilde{\psi}_{\mathrm{MH}} \xrightarrow{P} \psi$, if $\sum_k \tilde{\phi}_k(\psi)/K$ is a smooth function of ψ, then $\sum_k \tilde{\phi}_k(\tilde{\psi}_{\mathrm{MH}})/K$ will be consistent as well.

$\tilde{\phi}_{\mathrm{C}}(\psi) \equiv \mathrm{var}(R - \psi S \mid t)$ is the exact variance of $R - \psi S$ under the noncentral hypergeometric distribution for X with parameters (n, m, t) and noncentrality parameter ψ. It is a smooth function of ψ, and $\mathrm{E}[\tilde{\phi}_{\mathrm{C}}(\psi)] = \mathrm{var}(R - \psi S)$ since $\mathrm{E}(R - \psi S \mid t) = 0$ (Breslow, 1981). Similarly, $\tilde{\phi}_{\mathrm{E}}(\psi) \equiv (R - \psi S)^2$ is unbiased for $\mathrm{var}(R - \psi S)$ since $\mathrm{E}(R - \psi S) = 0$. Furthermore, $\mathrm{E}(\tilde{\phi}_{\mathrm{E}}(\psi) \mid t] = \mathrm{var}(R - \psi S \mid t)$. Since $\mathrm{E}(S \mid t)$ and S are unbiased for $\mathrm{E}(S)$, it follows that $K\ \tilde{\mathrm{var}}_{\mathrm{C}}(\tilde{\psi}_{\mathrm{MH}})$ and $K\ \tilde{\mathrm{var}}_{\mathrm{E}}(\tilde{\psi}_{\mathrm{MH}})$ are consistent.

Now define $A = (1/N^2)(\psi - 1)^2 n p_1(1 - p_1) m p_0(1 - p_0)$ and

$$B = \frac{mn}{N^2}\{m[1 + (\psi - 1)p_0]\psi p_0(1 - p_1) + n[\psi - (\psi - 1)p_1]p_1(1 - p_0)\}.$$

$\tilde{\phi}_{\mathrm{U}}(\psi)$ was constructed by replacing p_1, p_0 by X/n, Y/m, respectively, whenever they occur in B.

Lemma 1. $\mathrm{E}[\tilde{\phi}_{\mathrm{U}}(\psi)] = \mathrm{var}(R - \psi S)$.

Proof.

$$\mathrm{var}(R - \psi S) = \mathrm{var}\{(1/N)[(\psi - 1)XY + mX - \psi nY]\} = A + B,$$

where we have used the independence of X and Y. Since $\tilde{\phi}_{\mathrm{U}}(\psi)$ is a Fisher-consistent estimator of B, it may appear moderately surprising that $\mathrm{E}[\tilde{\phi}_{\mathrm{U}}(\psi)] = A + B$. Fortuitously, the bias of $\tilde{\phi}_{\mathrm{U}}(\psi)$ in estimating B is exactly A, and this proves the lemma.

It follows that $K\ \tilde{\mathrm{var}}_{\mathrm{U}}(\tilde{\psi}_{\mathrm{MH}})$ and thus $K\ \tilde{\mathrm{var}}_{\mathrm{US}}(\tilde{\psi}_{\mathrm{MH}})$ are consistent.

The same arguments show that the four estimators are also consistent estimators of the Mantel-Haenszel variance given in equation (1) under a limiting Model III in which n_k is bounded but $m_k \to \infty$ as $K \to \infty$.

3. Large Strata

Under Model I, the number of tables K remains fixed but, for each k, $N_k \to \infty$ in such a way that n_k/N_k and N_k/N_+ approach nonzero limits l_k and r_k, where $N_+ = \sum_{k=1}^{K} N_k$.

Lemma 2. $N_+ \tilde{\mathrm{var}}_{\mathrm{U}}(\tilde{\psi}_{\mathrm{MH}})$ and $N_+ \tilde{\mathrm{var}}_{\mathrm{C}}(\tilde{\psi}_{\mathrm{MH}})$ consistently estimate $N_+ \mathrm{var}^{\mathrm{A}}(\tilde{\psi}_{\mathrm{MH}})$ under Model I but $N_+ \tilde{\mathrm{var}}_{\mathrm{E}}(\tilde{\psi}_{\mathrm{MH}})$ does not.

Proof. Hauck (1979) showed that

$$N_+ \mathrm{var}^{\mathrm{A}}(\tilde{\psi}_{\mathrm{MH}}) = \frac{\sum_{k=1}^{K} N_+ \mathrm{var}^{\mathrm{A}}[(R_k - \psi S_k)/N_+)]}{[\sum_{k=1}^{K} \mathrm{E}^{\mathrm{A}}(S_k/N_+)]^2} = \frac{\psi^2 \sum_k \mu_k^2/\omega_k}{(\sum_k \mu_k)^2}, \tag{2}$$

where the asymptotic expectations and variances are defined under Model I and where

$$\mu_k = r_k l_k(1 - l_k) q_{1k} p_{0k}, \quad \omega_k^{-1} = (r_k l_k p_{1k} q_{1k})^{-1} + [r_k(1 - l_k) p_{0k} q_{0k}]^{-1}, \quad \text{and} \quad q_{ik} = 1 - p_{ik}.$$

But

$$\frac{\psi^2\mu^2}{\omega} = \lim_{N_+\to\infty} \frac{B}{N_+}. \tag{3}$$

We should have expected equation (3) to hold since A/N_+ is $O(1/N_+)$ and B/N_+ is $O(1)$, and thus B/N_+ is the asymptotically nonnegligible portion of the exact $\mathrm{var}(R - \psi S)/N_+$. Since X/n, Y/m are consistent for p_1, p_0, it follows by the construction of $\tilde{\phi}_{\mathrm{U}}(\psi)$ that $\tilde{\phi}_{\mathrm{U}}(\psi)/N_+$ converges to $\lim_{N_+\to\infty} B/N_+$. Furthermore, since S_k/N_+ converges to μ_k and $\tilde{\phi}_{\mathrm{U}}(\psi)$ is a smooth function of ψ, $N_+\widetilde{\mathrm{var}}_{\mathrm{U}}(\tilde{\psi}_{\mathrm{MH}})$ and thus $N_+\widetilde{\mathrm{var}}_{\mathrm{US}}(\tilde{\psi}_{\mathrm{MH}})$ are consistent estimators of the asymptotic variance given in equation (2).

To show that $N_+\widetilde{\mathrm{var}}_{\mathrm{C}}(\tilde{\psi}_{\mathrm{MH}})$ is consistent under Model I, we proceed as follows. Since $\lim_{N_+\to\infty} B/N_+ = \lim_{N_+\to\infty} N_+\mathrm{var}[(R - \psi S)/N_+]$ and $\mathrm{E}[\tilde{\phi}_{\mathrm{C}}(\psi)] = \mathrm{var}(R - \psi S)$, it follows from the moment convergence theorem that, provided $\tilde{\phi}_{\mathrm{C}}(\psi)/N_+$ is bounded, a necessary and sufficient condition that $\tilde{\phi}_{\mathrm{C}}(\psi)/N_+$ consistently estimates $\lim_{N_+\to\infty} (B/N_+)$ is that $\tilde{\phi}_{\mathrm{C}}(\psi)/N_+$ converges to some constant in probability as $N_+ \to \infty$. Using the fact that $t(n + m) \xrightarrow{P} p_1 l + p_0(1 - l)$, one can show that $\tilde{\phi}_{\mathrm{C}}(\psi)/N_+$ converges to a constant in probability. It is shown to be bounded in Corollary 1 of the Appendix. Similar arguments applied to the denominator show that $N_+\widetilde{\mathrm{var}}_{\mathrm{C}}(\tilde{\psi}_{\mathrm{MH}})$ is consistent. On the other hand, $\tilde{\phi}_{\mathrm{E}}(\psi)/N_+$ fails to converge to a constant in probability as $N_+ \to \infty$, essentially because it does not take advantage of the fact that X, Y are sums of independent Bernoulli random variables. If $K = 1$, $N_+\widetilde{\mathrm{var}}_{\mathrm{E}}(\tilde{\psi}_{\mathrm{MH}}) = 0$.

4. Monte Carlo Experiments

Several Monte Carlo experiments were conducted to compare the performance of various estimators of the variance of $\ln(\tilde{\psi}_{\mathrm{MH}})$. As noted by Breslow and Liang (1982), the distribution of $\tilde{\psi}_{\mathrm{MH}}$ can be quite skew in moderate size samples and inference on ψ is better based on approximating the distribution of $\ln(\tilde{\psi}_{\mathrm{MH}})$ by the normal. We included in the experiment the four variance estimators defined earlier and V_{H}, V_{BL}, V_{HG} as defined below. First,

$$V_{\mathrm{H}} = \frac{(\sum_k S_k^2/w_k)}{(\sum_k S_k)^2}$$

with

$$w_k = \left(\frac{1}{X_k + .5} + \frac{1}{n_k - X_k + .5} + \frac{1}{m_k - Y_k + .5} + \frac{1}{Y_k + .5}\right)^{-1}$$

is the estimator originally proposed by Hauck, except Hauck did not add .5 to each cell when computing w. Since V_{H} is not invariant under interchange of rows or columns (Ury, 1982), Breslow and Liang (1982) proposed that a symmetrized version

$$V_{\mathrm{HS}} = \left[\frac{(\sum_k S_k^2/w_k)(\sum_k R_k^2/w_k)}{(\sum_k S_k)^2(\sum_k R_k)^2}\right]^{1/2}$$

be used in their "combined" estimator

$$V_{\mathrm{BL}} = \frac{N_+V_{\mathrm{HS}} + K^2V_{\mathrm{E}}}{N_+ + K^2}.$$

V_{H} and V_{HS} are inconsistent under limiting Models II and III (Breslow and Liang, 1982). V_{BL} is consistent under limiting Models I and II but will be inconsistent under limiting Model III if m_k/K is bounded away from 0 for $k \in \{1, \ldots, K\}$.

Gart (reported by Ury, 1982) suggested another symmetrized version of the original Hauck estimator wherein the data X_k, Y_k in the kth table are replaced by fitted values whose marginal totals agree with those observed and whose odds ratio equals $\tilde{\psi}_{\mathrm{MH}}$. Since both V_{U} and Hauck's original estimator are Fisher-consistent estimators of $\mathrm{var}^{\mathrm{A}}(\tilde{\psi}_{\mathrm{MH}})$ under Model I and, by definition, all Fisher-consistent estimators must agree for data outcomes in which $\tilde{\psi}$ is constant across tables, it follows that calculating either Hauck's estimator or V_{U} from these fitted values results in the same statistic, which we denote V_{HG}. Since $\tilde{\phi}_{\mathrm{U}}(\psi)$ is a biased estimate of $\mathrm{var}(R - \psi S)$ when X and Y are replaced by their fitted values, V_{HG} is inconsistent under Model II. Nonetheless, as we show below, it is consistent under limiting Model III.

We performed two sets of Monte Carlo experiments. Each set determined an incomplete factorial design with 1000 Monte Carlo trials per experiment. We chose 1000 trials in order to estimate the actual coverage rate of 90% and 95% confidence intervals to within several percent. The first set, summarized in Tables 1–4, was composed of component experiments using values of p_{1k} and p_{0k} nearly identical to those reported in Breslow and Liang (1982), namely $p_{1k} = .3 + .005k(100/K)$ for $K \in \{1600, 400, 100, 50, 25\}$, $k \in \{1, \ldots, K\}$, and $p_{1k} = .05 + .04k(20/K)$ for $K \in \{20, 10, 5\}$ and with the values of ψ, n, m shown in the tables. The second set of experiments used the following levels of each factor: $\psi = 1, 3, 6.5$; $n = 1, 2, 8$; $m = 2, 8, 80$ with $n \leqslant m$; $K \in \{50, 25, 15, 5\}$; $p_{0k} = .2 + .1[k \bmod 5]$, $k \in \{1, \ldots, K\}$. In order to save space, we report results in Tables 1–4 only for those experiments in the second set that provide qualitatively different information from that obtained from the first set. The Monte Carlo results reported in Breslow and Liang (1982) were in error (Breslow and Liang, 1984). The results reported here serve to correct their Tables 2 and 3.

When $m = 80$ we did not compute V_{C} but considered V_{HG} as an approximation to V_{C} in order to save computing costs. This is justified theoretically by the fact that as m and thus t, $N \to \infty$, n bounded, $\tilde{\phi}_{\mathrm{HG}}(\psi)$ and $\tilde{\phi}_{\mathrm{C}}(\psi)$ converge in probability to $\lim_{N\to\infty} B$ since A/B is $o(1)$, where $\phi_{\mathrm{HG}}(\psi)$ is $\phi_{\mathrm{U}}(\psi)$ with X and Y replaced by their fitted values. It follows that $|\tilde{\phi}_{\mathrm{HG}}(\tilde{\psi}_{\mathrm{MH}}) - \tilde{\phi}_{\mathrm{C}}(\tilde{\psi}_{\mathrm{MH}})|$ converges to 0 in probability. Similar remarks apply to the denominators of V_{HG} and V_{C}. The above argument implies that V_{HG} is consistent under limiting Model III. Several numerical examples showed that V_{C} and V_{HG} differed by no more than 1% to 2% when $m = 80$, $k = 1$, $n = 1$, $p_0 = .2$, $\psi = 1$. Similarly, when $n = 30$ and $m = 30$ we did not compute V_{C}.

Binomial random variables were generated using IMSL subroutine GGBN on an IBM 4341. When $\sum_k R_k = 0$ or $\sum_k S_k = 0$ in a given trial, we completely ignored the trial for estimation of $\mathrm{var}(\ln \tilde{\psi}_{\mathrm{MH}})$. Nonetheless we considered such a trial to represent a failure of the confidence interval to cover the true parameter. Finally, we accumulated the 1000 trials in 20 batches of 50 trials so that standard errors of empirical standard errors could be estimated.

For each estimator V_i we computed both the average of $\sqrt{V_i}$, i.e., $\overline{\sqrt{V_i}}$, over the 1000 trials, and the Monte Carlo variance, $\widetilde{\mathrm{var}}(\sqrt{V_i})$, as an indicator of the efficiency of $\sqrt{V_i}$. In addition, we computed the actual coverage rates of nominal 90% $[\ln(\tilde{\psi}_{\mathrm{MH}}) \pm 1.64 \sqrt{V_i}]$ and 95% $[\ln(\tilde{\psi}_{\mathrm{MH}}) \pm 1.96 \sqrt{V_i}]$ confidence intervals for $\ln \psi$. The Monte Carlo standard error of the simulated standard error of $\ln(\tilde{\psi}_{\mathrm{MH}})$, i.e., $\widetilde{\mathrm{se}}\{\widetilde{\mathrm{se}}[\ln(\tilde{\psi}_{\mathrm{MH}})]\}$, was computed from the empirical between-batch variability of the within-batch empirical standard error of $\ln(\tilde{\psi}_{\mathrm{MH}})$. Since $\widetilde{\mathrm{se}}\{\widetilde{\mathrm{se}}[\ln(\tilde{\psi}_{\mathrm{MH}})]\}$ exceeded the Monte Carlo standard error of $\overline{\sqrt{V_{\mathrm{US}}}}$ or $\overline{\sqrt{V_{\mathrm{C}}}}$ by nearly tenfold in all experiments, we report only $\widetilde{\mathrm{se}}[\ln(\tilde{\psi}_{\mathrm{MH}})] \pm 2\,\widetilde{\mathrm{se}}\{\widetilde{\mathrm{se}}[\ln(\tilde{\psi}_{\mathrm{MH}})]\}$ as an indication of the uncertainty remaining when assessing the bias of $\sqrt{V_i}$.

Tables 1–4 summarize the results. V_{U} and V_{US} gave similar results. Therefore, only V_{US} is reported. With $\psi \leqslant 3.5$, $\ln(\tilde{\psi}_{\mathrm{MH}})$ was essentially unbiased (data not shown). In these instances V_{US} and V_{C} were unbiased and gave nominal coverage rates to within Monte

Table 1
Mean and variance of six estimators of the variance of the log Mantel–Haenszel estimator, $\theta = \ln(\psi)$

n	m	K	$\widetilde{\mathrm{se}}(\tilde{\theta}_{MH}) \pm 2\ \widetilde{\mathrm{se}}[\widetilde{\mathrm{se}}(\tilde{\theta}_{MH})]$	$\overline{\sqrt{V_E}}$	$10^3 \times \widetilde{\mathrm{var}}(\sqrt{V_E})$	$\overline{\sqrt{V_H}}$	$10^3 \times \widetilde{\mathrm{var}}(\sqrt{V_H})$	$\overline{\sqrt{V_{BL}}}$	$10^3 \times \widetilde{\mathrm{var}}(\sqrt{V_{BL}})$	$\overline{\sqrt{V_{HG}}}$	$10^3 \times \widetilde{\mathrm{var}}(\sqrt{V_{HG}})$	$\overline{\sqrt{V_{US}}}$	$10^3 \times \widetilde{\mathrm{var}}(\sqrt{V_{US}})$	$\overline{\sqrt{V_C}}$	$10^3 \times \widetilde{\mathrm{var}}(\sqrt{V_C})$
								$\psi = 1$, **Set 1**							
1	1	100	.307 ± .019	.301	.340	.493	2.450	.306	.349	.413	2.490	.301	.340	.301	.340
1	2	100	.263 ± .016	.259	.108	.403	.988	.264	.112	.316	.294	.259	.137	.259	.109
		50	.372 ± .023	.369	.409	.573	3.490	.382	.431	.449	.639	.370	.555	.370	.417
1	4	50	.343 ± .021	.335	.145	.480	2.090	.349	.156	.374	.157	.337	.257	.336	.114
		25	.503 ± .031	.478	.810	.683	9.400	.512	.869	.534	.821	.484	1.390	.481	.878
1	8	50	.317 ± .019	.317	.111	.421	1.731	.332	.116	.337	.106	.319	.223	.318	.116
		25	.465 ± .028	.453	.386	.600	5.392	.487	.379	.481	.376	.457	.807	.455	.444
5	5	20	.335 ± .021	.309	1.735	.405	2.379	.344	1.483	.339	.181	.323	.217	.322	.174
		10	.474 ± .029	.422	7.007	.560	6.756	.496	5.001	.482	.792	.462	.910	.459	.776
		20	.179 ± .011	.178	1.139	.203	.493	.194	.512	.187	.054	.184	.060	.174	.044
15	15	10	.256 ± .015	.241	2.875	.284	.551	.274	.662	.265	.061	.261	.064	.244	.021
		5	.369 ± .022	.311	12.06	.397	1.885	.387	1.975	.376	.278	.372	.286	.340	.066
		20	.129 ± .008	.125	1.047	.136	.112	.134	.152	.131	.002	.130	.002	—	—
30	30	10	.186 ± .011	.170	1.758	.192	.008	.189	.132	.185	.001	.184	.001	—	—
		5	.251 ± .015	.222	6.960	.271	.336	.268	.359	.263	.006	.261	.006	—	—
								$\psi = 1$, **Set 2**							
2	8	25	.353 ± .014	.333	1.136	.421	1.223	.366	1.044	.358	.142	.342	.231	.340	.154
2	80	25	.315 ± .014	.304	2.320	.354	.831	.353	1.012	.310	.131	.309	.171	—	—

Table 2
Mean and variance of six estimators of the variance of the log Mantel–Haenszel estimator, $\theta = \ln(\psi)$

n	m	K	$\widetilde{\mathrm{se}}(\tilde{\theta}_{\mathrm{MH}}) \pm 2\,\widetilde{\mathrm{se}}[\widetilde{\mathrm{se}}(\tilde{\theta}_{\mathrm{MH}})]$	$\overline{\sqrt{V_{\mathrm{E}}}}$	$10^3 \times \widetilde{\mathrm{var}}(\sqrt{V_{\mathrm{E}}})$	$\overline{\sqrt{V_{\mathrm{H}}}}$	$10^3 \times \widetilde{\mathrm{var}}(\sqrt{V_{\mathrm{H}}})$	$\overline{\sqrt{V_{\mathrm{BL}}}}$	$10^3 \times \widetilde{\mathrm{var}}(\sqrt{V_{\mathrm{BL}}})$	$\overline{\sqrt{V_{\mathrm{HG}}}}$	$10^3 \times \widetilde{\mathrm{var}}(\sqrt{V_{\mathrm{HG}}})$	$\overline{\sqrt{V_{\mathrm{US}}}}$	$10^3 \times \widetilde{\mathrm{var}}(\sqrt{V_{\mathrm{US}}})$	$\overline{\sqrt{V_{\mathrm{C}}}}$	$10^3 \times \widetilde{\mathrm{var}}(\sqrt{V_{\mathrm{C}}})$
							ψ = 3.5, Set 1								
1	1	100	.374 ± .023	.356	2.426	.729	16.38	.360	2.393	.412	3.860	.356	2.426	.356	2.426
1	2	100	.286 ± .017	.288	.602	.518	4.333	.293	.594	.331	.885	.289	.640	.290	.632
		50	.434 ± .026	.414	2.920	.741	18.79	.426	2.816	.473	2.365	.417	3.278	.418	3.399
1	4	50	.360 ± .022	.361	1.110	.567	5.660	.374	.743	.399	.636	.364	.887	.363	.812
		25	.535 ± .033	.516	3.899	.808	25.38	.548	3.535	.573	3.406	.526	5.050	.526	5.335
1	8	50	.340 ± .021	.333	.489	.474	3.649	.347	.445	.355	.405	.335	.642	.334	.572
		25	.504 ± .031	.473	1.500	.672	10.64	.508	1.262	.507	1.498	.480	2.178	.479	2.048
5	5	20	.358 ± .022	.347	2.899	.462	5.115	.379	2.340	.373	.601	.362	.826	.362	.786
		10	.514 ± .031	.471	10.79	.637	12.88	.540	7.247	.529	2.609	.517	3.543	.516	3.514
		20	.196 ± .012	.196	1.609	.220	.868	.213	.754	.206	.178	.204	.230	.199	.215
15	15	10	.238 ± .017	.263	4.187	.306	.806	.300	1.118	.290	.220	.287	.248	.278	.177
		5	.410 ± .025	.332	13.89	.421	1.918	.416	2.731	.407	.936	.404	1.042	.386	.642
		20	.139 ± .008	.136	1.236	.148	.119	.147	.219	.144	.057	.143	.063	—	—
30	30	10	.199 ± .012	.183	2.061	.207	.082	.207	.209	.202	.045	.202	.049	—	—
		5	.283 ± .017	.236	7.258	.288	.282	.289	.573	.284	.193	.283	.212	—	—
							ψ = 6.5, Set 1								
1	1	400	.154 ± .019	.158	.057	.326	.426	.158	.057	.189	.012	.158	.057	.158	.057
1	1	1600	.096 ± .017	.102	.013	.220	.087	.102	.013	.105	.203	.102	.013	.102	.013

Table 3
Empirical coverage rates of nominal 95% [$\ln(\tilde{\psi}_{MH}) \pm 1.96\sqrt{V_i}$] and 90% [$\ln(\tilde{\psi}_{MH}) \pm 1.64\sqrt{V_i}$] confidence intervals for six variance estimators of the log Mantel–Haenszel estimator

			V_E		V_H		V_{BL}		V_{HG}		V_{US}		V_C	
n	m	K	95	90	95	90	95	90	95	90	95	90	95	90
							$\psi = 1$, Set 1							
1	1	100	.951	.905	.992	.982	.951	.913	.979	.963	.951	.905	.951	.905
1	2	100	.951	.897	.996	.985	.953	.906	.983	.951	.950	.901	.951	.898
		50	.951	.906	.990	.976	.959	.915	.983	.950	.956	.910	.952	.906
1	4	50	.946	.906	.991	.969	.954	.916	.969	.927	.946	.904	.947	.906
		25	.954	.898	.984	.963	.964	.922	.972	.933	.957	.904	.959	.901
1	8	50	.955	.908	.988	.966	.964	.915	.966	.917	.958	.908	.957	.906
		25	.955	.906	.977	.959	.966	.925	.963	.923	.958	.911	.956	.909
5	5	20	.921	.871	.975	.943	.944	.900	.952	.905	.941	.882	.941	.879
		10	.902	.841	.983	.938	.947	.897	.962	.913	.951	.901	.945	.898
		20	.940	.885	.977	.930	.966	.920	.960	.908	.958	.900	.942	.891
15	15	10	.913	.841	.970	.927	.964	.923	.959	.914	.957	.907	.940	.885
		5	.851	.795	.960	.926	.957	.919	.956	.907	.952	.905	.925	.878
		20	.931	.869	.963	.921	.961	.912	.959	.909	.958	.907	—	—
30	30	10	.894	.833	.960	.907	.955	.900	.951	.898	.950	.894	—	—
		5	.844	.785	.967	.918	.963	.910	.963	.915	.962	.910	—	—
							$\psi = 1$, Set 2							
2	8	25	.942	.871	.993	.981	.961	.911	.960	.904	.951	.893	.950	.884
2	80	25	.951	.882	.972	.931	.971	.937	.953	.902	.949	.894	—	—

Table 4

Empirical coverage rates of nominal 95% $[\ln(\tilde{\psi}_{MH}) \pm 1.96\sqrt{V_i}]$ *and* 90% $[\ln(\tilde{\psi}_{MH}) \pm 1.64\sqrt{V_i}]$ *confidence intervals for six variance estimators of the log Mantel–Haenszel estimator*

			V_E		V_H		V_{BL}		V_{HG}		V_{US}		V_C	
n	m	K	95	90	95	90	95	90	95	90	95	90	95	90
							$\psi = 3.5$, **Set 1**							
1	1	100	.955	.894	.999	.999	.958	.899	.966	.918	.955	.894	.955	.894
1	2	100	.962	.906	1.000	.998	.966	.909	.980	.943	.966	.903	.966	.905
		50	.946	.902	.998	.994	.955	.907	.972	.927	.954	.903	.956	.905
1	4	50	.953	.908	.995	.990	.963	.918	.974	.938	.959	.915	.960	.911
		25	.948	.903	.989	.977	.963	.918	.967	.930	.960	.915	.961	.917
1	8	50	.950	.897	.993	.975	.961	.908	.966	.916	.956	.900	.953	.899
		25	.945	.888	.983	.964	.960	.911	.958	.913	.951	.900	.949	.901
5	5	20	.935	.873	.985	.959	.952	.909	.963	.917	.959	.905	.958	.906
		10	.927	.873	.986	.962	.964	.921	.972	.928	.968	.917	.968	.919
		20	.936	.887	.972	.932	.963	.923	.967	.909	.964	.907	.964	.907
15	15	10	.914	.845	.969	.919	.962	.907	.952	.898	.952	.896	.944	.883
		5	.839	.765	.965	.908	.959	.898	.955	.897	.955	.893	.944	.878
		20	.930	.876	.960	.913	.955	.908	.953	.906	.952	.905	—	—
30	30	10	.910	.847	.964	.913	.963	.911	.957	.902	.956	.903	—	—
		5	.845	.793	.955	.904	.955	.902	.952	.897	.952	.896	—	—
							$\psi = 6.5$, **Set 1**							
1	1	400	.137	.083	.775	.579	.137	.083	.233	.125	.137	.083	.137	.083
1	1	1600	.950	.891	.999	.999	.950	.891	.933	.891	.950	.891	.950	.891

Carlo sampling variation. When $K < 20$, V_E was biased downward and failed to reach nominal coverage rates. In sparse data V_H was biased upward and gave excessive coverage rates. In sparse data when m was small, V_{HG} was biased upward and gave excessive coverage rates. In sparse data when m was large, V_{BL} was biased upward and gave excessive coverage rates. It follows that V_{BL} recomputed with V_{HG} substituted for V_{HS} in the expression for V_{BL} would have performed nearly as well as V_C or V_{US}.

Although V_C was in most experiments more efficient than V_U or V_{US}, in other experiments (e.g., $n = 1$, $m = 2$, $K = 50$, $\psi = 3.5$), the efficiency rankings were reversed. It may be somewhat surprising that V_U can ever be as efficient as V_C since, with ψ known, $\tilde{\phi}_C(\psi)$ and $E(S \mid t)$ are, respectively, the minimum variance unbiased estimators of $\text{var}(R - \psi S)$ and $E(S)$. Presumably in sparse data the variability in $\tilde{\psi}_{MH}$ and correlations between numerator and denominator can destroy any significant benefit that might accrue from using V_C, which depends only on $\tilde{\psi}_{MH}$ and t, in lieu of V_U.

When $\psi = 6.5$, $\ln(\tilde{\psi}_{MH})$ can be biased in sparse data unless K is quite large (Jewell, 1984). For instance, with $n = 1$, $m = 1$, $K = 400$, the bias of $\ln(\tilde{\psi}_{MH})$ was .40. When $K = 1600$, $\ln(\tilde{\psi}_{MH})$ was unbiased. When all values of p_{1k} and p_{0k} were between .1 and .9, $\ln(\tilde{\psi}_{MH})$ was nearly unbiased for $n = 1$, $m = 1$, $K = 100$, $\psi = 6.5$.

5. Estimation of $\text{var}^A(\tilde{\psi}_{MH})$ in the Cox Proportional Hazards Model

Assume (n_k, m_k, t_k) are fixed constants for $k \in \{1, \ldots, K\}$ and X_k is distributed noncentral hypergeometric with noncentrality parameter ψ such that n_k is bounded as $K \to \infty$. $\tilde{\psi}_{MH}$ is still asymptotically normal with the asymptotic mean ψ (Breslow, 1981). Previously we noted that $E[\tilde{\phi}_i(\psi) \mid t] = \text{var}(R - \psi S \mid t)$, $i \in \{C, E\}$. Surprisingly, $E[\phi_U(\psi) \mid t] = \text{var}(R - \psi S \mid t)$ even though t is fixed rather than being the sum of independent binomial random variables. We established this result for the special case $t = 1$. A referee kindly supplied the proof given in Theorem 1 of the Appendix for arbitrary t. By arguments essentially identical to those given in Section 2, the variance estimators defined there will be consistent in the above hypergeometric limiting model for all t. We now apply these results to the evaluation of estimators of $\text{var}^A(\ln \tilde{\psi}_{MH})$ in the Cox proportional hazards model.

Consider a cohort survival study in continuous time composed of J strata with n_s exposed (E) and m_s unexposed $(\bar{E})$ individuals in stratum j. With $\lambda_j(u \mid \cdot)$ being the hazard in stratum j at time u, assume $\lambda_j(u \mid E) = \psi\lambda_j(u \mid \bar{E})$ and that an independent censoring mechanism is operating. Let $u_{1j} < u_{lj} < u_{L_j j}$ be the ordered failure times in stratum j. Consider the set of 2×2 tables with n_{lj}, m_{lj} the number of exposed and unexposed individuals at risk just prior to u_{lj} and with $X_{lj} + Y_{lj} = 1 = t_{lj}$ where $X_{lj} = 1$ if an exposed individual failed at u_{lj} and $X_{lj} = 0$ otherwise. Since correct large-sample inference concerning ψ is obtained by treating $(X_{lj} \mid n_{lj}, m_{lj}, t_{lj})$ as independent noncentral hypergeometric random variables with noncentrality parameter ψ (Crowley, Liu, and Voelkel, 1982), it follows that regardless of the orientation of the 2×2 tables, V_U, V_{US}, V_C, and V_E will be consistent for $\text{var}^A(\ln \tilde{\psi}_{MH})$ both in a limiting model in which $J \to \infty$, but n_j, m_j remain bounded, and in a model in which J remains bounded but n_j, $m_j \to \infty$. These results continue to hold under a sampling scheme in which the case and a random sample of noncases are sampled without replacement from each risk set (Oakes, 1981). In the limiting model in which J remains bounded, if the entire risk set is sampled, $m_{lj} \to \infty$ for all l, j and therefore V_{HG} will be consistent but V_{BL} will not.

If ties occur (i.e., if t_{lj} can exceed 1), then under the discrete time logistic model discussed by Cox (1972), $\ln \tilde{\psi}_{MH}$ is consistent for the logistic parameter. Furthermore, under either risk set sampling plan defined above and regardless of table orientation, V_U, V_{US}, and V_C will be consistent.

6. Some Examples

Table 5 presents an empirical comparison of the four variance estimators that we believe will have continuing value for applications. The first two data sets are examples of matched sets to which the estimators V_C, V_{US}, or V_E would ordinarily be applied. Note that all three of these variances are identical for 1:1 pair matching. Even V_{HG}, which we know to be inconsistent with such sparse data, yields numerical results that are reasonably close in this example. Simplified computing formulas are available when V_E is applied to matched sets (Connett et al., 1982; Fleiss, 1984).

Table 5
Empirical comparison of four standard errors for the log Mantel–Haenszel estimator

Data set	K	N_+	$\log \tilde{\psi}_{MH}$	$\sqrt{V_C}$	$\sqrt{V_{US}}$	$\sqrt{V_E}$	$\sqrt{V_{HG}}$
Leisure World 1:1 matched pairs[a]	63	126	2.2687	.6065	.6065	.6065	.5826
Leisure World 1:4 matched sets[b]	63	315	2.1355	.4516	.4635	.4673	.4408
Ille-et Vilaine[c]	6	975	1.6405	.1892	.1888	.2052	.1897
Avadex[d]	4	317	1.1246	.4028	.4057	.1977	.4066

[a] Breslow and Day (1980, p. 167).
[b] Ibid (p. 174). See also Connett et al. (1982).
[c] Ibid (p. 137).
[d] Gart (1971, Table 2).

The next two data sets illustrate the type of large strata to which V_C, V_{US}, or V_{HG} would ordinarily be applied. The Ille-et-Vilaine data consist of 975 subjects classified into six 2×2 tables defined by age, while the Avadex data consist of 317 mice grouped into four 2×2 tables on the basis of sex and strain. The bias in V_E is quite evident, especially with the Avadex data, but the other three are all close.

In summary, one could recommend the use of V_U, V_{US}, or V_C for all types of data; V_{HG} when $m > 8$ or so in all tables; and V_E for matched sets or finely stratified data. V_{US}, in particular, is robust, symmetric, and easily computed. Variance estimators analogous to V_{US} for the Mantel–Haenszel risk and rate ratios and risk and rate differences are given by Greenland and Robins (1985). A variance estimator analogous to V_C for the Mantel–Haenszel rate ratio is given by Breslow (1984).

[*Note added in proof:* It has come to our attention that Flanders (1985, *Biometrics* **41,** 637–642) has published an estimator of the Mantel–Haenszel variance similar to V_U. Given this similarity, one would expect there to be little difference in performance. This conjecture appears to be borne out by Flanders' simulations.]

Résumé

Cet article propose un nouvel estimateur facile à calculer et convergent à la fois dans les modèles limites à "données rares" ou à "fortes strates" de la variance de l'estimateur de Mantel–Haenszel du classique rapport des probabilités. Ses performances sont comparées, à l'aide d'une méthode de Monte-Carlo, à celles d'autres estimateurs déjà proposés.

References

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 Day, N. E. (1980). *Statistical Methods in Cancer Research* I: *The Analysis of Case–Control Studies.* Lyon: International Agency for Research on Cancer.

Breslow, N. E. and Liang, K. Y. (1982). The variance of the Mantel–Haenszel estimator. *Biometrics* **38,** 943–952.
Breslow, N. E. and Liang, K. Y. (1984). Correction to "The variance of the Mantel–Haenszel estimator." *Biometrics* **40,** 1217.
Connett, J., Ejigou, A., McHugh, R., and Breslow, N. (1982). The precision of the Mantel–Haenszel estimator in case–control studies with multiple matching. *American Journal of Epidemiology* **116,** 875–877.
Cox, D. R. (1972). Regression models and life-tables (with Discussion). *Journal of the Royal Statistical Society, Series B* **34,** 187–220.
Crowley, J. C., Liu, P. Y., and Voelkel, J. G. (1982). Estimation of the ratio of hazard functions. In *Survival Analysis,* IMS Lecture Notes-Monograph Series II, R. A. Johnson and J. C. Crowley (eds), 56–73. Hayward, California: The Institute of Mathematical Statistics.
Davis, L. J. (1985). Generalization of the Mantel–Haenszel estimator to nonconstant odds ratios. *Biometrics* **41,** 487–495.
Fleiss, J. L. (1984). The Mantel–Haenszel estimator in case–control studies with varying numbers of controls matched to each case. *American Journal of Epidemiology* **120,** 1–3.
Gart, J. J. (1971). The comparison of proportions: A review of significance tests, confidence intervals, and adjustments for stratification. *Review of the International Statistical Institute* **39,** 148–169.
Greenland, S. and Robins, J. M. (1985). Estimation of a common effect parameter from sparse follow-up data. *Biometrics* **41,** 55–68.
Hauck, W. W. (1979). The large-sample variance of the Mantel–Haenszel estimator of a common odds ratio. *Biometrics* **35,** 817–819.
Jewell, N. P. (1984). Small-sample bias of point estimators of the odds ratio from matched sets. *Biometrics* **40,** 421–435.
Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* **22,** 719–748.
Oakes, D. (1981). Survival times: Aspects of partial likelihood. *International Statistical Review* **49,** 235–264.
Ury, H. K. (1982). Hauck's approximate large-sample variance of the Mantel–Haenszel estimator. *Biometrics* **38,** 1094–1095.

Received November 1984; *revised September* 1985.

Appendix

Define $a_{11} = X$, $a_{12} = Y$, $a_{21} = n - X$, $a_{22} = m - Y$.

Lemma. For any function g, defined on the integers,

$$E\{g(a_{11}, a_{12}, a_{21}, a_{22})a_{11}a_{22} \mid t\} = \psi E\{g(a_{11} + 1, a_{12} - 1, a_{21} - 1, a_{22} + 1)a_{12}a_{21} \mid t\}.$$

Proof. A proof is given in the Appendix of Davis (1985).

Theorem 1. $E(\tilde{\phi}_U(\psi) \mid t) = \operatorname{var}(R - \psi S \mid t)$.

Proof. Now, from the above lemma,

$$\begin{aligned}
&E\{\tilde{\phi}_U(\psi) \mid t\}\\
&= N^{-2}[m\psi E(a_{12}a_{21} \mid t) + (\psi - 1)\psi E(a_{12}^2 a_{21} \mid t) + n\psi E(a_{11}a_{22} \mid t) - (\psi - 1)E(a_{11}^2 a_{22} \mid t)]\\
&= N^{-2}\{mE(a_{11}a_{22} \mid t) + (\psi - 1)E[(a_{12} + 1)a_{11}a_{22} \mid t] + n\psi E(a_{11}a_{22} \mid t) - (\psi - 1)E(a_{11}^2 a_{22} \mid t)\}\\
&= N^{-2}E\{a_{11}a_{22}[m + (\psi - 1)(m - a_{22} + 1) + n\psi - (\psi - 1)a_{11}] \mid t\}\\
&= N^{-2}E\{a_{11}a_{22}[N\psi + (1 - \psi)(a_{11} + a_{22} - 1)] \mid t\}.
\end{aligned}$$

Also, using the lemma, we have

$$\begin{aligned}
\operatorname{var}(R - \psi S \mid t) &= \mathrm{E}[(R - \psi S)^2 \mid t] \\
&= N^{-2}[\mathrm{E}(a_{11}^2 a_{22}^2 \mid t) - 2\psi \mathrm{E}(a_{11}a_{12}a_{21}a_{22} \mid t) + \psi^2 \mathrm{E}(a_{12}^2 a_{21}^2 \mid t)] \\
&= N^{-2}\{\mathrm{E}(a_{11}^2 a_{22}^2 \mid t) - 2\psi \mathrm{E}(a_{11}a_{12}a_{21}a_{22} \mid t) + \psi \mathrm{E}[(a_{12} + 1)(a_{21} + 1)a_{11}a_{22} \mid t]\} \\
&= N^{-2}\{\mathrm{E}(a_{11}^2 a_{22}^2 \mid t) - \psi \mathrm{E}(a_{11}a_{12}a_{21}a_{22} \mid t) + \psi \mathrm{E}[(a_{12} + a_{21} + 1)a_{11}a_{22} \mid t]\} \\
&= N^{-2}\{\mathrm{E}(a_{11}^2 a_{22}^2 \mid t) - \mathrm{E}\{(a_{11} - 1)(a_{22} - 1)a_{11}a_{22} \mid t\} + \psi \mathrm{E}[(m - a_{22} + n - a_{11} + 1)a_{11}a_{22} \mid t]\} \\
&= N^{-2}\mathrm{E}\{a_{11}a_{22}[a_{11} + a_{22} - 1 + N\psi - \psi(a_{11} + a_{22}) + \psi] \mid t\} \\
&= N^{-2}\mathrm{E}\{a_{11}a_{22}[N\psi + (1 - \psi)(a_{11} + a_{22} - 1)] \mid t\}.
\end{aligned}$$

Therefore, $\mathrm{E}(\tilde{\phi}_{\mathrm{U}}(\psi) \mid t) = \operatorname{var}(R - \psi S \mid t)$.

Corollary 1. $\operatorname{var}(R - \psi S \mid t)/N_+$ is bounded for all N_+.

Proof. It is straightforward to check from its definition that $\tilde{\phi}_{\mathrm{U}}(\psi)/N_+$ is bounded and thus so is its expectation.