American Journal of Epidemiology

Vol. 147, No. 8
Printed in U.S.A.

Confidence Limits Made Easy: Interval Estimation Using a Substitution Method

Leslie E. Daly

The use of confidence intervals has become standard in the presentation of statistical results in medical journals. Calculation of confidence limits can be straightforward using the normal approximation with an estimate of the standard error, and in particular cases exact solutions can be obtained from published tables. However, for a number of commonly used measures in epidemiology and clinical research, formulae either are not available or are so complex that calculation is tedious. The author describes how an approach to confidence interval estimation which has been used in certain specific instances can be generalized to obtain a simple and easily understood method that has wide applicability. The technique is applicable as long as the measure for which a confidence interval is required can be expressed as a monotonic function of a single parameter for which the confidence limits are available. These known confidence limits are substituted into the expression for the measure—giving the required interval. This approach makes fewer distributional assumptions than the use of the normal approximation and can be more accurate. The author illustrates his technique by calculating confidence intervals for Levin's attributable risk, some measures in population genetics, and the "number needed to be treated" in a clinical trial. Hitherto the calculation of confidence intervals for these measures was quite problematic. The substitution method can provide a practical alternative to the use of complex formulae when performing interval estimation, and even in simpler situations it has major advantages. *Am J Epidemiol* 1998;147:783–90.

binomial distribution; confidence intervals; epidemiologic methods; Poisson distribution; statistics

Confidence intervals are now required by most medical journals for the presentation of statistical results. A confidence interval is a range of likely values for an unknown population parameter at a given confidence level. The endpoints of this range are called the confidence limits.

A number of different methods can be used to estimate confidence limits. Exact limits can be obtained using published tables (1–3) or appropriate software (4, 5) for a single proportion, percentage, or risk (binomial limits), as well as for a count (Poisson limits). However, the most commonly used method of calculating confidence limits involves the normal approximation, in which a multiple of the standard error (SE) is added to and subtracted from the sample value for the measure. For 95 percent confidence limits, the general expression is

$$\text{statistic} \pm 1.96\ \text{SE(statistic)}, \qquad (1)$$

where SE(statistic) is the standard error of the relevant quantity and 1.96 is the appropriate percentile of the normal distribution. Confidence limit estimation is relatively straightforward using this approach, and methods for use with single means, proportions, or counts, for differences between these, and for relative risk-type measures are well known (1, 2, 6). Several commonly used standard error formulae are given in the Appendix.

Although the normal approximation (expression 1) can often be used directly for confidence interval estimation, sometimes it must be used on a transformation of the measure of interest. For instance, 95 percent confidence limits for the relative risk (RR) can be based on the limits for $\log_e$ RR:

$$\log_e \text{RR} \pm 1.96\ \text{SE}(\log_e \text{RR}), \qquad (2)$$

where SE($\log_e$ RR) is the standard error of the natural logarithm of RR (expression A2 in the Appendix). Transforming back to the original scale, the exponential of these limits gives the limits for the relative risk itself. A similar approach can also be used for the odds ratio. It is important to realize that it is the actual limits of the transformed quantity that must be back-transformed. When the limits are transformed in this

Received for publication October 13, 1994, and in final form October 3, 1997.

Abbreviations: LAR, Levin's attributable risk; NNT, number needed to be treated; RR, relative risk; SE, standard error.

From the Department of Public Health Medicine and Epidemiology, University College Dublin, Earlsfort Terrace, Dublin 2, Ireland. (Reprint requests to Prof. Leslie E. Daly at this address).

way, the confidence limits are not symmetrical around the point estimate. A common error for the unwary is to back-transform the "plus-and-minus" part of the expression, which gives a symmetrical but incorrect interval.

Unfortunately, however, for a number of measures used in epidemiology or clinical research, either no standard error formulae are available or the formulae are complex and tedious to use. In addition, these more complex formulae often can only be found in specialist articles or books, and they are rarely implemented in computer software packages.

Other techniques for confidence limit estimation have also been proposed, the best known of which include Miettinen's test-based limits (1, pp. 197–200) and Thomas and Gart's (7) computationally difficult procedure for parameters of 2 × 2 tables. These techniques are of limited applicability.

This paper describes a particular approach to confidence limit estimation which, though previously described for certain simple situations, gives rise to a hitherto unrecognized general method. It is easily understood and simple to apply, makes fewer assumptions than the normal approximation approach, is inherently more accurate, and is applicable in many situations that were previously intractable.

THE SUBSTITUTION METHOD

The following example, pertaining to an incidence rate, illustrates a simple application of what might be called "the substitution method" for estimating confidence limits. Seventeen cases of Wilson's disease were detected in 1,240,091 births in Ireland (8), giving a birth incidence of

$$I = x/N = 17/1{,}240{,}091 = 13.71 \text{ per million.} \quad (3)$$

In this situation, the number of cases (x) can be assumed to have a Poisson distribution, and the denominator (N), which is large and based on census data, can be considered fixed and without sampling variation. Confidence limits for I can then be based on confidence limits for x (1, pp. 67–8). If x_l and x_u are the lower and upper confidence limits for x, respectively, then the lower and upper confidence limits for I are simply

$$I_l = x_l/N; \quad (4)$$

$$I_u = x_u/N. \quad (5)$$

In Poisson confidence interval tables (1, pp. 393–5), the 95 percent limits for this number of cases ($x = 17$) are 9.903 and 27.219. Substitution into expressions 4 and 5 then gives a 95 percent confidence interval for the incidence rate of 7.99–21.95 per million.

Although the substitution of known confidence limits into an expression for a quantity has also been described for interval estimation of a standardized mortality ratio (1, p. 279) and the ratio of two rates (1, p. 200), the approach has never been considered as a method for confidence limit estimation in its own right, and its general applicability has never been exploited. In each of these examples, the measure for which a confidence interval is required is expressed as a function of a *single* quantity for which limits are easy to calculate. The confidence limits for this single quantity are then substituted into the formula for the measure of interest to obtain the required interval.

It is interesting to note that confidence limit estimation based on a transformation of a particular quantity can also be considered an application of this substitution method. For instance, any quantity is a function (the exponential or back-transformation) of its logarithm. Taking the relative risk as an example (see above), the confidence limits for $\log_e$ RR are readily computed, and their exponentiation, giving the confidence limits for the RR itself, is equivalent to their substitution into the formula for the RR.

The general applicability of the substitution method is illustrated below by extending its application to three situations which hitherto were quite problematic. Other examples can easily be found. In each situation, the use of the method is illustrated using previously published data, and further calculations are performed on example data sets. Twenty-one 2 × 2 contingency tables were chosen for illustrating two of the applications. These example tables (table 1), each of which shows a statistically significant (uncorrected chi-square, $p < 0.05$) difference in mortality between two groups, cover a range of sample sizes, baseline risks, relative risks, and risk differences.

Levin's attributable risk

Several attributable risk-type measures are suggested in the literature (9). One in particular is called Levin's attributable risk. Table 2 gives infant mortality by birth weight for 72,730 births among whites in New York City for 1974 (10, p. 77). If low birth weight births did not occur in the population, the population risk would be reduced to that observed for infants with normal birth weight. Letting $R_T = 1{,}040/72{,}730 = 0.0143$ and $R_2 = 422/67{,}515 = 0.0063$ represent, respectively, the total observed risk in the population and the risk in normal birth weight infants, the difference between these is the amount of risk in the population that is attributable to low birth weight. Levin's

TABLE 1. Structure of example data sets used to illustrate the application of the substitution method for estimation of confidence limits*

Example	Total sample size	Risk in group 1 (R_1)	Relative risk		Risk difference ($R_1 - R_2$)	Cell			
						Group 1		Group 2	
			R_1/R_2	R_2/R_1		Dead	Alive	Dead	Alive
A	50	0.72	1.5	0.67	0.24	36	14	24	26
B	50	0.72	2.0	0.50	0.36	36	14	18	32
C	50	0.72	3.0	0.33	0.48	36	14	12	38
D	100	0.72	1.5	0.67	0.24	72	28	48	52
E	100	0.72	2.0	0.50	0.36	72	28	36	64
F	100	0.72	3.0	0.33	0.48	72	28	24	76
G	100	0.48	1.5	0.67	0.16	48	52	32	68
H	100	0.48	2.0	0.50	0.24	48	52	24	76
I	100	0.48	3.0	0.33	0.32	48	52	16	84
J	500	0.72	1.5	0.67	0.24	360	140	240	260
K	500	0.72	2.0	0.50	0.36	360	140	180	320
L	500	0.72	3.0	0.33	0.48	360	140	120	380
M	500	0.48	1.5	0.67	0.16	240	260	160	340
N	500	0.48	2.0	0.50	0.24	240	260	120	380
O	500	0.48	3.0	0.33	0.32	240	260	80	420
P	500	0.24	1.5	0.67	0.08	120	380	80	420
Q	500	0.24	2.0	0.50	0.12	120	380	60	440
R	500	0.24	3.0	0.33	0.16	120	380	40	460
S	500	0.12	1.5	0.67	0.04	60	440	40	460
T	500	0.12	2.0	0.50	0.06	60	440	30	470
U	500	0.12	3.0	0.33	0.08	60	440	20	480

* Each data set is a 2 × 2 table relating mortality in two groups of equal sample sizes. (Group 1 has the higher risk.)

attributable risk (LAR) is this quantity expressed as a proportion of the total population risk:

$$\text{LAR} = \frac{R_T - R_2}{R_T} = 0.563. \quad (6)$$

A complex standard error formula for LAR was proposed by Walter (9), and an alternative for the logarithm of LAR was proposed by Fleiss (10, pp. 76–7), both of which enable estimation of confidence limits. However, application of the substitution method provides a far easier solution. First, LAR must be expressed as a function of a single parameter for which confidence limits are easy to obtain. A small amount of algebraic manipulation gives the following expression for LAR in terms of the relative risk of infant death among low birth weight infants compared with normal birth weight infants (RR = R_1/R_2 = 18.959, where R_1 = 618/5215 = 0.119) and the prevalence of low birth weight in the population (Prev = 5,215/72,730 = 0.0717):

$$\text{LAR} = \frac{\text{Prev}(\text{RR} - 1)}{1 + \text{Prev}(\text{RR} - 1)}. \quad (7)$$

This expression for LAR is in common use, even though its interpretation is not intuitively obvious. If the prevalence of low birth weight is assumed to be free of sampling variation (equivalent to assuming that one of the margins of table 2 is fixed), LAR is seen to be expressed in terms of the relative risk, for which confidence limits are easily obtained. These limits can then be substituted into expression 7 to obtain limits for LAR. The lower and upper 95 percent confidence limits for this RR, RR_l and RR_u, estimated using

TABLE 2. Infant mortality among whites in New York City, by birth weight, 1974*

Birth weight (g)	Outcome at 1 year		Total
	Dead	Alive	
≤2,500	618 (*a*)†	4,597 (*b*)	5,215 (*a* + *b*)
>2,500	422 (*c*)	67,093 (*d*)	67,515 (*c* + *d*)
Total	1,040	71,690	72,730

* Data from Fleiss (10, p. 77).
† *a*, *b*, *c*, and *d* are cell entries.

expressions 2 and A2, are 16.807 and 21.387, respectively. By substitution, the limits for LAR are

$$LAR_l = \frac{Prev(RR_l - 1)}{1 + Prev(RR_l - 1)} = \frac{0.0717(16.807 - 1)}{1 + 0.0717(16.807 - 1)} = 0.531; \quad (8)$$

$$LAR_u = \frac{Prev(RR_u - 1)}{1 + Prev(RR_u - 1)} = \frac{0.0717(21.387 - 1)}{1 + 0.0717(21.387 - 1)} = 0.594. \quad (9)$$

These are almost identical to the values of 0.530 and 0.594 calculated using Fleiss' standard error formula (10) and to the limits of 0.532 and 0.594 given by Walter's more complex approach (9). Of course, the limits obtained by the substitution method depend on which formula is employed for the relative risk limits.

Table 3 compares the 95 percent confidence limits obtained by means of Walter's method, Fleiss' method, and the substitution method (used as described above) for the 21 example tables. For this application, group 1 is taken as the exposed group and group 2 as the nonexposed. There is good agreement between the three approaches, but in general the substitution limits (lower and upper) tend to be lower than those given by Walter's method and higher than those given by Fleiss'. Thus, the substitution limits would seem to be a better approximation of Walter's limits than the limits proposed by Fleiss. In example table S, for instance, Fleiss' lower limit is less than zero, which would correspond to a statistically nonsignificant (5 percent level) association between exposure and mortality. Based on the (conservative) continuity-corrected chi-square, the association is statistically significant, and both the substitution method and Walter's method give the required nonnegative limits.

Population genetics

Under Hardy-Weinberg equilibrium, the frequency (the proportion or percentage) of a rare recessive gene

TABLE 3. Lower and upper 95 percent confidence limits for Levin's attributable risk (LAR_l and LAR_u), calculated using two published methods and the substitution method in 21 example data sets*

Example	Levin's attributable risk (LAR)	Confidence limit calculation method					
		Method of Fleiss (10)		Method of Walter (9)		Substitution method	
		LAR_l	LAR_u	LAR_l	LAR_u	LAR_l	LAR_u
A	0.200	0.017	0.349	0.039	0.361	0.035	0.355
B	0.333	0.118	0.496	0.152	0.515	0.142	0.501
C	0.500	0.251	0.666	0.304	0.696	0.280	0.670
D	0.200	0.075	0.308	0.086	0.314	0.084	0.311
E	0.333	0.188	0.453	0.205	0.462	0.200	0.455
F	0.500	0.335	0.624	0.361	0.639	0.349	0.626
G	0.200	0.011	0.353	0.031	0.369	0.027	0.361
H	0.333	0.124	0.493	0.154	0.513	0.144	0.499
I	0.500	0.272	0.657	0.315	0.685	0.294	0.662
J	0.200	0.146	0.250	0.149	0.251	0.148	0.250
K	0.333	0.272	0.390	0.276	0.391	0.275	0.389
L	0.500	0.432	0.560	0.438	0.562	0.436	0.559
M	0.200	0.120	0.273	0.125	0.275	0.124	0.274
N	0.333	0.247	0.410	0.253	0.414	0.251	0.411
O	0.500	0.408	0.577	0.417	0.583	0.413	0.578
P	0.200	0.068	0.314	0.078	0.322	0.075	0.318
Q	0.333	0.193	0.449	0.207	0.460	0.202	0.453
R	0.500	0.356	0.612	0.374	0.626	0.364	0.615
S	0.200	−0.005	0.363	0.017	0.383	0.013	0.374
T	0.333	0.117	0.497	0.147	0.520	0.136	0.506
U	0.500	0.277	0.654	0.316	0.684	0.295	0.661

* See table 1.

(q) in the population can be estimated from the square root of the birth incidence of homozygotes (I) (11, p. 5):

$$q = \sqrt{I}. \quad (10)$$

An approximation for the standard error of this estimate of q is usually given by (11, p. 5):

$$SE(q) = \sqrt{(1 - q^2)/4N}, \quad (11)$$

where N is the number of births on which the birth incidence is based. In the study described above, the 17 (homozygous) cases of Wilson's disease in 1,240,091 births gave an incidence of 13.71 per million and a gene frequency of 0.37 percent. The 95 percent confidence interval for the latter figure using the normal approximation (expressions 1 and 11) is 0.28–0.46 percent.

The substitution method offers an alternative to this approach. The 95 percent confidence limits for the incidence rate were previously determined from expressions 4 and 5 (using the substitution method with the Poisson distribution) to be 7.99 per million and 21.95 per million. Applying the substitution method again by taking the square roots of these limits (expression 10), the lower and upper 95 percent confidence limits for the gene frequency are 0.28 percent and 0.47 percent—almost identical to those obtained using the standard error method.

Table 4 compares the substitution method with the usual approach for a series of birth incidence examples covering a range of gene frequencies and total births. Agreement is close, particularly with large numbers of affected births. When there was only one affected birth (examples AA and EE), however, the lower 95 percent limits for the gene frequency were much higher using the substitution method. It should be realized, of course, that use of expressions 1 and 11 does not give exact limits for the gene frequency, since the calculation is based on an approximate formula and on the assumption that the sampling distribution of q is normal. The substitution limits, on the other hand, based on a transformation of the exact Poisson limits for the incidence rate, can be considered exact in this situation.

Although the substitution method is more accurate, the simplicity of the standard error formula in this example gives no advantage to the new approach in terms of ease of use. If, however, the estimation were to allow for inbreeding, the formula relating the gene frequency to the incidence is more intricate and in-

TABLE 4. Lower and upper 95 percent confidence limits for gene frequency (q_l and q_u), calculated using the usual method and the substitution method in a series of 22 birth incidence examples

Example	No. of affected births	Total no. of births	Gene frequency (q)	Confidence limit calculation method			
				Usual method		Substitution method	
				q_l	q_u	q_l	q_u
AA	1	1,000	3.16	0.06	6.26	0.50	7.46
BB	2	1,000	4.47	1.38	7.57	1.56	8.50
CC	5	1,000	7.07	3.98	10.16	4.03	10.80
DD	10	1,000	10.00	6.92	13.08	6.92	13.56
EE	1	5,000	1.41	0.03	2.80	0.23	3.34
FF	2	5,000	2.00	0.61	3.39	0.70	3.80
GG	5	5,000	3.16	1.78	4.55	1.80	4.83
HH	10	5,000	4.47	3.09	5.86	3.10	6.06
II	20	5,000	6.32	4.94	7.71	4.94	7.86
JJ	50	5,000	10.00	8.62	11.38	8.62	11.48
KK	5	10,000	2.24	1.26	3.22	1.27	3.42
LL	10	10,000	3.16	2.18	4.14	2.19	4.29
MM	25	10,000	5.00	4.02	5.98	4.02	6.07
NN	50	10,000	7.07	6.09	8.05	6.09	8.12
OO	100	10,000	10.00	9.02	10.98	9.02	11.03
PP	5	50,000	1.00	0.56	1.44	0.57	1.53
QQ	10	50,000	1.41	0.98	1.85	0.98	1.92
RR	25	50,000	2.24	1.80	2.67	1.80	2.72
SS	50	50,000	3.16	2.72	3.60	2.72	3.63
TT	100	50,000	4.47	4.03	4.91	4.03	4.93
UU	250	50,000	7.07	6.63	7.51	6.63	7.52
VV	500	50,000	10.00	9.56	10.44	9.56	10.45

cludes a population inbreeding coefficient (11, p. 20). The standard error for this corrected estimate of the gene frequency is not given in standard textbooks, but the substitution method allows for easy estimation of gene frequency confidence limits. In addition, many genetic parameters, such as the population proportion of heterozygotes, are functions of the gene frequency. Although standard error formulae are not easily found, further repeated application of the substitution method again allows for easy confidence interval estimation.

Number needed to be treated

A measure summarizing the results of a clinical trial was described several years ago by Laupacis et al. (12), without any explicit formulation for estimating its confidence limits. The "number needed to be treated" (NNT) is the number of patients that would have to be treated with a trial therapy to prevent one adverse event, and thus it gives clinicians and patients a measure of the effort required to achieve a particular result.

The effect of an insulin-glucose infusion followed by intensive subcutaneous insulin in diabetic patients with myocardial infarction was examined in a randomized controlled trial (13). After 1 year of follow-up, there were 58 deaths in the 306 patients receiving the new therapy (19.0 percent) as compared with 82 deaths in the 314 control patients on standard therapy (26.1 percent). On the basis of these figures, one would expect 261 deaths in 1,000 patients on standard therapy. However, if these patients had received the new treatment, there would have been just 190 deaths. Thus, treatment of 1,000 patients would have prevented 71 deaths (261 − 190 = 71), meaning that 14 patients (1,000/71) would have had to be treated to prevent one death. NNT is then 14. It is easy to see that, in fact, NNT is simply the reciprocal of the absolute risk reduction (the difference in the risk of an event between the treated and control groups).

$$\text{NNT} = \frac{1}{R_1 - R_2} = \frac{1}{0.261 - 0.190} = \frac{1}{0.071} = 14.08, \quad (12)$$

where R_1 is the event risk in the control group and R_2 is the risk in the treated group.

In the original paper describing this measure, published in 1988 (12), it was suggested that confidence limits might be obtained using a complex technique requiring a special computer program (7). Applying this to the above data gives 95 percent confidence limits for NNT of 7.2 and 378.5. The substitution method, however, provides a very simple solution for this problem. Using expressions 1 and A3, the 95 percent confidence limits for the difference between the proportions of events in the treatment and placebo groups ($R_1 - R_2 = 0.071$) are 0.006 and 0.137, respectively. Using the substitution principle on expression 12, the reciprocal of these limits gives the 95 percent limits for the NNT as 7.3 and 166.7. The upper limit is considerably lower than that obtained using the Laupacis et al. method, and this is due to the fact that the lower confidence limit for the risk difference is close to zero. In cases like this, where the significance level is not very high, the stability of the upper limit for NNT may be in question. Since a zero risk difference corresponds to an NNT of infinity, small changes in the lower limit for the risk difference close to zero can result in very large changes in the NNT estimate.

In 1992, Chatellier et al. (14) published a nomogram for estimating NNT from the relative risk and the risk in the control group, also showing how to estimate its confidence limits. Their application is actually equivalent to substitution of the limits for the relative risk into a formula for NNT expressed in terms of RR and R_1. This approach must be less accurate than the method described above, which substituted limits for the absolute risk reduction: Chatellier et al.'s method assumes that R_1 is without sampling variation, and estimating the limits for RR requires more stringent assumptions and a greater degree of approximation than the estimation of limits for the risk reduction.

Table 5 compares the 95 percent confidence limits for NNT calculated on the 21 example tables using the suggestions of Laupacis et al. (12), the suggestions of Chatellier et al. (14), and the substitution method. Here group 1 is taken as the controls, with group 2 representing the treated patients. Expression A2 was employed to estimate the relative risk limits for Chatellier et al.'s method, and the limits for the absolute risk reduction based on expression A3 were employed for the substitution method.

In general, the substitution limits were wider than those of Chatellier et al., which can be explained by the more stringent assumptions underlying the latter. The substitution method agrees well with the complex approach of Laupacis et al., especially for larger sample sizes, though it seems to give a consistently smaller upper limit. As was noted above, a large discrepancy can be expected for results that are close to significance, as in example S.

DISCUSSION

The substitution method of estimating confidence limits described in this paper does not seem to have been proposed explicitly before, although some specific applications are well known. The kernel of the method is expressing the measure for which confi-

TABLE 5. Lower and upper 95 percent confidence limits for the number needed to be treated (NNT_l and NNT_u), calculated using two published methods and the substitution method in 21 example data sets*

Example	No. needed to be treated (NNT)	Confidence limit calculation method					
		Method of Laupacis et al. (12)		Method of Chatellier et al. (14)		Substitution method	
		NNT_l	NNT_u	NNT_l	NNT_u	NNT_l	NNT_u
A	4.2	2.3	35.1	2.7	20.8	2.3	18.6
B	2.8	1.8	6.8	2.1	5.6	1.8	5.6
C	2.1	1.5	3.7	1.7	3.2	1.5	3.2
D	4.2	2.7	10.5	2.9	9.0	2.7	9.2
E	2.8	2.0	4.6	2.2	4.2	2.0	4.3
F	2.1	1.7	2.9	1.8	2.7	1.7	2.8
G	6.3	3.4	69.7	3.9	39.4	3.4	38.4
H	4.2	2.7	10.2	3.1	8.3	2.7	9.0
I	3.1	2.3	5.4	2.6	4.6	2.3	5.0
J	4.2	3.3	5.6	3.5	5.4	3.3	5.5
K	2.8	2.4	3.3	2.5	3.2	2.4	3.3
L	2.1	1.9	2.4	1.9	2.3	1.9	2.3
M	6.3	4.5	10.3	4.8	9.5	4.5	10.0
N	4.2	3.4	5.6	3.6	5.2	3.4	5.5
O	3.1	2.7	3.8	2.8	3.6	2.7	3.8
P	12.5	7.7	34.9	8.6	29.7	7.7	32.6
Q	8.3	6.0	14.0	6.7	12.4	6.0	13.7
R	6.3	5.0	8.7	5.5	7.8	4.9	8.6
S	25.0	13.0	1,105.6	15.3	336.6	13.0	345.5
T	16.7	10.7	43.5	12.4	34.9	10.5	40.5
U	12.5	9.2	21.7	10.5	18.3	8.8	21.4

* See table 1.

dence limits are required as a function of a *single* quantity for which limits are easily obtained. This is often not difficult, and in many cases the usual formula for the measure will be sufficient. It is important to note, however, that the measure must be a function of a *single* parameter in order for the substitution method to work. For example, it is not possible to obtain a confidence interval for a relative risk by using the confidence limits for the two component absolute risks.

In some cases it is necessary to assume that some of the quantities that make up the relevant formula are without sampling variation and are thus essentially constant. If the measure is derived from a contingency table, this will often be equivalent to the assumption that one or both of the margins of the table are fixed, making the analysis conditional on those margins. Thus, for Levin's attributable risk, the prevalence of the condition was taken as constant. This is a common assumption in contingency table analysis. The conditional assumption can sometimes be avoided by judicious choice of the parameter to be substituted, as in the case of the NNT discussed above.

The substitution method will be applicable as long as the relation between the measure and the parameter for which limits are available is fairly simple. Technically, the requirement is that the functional relation be monotonically increasing (or decreasing). This means that if the parameter increases, the measure must always either remain the same or increase (or decrease). (In the latter case, the lower limit for the parameter will give the upper limit for the measure and vice versa.) It is difficult to imagine a practically useful measure in medical or epidemiologic applications for which this condition will not hold.

Although the examples presented in this paper cannot be taken as a formal comprehensive numerical evaluation of the substitution method for confidence interval estimation, there is good agreement with the more established procedures in the cases considered. Even though there is no explicit formula for the confidence limits, the substitution method is without doubt easier to explain and to use. A suitable formula for the measure of interest is all that is required, and the usually incomprehensible expressions for standard errors are avoided entirely.

Another major advantage of the method is that no distributional assumptions are necessary for the sampling distribution of the measure for which the confidence limits are required. If exact confidence limits are known for the underlying parameter (as in the binomial or Poisson cases), the limits for a function of the parameter will also be exact. Thus, there is a distinct advantage to the substitution method even when an alternative exists using known standard error formulae.

For measures that are a function of a single parameter, a Taylor series expansion is often used for interval estimation (6, pp. 91–2). The standard error of the function of a parameter $f(x)$ is given by

$$\mathrm{SE}[f(x)] \approx \left(\frac{df}{dx}\right)\mathrm{SE}(x),$$

where the derivative of f is evaluated at the mean value of x.

Not only is this standard error an approximation but the additional assumption of normality is required to derive the confidence limits for the function using expression 1. The standard error formula for the gene frequency (expression 11) can be derived in this way. The requirements for valid use of the Taylor series expansion method are also more stringent than those for the substitution method, in that the functional relation must be *strictly* monotonically increasing (or decreasing) and must have a nonzero first derivative. (A strictly monotonic function requires that the function always changes as the parameter changes.) Thus, the substitution method can and should always be used instead of a Taylor series expansion.

For the end user, the general approach of the substitution method and its lack of reliance on complex formulae make it clearer what the confidence limits are measuring. It is particularly suitable for "hand" calculations when specialized computer software is not available. The substitution method should be adopted as a practical alternative to complex formulae when performing interval estimation. Even in simple cases, the inherent accuracy of the method suggests that it should replace some standard approaches.

ACKNOWLEDGMENTS

The author thanks Dr. Douglas G. Altman of the Imperial Cancer Research Fund (Oxford, England) and Prof. Marcello Pagano of the Department of Biostatistics, Harvard School of Public Health (Brookline, Massachusetts), for valuable suggestions and encouragement.

REFERENCES

1. Daly LE, Bourke GJ, McGilvray J. Interpretation and uses of medical statistics. 4th ed. London, England: Blackwell Scientific Publications, 1991.
2. Gardner MJ, Altman DG, eds. Statistics with confidence: confidence intervals and statistical guidelines. London, England: British Medical Journal, 1989.
3. Lentner C, ed. Geigy scientific tables. 8th ed. Version 2. Basel, Switzerland: Ciba-Geigy Ltd, 1982.
4. Daly L. Simple SAS macros for the calculation of exact binomial and Poisson confidence limits. Comput Biol Med 1992;22:351–61.
5. Gardner MJ, Gardner SB, Winter PD. Confidence interval analysis microcomputer program manual. London, England: British Medical Journal, 1989.
6. Armitage P, Berry G. Statistical methods in medical research. 3rd ed. Oxford, England: Blackwell Scientific Publications, 1994.
7. Thomas DG, Gart JJ. A table of exact confidence limits for differences and ratios of two proportions and their odds ratios. J Am Stat Assoc 1977;72:73–6.
8. Reilly M, Daly L, Hutchinson M. An epidemiological study of Wilson's disease in the Republic of Ireland. J Neurol Neurosurg Psychiatry 1993;56:298–300.
9. Walter SD. The estimation and interpretation of attributable risk in health research. Biometrics 1976;32:829–49.
10. Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York, NY: John Wiley and Sons, Inc, 1981.
11. Emery AEH. Methodology in medical genetics: an introduction to statistical methods. Edinburgh, Scotland: Churchill Livingstone, 1976.
12. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med 1988;318:1728–33.
13. Malmberg K. Prospective randomised study of intensive insulin treatment on long term survival after acute myocardial infarction in patients with diabetes mellitus. BMJ 1997;314:1512–15.
14. Chatellier G, Zapletal E, Lemaitre D, et al. The number needed to treat: a clinically useful nomogram in its proper context. BMJ 1996;312:426–9.

APPENDIX

Some common standard error formulae which were employed in the derivation of results presented in this paper are shown below.

For a binomial proportion (p) in a sample size of n, the standard error is calculated as

$$\sqrt{p(1-p)/n}. \qquad \text{(A1)}$$

For the natural logarithm of the relative risk ($\log_e$ RR) (a, b, c, and d are table entries—see table 1 in the text), the standard error is calculated as

$$\sqrt{\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}}. \qquad \text{(A2)}$$

For the difference between two risks ($R_1 - R_2$) in sample sizes of n_1 and n_2, the standard error is calculated as

$$\sqrt{(R_1(1-R_1)/n_1) + (R_2(1-R_2)/n_2)}. \qquad \text{(A3)}$$