Volume 13 • Number 4 • 2010
VALUE IN HEALTH

The Effect Size in Uncertainty Analysis

Jan J. Barendregt, MA, PhD

School of Population Health, University of Queensland, Herston, QLD, Australia

ABSTRACT

Objective: In model-based health economic evaluation, uncertainty analysis is often done using parametric bootstrapping. This requires specifying probability distributions for the model variables that are uncertain.
Methods: The effect size of the intervention is often expressed as a relative risk, and the standard assumption for a relative risk is that it has a lognormal distribution with the natural log of the relative risk and its standard error as parameters. The problem with this assumption is that the mean of the bootstrap draws from the lognormal distribution is always higher than the relative risk.
Results: This article looks at two ways to correct for this effect and discusses their advantages and drawbacks. Both methods return a bootstrap mean equal to the relative risk, but the first returns an uncertainty interval that is narrower than the corresponding confidence interval, although the second method retains the corresponding width.
Conclusions: The article concludes that the second correction method is preferred.
Keywords: parametric bootstrap, relative risk, uncertainty analysis.

Introduction

Uncertainty analysis (also known as probabilistic sensitivity analysis) is fast becoming a mandatory requirement for modeled health economic evaluation. When there is uncertainty in both costs and health benefits, it is not possible to derive an exact analytical expression for the uncertainty interval around an incremental cost-effectiveness ratio [1].

Several approximation methods exist, but they leave rather a lot to be desired. The Taylor series and confidence ellipse methods overestimate the uncertainty interval, while of the two varieties of the confidence box method, one overestimates it and the other underestimates it [1]. Fieller's theorem does produce an analytical expression, but at the cost of assuming that both numerator and denominator are normally distributed, a questionable assumption in particular for costs [1].

With analytical methods found wanting, numerical methods are an alternative. When patient level data are available, nonparametric bootstrapping is the gold standard [1]. Many economic evaluations, however, are partially or wholly model based, and in that case, the uncertainty analysis employs parametric bootstrapping.

In parametric bootstrapping, pivotal variables in the model are replaced by appropriate distributions and parameters, and the model is recalculated many times, while each time a value is randomly drawn from each distribution. This results in a distribution of the outcomes, which allows to derive various ways to quantify the uncertainty, such as uncertainty intervals and cost-effectiveness acceptability curves.

The main issue with parametric bootstrapping is to decide what, given a specific model variable, constitutes an appropriate distribution and parameters. A criticism of the method is that these choices are essentially arbitrary, which, if true, would render it rather less useful. And in fairness, it must be said that for many early applications, often characterized by a proliferation of triangular distributions, the critique definitely holds water.

Address correspondence to: Jan J. Barendregt, School of Population Health, University of Queensland, Herston Road., Herston, QLD 4006, Australia. E-mail: j.barendregt@sph.uq.edu.au
10.1111/j.1524-4733.2009.00686.x

Recently, Briggs et al. have argued, however, that the choice is (and should) not be arbitrary and that the type of model variable and the way it is estimated gives strong guidance [2]. By using this guidance, they argue, the quality and credibility of the uncertainty analysis is enhanced.

Taking their cue, I propose the following three desirable properties for the distribution that represents the uncertainty of a model variable:

1. The type of distribution is based on the kind of variable, and the way the point estimate and confidence interval (CI) were obtained. This follows the Briggs et al. recommendation.
2. The distribution returns a mean that is equal to the point estimate. This property ensures consistency of the central model outcome and the results from the uncertainty analysis.
3. The distribution returns an uncertainty interval that replicates the CI of the point estimate. This property ensures that the modeled uncertainty neither underrepresents nor overrepresents the uncertainty implied by the CI.

In most, but not in all, cases, the first property will imply the other two. An important exception is the effect size, which is usually expressed as a relative risk. An appropriate choice of probability distribution (and its parameters) for this variable follows from the way epidemiologists calculate the CI for the relative risk. To obtain approximate large-sample CIs, the standard assumption in the epidemiological literature is that the natural log of the relative risk has a normal distribution [3], which is equivalent to saying the relative risk has a lognormal distribution.

The lognormal distribution has two parametrizations: one with parameters mean and standard deviation, the second with parameters μ and σ. The epidemiological formulation corresponds to a lognormal distribution of the second kind. Using this distribution for parametric bootstrapping has a well-known disadvantage, however. As the lognormal distribution is skewed, its mean, given a relative risk as the first parameter, is higher than that relative risk [2]. The consequence is that the mean effect size of the uncertainty analysis is not equal to the point estimate, and systematically so: it is always bigger. When the effect size is

1098-3015/10/388 388–391

Table 1 Two-by-two table with exposed, unexposed, cases, and people at risk

	Exposed	Unexposed
Cases	a	b
People at risk	N_1	N_0

expressed as a relative risk <1, the mean of the randomly drawn effect sizes will therefore be closer to 1, producing a less favorable mean outcome than the result from the point estimate. This article examines two ways to correct for this effect, gives an example, and discusses the pros and cons of both corrections.

Methods

The estimation of the CI of a relative risk (RR) uses the natural log of the RR, denoted by ln(RR) and its standard error (SE), denoted by SE[ln(RR)]. Both RR and the SE[ln(RR)] are estimated by considering the two-by-two table (Table 1).

The RR is estimated by

$$RR = \frac{\frac{a}{N_1}}{\frac{b}{N_0}} \tag{1}$$

The estimate of the SE[ln(RR)] depends on the definition of the RR, which depends on the kind of data in the two-by-two table. If the RR is a rate ratio (with a and b assumed to have a Poisson distribution, N_0 and N_1 person years at risk, and a/N_1 and b/N_0 rates) the SE is obtained by [3]

$$\mathrm{SE}[\ln(RR)] = \sqrt{\frac{1}{a} + \frac{1}{b}} \tag{2}$$

If the RR is defined as a risk ratio (with a and b assumed to have a binomial distribution, N_0 and N_1 the number of people at risk at the start of the observation interval, and a/N_1 and b/N_0 probabilities) the following equation holds [3]

$$\mathrm{SE}[\ln(RR)] = \sqrt{\frac{1}{a} - \frac{1}{N_1} + \frac{1}{b} - \frac{1}{N_0}} \tag{3}$$

The CI is then calculated using

$$\begin{aligned} \mathrm{CI}_\gamma &= \exp(\mu \pm Z_\gamma \sigma) \\ &= \exp(\ln(RR) \pm Z_\gamma \mathrm{SE}[\ln(RR)]) \end{aligned} \tag{4}$$

where Z_γ denotes the appropriate factor from the standard normal distribution for the desired confidence percentage (e.g., $Z_{95} = 1.96$) [4].

From Eq. 4, it follows that the natural candidate for the distribution of RR in an uncertainty analysis is

$$RR \sim \exp(N(\ln(RR), \mathrm{SE}[\ln(RR)])) \tag{5}$$

or equivalently

$$RR \sim L(\ln(RR), \mathrm{SE}[\ln(RR)]) \tag{6}$$

where N denotes the normal and L the lognormal distribution.

When X has a lognormal distribution with parameters $\mu = \ln(RR)$ and $\sigma = \mathrm{SE}[\ln(RR)]$ then the following holds [5]:

$$\begin{aligned} \mathrm{Mean}[X] &= \exp\left(\mu + \frac{1}{2}\sigma^2\right) \\ &= \exp\left(\ln(RR) + \frac{1}{2}\mathrm{SE}[\ln(RR)]^2\right) \end{aligned} \tag{7}$$

and

$$\begin{aligned} \mathrm{SD}[X] &= \sqrt{\exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2)} \\ &= \sqrt{\begin{array}{l}\exp\left(2\ln(RR) + 2\mathrm{SE}[\ln(RR)]^2\right) - \\ \exp\left(2\ln(RR) + \mathrm{SE}[\ln(RR)]^2\right)\end{array}} \end{aligned} \tag{8}$$

From Eq. 7 it is clear that it is the second term that causes the mean of this distribution to be systematically higher than RR, with the effect depending on the size of SE[ln(RR)]. Briggs et al. (in the context of lognormally distributed costs) suggest a solution that can be readily inferred from Eq. 7: use an adjusted μ' [2]

$$\mu' = \ln(RR) - \frac{1}{2}\mathrm{SE}[\ln(RR)]^2 \tag{9}$$

Substituting Eq. 9 into Eqs. 7 and 8, it can be worked out that the following holds: when X has a lognormal distribution with parameters μ' from Eq. 9 and $\sigma = \mathrm{SE}[\ln(RR)]$, then

$$\mathrm{Mean}[X] = RR \tag{10}$$

and

$$\mathrm{SD}[X] = \sqrt{\exp(2\ln(RR))\left(\exp\left(\mathrm{SE}[\ln(RR)]^2\right) - 1\right)} \tag{11}$$

From Eq. 10, it follows that this correction would indeed make the mean of the draws from the lognormal equal to the estimated RR. However, the drawback is that the standard deviation from Eq. 11 is smaller than what it should be according to Eq. 8, with, as a consequence, a too narrow uncertainty range that does not reproduce the CI of the epidemiological input data. Put differently, although this correction complies with the second desirable property, it does at the price of noncompliance with the third one.

So the question now is how to preserve the standard deviation, while at the same time obtaining the result of Eq. 10. Clearly, as the standard deviation is a function of RR and SE[ln(RR)], this implies a correction of the SE[ln(RR)] to offset the change in RR.

It is possible to write SE[ln(RR)] as a function of the mean and standard deviation of the lognormal distribution. Because we know the mean (RR) and can calculate the standard deviation from Eq. 8, we can thus obtain a corrected $\sigma' = \mathrm{SE}[\ln(RR)]'$:

$$\sigma' = \sqrt{\ln(s^2 + \exp(2\ln(RR))) - 2\ln(RR)} \tag{12}$$

where s is the standard deviation from Eq. 8.

With this adjusted SE[ln(RR)]′, we then calculate a correspondingly adjusted μ'':

$$\mu'' = \ln(RR) - \frac{1}{2}\left(\mathrm{SE}[\ln(RR)]'\right)^2 \tag{13}$$

From this derivation it follows that when X has a lognormal distribution with parameters μ'' from Eq. 13 and σ' from Eq. 12 then Mean[X] = RR and the SD[X] is as obtained from Eq. 8. The same result can be obtained by using a lognormal distribution parametrized with mean and standard deviation, and using RR and the standard deviation from Eq. 8 as its parameters.

A Hypothetical Example

Consider the hypothetical data in Table 2. In both the intervention and control arm are 100 person years at risk, with 20 observed cases in the intervention arm, and 40 in the control arm.

This leads to the RR and SE[ln(RR)] that are shown in Table 3. Because Table 2 reports person years at risk, Eq. 2 applies to obtain an estimate of the SE[ln(RR)].

Table 2 Hypothetical example of trial data

	Active	Control
Cases	20	40
Person years at risk	100	100

That same Table 3 also shows what the mean and standard deviation is of a lognormal distribution with parameters ln(RR) and SE[ln(*RR*)], using Eqs. 7 and 8, respectively, and the original and corrected parameters as obtained from Eqs. 9, 12, and 13. For the numbers in Table 2, the mean of the lognormal is 0.519, as against the estimated RR of 0.5.

Table 4 (top panel) shows the 95% approximate CI as calculated using Eq. 4, and by way of comparison the exact 95% confidence limits using the methods for person-time data as outlined by Greenland and Rothman [4]. This method relies on calculating *P*-values directly from a binomial probability model. The range column is the difference between the higher and the lower uncertainty interval, and the last column contains the ratio of this range over the range of the exact confidence limits. As can be seen, for these numbers, the approximation of Eq. 4, including the range of the CI, is quite good.

I implemented an uncertainty analysis in Excel using Ersatz (version 1.0, Brisbane, Australia; available at: http://www.epigear.com), an add-in that allows bootstrapping in Excel (Microsoft Corporation, Redmond, WA) [6]. Ersatz adds a large number of functions to Excel that allow drawing random deviates from specific distributions. Ersatz makes Excel recalculate repeatedly, each time drawing random values from these functions. It then calculates, among other things, means, medians, and uncertainty intervals (using percentiles) from the realized values for designated output variables.

Median and mean output RR and 95% uncertainty intervals (UI_{95}) were calculated using 50,000 draws from three Ersatz lognormal random functions, one with the standard parameters μ (= ln(*RR*)) and σ (= SE[ln(*RR*)]), the second with parameters μ' (Eq. 9) and σ (= SE[ln(*RR*)]), and the third with parameters μ'' (Eq. 13) and σ' (Eq. 12).

In Table 4 (lower panel) the results of the standard and the two correction methods are compared, both for the bootstrap and for analytical results obtained using Eqs. 7, 8, and 4. In all cases, the results from the bootstrap quite closely reproduce the analytical results.

For the standard parameters, the mean shows the expected difference; the uncertainty interval and its range are all quite close, with the ratio of the range over the exact solution very near 1.

The first correction method achieves its aim with a mean of 0.5, but as predicted by Eq. 11, the standard deviation is smaller, and consequently, the uncertainty interval is narrower with a ratio of its range over the exact solution decidedly below 1.

Table 3 Standard and corrected parameter values and their sources

	Source	Value
RR	Eq. 1	0.500
SE[ln(RR)]	Eq. 2	0.274
Mean[RR]	Eq. 7	0.519
SD[RR]	Eq. 8	0.145
μ	ln(RR)	−0.693
μ'	Eq. 9	−0.731
μ''	Eq. 13	−0.733
σ'	Eq. 12	0.284

The second correction method on the other hand achieves a mean of 0.5 while retaining the size of the standard deviation and the range of its uncertainty interval. The effect of this correction is basically that, as compared with the standard solution, the mean and the uncertainty interval are shifted by the size of the difference between mean effect size and the point estimate of RR.

Discussion

An objection that has been raised against the use of parametric bootstrapping for uncertainty analysis is that the choice of probability distribution and its parameters is arbitrary. However, in most cases appropriate choices of distributions can be logically deduced from theoretical considerations such as constraints on the parameter, or from the estimation method used for the parameter [2]. For the effect size expressed as a relative risk, the logical choice from the estimation method is that the relative risk follows a lognormal distribution with ln(RR) and SE[ln(RR)] as its parameters.

The RR is the ratio of two means, and not a mean itself, but the point estimate is being used as a mean when it is applied to calculate the risk under exposure of a target population. When modeling the uncertainty of this effect size, we would like it 1) to be aligned with the standard method used to construct a CI, i.e., to follow a lognormal distribution with ln(RR) and SE[ln(RR)] as its parameters; 2) to return a mean equal to the point estimate; and 3) to return the same CI.

These properties may seem simple and desirable, but from a mathematical point of view, they are incompatible: the point estimate of the relative risk is the median and not the mean of the lognormal distribution used to obtain the CI and model the uncertainty. So, the researcher must choose which of the three properties is most desirable. There are basically three options:

1. Use the unmodified lognormal distribution, reproduce the CI, but accept that the mean effect size is bigger than the point estimate.

Table 4 Confidence intervals according to the exact and approximate methods (top panel) and results from the uncertainty analysis (bottom panel): comparison of bootstrap (50,000 draws) and analytical solutions from the lognormal with standard parameters and with two corrected sets of parameters

	Source	Median	Mean	SD	Low 95% limit	High 95% limit	Range	Range ratio over exact
Exact 95% confidence limits	*	0.502	n/a	n/a	0.287	0.850	0.563	1.000
Approximate confidence limits	Eqs. 1 and 4	n/a	0.500	n/a	0.292	0.855	0.563	1.001
Standard parameters	Bootstrap	0.501	0.520	0.145	0.293	0.857	0.564	1.003
	Analytical	0.500	0.519	0.145	0.292	0.855	0.563	1.001
μ' and SE[ln(RR)]	Bootstrap	0.482	0.500	0.140	0.282	0.825	0.543	0.966
	Analytical	0.482	0.500	0.140	0.282	0.824	0.542	0.964
μ'' and σ'	Bootstrap	0.480	0.500	0.145	0.276	0.841	0.565	1.004
	Analytical	0.480	0.500	0.145	0.275	0.838	0.563	1.000

*Calculated using a hypergeometric model [4].

2. Use the first correction method, reproduce the point estimate effect size, but accept that the uncertainty interval is narrower than the CI.
3. Use the second correction method, reproduce the point estimate effect size and the width of the CI, but accept that the uncertainty interval is shifted somewhat.

Recently, Boshuizen and van Baal looked at this problem from a Bayesian point of view [7]. They show that a lognormal prior with a conservative mean of 1 and a sufficiently large σ leads to a posterior lognormal that is identical to the one derived here as correction method 1.

Of course their solution has the same drawback: the uncertainty interval will be narrower than the related CI. They do not discuss this issue, presumably because from a Bayesian point of view, this is not a drawback: the prior distribution contains information, even with a large σ, leading to a higher precision in the posterior estimate and thus to a narrower uncertainty interval.

But non-Bayesians who dislike this can easily see from their equations that the same prior with an estimated σ' corrected as per Eq. 12 would lead to a posterior lognormal identical to the result of correction method 2.

The first correction method has the advantage of being very simple. The second correction is somewhat more involved, using Eqs. 8, 12, and 13, but has the advantage of retaining the width of the CI. For ease of use, the second correction method has been implemented in the Ersatz ErRelativeRisk random function that takes RR and SE[ln(RR)] as parameters, and recalculates them according to the equations given above to produce a mean effect size equal to the point estimate of RR in the uncertainty analysis [6].

Source of financial support: NHMRC grant 351558.

References

1 Briggs A, Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. Health Econ 1998;7:723–40.
2 Briggs A, Sculpher M, Claxton K. Decision Modelling for Health Economic Evaluation. Oxford: Oxford University Press, 2006.
3 Rothman KJ. Epidemiology, an Introduction. Oxford: Oxford University Press, 2002.
4 Greenland S, Rothman KJ. Introduction to categorical statistics. In: Rothman KJ, Greenland S, eds. Modern Epidemiology (2nd ed.). Philadelphia: Lippincott-Raven; 1998;231–52.
5 Mood AM, Graybill FA, Boes DC. Introduction to the Theory of Statistics. (3rd ed.). New York: McGraw-Hill, 1974;
6 Epigear. Ersatz. version 1.0. Brisbane, Australia: Available from: http://www.epigear.com 2009 [Accessed December 14, 2009].
7 Boshuizen HC, van Baal PH. Probabilistic sensitivity analysis: be a Bayesian. Value Health 2009;12:1210–14.