Asymptotic efficiency of general noniterative estimators of common relative risk

BY MARKKU NURMINEN

Department of Epidemiology and Biometry, Institute of Occupational Health, Helsinki, Finland

SUMMARY

Gart (1962) presented consistent and efficient estimators of the common relative risk for several $2 \times 2$ tables, together with their associated confidence intervals. A theorem on the asymptotic efficiency of general, weighted noniterative point estimators is formulated and proved. A modification of interval estimators based on log normality is given.

Some key words: Combination of $2 \times 2$ tables; Consistency; Homogeneity of relative risk; Interval estimation; Log normal distribution; Minimum variance.

1. INTRODUCTION

In practice, we are often faced with the problem of amalgamating independent estimates of a parameter obtained from different sources. Generally, each estimate is reported as a number with an estimated confidence interval. In such cases two problems are to be considered. First, if the estimates are estimating the same quantity, what is the best way of combining them to obtain a single estimate? Secondly, what is the most advantageous way to determine an associated confidence interval for this estimate? With these objects in view, consider the particular case of two series of independent unmatched binomial variables $X_j$ and $Y_j$ with corresponding parameters $p_{1j}$ and $p_{0j}$ and sample sizes $n_{1j}$ and $n_{0j}$ ($j = 1, \ldots, k$). Assume further that the relative risks $\psi_j$ parameterized either as rate ratios $p_{1j}/p_{0j}$ or odds ratios $(p_{1j}(1 - p_{0j}))/p_{0j}(1 - p_{1j})$, are finite for each $j$.

Gart (1962) considered different methods of estimating the combined relative risk parameter $\hat{\psi}$ for $k$ two by two tables with nonrandom sample sizes and the other marginal totals unconstrained, i.e. if we let $x_j = \hat{p}_{1j} n_{1j}$ and $y_j = \hat{p}_{0j} n_{0j}$, then $x_j + y_j$ is random. He showed how the method of maximum likelihood may be applied to the problem and that it yields both consistent and efficient estimators of $\hat{\psi}$ under certain asymptotic conditions; see, for instance, Bradley & Gart (1962). Another issue dealt with in Gart’s paper is the asymptotic efficiency of point estimators of $\hat{\psi}$ not requiring an iterative computation of the likelihood equation. Gart presented three simple weighted estimators, namely a weighted arithmetic, geometric and harmonic mean of the individual estimators $\hat{\psi}_j$. All these estimators were said to be asymptotically efficient.

The present paper proceeds to show, as anticipated by Gart, that under general regularity conditions a weighted function of the $\hat{\psi}_j$’s is an asymptotically efficient estimator of $\hat{\psi}$. The asymptotic case considered here is where the number of tables, $k$, remains fixed but the sample sizes, $n_{1j}$ and $n_{0j}$, within each table increase without limit. Throughout, sums run from $j = 1$ to $k$. 

2. ASYMPTOTIC EFFICIENCY OF GENERAL WEIGHTED NONITERATIVE ESTIMATORS

Our object is to study asymptotic conditions of a weighting system producing estimates with desirable statistical and computational properties. We first formulate the following proposition.

**Theorem.** Let \( f(\hat{W}; \hat{\Psi}) = f(\hat{w}_1, \ldots, \hat{w}_k; \hat{\psi}_1, \ldots, \hat{\psi}_k) \) be a nonnegative real valued function, where \( \hat{\Psi} \) is an estimator of a vector parameter in a \( k \)-dimensional parameter space, that is \( \Psi \in \Theta \subset \mathbb{R}_+^k \), and \( \hat{w}_j \) is the weight associated with the \( j \)th sample. We also assume that:

(i) the function \( f(\hat{W}; \hat{\Psi}) \) is continuous;

(ii) continuous first partial derivatives \( D_j f(\hat{W}; \hat{\Psi}) \) with respect to \( \psi_j \) exist in the neighbourhood of the point \( (W; \psi, \ldots, \psi) \);

(iii) the quantities \( \hat{\psi}_1, \ldots, \hat{\psi}_k \), calculated from separate related samples, are mutually independent, so that \( \hat{\sigma}_{i \hat{j}} = \text{cov}(\hat{\psi}_j, \hat{\psi}_i) = 0 \) \( (j \neq i; j, i = 1, \ldots, k) \);

(iv) the estimators \( \hat{\psi}_1, \ldots, \hat{\psi}_k \) have a common asymptotic expectation \( E(\hat{\psi}_j) = \psi \), where \( 0 < \psi < \infty \), and unequal asymptotic variances

\[
\sigma_{jj}^2 = \text{var}(\hat{\psi}_j) = \psi^2/w_j \quad (j = 1, \ldots, k);
\]

(v) under homogeneity of relative risks \( f(W; \psi, \ldots, \psi) = \psi \), and \( D_j f(W; \hat{\psi}_1, \ldots, \hat{\psi}_k) \) converges in probability to \( D_j f(W; \psi, \ldots, \psi) = w_j/w \), where \( w = \Sigma w_j \).

Then the function \( f(\hat{W}; \hat{\Psi}) \) is an asymptotically efficient estimator of \( \psi \) with asymptotic variance \( \psi^2/w \).

Condition (i) ensures that function \( f \) is 'reasonably well behaved'. This condition is not really needed for the local properties of the weighted estimates of \( \psi \), but it is satisfied by all the functions that usually arise in applications. In general the weight \( \hat{w}_j \) attributed to \( \hat{\psi}_j \) is a function of the value of \( \hat{\psi}_j \), and the sample sizes \( n_{1j} \) and \( n_{0j} \) which are nonrandom, either directly, or indirectly via the contrasted occurrence/exposure rates \( p_{1j} \) and \( p_{0j} \); see for example Hoem (1976). A more realistic assumption for the weights might be \( \hat{w}_j = w_j(\hat{\psi}) \), but this would involve iteration to find \( \hat{\psi} \) and recomputation of the weights.

Condition (ii) is an additional smoothness condition on \( f \) and could be replaced by a slightly weaker, but harder to verify, condition that the partial derivatives of \( f \), with respect to each \( \psi_j \), exist at \( \psi \) and \( f \) has a linear approximation at \( \psi \). This expansion is the crucial device needed to prove the theorem.

The independence of the individual relative risks, condition (iii), is met in practical applications that combine results from several subgroups or different studies performed independently of each other. Relatedness here refers to the sameness of the subject matter content of the separate studies.

Assumption (iv) ensures that, with increasing sample sizes, we would expect all the individual studies to yield increasingly unbiased and equal estimates of the studied relationship, with maximum attainable precision.

Condition (v) was the only one explicitly stated by Gart (1962) for the fixed case of unvarying weights \( w_j \), whence the partial derivative of \( f \) with respect to \( \psi_j \) equalled the proportion \( w_j/w \). In the more general situation in which the change in \( f \) reflects the amount of sampling variability in the \( w_j \)'s, it may be defined as

\[
D_j f = \partial f/\partial \psi_j + (\partial f/\partial w_j)(dw_j/\partial \psi_j),
\]

when \( \hat{w}_j = w_j(\hat{\psi}_j) \). The convergence of \( D_j f \) requires the sequence to be bounded by
some constant if sample sizes are sufficiently large. Gart also noted that the assumption of the homogeneity of relative risks is not sufficient for the pooled estimator, obtained from the pooling of the data from individual studies, to be consistent.

The efficiency of \( \hat{\psi} \) as an estimator of \( \psi \) is shown in the Appendix.

As an example of the choice of the set of weights and the application of the theorem consider the Mantel–Haenszel estimator of the common relative risk which was introduced with reference to the exposure-odds ratio in case-referent, case-control, studies (Mantel & Haenszel, 1959). Its extension to the rate ratio in follow-up studies, with no restriction imposed on the overall occurrence rates, \( (x_i + y_i)/(n_{ij} + n_{0j}) \), takes the form

\[
\hat{\psi}_{MH} = \frac{\sum \{x_i n_{0j}/(n_{ij} + n_{0j})\}}{\sum \{y_i n_{1j}/(n_{ij} + n_{0j})\}}.
\]

Formally this can be interpreted as a weighted average of the individual rate ratios (Hauck, 1979), \( \hat{\psi}_{MH} = \sum \hat{\psi}_j \hat{w}_j / \hat{w} \), where

\[
\hat{\psi}_j = \hat{\phi}_{1j}/\hat{\phi}_{0j}, \quad \hat{w}_j = \hat{\phi}_{0j}/(n_{1j}^{-1} + n_{0j}^{-1}).
\]

The weighting can thus be said to weight the study-specific rate ratios according to their precision and importance, importance being measured by the occurrence rate in the population selected for reference, \( \hat{\phi}_{0j} \). It is easily verified that the function \( \hat{\psi}_{MH} \) fulfills the assumption (v). Obviously, under the uniformity of the rate ratios \( \hat{\psi}_{MH} = \sum \hat{\psi}_j \hat{\psi} / \hat{w} = \psi \), guaranteeing the asymptotic consistency of \( \hat{\psi}_{MH} \). Then differentiation yields

\[
D_j \hat{\psi}_{MH} = \frac{\partial \hat{\psi}_{MH}}{\partial \hat{\psi}_j} + \frac{\partial \hat{\psi}_{MH}}{\partial \hat{w}_j} \frac{\partial \hat{w}_j}{\partial \hat{\phi}_{0j}} = \frac{w_j \hat{\phi}_j - \hat{\psi}_{MH}}{n_{1j}^{-1} + n_{0j}^{-1}}.
\]

When the sample sizes tend to infinity, also the estimate of \( \hat{\psi}_j \) tends to \( \psi \) by (iv), so that \( D_j \hat{\psi}_{MH} \to w_j / w \) in probability. Thus, the large sample variance of \( \hat{\psi}_{MH} \) approaches the Cramér–Rao lower bound, \( \psi^2 / w \).

Hauck (1979) noted

that a sufficient condition for \( \hat{\psi}_{MH} \) to be asymptotically efficient . . . occurs whenever \( \psi = 1 \).

Here our concern was not with testing whether a relationship exists between the occurrence and exposure at issue, but rather to arrive at a quantitative estimate of the assumed common value of the effect parameter \( \psi \). Thus, in the derivation of the asymptotic variance we directly aimed at a system of weights that would converge to that of the efficient maximum likelihood estimator of \( \psi \).

3. INTERVAL ESTIMATION

In practical applications the sampling distribution of \( \hat{\psi} \) is often skewed to the right and ranging from 0 to +\( \infty \). Thus, with sufficiently large sample sizes, it is justified to regard \( \log \hat{\psi} \) instead of \( \hat{\psi} \) as approximately normally distributed. Thus, if the interval estimation is based, for example, on the efficient geometric mean estimator, \( \hat{\psi}_g = \exp (\hat{w}^{-1} \sum \hat{w}_j \log \hat{\psi}_j) \), then approximately (Chiang, 1968, p. 18)

\[
E (\log \hat{\psi}_g) \approx \log \psi, \quad \text{est} \{\text{var} (\log \hat{\psi}_g)\} \approx \text{est} \{\text{var} (\hat{\psi}_g)\} / (E (\hat{\psi}_g))^2 \approx \hat{w}^{-1},
\]

where est \((F)\) denotes an estimator of \( F \).

The simplest way to construct a confidence interval based on the foregoing
approximations is to solve for $\psi$ from the equation

$$\hat{\psi}(\log \hat{\psi} - \log \psi)^2 = \chi^2_1(1),$$

where $\chi^2_1(1)$ is the $(1-\alpha)$-point of the $\chi^2$ distribution with one degree of freedom. The equation yields

$$\psi_l = \hat{\psi}_l \exp \left(-\chi^* \hat{\omega}^{-\frac{1}{2}}\right), \quad \psi_u = \hat{\psi}_u \exp \left(\chi^* \hat{\omega}^{-\frac{1}{2}}\right)$$

as the lower and upper confidence limits (Woolf, 1955; Gart, 1962).

We can modify these limits by more fully utilizing the assumption that whichever efficient estimator $\hat{\psi}$ is log normally distributed; this implies

$$\text{est} \{\text{var}(\log \hat{\psi})\} = \log \left[1 + \text{est} \{\text{var}(\log \hat{\psi})\}/\{E(\hat{\psi})\}^2\right] = \log (1 + \hat{\omega}^{-1}).$$

Hence the lower and upper limits of the confidence interval become (Nurminen, 1974, p.117)

$$\psi_l = \hat{\psi} \exp \left[-\chi^* \{\log (1 + \hat{\omega}^{-1})\}^\frac{1}{2}\right], \quad \psi_u = \hat{\psi} \exp \left[\chi^* \{\log (1 + \hat{\omega}^{-1})\}^\frac{1}{2}\right].$$

The two intervals given by the two approaches are asymptotically equivalent to the order $O_p(\hat{\omega}^{-2})$ in their variance estimates.

4. DISCUSSION

While differences between separate studies could be of interest on their own, the present paper is concerned with the method of combining homogeneous relative risk estimates from independent samples. Gart (1962) questioned the applicability of the noniterative weighting method only to data which either are assumed to have equal relative risks or have been selected by having passed tests as to the homogeneity of the relative risk; Gart remarked that the use of the type of estimators in § 2 is a valid but perhaps nonoptimal procedure.

The same condition of homogeneity of relative risk is explicitly or implicitly assumed in almost all test procedures for testing independence in a set of $2 \times 2$ tables with fixed marginal totals. Mantel, Brown & Byar (1977) advise that tests for homogeneity be conducted with care and that primary concern should be given to trying to determine, by inspection of the data, how to make studies homogeneous rather than to demonstrate their heterogeneity.

A choice among the various point estimators of $\psi$ can be made in two stages. The first choice is determined by the study design, i.e. whether all the marginal totals of the individual tables can be regarded as random or fixed. This paper deals only with the problem of estimation in the mathematically unrestricted sample space. But, for example, making the studies homogeneous creates implicit restrictions for the margins. Computational aspects and applicability to small samples affect the final choice. For example, on investigating the assumption of zero second-order interaction using simulation, McKinlay (1978) found that Woolf’s point estimator, $\hat{\psi}_p$, although usually precise, shows instability in its mean value to sample size and marked increases in bias for larger numbers of strata. An addition of $\frac{1}{2}$ to the table entries reduces the bias in small samples and prevents $\hat{\psi}_p$ from becoming indeterminate should any of the $\hat{\psi}_j$’s be zero, or infinity; see assumption (iv). She concluded that for large samples and for $k > 2$ the Mantel–Haenszel statistic should be chosen, whether or not the relative risk is homogeneous.

Likewise, preference among the interval estimators cannot be founded solely on their asymptotic properties. If, for example, the weights $\hat{\omega}_j$’s are small, then both of the logit
type limits of § 3 yield very unstable estimates. In the case of ample numbers with appreciable asymmetry, the interval estimates which arise from the maximum utilization of the assumption of log normality of the distribution of $\hat{\psi}$ are expected to achieve a slightly better overall accuracy for the two approaches. But there is nothing to prevent us from computing the upper limit using one approach and the lower limit using another, if a more accurate interval estimate is obtained than would be the case if both limits were estimated by the same method.

Finally, the theorem of § 2 is general in the sense that it does not make use of the particular parametric structure of the relative risk or the special system of weights. Thus, the theorem is valid for any weighted statistic, $f(W; \hat{\Psi})$, for which the stated conditions hold.

The author is indebted to Juni Palmgren and Erkki Järvinen for their helpful comments.

**APPENDIX**

**Efficiency of $f(W; \hat{\Psi}$ as an estimator of $\psi$**

Since the function $f(W; \hat{\Psi})$ has continuous first derivatives in the neighbourhood of point $(W; \psi,\ldots,\psi)$, we obtain a linear Taylor series expansion

$$f(W; \hat{\Psi}) = f(W; \psi,\ldots,\psi) + \sum (\hat{\psi}_j - \psi) D_j f(W; \psi,\ldots,\psi) + O_p(||\hat{\Psi} - \Psi||).$$  \hfill (A1)

Using the homogeneity we can write (A1) as

$$f(W; \hat{\Psi}) - \psi - \sum (\hat{\psi}_j - \psi) D_j f(W; \psi,\ldots,\psi) \rightarrow 0$$  \hfill (A2)

in probability, since $D_j f$ remains bounded by assumption (v), and the asymptotic expectation

$$E\{\sum (\hat{\psi}_j - \psi) D_j f(W; \psi,\ldots,\psi)\} = \sum E(\hat{\psi}_j - \psi) D_j f(W; \psi,\ldots,\psi) \rightarrow 0.$$  

Thus also the asymptotic expectation of the estimator $f(W; \hat{\Psi})$ is $\psi$. Furthermore, the stochastic convergence $\hat{\psi}_j \rightarrow \psi$ implies that the distribution of $f$ concentrates entirely around $\psi$.

To find the minimum variance of $f$, we apply a lemma of Rao (1952, Corollary 1, p. 146) to obtain the approximation:

$$\text{var} \{f(W; \hat{\Psi})\} \geq \sum_{j=1}^{k} \sum_{i=1}^{k} \sigma_{ij} D_j f(W; \psi,\ldots,\psi) D_i f(W; \psi,\ldots,\psi),$$

which, recalling uncorrelatedness, can be reformulated as

$$\text{var} \{f(W; \hat{\Psi})\} \geq \sum \sigma_{jj}^2 (D_j f(W; \psi,\ldots,\psi))^2,$$  \hfill (A3)

where covariance terms of the form $\sigma_{ii} (j \neq i)$ are equal to zero and, with the additional use of assumptions (iv) and (v), can be put into the limiting form

$$\text{var} \{f(W; \hat{\Psi})\} \geq \sum (\psi^2/w_j) (w_j/w)^2 = \psi^2/w.$$  \hfill (A4)

Through a calculation starting from expression (A2) it is readily seen that

$$\text{var} \{(f(W; \hat{\Psi}) - \psi)\} = \text{var} \{f(W; \hat{\Psi})\} = \text{var} \{\sum (\hat{\psi}_j - \psi) D_j f\} = \sum \text{var}(\hat{\psi}_j) (D_j f)^2.$$  

Since this is the same as the second member of (A3), $\text{var} \{f(W; \hat{\Psi})\}$ attains the expression for the minimum variance defined by the inequality (A4).
REFERENCES


[Received November 1979. Revised November 1980]