Estimating confidence limits on a standardised mortality ratio when the expected number is not error free

Paul Silcocks

Abstract

Objective – The aim was to demonstrate how the beta distribution may be used to find confidence limits on a standardised mortality ratio (SMR) when the expected number of events is subject to random variation and to compare these limits with those obtained with the standard exact approach used for SMRs and with a Fieller-based confidence interval.

Design – The relationship of the binomial and the beta distributions is explained. For cohort studies in which deaths are counted in exposed and unexposed groups exact confidence limits on the relative risk are found conditional on the total number of observed deaths. A similar method for the SMR is justified by analogy between the SMR and the relative risk found from such cohort studies, and the fact that the relevant (beta) distribution does not require integer parameters.

Source of data – Illustrative examples of hypothetical data were used, together with a MINITAB macro (see appendix) to perform the calculations.

Main results – Exact confidence intervals that include error in the expected number are much wider than those found with the standard exact method. Fieller intervals are comparable with the new exact method provided the observed and expected numbers (taken to be means of Poisson variates) are large enough to approximate normality. As the expected number is increased, the standard method gives results closer to the new method, but may still lead to different conclusions even with as many as 100 expected.

Conclusions – If there is reason to suppose the expected number of deaths in an SMR is subject to sampling error (because of imprecisely estimated rates in the standard population) then exact confidence limits should be found by the methods described here, or approximate Fieller-based limits provided enough events are observed and expected to approximate normality.

\[ \text{SMR} = \frac{D}{E} = R \text{ (say)} \] (1)

Then if D and E are normal, so will \((D - E^*R)\) be normal and

\[ \frac{(D - E^*R)^2}{\text{Var}(D - E^*R)} \sim \text{approximately as } \chi^2 \text{ on} \]

1 degree of freedom

95% Confidence intervals for R are solutions to the quadratic equation in R:

\[ R^2E^*(E - \chi^2) - R^*(2D^*E - 2D^*E^*q^*\chi^2) + D^*(D - \chi^2) = 0 \] (3)

Where we use the fact that \(\text{var}(x) = x\) for a Poisson variate and \(\chi^2\) is the appropriate critical value for \(\chi^2: 3.8416\) in this instance. Here q
is the mean proportion of the standard population contributed by the index population, averaged across age bands and weighted by the number of observed events in each age band. This would apply, for example, when the mortality rate for an electoral ward is standardised on the health district of which it forms a part. As indicated above, however, this approach will not apply if we cannot assume normality for D and E (for example when we have only small numbers of events) so some other method is needed. Equation (3) is derived in appendix 1.

**Method**

In the context of a cohort study or randomised controlled trial where we want confidence limits on a relative risk derived from an exposed and an unexposed group, one method is to condition on the total number of observed events and find confidence limits on the proportion \( p \) of events that occur in the exposed group. We have the following relations:

\[
p = \frac{r_1}{r_1 + r_2}
\]

(4)

\( r_1, r_2 \) are observed events in exposed and unexposed groups.

Relative risk = \[
\frac{(r_1/r_2)^*(n_2/n_1)}{[p/(1-p)]^*(n_2/n_1)}
\]

(5)

\( n_1, n_2 \) are person-years at risk or population sizes for exposed and unexposed.

Confidence intervals on the relative risk are obtained by finding lower and upper confidence limits for \( p \) and substituting these in (5). Since \( p \) is a binomial proportion, exact confidence limits can be found by exploiting the well-known relation between the beta distribution and the binomial. Specifically the integral:

\[
G(k) = \int_{z = -\infty}^{z = \infty} z^{k-1}(1-z)^{(n-k)}dz = \text{Pr}(z^* p)
\]

(6)

is equal to the binomial expansion:

\[
\sum_{j = 0}^{j = k-1} \binom{n}{j} p^j(1-p)^{n-j}
\]

(7)

giving \( \text{Pr}(r < k) \), where \( r \) is a count of events.

Now \( G(k) = 1 - \text{Be}(p,k,n-k+1) \) where \( \text{Be} \) is the incomplete beta function (that is, the beta cumulative distribution function for \( z = p \)), and so we have:

\[
\text{Pr}(z \leq p) = 1 - \text{Pr}(z > p) = 1 - \text{Pr}(r < k) = \text{Be}(p,k,n-k+1)
\]

(8)

and since \( \text{Pr}(r > k) = 1 - \text{Pr}(r < k+1) \),

\[
\text{Pr}(r > k) = 1 - G(k+1) = \text{Be}(p,k+1,n-k)
\]

(9)

So, the lower 95% confidence limit for \( p \) is the value of \( p \) such that \( \text{Be}(p,k,n-k+1) = 0.025 \) while the upper limit satisfies \( \text{Be}(p,k+1,n-k) = 0.975 \). In the past these values have been found by further exploiting the relation between \( \text{Be} \) and \( F \) distributions, but this is no longer needed as packages like MINITAB\(^6\) give the \( \text{Be} \) cumulative distribution function and inverse cumulative distribution function directly.

Now an SMR can be thought of as the relative risk obtained by comparison of an “exposed” group (the index population) with a synthetic “unexposed” group derived from the standard population. If the standard population is distinct from the index, or if the index population is only a small fraction of the standard so that the covariance of \( D \) and \( E \) is negligible, then the position is just as described above, except that the total number of events need not be an integer. The \( \text{Be} \) function parameters do not need to be integers, however, and so we can find confidence limits on \( p \) exactly as if we had followed up two real groups. This argument gives confidence limits that preserve the integer character of the observed number of events.

A MINITAB macro is listed in appendix 2 that finds these confidence limits:

(a) Assuming no error in the expected number of events;

(b) By the method described above (which assumes no covariance between \( D \) and \( E \));

(c) Using Fieller’s theorem (different amounts of overlap between index and standard populations can be allowed for by varying the quantity \( q \)).

For those unhappy with the use of a non-integer value for the denominator of a proportion, virtually the same results will be found with the following approach. Take the expected number of events to be an estimate of the (integer) number that would have been found with a real comparison group. Suppose 2-8 events are expected – in reality we would have obtained 2 or 3. We can find an SMR (and confidence limits) separately for 2 and then for 3 events, and obtain the actual SMR as the appropriately weighted harmonic mean of the separate SMRs:

\[
D_{0.2*E1+0.8*E2} = \frac{1}{(0.2/\text{SMR1})+(0.8/\text{SMR2})}
\]

(10)

with SMR1 and SMR2 being the SMRs found using the integer expected values. Confidence limits are obtained by substituting the corresponding limits on SMR1 and SMR2.

**Results**

For seven observed events with 2-8 expected: the SMR is 2.5 with exact 95% confidence limits of 1.005 to 5.151, assuming no error in the expected value, and limits of 0.5548 to 16.0921 assuming error, found using the \( \text{Be} \) distribution.
Using interpolation: \[ 2.8 = 0.2 \times 2 + 0.8 \times 3 \]

If we assume that a Poisson distribution with a mean of 10 or more will be satisfactorily “normal” (based on the relation \( x > 3^*\sqrt{x} > 0 \)), then we can compare the confidence limits found using Fieller’s theorem, for an SMR of 1.4, say, with different observed and expected values (Table 1).

When the index population (for which the SMR is to be calculated) forms part of the standard, then the observed and expected numbers of deaths must be correlated. To assess the effect of this, we require the quantity \( q \), which is a weighted average of the proportions in each age band of the standard population, contributed by the index population. Calculation of \( q \) using routine data, and its effect are illustrated in Table 2. The effect of different values of \( q \) is illustrated in Table 3, using Fieller confidence limits. When \( q = 0 \) then observed and expected numbers are independent, but when \( q > 0 \) the correlation between them narrows the width of the confidence interval.

### Table 1

**Comparison of exact, Fieller, and standard confidence limits**

<table>
<thead>
<tr>
<th>Method</th>
<th>Observed/expected (95% confidence limits)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14/10</td>
</tr>
<tr>
<td></td>
<td>140/100</td>
</tr>
<tr>
<td>“Exact” limits</td>
<td>0.382 to 5.595</td>
</tr>
<tr>
<td>Fieller limits</td>
<td>0.578 to 3.523</td>
</tr>
<tr>
<td>Standard limits</td>
<td>0.765 to 2.349</td>
</tr>
<tr>
<td></td>
<td>1.075 to 1.828</td>
</tr>
<tr>
<td></td>
<td>1.085 to 1.827</td>
</tr>
<tr>
<td></td>
<td>1.178 to 1.652</td>
</tr>
</tbody>
</table>

That the Fieller approach breaks down completely with small numbers is seen from the Fieller confidence limits obtained for 7 observed and 2.8 expected, which are 0.542 to (−)13.983.

### Discussion

It is clear that it is very easy to obtain false impressions both of statistical significance and of precision using the standard approach.

If we assume that a 10% coefficient of variation is the maximum tolerable for an estimate of some “baseline” value (even tighter coefficients of variation are common in laboratory medicine) then we might assume the expected number to include random variation if the average coefficient of variation of the rates from which the expected numbers will be estimated exceeds 10%.

Alternatively, we might assume that the denominator of any SMR will be subject to random variation. The justification for this view is consistency: when comparing the mortality in a health district with that in the parent region, the expected number of deaths derived from regional rates is assumed to be error-free; yet if regional observed deaths (which go to form the regional rates) are compared with those expected from national rates, somehow they are now assumed to be subject to error. Hence when the denominator is less than about 100, so that the expected number has coefficient of variation more than 10% “exact” limits should be used. Even with this cut off our conclusions may be less well justified than we think. For example, with 125 observed events and 100 expected, the confidence limits assuming no error in the denominator are: 1.04 to 1.49, but these are 0.95 to 1.64 if we allow for error.

Fieller-based limits are similar to the exact values once approximate normality holds, but break down with small numbers of events. On the other hand these limits can easily accommodate covariance between observed and expected values, unlike the exact method.

Random variation (taken to be Poisson) in the numbers at risk in the index population can reasonably be ignored. In England and Wales the all-causes mortality rate is of the order of 280/100,000/y so to get a “reasonable” number of observed deaths such as 5, one would need of the order of 2000 person-years, which for a 20 year period gives \( n = 100 \) with coefficient of variation at 10%. For particular causes of death with lower mortality rates, of course, the numbers at risk would need to be greater still, with correspondingly less sampling error.

An additional potentially rather interesting
Appendix 1
DERIVATION OF EQUATION 3
Consider two subpopulations. One is the index population, while the standard population is the sum of the subpopulations. For simplicity consider only age-standardisation.
Let D = observed deaths in index population;
Let E = expected deaths in index population;
Let the estimated death rate in the ith age band in subpopulation j be \( r_{ij} = \frac{n_{ij}}{n_j} \);

Then \( D = \sum_i r_{ij} \)

\&

\[ E = \sum_i \frac{n_{ij}(r_{ij} + r_{ij})}{(n_i + n_j)} \]

\[ = \sum_i p_i (r_{ij} + r_{ij}) \]

where \( p_i \) is the proportion of the ith age band in the combined (standard) population.

Now

\[ \text{covar}(D,E) = \text{covar}\{ \sum_i r_{ij}, \sum_i p_i (r_{ij} + r_{ij}) \} \]

\[ = \text{covar}\{ \sum_i r_{ij} p_i, (r_{ij} + r_{ij}) \} + \sum_i \sum_{i 
eq j} r_{ij} p_i (r_{ij} + r_{ij}) \]

\[ = \sum_i \text{covar}(r_{ij}, r_{ij}) + \sum_i \sum_{i 
eq j} r_{ij} p_i (r_{ij} + r_{ij}) \]

\[ = \sum_i p_i \text{var}(r_{ij}) \]

(because \( r_{ij} \) & \( r_{ij} \) from disjoint populations)

\[ = \sum_i p_i r_{ij} \]

(assuming \( r \) are Poisson)

That is, an average of the \( p_i \), weighted by the number of observed events in each age band. Hence Covar \((D,E) = q^2 D\) where \( q \) is given as above.

The method based on Fieller’s approach is as follows:

\[ \text{SMR} = \frac{D/E}{R} = R \quad \text{say} \]

Then if \( D \) and \( E \) are “normal”, so will \((D-E*R)\) be “normal” and

\[ (D-E*R)^2 \sim \text{approximately as } \chi^2 \]

on \( \text{Var}(D-E*R) \) 1 degree of freedom

Now \( \text{Var}(D-E*R) = \text{Var}(D) - 2q^2 R^2 \text{Covar}(D,E) + R^2 \text{Var}(E) = D - 2q^2 D*R + R^2 E \)

on using the above result for the covariance.

Confidence limits for \( R \) are solutions to the quadratic equation in \( R \):

\[ R^2 E + (E - X^2) - R^2 (2*D + E - 2*D^*q^2 + E^2 + D^*(D + X^2)) = 0 \]

Where we use var(\(x\)) for a Poisson variate, and \( X^2 \) is the critical value for \( \chi^2 \) appropriate for the desired confidence limits.

Confidence limits

Appendix 2

# macro for 95% confidence limits on an SMR
#
# --------------------------------------------
data input-------------------------------------
ncecho
erase c1-c52
# housekeeping)
#
# Enter observed frequency
set 'terminal' c50;
nobs = 1.
copy c50 k1
#
# Enter expected frequency
set 'terminal' c50;
nobs = 1.
copy c50 k21
#
# Enter average fraction of standard population
note that is the index (= q)
set 'terminal' c50;
nobs = 1.
copy c50 k22
#
# --------------------------------------------

calculate 95% limits on observed no., using relation
between Poisson and chi^2 distribution
# --------------------------------------------

let k2 = 2*k1
let k3 = 2*(k1 + 1)
#
invcdf 0.025 k4;
chisq k2.
let k4 = 0.5*k4/k21
#
invcdf 0.975 k5;
chisq k3.
let k5 = 0.5*k5/k21
let k6 = k1/k21
#
# k6 = smr
#
# --------------------------------------------

# Now find confidence limits for the proportion O/(O + E) This
# treats E as if from a separate, control group. The confidence
# limits are conditioned on the sum (O + E) and uses the beta
# distribution, for which E does not need to be an integer
#
# --------------------------------------------

let k8 = k1 + 1
let k9 = k21 + 1
#
invcdf 0.975 k10;
beta k8 k21.
#
invcdf 0.025 k11;
beta k1 k9.
#
let k10 = k10/(1 - k10)
let k11 = k11/(1 - k11)
#
# --------------------------------------------

copy k10 c51
stack k11 c51 c51
stack k6 c51 c51
#
# --------------------------------------------

# Now find approximate limits with Fieller's result
#
# --------------------------------------------

let k13 = k1/k21
let k14 = k21*(k21 - 3.8416)
let k15 = -2 * k1*(k21-k22*3.8416)
let k16 = k1*(k1 - 3.8416)
#
let k17 = sqrt(k15*k15 - 4*k14*k16)
let k18 = 0.5*(+k15 - k17)/k14
let k19 = 0.5*(-k15 - k17)/k14
#
# --------------------------------------------

copy k18 c52
stack k19 c52 c52
stack k6 c52 c52
#
# --------------------------------------------

# Note
# --------------------------------------------

# Now find confidence limits for the proportion O/(O + E) This
# treats E as if from a separate, control group. The confidence
# limits are conditioned on the sum (O + E) and uses the beta
# distribution, for which E does not need to be an integer
#
# --------------------------------------------

# Now find approximate limits with Fieller's result
#
# --------------------------------------------
LETTERS TO THE EDITOR

Prevalence of total hip replacement: how much demand has been met?

Sir – In their recent study, Williams et al found a substantial increase in the prevalence of total hip replacement compared with earlier years, and concluded that there was a greater level of satisfied demand. Before becoming compliant and basing policy decisions on this conclusion, however, cash-strapped purchasing authorities, struggling to meet the demand for total hip replacement in the UK, need to consider several other factors which are likely to increase demand still further.

Firstly, the study of Williams et al was confined to people ≥65 years of age and one of the main contributory factors to the increased demand for total hip replacement is its application to younger patients. In our recent population based study of total hip replacement in Olmsted County, Minnesota (characterised inaccurately by Williams et al as an elective case series), people aged under 65 comprised 48% of all primary total hip procedures in men and 30% in women between 1987 and 1990. Moreover, though the average age at surgery has remained relatively constant through the years, the range has increased. For example, the age range of women in 1969 was 62–82 (mean 69) years and in 1990 it was 29–91 (mean 70) years. This substantial use of total hip replacement in younger people is a strain on resources, a fact that is underemphasized by Williams et al. It also entails a substantial liability for revision surgery since the average life expectancy of these younger, more active people may exceed the expected useful life of the implants.

Secondly, and following from above, the size of the elderly population is growing rapidly and their composition within the broad age groups that were studied (for example ≥75 years) is changing: this has implications for provision for total hip replacement. Recent analysis of the data from 400 hospitals in the United States showed that while the number of people aged 85 years and over increased by 28%, the proportion within that age group who received total hip replacement increased from 1-4 per thousand to 3-1 per thousand between 1980 and 1987. It would therefore be interesting to see a finer breakdown in ages and an assessment of the increased number of total hip replacement that might be needed in the future, even at current per capita surgery rates, solely because of the demographic changes in the UK.

Finally, with technological and medical advances, the indications for total hip replacement have increased. Primary degenerative joint disease in our study accounted for only 63% of the primary procedures, and not 76% as quoted by Williams et al. The other major category was hip fractures. Even though only a small proportion of hip fractures are treated with total hip replacement, these operations still accounted for nearly 20% of the total. With continued ageing of the population, the number of people with hip fractures needing surgery will increase further. In addition, people with serious co-morbidities such as cancer and cardiovascular disease who may not have been considered good candidates for total hip replacement in the past are now being offered the procedure.

The need for total hip replacement, or any other medical intervention, is a complex interplay of population demographics, patient and surgeons’ attitudes, and economic and sociopolitical factors. We do not believe that the plateau for total hip replacement has been reached yet and support the call for further research into the needs... but for all age groups and for all indications.

RAJAN MADHOK
DAVID G LEWALLON
L JOSEPH MELTON, III
Departments of Health Sciences Research and Orthopedic Surgery,
Mayo Clinic and Foundation, Rochester, MN 55905, USA


Aetiology of peptic ulcer in Norway

Sir – Johnsen et al report an increased risk of duodenal ulcer in men with a high milk consumption. This factor was independent of age, cigarette smoking, and inheritance and the authors suggest that milk might increase duodenal ulcer risk by stimulating acid secretion through its calcium content.

In our analysis of the association between duodenal ulcer and calcium intake we found that the risk of duodenal ulceration increased with increasing intake of calcium in a dose-response manner. In contrast to Johnsen's report this relationship was observed irrespective of sex. There was no correlation between calcium or milk intake and the duration of dyspeptic symptoms and none of the subjects reported a change in consumption of dairy products because of dyspepsia. These results make it less likely that increased milk consumption was a consequence of dyspepsia.

In summary, the findings of Johnsen's study and our own support a causal relationship between milk or calcium intake and duodenal ulcer risk.

BREUER-KATSCHINSKI B
Zentrum fur Innere Medizin
Universitaetsklinikum Essen,
Hufelandstr 55, 4300 Essen I, Germany

Corrigendum

Estimating confidence limits on a standardised mortality ratio when the expected number is not error free. 1994;48:313-7.

The author wishes to correct some minor errors in the above article.

Firstly, in table 3 the numbers in the body of the table should be as follows:

\[ q = 0.2, 0.1, 0.05, 0.00 \]

The lower 95% limit is therefore 0-602 0-572 0-559 0-546 0-546 and the upper 95% limit 2-840 2-987 3-059 3-131 2-925.

Secondly, the quadratic equation which must be solved for R in the appendix has a plus sign where there should be a minus sign, and should be:

\[ R^2 = (E - X^2) - (2.D.E. - 2.D.q.X^2) + (D - D^2) = 0 \]

NOTICE

Conference '95 – Canadian Society for Epidemiology and Biostatistics will be held on August 16–19, 1995 at St John's, Newfoundland. For information regarding registration and submission of abstracts, contact the conference office:

CSEB – Conference '95 Office, c/o Health Research Unit, PO Box 23068, St John's, Newfoundland, Canada A1B 4J6. Tel: 709 737-6720. Fax: 709 737-7382.