HEALTH ECONOMICS, VOL. 4: 373–381 (1995)

ECONOMIC EVALUATION

CONFIDENCE INTERVALS FOR COST/EFFECTIVENESS RATIOS

PETER WAKKER AND MARC P. KLAASSEN
Medical Decision Making Unit, University of Leiden, The Netherlands

SUMMARY

The reduction of costs is becoming increasingly important in the medical field. The relevant topic of many clinical trials is not effectiveness *per se*, but rather cost-effectiveness ratios. Surprisingly, no statistical tools for analyzing cost-effectiveness ratios have been provided in the medical literature yet. This paper explains the gap in the literature, and provides a first technique for obtaining confidence intervals for cost-effectiveness ratios. The technique does not use sophisticated tools to achieve maximal optimality, but seeks for tractability and ease of application while still satisfying all formal statistical requirements.

KEY WORDS—cost-effectiveness ratio; healthy economics; confidence interval

The reduction of costs has become increasingly important in the medical field.[1–6] The relevant question is no more exclusively: 'Which treatment yields most effects?' but, more often, 'Which treatment takes the least amount of money to yield an appropriate level of effectiveness?'[7] That is, the decisive quantity is not effectiveness *per se*, but rather the cost-effectiveness (C/E) ratio.

Traditional investigations of effectiveness yield point estimates, surrounded by an assessment of the uncertainty comprised in the estimation. For uncertainty assessment, two techniques are relevant. The first is sensitivity analysis, demonstrating how sensitive the relevant quantities are to variations in inputs. Secondly, statistical analyses are adopted to assess the probabilistic uncertainty inherent in the data and thus in the conclusions.

Recent studies of C/E ratios similarly provide point estimates, and assessments of uncertainty through sensitivity analysis; the latter has been recommended in several studies.[1,7–10] However, almost no statistical analyses of C/E ratios have been reported in the literature to date. There obviously is a need for such analyses. A description of the level of confidence with which one can conclude that some critical level for the C/E ratio will not be exceeded, is extremely relevant to optimal medical decision making and to the proper control of medical expenses. Further motivations for statistical analyses of C/E ratios have been provided in the literature.[6,11–13]

Let us next suggest an explanation for the almost complete absence of statistical analyses of C/E ratios. This absence can be due to the complicated nature of the probability calculus of C/E ratios. Under the common assumption that costs and effects are normally distributed, C/E ratios are 'Cauchy' distributed. Cauchy distributed variables can take very extreme values, i.e., 'they have heavy tails'. This phenomenon is mainly caused by the possibility of the denominator E getting close to 0. Therefore, means of Cauchy distributions turn out not to exist. Whereas for most common distributions the averages converge more and more

Address for correspondence: Dr. P. Wakker, Medical Decision Making Unit, University of Leiden (AZL #43), P.O. Box 9600, Leiden, 2300 RC, The Netherlands.

CCC 1057–9230/95/050373–09

Received 20 October 1994
Accepted 18 May 1995

to a point value as the number of observations increases, and this is the everyday experience, such a convergence does not occur for averages of Cauchy distributed variables. Observed average values will not be stable but continue to exhibit extreme variations. It can be demonstrated that, paradoxically, the average C/E ratio over any number of observations exhibits the same degree of variation as one single observation![14] Thus the estimation of average C/E ratios is not possible. As the traditional methods for hypothesis testing and confidence intervals are based on average estimations, surrounded by a number of standard-deviations, these traditional methods cannot be invoked here.

At this stage it might be concluded that C/E ratios are simply not suited to statistical analyses and indeed this idea may underly the absence of such analyses in the literature. This conclusion would imply a most inconvenient state of affairs, given the importance of statistical analyses. Hence this paper provides tools that do permit a statistical analysis of C/E ratios. While the estimation of C/E ratios is problematic indeed, we show that confidence intervals (and hence hypothesis tests) can still be developed. Even more, it is our intention to demonstrate that these tools can be obtained by elementary means, requiring few mathematical calculations. Thus, in a sense, our approach will be 'satisficing'[15] in providing useful tools that can readily be applied, without yet seeking for maximal optimality that will be at the expense of accessibility. The search for optimal but sophisticated techniques is a topic for future research. Let us emphasize that the confidence intervals derived in this paper do satisfy all formal statistical requirements for confidence intervals, i.e. they are confidence intervals in every formal mathematical sense.

Our paper builds on O'Brien, Drummond, and Labelle,[11] who initiated the research on confidence intervals for C/E ratios, and Van Hout *et al.*,[13] abbreviated [H] hereafter, who pointed out the problems in the probability calculus for C/E ratios and proposed a pragmatic alternative to confidence intervals. These studies are discussed in more detail below.

Let us end this introduction by mentioning a remarkable list of papers from the forties and fifties, i.e. Fieller[16] and the references therein. (These references were pointed out to us by a referee.) These papers assume the 'fiducial' approach to statistics, an interesting alternative to the Bayesian approach and the nowadays customary 'classical' approach. The fiducial approach was initiated by Fisher[17] but is rarely used today. For that approach, Fieller derives 'fiducial regions' for quotients of means of normal distributions, for the special case where the distributions are independent. Fiducial regions are analogs of confidence intervals.

THE BASIC QUANTITIES IN OUR ANALYSIS

Because we want to convince the reader, beyond any ambiguity, that our confidence intervals satisfy all formal statistical requirements, we give a precise analysis. To that end it is useful to introduce some helpful abbreviations and notations to simplify the exposition. The system is simple. The most important notions, treatments, costs, effects, numbers (of patients in the randomized clinical trial), ratios (of costs and effects) are abbreviated by their first letter. Observable values are denoted by regular Roman letters, the corresponding unknown parameters ('population means') that are not observable but one wants to estimate on the basis of the observed values, are denoted by the corresponding Greek letters. Finally, the values that were just introduced appear both in a traditional treatment, denoted by a subscript 0, and in a new alternative treatment that is indexed by a subscript 1.

Let us now describe these quantities in more detail. We assume that a decision must be made whether a new treatment T_1 can replace a customary treatment T_0. To this effect, for N_0 patients the costs and effects of treatment T_0 have been observed, and for N_1 patients the costs and effects of treatment T_1.

C_0: average cost of T_0;
E_0: average effect of T_0;
C_1: average cost of T_1;
E_1: average effect of T_1.

Thus the total sum of observed effects over the group of patients that received the standard treatment T_0 is $N_0 * E_0$, the total sum of observed costs over the group of patients that received the alternative treatment is $N_1 * C_1$, etc. We make the common assumptions that all variables are normally distributed. The means are given below; they can be interpreted as means and averages per patient over the entire population and determine

the sampling probabilities.

γ_0: The mean cost of treatment T_0;
ε_0: The mean effect of treatment T_0;
γ_1: The mean cost of treatment T_1;
ε_1: The mean effect of treatment T_1.

These are the (unobserved) parameter values. Thus the expectation of the total effect over the group of patients that received T_1 is $N_1\varepsilon_1$, etc. The belonging standard deviations are also unknown and relevant for the probability distribution, but we need not introduce symbols to denote them. The distributions of all variables are mutually independent, with the exception of the costs and effects within one patient. It is obvious that between these two variables there will be dependencies. Following [H],[13] we assume that these two variables have a joint normal distribution with, possibly, nonzero correlations.

We assume that the relevant cost-effectiveness ratio in the analysis, denoted by ρ, concerns here the incremental costs and effects, i.e. it concerns

$$\rho = \frac{\gamma_1 - \gamma_0}{\varepsilon_1 - \varepsilon_0}.$$

This is the quantity we want to know as precisely as possible. A natural notation for incremental quantities is obtained by dropping subscripts, i.e.

$$C = C_1 - C_0;$$
$$E = E_1 - E_0;$$
$$\gamma = \gamma_1 - \gamma_0;$$
$$\varepsilon = \varepsilon_1 - \varepsilon_0.$$

Obviously, $\rho = \gamma/\varepsilon$. The *observed* ratio is:

$$R = C/E.$$

THE PROBLEM FOR C/E RATIO-CONFIDENCE INTERVALS, AND THE PRAGMATIC ALTERNATIVE SOLUTION BY VAN HOUT *ET AL.*

As the precise meaning of a confidence interval is not elementary, let us briefly repeat it here. A 95% confidence interval for the C/E ratio is a prescription that describes, for all possibility observed costs and effects, an interval of C/E values satisfying the following condition: For all population means γ_1, γ_0, ε_1, ε_0, (this similarly applies to all standard deviations and correlations between costs and effects), there is at least a 0.95 probability that the observed costs and effects are such that the interval, prescribed for these costs and effects, contains the true ratio $\rho = (\gamma_1 - \gamma_0)/(\varepsilon_1 - \varepsilon_0)$. Note here that the probability refers to the sampling variability in the observed costs and effects. The important point to note for the sequel is, however, that the probability refers to events that describe whether:

— The *true* ratio $\rho = (\gamma_1 - \gamma_0)/(\varepsilon_1 - \varepsilon_0)$ is contained in an interval.

Let us next, briefly, describe the traditional way for obtaining confidence intervals. We consider the case of interest in this paper, i.e. a one-sided confidence interval, bounded from above, for ρ. The traditional approach could be applied if expectation and variance of ρ were to exist, and if $R = C/E$ were a proper (e.g., unbiased) estimator for ρ. Then one would take the observed value R as a starting value, and next take an interval $(\leftarrow, R + x]$ (bounded from above, unbounded from below), where $x > 0$ has been chosen sufficiently large to ensure the right level of confidence; further x is chosen sufficiently small to guarantee that the confidence interval is not so wide as to provide no useful information. The size of x traditionally depends on the standard deviation of R, e.g., if R were normally distributed then x could be 1.65 times the (estimated) standard deviation of R to obtain a 95% confidence interval. In general, if R were not exactly normally distributed, but it at least had finite expectation and standard deviation, then by the 'central limit theorem' normal approximations could be invoked. After a sufficient number of observations, confidence intervals could then still be obtained in the way described above.

O'Brien, Drummond, and Labelle,[11] who initiated the discussion of C/E ratio confidence intervals, stay close to the classical approach. They use a method based on the Taylor approximation that, at least for many probability distributions, gives estimations for means and variances; based on these estimations, confidence intervals can be formulated in the usual manner. The method can be used as soon as means and variances of C/E ratios do exist.

[H][13] pointed out, however, that means and variances of C/E ratios do not exist under the common distributional assumptions for costs and effects, and introduced a pragmatic alternative approach for constructing intervals for C/E ratios.

We do not describe the pragmatic approach of [H][13] in detail here. In short, it is as follows. First, from the actually observed costs and effects, estimates are obtained for the true values of costs and effects (as well as standard deviations and correlations); these estimates can be obtained by well-known methods. Next an interval, a C/E acceptability curve, is constructed that satisfies the following condition: If the population means were identical to the estimated values, and the randomized clinical trial (RCT) were repeated with the same number of patients, then the probability that the newly sampled average cost C' and the newly observed effect E' would yield a ratio C'/E' contained within the constructed C/E-acceptability curve, is at least 95%.

The procedure proposed by [H][13] is a pragmatic alternative to the traditionally derived confidence intervals. It provides an index for the degree of precision with which the C/E ratio can be estimated, and thereby it provides a tool for deciding on acceptance or rejection of a new treatment. It should, however, be understood that acceptability curves are not confidence intervals in a formal statistical sense, and the 95% level is not a confidence level. For instance, the probability of 95% does not refer to the event that the *true* ratio is contained in the 'curve', but it refers to an event where an *observed* ratio in a repeated RCT would be contained in the 'curve'. Therefore, the 95% level must be interpreted differently than the traditional 95% confidence levels, and [H][13] use different formulations. They write: 'One may conclude that the C/E ratio ... is acceptable with 95% certainty when a limit to cost effectiveness is used of 42,000.'

Acceptability curves satisfy desirable conditions; for instance, the acceptability curve will be tighter around the true value as the number of observed patients increases. Therefore acceptability curves can be useful alternatives to confidence intervals. A detailed discussion, and comparison to confidence intervals, lies beyond the scope of this paper.

We propose another, more formal, approach. Our approach is less pragmatic and more conservative than the one by [H],[13] but in return provides intervals that do satisfy all formal statistical criteria; therefore, these intervals are confidence intervals in the traditional statistical sense, and the associated confidence levels can be given the according probabilistic interpretations. The basic idea of our approach is elementary, and easy to implement.

CONFIDENCE INTERVALS FOR C/E RATIOS

It was already observed in the literature that for a decision, based on C/E ratios, some cases should be distinguished, depending on whether costs and effects are positive or negative.[1,11,13] Following this observation, we propose that decisions are taken in the following manner, based on the estimates of γ, the true difference in costs between the two treatments, and of ε, the true difference in effect between the two treatments. Note here that the statistical analysis of the difference of two normal distributions (the observed effect E, estimating ε, is such a difference, and so is the observed cost C, estimating γ) is well-known; it is used below without further discussion.

Case 1

The data do not suffice to conclude if ε is positive, negative, close to 0, or remote from 0. Then a reliable decision is not available (unless γ would very clearly be extremely positive or negative). In particular: the decision should not be based on the C/E ratio.

Case 2

The data suffice to reliably conclude that ε is close to 0. Then, again, the decision should not be based on the C/E ratio. Now the decision should be based on the cost estimate, and the new treatment should be accepted only if it is clearly less expensive ($\gamma < 0$).

Case 3

The data suffice to reliably conclude that ε is positive and not very close to 0. Then the decision can be based on the C/E ratio. The new treatment should be accepted if ρ falls below some critical values.

Case 4

The data suffice to reliably conclude that ε is negative and not very close to 0. Then the decision can be based on the C/E ratio. The new treatment should be accepted if ρ exceeds some critical value.

We think that in Cases 1 and 2 above, if the effects cannot be clearly separated from 0, then

C/E ratios are not proper tools for decisions. For effects close to 0, the behaviour of the C/E ratio becomes unreliable and can be massively affected by minor errors. Therefore we consider only the Cases 3 and 4. Case 3 is analyzed in detail below. For Case 4, similar observations hold. Therefore we assume henceforth:

Assumption

The statistical null hypothesis H_0: $\varepsilon \leq 0$ can be rejected at a significance level lower than 2.5%.

In this case we propose a simple device for obtaining a 95% confidence interval for ρ. Formally, confidence intervals should be defined for all possibly observed values of costs and effects, hence also for those not satisfying the above assumption, i.e. also for cases 1, 2 above. For these, we formally define the associated confidence intervals as the 'noninformative confidence set', i.e., the entire real line. Such confidence intervals correspond to 'no decision' as it was described in the text, which seems the rational conclusion from C/E ratios if effects are small. This construction maximally meets any restriction to levels of confidence and does not affect the reasoning in the next section. A similar formal definition is used in the end of Section 5 in Fieller.[16] Existence of a positive e in Step 2 below is guaranteed by the above assumption, as will be explained shortly.

STEP 1. Take a 97.5% confidence interval $(\leftarrow, c]$ for γ.
STEP 2. Take a 97.5% confidence interval $[e, \rightarrow)$ for ε, where $e > 0$.
STEP 3. $(\leftarrow, c/e]$ is a 95% confidence interval for ρ.

DERIVATION OF CONFIDENCE LEVEL

This section proves the confidence claim of 95% made in Step 3 of the above section, and can be skipped without interrupting the flow of the paper. Note that by the above assumption there must exist a strictly positive value e such that H_0: $\varepsilon \leq e$ can be rejected at the 2.5% level. Then indeed $[e, \rightarrow)$ can be taken as a 97.5% confidence interval for ε. Next let us explain the claim in Step 3, that $(\leftarrow, c/e]$ is a 95% confidence interval.

(1a) For each set of population means, the ('sampling') probability that the confidence interval $[\leftarrow, c)$ does not contain γ is at most 2.5%: this property defines the 97.5% confidence interval $[\leftarrow, c)$.
(1b) For each set of population means, the probability that the confidence interval $[e, \rightarrow)$ does not contain ε is at most 2.5%: this property defines the 97.5% confidence interval $[e, \rightarrow)$.

Combining the above two statements we see:

(2) For each set of population means, the probability that either $c < \gamma$ or $e > \varepsilon$ or both, is at most $2.5\% + 2.5\% = 5\%$.
(3) For each set of population means, the probability that both $c \geq \gamma$ and $e \leq \varepsilon$ is at least 95%.
(4) For each set of population means, the probability that $c/e \geq \rho$ is at least 95%.
(5) $(\leftarrow, c/e]$ is a 95% confidence interval for ρ.

DISCUSSION

The above procedure has provided, in an elementary manner, a confidence interval that satisfies all the formal statistical requirements. The method can easily be extended to other distributions for costs and effects than the normal distribution, and is applicable irrespective of the correlation between costs and effects. In return for its simplicity and general validity, however, the method is conservative. That is, it 'gives up power' by taking large confidence intervals. If much data is available, little is lost. In other cases, however, refinements of the procedure may be warranted. This is a topic for future research.

One obvious way to refine our procedure is as follows. To obtain a 95% confidence interval for the C/E ratio, one can take other confidence levels than 97.5% for the $(\leftarrow, c]$ interval and for the $[e, \rightarrow)$ interval. For instance, one can, on the basis of prior knowledge, take a 99% confidence level for the $(\leftarrow, c]$ interval and a 96% confidence level for the $[e, \infty)$ interval. Thus one can find the optimal level of distributing the 5% 'unconfidence' over costs and effects, where optimal means that the most tight confidence interval with the lowest upper bound c/e is obtained. This is discussed further in the simulation below (see Table 1).

First we show that our method is conservative, in particular for the case of few data containing much variation. Such a case was analyzed by

[H].[13] [H][13] studied the cost-effectiveness ratio for data collected in a phase II trial addressing the value of IL-1ra in treating sepsis. As usual, costs and effects relate here to incremental costs and effects. Both treatments, T_0 the 'placebo treatment', and T_1 the 'high-dose treatment', concerned groups of 25 ($=N_0=N_1$) patients. These numbers are quite small, and in addition the distance of averages from 0, measured in standard deviations, were large (approximately 2.27 for costs and 0.24 for effects, i.e. number of lives saved). Although the number of patients is rather small, the results suffice to significantly distinguish effects from 0, so a C/E analysis can be meaningful. The 95% confidence interval that results from our method is $(\leftarrow, 224{,}514]$, which is much more prudent (i.e., wider) than the C/E acceptability curve $(\leftarrow, 42{,}000]$ obtained in [H].[13] Our method requires more data before high levels of accuracy are claimed.

SIMULATION

To further illustrate the features of our method, and show that it provides good accuracy if more data is available, we present the results of a simulation, illustrated in Tables 1 and 2.

The default values in the simulation are: $N_1=40$, $N_0=60$, $\gamma_1=40{,}000$, $\gamma_0=30{,}000$ (hence $\gamma=10{,}000$), $\varepsilon_1=60$, $\varepsilon_0=50$ (hence $\varepsilon=10$). Further we assume that $\sigma_1=5500/\sqrt{2}$ and $\sigma_0=5000/\sqrt{2}$ are default standard deviations for costs, $\tau_1=1.2/\sqrt{2}$ and $\tau_0=1/\sqrt{2}$ are default standard deviations for effects, and the correlations between costs and effects are $\lambda_1=-0.2$, $\lambda_0=-0.3$. In all cases considered, the true C/E ratio is $10{,}000/10=1{,}000$. Deviations from the default values have been indicated in the left columns of the tables.

In the tables, the '95%-CI-ub' column describes the upper bound of the confidence interval that our method provides for the C/E ratio, i.e. the interval is $(\leftarrow, c/e]$ where c/e is the value in the column. Obviously, for the true C/E ratio 1,000 to be contained in the confidence interval, c/e should be greater than 1,000. The closer c/e is to 1000, the more accurate is the estimate.

The columns $\%C$ and $\%E$ describe the allocation of unconfidence liberty over costs and effects. For instance, in the first row in Table 2, describing the default values, we used 96% as the level of confidence for $(\leftarrow, c]$, and for $[e, \rightarrow)$ the level of confidence was 99%. These numbers are further discussed below, and still yield a 95% level of confidence for the C/E ratio.

One more aspect must be explained. Our simulation results are not based on single simulations, but in each case on 10,000 independently repeated simulations, and the c/e values in the 95%-CI-ub column are averages over those 10,000 repetitions. We also give the standard deviations over those 10,000 repetitions of the c/e values in the adjacent column. Finally, the column Err.% ('error percentage') gives the percentage of cases in those 10,000 repetitions of the simulation where the confidence interval was distorted to such a degree that the true C/E ratio 1,000 was not contained in the confidence interval. For a 95% confidence

Table 1.

non-default	%C	%E	95%-CI-ub	S.D.	Inacc.	Err.%
	99.5%	95.5%	1232.12	87.09	1.232	0.31%
	99.0%	96.0%	1213.71	86.77	1.214	0.58%
	98.0%	97.0%	1195.00	86.57	1.195	1.02%
	97.5%	97.5%	1189.09	86.56	1.189	1.22%
	97.0%	98.0%	1184.75	86.60	1.185	1.40%
	96.0%	99.0%	1179.84	86.87	1.180	1.71%
	95.5%	99.5%	1180.34	87.22	1.180	1.71%
$\tau_1=5.5/\sqrt{2}$, $\tau_0=5/\sqrt{2}$	99.5%	95.5%	1386.73	170.28	1.387	0.34%
$\tau_1=5.5/\sqrt{2}$, $\tau_0=5/\sqrt{2}$	99.0%	96.0%	1371.64	169.87	1.372	0.44%
$\tau_1=5.5/\sqrt{2}$, $\tau_0=5/\sqrt{2}$	98.0%	97.0%	1363.53	171.27	1.364	0.61%
$\tau_1=5.5/\sqrt{2}$, $\tau_0=5/\sqrt{2}$	97.5%	97.5%	1364.81	172.68	1.365	0.62%
$\tau_1=5.5/\sqrt{2}$, $\tau_0=5/\sqrt{2}$	97.0%	98.0%	1369.59	174.71	1.370	0.61%
$\tau_1=5.5/\sqrt{2}$, $\tau_0=5/\sqrt{2}$	96.0%	99.0%	1392.62	181.59	1.393	0.45%
$\tau_1=5.5/\sqrt{2}$, $\tau_0=5/\sqrt{2}$	95.5%	99.5%	1420.96	188.98	1.421	0.36%

Table 2.

non-default	%C	%E	95%-CI-ub	S.D.	Err.%
	96.0%	99.0%	1179.84	86.87	1.71%
$\tau_1 = 5.5/\sqrt{2}$, $\tau_0 = 5/\sqrt{2}$	98.0%	97.0%	1363.53	171.27	0.61%
std/8	95.5%	99.5%	1021.57	10.36	1.85%
$N_1 = N_0 = 200$	96.0%	99.0%	1087.60	41.63	1.47%
$N_1 = 10$, $N_0 = 15$	96.0%	99.0%	1368.15	190.11	2.24%
$\lambda_1 = \lambda_0 = -0.95$	96.0%	99.0%	1180.23	100.85	3.30%
$\lambda_1 = \lambda_0 = 0$	96.0%	99.0%	1179.29	81.91	1.38%
$\lambda_1 = \lambda_0 = +0.95$	96.0%	99.0%	1177.31	63.04	0.29%

interval, this number should never exceed 5%; indeed it never does in the simulations.

The first row in Table 2 describes the results for the default parameters in our simulation. The error percentage is 1.71%, which is clearly below the upper bound of 5%. The interpretation and discussion of the percentage is somewhat subtle, and is as follows. The value of 1.71% may suggest that our method is too prudent, i.e. that a level of confidence higher than 95% could have been claimed, possibly even $100 - 1.71\% = 98.29\%$. Such a claim cannot, however, be based on the simulation. It should be kept in mind that in practice the confidence interval is derived solely from the observed costs and effects, without knowledge of the true parameters (population averages), in particular of the true C/E ratio 1000. The confidence restriction should not only be satisfied for one set of parameters as verified in our simulation, but for *all* possible values of the population averages. Therefore the error percentage 1.71% only yields an upper bound of 98.29% to the true level of confidence, and the latter may necessarily be lower. The error percentage 3.3% in the column with $\lambda_1 = \lambda_0 = -0.95$ shows that in general not much more confidence can be claimed for the confidence intervals than 95%.

Table 1 illustrates different allocations of the 5% 'unconfidence' over costs and effects. It shows that for the default values the optimal allocation, giving optimal accuracy, is obtained under a 96/99 allocation, where most of the unconfidence liberty is assigned to costs. The table also presents different allocations of 'unconfidence' if other, larger, standard deviations τ_1 and τ_0 are taken for effects. The 'true' average costs γ and average effects ε are now both at a distance of two standard deviations from 0. It turns out that the optimal allocation is now 98/97, i.e. somewhat more unconfidence liberty is allocated to the effects.

In general, it can be expected that most unconfidence liberty should be allocated to the variable (cost, effect) that is closest to 0 in terms of standard deviation units. The reason is that small absolute distortions have the largest relative effects for that variable, and the relative effect is relevant in the quotient C/E. This might suggest that 97.5/97.5 would be the optimal allocation for the τ values that deviate from the default values. However, it is optimal to assign some more liberty to effects, because their critical levels concern *under*estimations, and for those absolute errors generate larger relative errors than for overestimations as leading to critical values for costs. This explains the 98/97 finding above where it was optimal to allocate some more unconfidence liberty to effects.

In Table 2, only the optimal allocations of unconfidence liberty have been given, and these have been given for some deviations from the default parameter values. In the row indicated by std/8, all the standard deviations σ_1, σ_0, τ_1, τ_0 have been divided by 8, so that costs and effects are farther remote from 0, when measured in standard-deviation units. It means that there is less relative variation in the data, and the upper bound of the interval is more accurate (1,022 instead of 1,180). The effect of sampling variability is also reflected in confidence intervals based on different numbers of observed patients. This may be seen by taking $N_1 = N_0 = 200$ (see again Table 2). The greater number of observations can be expected to give more reliable information, and indeed the upper bound then is more accurate, i.e. it is 1,088. If we take a smaller number of observations ($N_1 = 10$ and $N_0 = 15$ in Table 2) then the upper bound is 1,368, so it is larger and hence worse. The change in standard deviation and the change in

number of observations have a smaller effect on the error percentage.

The last rows in Table 2 illustrate variations of correlations between costs and effects. The correlations chosen represent extremes. Results from other combinations in the range −0.95 to 0.95 would give results between those shown. The changes in correlation do not affect the accuracy, but determine the error percentage. The most critical case for the error percentage occurs for highly negative correlation. This can be explained as follows. If the correlation between costs and effects is negative, then high costs go together with low effects. Both higher costs and lower effects lead to an increase of the C/E ratio. That is, their effects on the C/E ratio reinforce each other. Thus sampling variability then favours an increase in the C/E variability due to negative correlation. This is reflected in the error percentages, that are larger as correlations are more negative. Note that for the very negative correlations, −0.95, the observed error percentage is 3.3%, which is not very far from 5%. This is relevant to the question of whether our method is overprudent. It illustrates that, in general, without further information about correlations etc., the maximal level of confidence that can be claimed is not much higher than our 95%. It also suggests that, based on estimations of the correlations, our method could be refined. If the estimates of correlations are positive, higher levels of confidence may be claimed, or lower upper bounds to the confidence intervals taken. Obviously, the described refinement would make our method less tractable, and is a topic for future research.

Summarizing:

— The optimal accuracy (most narrow interval) is obtained if most of the 'unconfidence' liberty is allocated to the variable that, in terms of units of standard deviations, is closest to 0 (where somewhat more unconfidence liberty can be given to effects).
— The accuracy is improved if:
  (i) The number of patients is increased;
  (ii) The standard deviations are decreased.
— The error probability decreases if:
  (iii) The correlation between costs and effects becomes more positive.

CONCLUSION

This paper has, to the best of our knowledge, given the first method for constructing confidence intervals for cost/effectiveness ratios. The method is conservative if few data are available, but is elementary and easy to apply. Our paper also suggests that cost/effectiveness ratios are useful quantities only if it can be reliably concluded from the data whether the effects of a new treatment are strictly more positive than the effects of a traditional treatment, or whether the effects are strictly more negative.

REFERENCES

1. Drummond, M. F., Stoddart, G. L., Torrance, G. W. *Methods for the economic evaluation of health care programmes*. Oxford: Oxford University Press, 1987.
2. Leaf, A. Cost effectiveness as a criterion for medicare coverage. *New England Journal of Medicine*, 1989; **321**: 898–900.
3. Commonwealth of Australia. *Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee*, Woden (ACT) Dept. of Health, Housing and Community Services, Canberra, AGPS 1992:13.
4. Ontario Ministry of Health. *Guidelines for the preparation of economic analysis to be included in submission to drug programs branch for listing in the Ontario Drug Benefit Formulary/Comparative Drug Index*, Ministry of Health, Toronto, 1991.
5. Rutten, F. F. H. and Bonsel, G. J. High cost technology in health care: a benefit or a burden? *Social Science and Medicine*, 1992; **4**: 567–77.
6. Briggs, A., Sculpher, M. and Buxton, M. Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis. *Health Economics*, 1994; **3**: 95–104.
7. Adams, M. E., McCall, N. T., Gray, D. T., *et al.* Economic analysis in randomised control trials. *Medical Care*, 1992; **30**: 231–43.
8. Eisenberg, J. M. A Guide to the economic analysis of clinical practices. *Journal of the American Medical Association*, 1989; **262**: 2879–86.
9. Luce, B. R. and Elixhauser, A. *Standards for socioeconomic evaluation of health care products and services*. Berlin: Springer, 1990.
10. Detsky, A. S. Guidelines for economic analysis of pharmaceutical products: A draft document for Ontario and Canada. *PharmacoEconomics*, 1993; **3**: 354–61.
11. O'Brien, B. J., Drummond, M. F. and Labelle, R. J. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Medical Care*, 1994; **32**: 150–63.
12. Drummond, M. F. and O'Brien, B. J. Clinical importance, statistical significance and the assess-

ment of economic and quality-of-life outcomes. *Health Economics*, 1992; **2**: 205–12.
13. Hout, B. A. van, Al, M. J., Gordon, G. S. and Rutten, F.F. H. Costs, effects and C/E-ratios alongside a clinical trial. *Health Economics*, 1994; **3**(5): 309–319.
14. Johnson, N. and Kotz, S. *Continuous univariate distributions 2*. New York: Wiley, 1970.
15. Simon, H. A. A behavioral model of rational choice. *Quarterly Journal of Economics*, 1955; **69**: 99–111.
16. Fieller, E. C. Some problems in interval estimation. *J.R. Statis. Soc. Ser. B*, 1954; **16**: 175–185.
17. Fisher, R. A. *Statistical methods for research workers* (10th ed.). London: Oliver & Boyd, 1946.