In: Biometrics: Methods, Applications and Analysis ISBN 978-1-60876-412-9
Editors: H. Schuster and W. Metzger, pp. 197-216 © 2010 Nova Science Publishers, Inc.

Chapter 11

A GENERALIZATION OF FIELLER'S THEOREM FOR RATIOS OF NON-NORMAL VARIABLES AND SOME PRACTICAL APPLICATIONS

Guang Yong Zou and Allan Donner
Department of Epidemiology and Biostatistics,
Robarts Clinical Trials of Robarts Research Institute,
University of Western Ontario,
London, Ontario, Canada N6A 5C1

Abstract

Fieller's Theorem is widely used in practice to construct a confidence interval for a ratio of two variables, under the assumption that both the numerator and denominator are normally distributed. However this assumption may not always be satisfied in practice. An important example arises in the case of the incremental cost-effectiveness ratio (ICER), defined in the context of a comparative study as the ratio of an estimated cost difference to the effect measure. Nevertheless, Fieller's Theorem is often adopted by researchers to obtain confidence limits about the ICER. In this article, we show how the method of variance estimates recovery (MOVER) can be applied to obtain confidence limits for this parameter without requiring the assumption of normality. We evaluate its performance using Monte Carlo simulation in the context of human-immunodeficiency virus (HIV) incidence and health cost-effectiveness evaluation.

Key words: bootstrap, confidence interval, delta method, Fieller's theorem, incidence, net health benefit.

1. Introduction

Many indices of practical interest are ratios. A classical example is the median lethal dose (LD_{50}), also known as median effective dose, which is defined by the ratio of the intercept and the slope in a generalized linear model describing the relationship between the probability of death as a function of dose. Estimation of indices such as LD_{50} in bio-assay may have actually been the motivation for Fieller (1940; 1944; 1947; 1954) to develop his now famous theorem for setting confidence limits about a ratio of normal random variables.

A second example is the ϕ-value in protein-folding research (Goldenberg et al., 1989; Ruczinski et al., 2006). Since amino acid substitutions often affect the stability of a protein's native state and the transition states, ϕ-value, the ratio of the variation-induced changes in relative free energies, was proposed to quantify these effects.

In epidemiology, the relative excess risk can also be seen as a ratio of net effects (Suissa, 1999). Consider a study assessing the risks (R_1, R_2) of two drugs, where an unexposed reference group is also of interest. As an effect measure, one could adopt the ratio of the two risk ratios using the unexposed risk (R_0) as background, but this approach conceals the background risk. In contrast, the ratio of the excess risks, $(R_1 - R_0)/(R_2 - R_0)$, is seen to remove the background risk. This measure has now also been recommended for the 'gold standard' design (Hauschke and Pigeot, 2005).

A further example of practical interest arises in the estimation of human immunodeficiency virus (HIV) incidence using prevalent biologic specimens with a 2-stage procedure (Janssen et al., 1998; Des Jarlais et al., 2005; Cole et al., 2007). First, each specimen is tested by a highly sensitive enzyme immunoassay for HIV. Persons tested negative are considered to be HIV-uninfected, while those tested positive are considered HIV-infected and are subjected to a second test that is less sensitive or 'detuned'. Persons testing positive in the second stage are considered to have established HIV infection, while those testing negative are considered as new cases of HIV infection. The incidence of HIV is then defined by the number of new cases divided by the time T (in days) between which the highly sensitive and less sensitive tests first turn positive. The estimate of T is usually obtained using an incubation study.

Ratios are probably most prevalent in economic analysis where elasticity is defined as a ratio of relationships (Miller et al., 1984; Hirschberg et al., 2008). As an integrated discipline from economics and health research, health economics focuses on quantities such as the cost-effectiveness ratio and the incremental cost-effectiveness ratio, i.e., the extra cost divided by the extra benefit (Laska et al., 1997). This helps health policymakers make decisions on the grounds of whether the ratio is lower than society's willingness to pay for that particular benefit (Wong et al., 2009).

The main purpose of this article is to provide a general confidence interval procedure for ratios in the context of examples similar to those described above. Aside from the delta method, which is known to perform poorly due to its enforced symmetry (Efron and Tibshirani, 1993, p. 180), the primary alternative is Fieller's theorem (Fieller, 1954). In contrast to many authors who rely on the central limit theorem to normalize the distributions of the numerator and denominator (e.g., Briggs et al., 1999; Walter et al., 2008), we present an approach that relaxes this assumption. This is important in practice because many examples listed above involve highly skewed data for which even the bootstrap will fail (Dinh and Zhou, 2006; Zou and Donner, 2008).

This article is organized as follows. Section 2. reviews confidence interval methods for a ratio, focusing on the properties of Fieller's theorem. We also provide an intuitive link to the bias-corrected accelerated bootstrap approach to constructing confidence intervals (Efron, 1987), which sheds light on why it may be unnecessary to use the bootstrap in setting confidence limits for a ratio of two normal random variables. In section 3., we derive a general confidence interval procedure for ratios that does not require the assumption of normality. Since our approach relies on recovering variance estimates from confidence

intervals for the numerator and denominator separately, it is referred to as the method of variance estimates recovery (MOVER). Section 4. applies the MOVER to inference concerning the incidence of human immunodeficiency virus and to cost-effectiveness ratios. We also provide the results of simulation studies pertinent to these examples. The article finishes with a discussion.

2. Confidence Interval for a Ratio: From Box Method to Fieller's Theorem

Suppose that θ_1 and θ_2 are two parameters and interest is in their ratio

$$R = \theta_1/\theta_2.$$

Note that for a given value R to be a ratio, we must have

$$\theta_1 - R\theta_2 = 0. \tag{1}$$

For inferences about R, one must estimate θ_1 and θ_2 by $\widehat{\theta}_1$ and $\widehat{\theta}_2$ respectively, which leads to uncertainty in the subsequent estimate $\widehat{R}$. The objective of confidence interval construction is to quantify this uncertainty. A naive confidence interval for R, commonly referred to as the box method (O'Brien et al., 1994; Briggs et al., 1999), is given by (R_l, R_u), with

$$R_l = \frac{l_1}{u_2},$$

and

$$R_u = \frac{u_1}{l_2},$$

where (l_1, u_1) and (l_2, u_2) are confidence intervals for θ_1 and θ_2, respectively. In light of constraint (1), the box method is equivalent to obtaining a confidence interval (L, U) for $\theta_1 - R\theta_2$ based on whether the intervals for θ_1 and $R\theta_2$ overlap, i.e., setting both $L = l_1 - Ru_2$ and $U = u_1 - Rl_2$ to zero. Schenker and Gentleman (2001) have shown how this method has plagued statistical inferences in multiple disciplines. It is interesting to also note that this method has received wide attention in the context of health economics (Laska et al., 1997; Briggs et al., 1999; Drummond et al., 2005). The conservative nature of the resulting interval was heightened by Wakker and Klaassen (1995) who suggested using the Bonferroni adjustment for l_i, u_i $(i = 1, 2)$. Application of this very conservative approach has appeared in high profile medical journals (Janssen et al., 1998).

Given a value of R, one may also assume that the variance estimate for $\widehat{\theta}_1 - R\widehat{\theta}_2$ is independent of this parameter and given by

$$\widehat{\text{var}} = \widehat{\text{var}}(\widehat{\theta}_1) + \widehat{R}^2\widehat{\text{var}}(\widehat{\theta}_2) - 2\widehat{R}\widehat{\text{cov}}(\widehat{\theta}_1, \widehat{\theta}_2)$$

Applying the central limit theorem, we have

$$L = \widehat{\theta}_1 - R\widehat{\theta}_2 - z\sqrt{\widehat{\text{var}}}$$

and

$$U = \widehat{\theta}_1 - R\widehat{\theta}_2 + z\sqrt{\widehat{\text{var}}},$$

where z is the $\alpha/2$ upper quantile of the standard normal distribution. Setting both L and U to zero gives

$$\begin{cases} R_l = \widehat{\theta}_1/\widehat{\theta}_2 - z/\widehat{\theta}_2\sqrt{\widehat{\mathrm{var}}(\widehat{\theta}_1) + \widehat{R}^2\widehat{\mathrm{var}}(\widehat{\theta}_2) - 2\widehat{R}\widehat{\mathrm{cov}}(\widehat{\theta}_1,\widehat{\theta}_2)} \\ R_u = \widehat{\theta}_1/\widehat{\theta}_2 + z/\widehat{\theta}_2\sqrt{\widehat{\mathrm{var}}(\widehat{\theta}_1) + \widehat{R}^2\widehat{\mathrm{var}}(\widehat{\theta}_2) - 2\widehat{R}\widehat{\mathrm{cov}}(\widehat{\theta}_1,\widehat{\theta}_2)} \end{cases}. \tag{2}$$

This method may be recognized as a Wald interval, and sometimes referred to as the delta method. The underlying assumption here is that the variance estimates are independent of θ_i.

Fieller's theorem for a ratio (Fieller, 1940, 1944, 1947, 1954) can be derived using an approach similar to that of Wilson (1927), who obtained confidence limits for a proportion. Specifically, the confidence interval for R can be regarded as the null hypothesis values R_0 that cannot be rejected at the α level, where R_l and R_u are values of R_0 that satisfy

$$\frac{(\theta_1 - R_0\theta_2)^2}{\widehat{\mathrm{var}}(\widehat{\theta}_1) + R_0^2\widehat{\mathrm{var}}(\widehat{\theta}_2) - 2R_0\widehat{\mathrm{cov}}(\widehat{\theta}_1,\widehat{\theta}_2)} = \chi^2_{1,\alpha} \tag{3}$$

Solving the quadratic equation in R_0 gives

$$R_l^F = \frac{\widehat{\theta}_1\widehat{\theta}_2 - z^2\widehat{\mathrm{cov}}(\widehat{\theta}_1,\widehat{\theta}_2) - \sqrt{\left[\widehat{\theta}_1\widehat{\theta}_2 - z^2\widehat{\mathrm{cov}}(\widehat{\theta}_1,\widehat{\theta}_2)\right]^2 - \left[\widehat{\theta}_1^2 - z^2\widehat{\mathrm{var}}(\widehat{\theta}_1)\right]\left[\widehat{\theta}_2^2 - z^2\widehat{\mathrm{var}}(\theta_2)\right]}}{\widehat{\theta}_2^2 - z^2\widehat{\mathrm{var}}(\widehat{\theta}_2)}$$

and

$$R_u^F = \frac{\widehat{\theta}_1\widehat{\theta}_2 - z^2\widehat{\mathrm{cov}}(\widehat{\theta}_1,\widehat{\theta}_2) + \sqrt{\left[\widehat{\theta}_1\widehat{\theta}_2 - z^2\widehat{\mathrm{cov}}(\widehat{\theta}_1,\widehat{\theta}_2)\right]^2 - \left[\widehat{\theta}_1^2 - z^2\widehat{\mathrm{var}}(\widehat{\theta}_1)\right]\left[\widehat{\theta}_2^2 - z^2\widehat{\mathrm{var}}(\theta_2)\right]}}{\widehat{\theta}_2^2 - z^2\widehat{\mathrm{var}}(\widehat{\theta}_2)}.$$

By assuming the $\widehat{\theta}_i$ to be normal variates, Guiard (1989) obtained these limits by inverting the likelihood ratio test of $H_0 : \theta_1 - R\theta_2 = 0$. Without substituting $\widehat{\theta}_1/\widehat{\theta}_2$ for R_0 in the denominator of Equation (3), Fieller's theorem does not enforce symmetry as in the case of the Wald method. Application of the latter method has been regarded as the 'most severe error' in confidence interval construction (Efron and Tibshirani, 1993, p. 180).

In an analytic comparison of Fieller's theorem to the Wald method, Sitter and Wu (1954) made the interesting observation that the former involves a shift factor as applied to $\widehat{\theta}_1/\widehat{\theta}_2$ and an inflation factor as applied to the standard error (p. 1022). A similar observation was made by Agresti and Coull (1998) in comparing the Wilson method to the Wald method for constructing a confidence interval about a proportion. We further note that this is exactly the goal of the bias-corrected accelerated (Bca) bootstrap (Efron, 1987; Efron and Tibshirani, 1993, Chapter 22) and may be the reason that several simulation studies have shown similar performance between the bootstrap and Fieller's theorem (Severens et al., 1999; Briggs et al., 1999; Dinh and Zhou, 2006). In particular these studies have shown that both approaches result in reliable confidence intervals when the sampling distributions of $\widehat{\theta}_i$ are approximately normal, and that both fail when the sampling distributions for $\widehat{\theta}_i$ are not normal, e.g., in the case of a lognormal mean (Dinh and Zhou, 2006). We also point out that even though the bootstrap has the intention of shifting the point estimate and

inflating the variance, it falls short of effectively achieving this goal when the data are normally distributed (Efron and Tibshirani, 1993, Ch 25). This is because as a non-parametric procedure, the bootstrap does not take into account all available information.

3. An Extension of Fieller's Theorem for a Ratio of Non-normal Random Variables

By definition

$$R = \theta_1/\theta_2,$$

which is equivalent to

$$\theta_1 - R\theta_2 = 0. \tag{4}$$

Now, suppose θ_1 and θ_2 are estimated from independent samples as $\widehat{\theta}_1$ and $\widehat{\theta}_2$, with corresponding $100(1-\alpha)\%$ confidence intervals (l_1, u_1) and (l_2, u_2), respectively. Among the plausible values in (l_1, u_1) for θ_1 and those in (l_2, u_2) for θ_2, the values of R satisfying the relationship in (4) must then be the plausible values for R. Thus, we can obtain a confidence interval for R by setting confidence limits l and u for $\theta_1 - R\theta_2$. Furthermore, we note that, for any given R, we must have $l = 0$ and $u = 0$. Therefore, the lower limit for R is the smallest value of R that satisfies $l = 0$, while the upper limit for R is the largest value of R that satisfies $u = 0$.

We now discuss briefly how to obtain l and u by recovering variance estimates from $\widehat{\theta}_1(l_1, u_1)$ and $\widehat{\theta}_2(l_2, u_2)$. More details can be found in Zou and Donner (2008), Zou (2008), and Zou et al. (2009a). Because $\theta_1 - R\theta_2 = \theta_1 + (-R\theta_2)$, we begin by constructing a confidence interval for $\theta_1 + \theta_2$.

By the central limit theorem, the lower limit (l) is given by

$$l = \widehat{\theta}_1 + \widehat{\theta}_2 - z\sqrt{\text{var}(\widehat{\theta}_1) + \text{var}(\widehat{\theta}_2)}, \tag{5}$$

where z is the upper $\alpha/2$ quantile of the standard normal distribution. Inspired by the score method for interval estimation (Bartlett, 1953; Efron, 1987), we can estimate the variance corresponding to l at $\theta_1 + \theta_2 = l$. This in general will require an iterative procedure, which may become an obstacle to its wide application in practice. Therefore we proceed with estimating the variance in the neighborhood of l given by $l_1 + l_2$ as follows.

By the central limit theorem, we have

$$l_i = \widehat{\theta}_i - z\sqrt{\widehat{\text{var}}(\widehat{\theta}_i)},$$

which gives a variance estimate for $\widehat{\theta}_i$ at $\theta_i = l_i$ as

$$\widehat{\text{var}}(\widehat{\theta}_i) = (\widehat{\theta}_i - l_i)^2/z^2.$$

Substituting the recovered variance estimates for $\text{var}(\widehat{\theta}_1)$ in (5), we have

$$\begin{aligned} l &= \widehat{\theta}_1 + \widehat{\theta}_2 - z\sqrt{\text{var}(\widehat{\theta}_1) + \text{var}(\widehat{\theta}_2)} \\ &= \widehat{\theta}_1 + \widehat{\theta}_2 - z\sqrt{(\widehat{\theta}_1 - l_1)^2/z^2 + (\widehat{\theta}_2 - l_2)^2/z^2} \\ &= \widehat{\theta}_1 + \widehat{\theta}_2 - \sqrt{(\widehat{\theta}_1 - l_1)^2 + (\widehat{\theta}_2 - l_2)^2}. \end{aligned}$$

Analogous steps, noting that $u_1 + u_2$ is in the vicinity of u, yield

$$\widehat{\text{var}}(\widehat{\theta}_i) = (u_i - \widehat{\theta}_i)^2/z^2,$$

resulting in the upper limit u given by

$$u = \widehat{\theta}_1 + \widehat{\theta}_2 + \sqrt{(u_1 - \widehat{\theta}_1)^2 + (u_2 - \widehat{\theta}_2)^2}.$$

Note that because the recovered variance estimates $\widehat{\text{var}}(\widehat{\theta}_i)$ at $\theta_i = l_i$ and $\theta_i = u_i$ are in general different, l and u are not symmetric with respect to $\widehat{\theta}_1 + \widehat{\theta}_2$. This is not unexpected since as mentioned above symmetric intervals are known to perform poorly in finite samples.

We note that the form of the interval (l, u), appearing first in Howe (1974), has been applied by Newcombe (1998) to inferences concerning two binomial proportions. However, there has been no analytic justification for its general applicability until recently (Donner and Zou, 2002; Zou and Donner, 2008; Zou, 2008).

Recognizing that the confidence interval for $-R\theta_2$ is given by $(-Ru_2, -Rl_2)$, we can obtain the confidence limits for $\theta_1 - R\theta_2$ as

$$\begin{cases} l = \widehat{\theta}_1 - R\widehat{\theta}_2 - \sqrt{(\widehat{\theta}_1 - l_1)^2 + R^2(u_2 - \widehat{\theta}_2)^2} \\ u = \widehat{\theta}_1 - R\widehat{\theta}_2 + \sqrt{(u_1 - \widehat{\theta}_1)^2 + R^2(\widehat{\theta}_2 - l_2)^2} \end{cases}.$$

In the event that $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are obtained from dependent samples with $r = \widehat{\text{corr}}(\widehat{\theta}_1, \widehat{\theta}_2)$, the equations for l and u can be extended as

$$\begin{cases} l = \widehat{\theta}_1 - R\widehat{\theta}_2 - \sqrt{(\widehat{\theta}_1 - l_1)^2 + R^2(u_2 - \widehat{\theta}_2)^2 - 2r(\widehat{\theta}_1 - l_1)R(u_2 - \widehat{\theta}_2)} \\ u = \widehat{\theta}_1 - R\widehat{\theta}_2 + \sqrt{(u_1 - \widehat{\theta}_1)^2 + R^2(\widehat{\theta}_2 - l_2)^2 - 2r(u_1 - \widehat{\theta}_1)R(\widehat{\theta}_2 - l_2)} \end{cases}. \quad (6)$$

Solving R from the equation $l = 0$ results in the lower limit (the smaller root) for R given by

$$R_l^M = \frac{\widehat{\theta}_1\widehat{\theta}_2 - \widehat{\rho}(\widehat{\theta}_1 - l_1)(u_2 - \widehat{\theta}_2) - \sqrt{[\widehat{\theta}_1\widehat{\theta}_2 - \widehat{\rho}(\widehat{\theta}_1 - l_1)(u_2 - \widehat{\theta}_2)]^2 - l_1u_2(2\widehat{\theta}_1 - l_1)(2\widehat{\theta}_2 - u_2)}}{u_2(2\widehat{\theta}_2 - u_2)}. \quad (7)$$

Solving R from the equation $u = 0$ results in the upper limit (the large root) for R given by

$$R_u^M = \frac{\widehat{\theta}_1\widehat{\theta}_2 - \widehat{\rho}(u_1 - \widehat{\theta}_1)(\widehat{\theta}_2 - l_2) + \sqrt{[\widehat{\theta}_1\widehat{\theta}_2 - \widehat{\rho}(u_1 - \widehat{\theta}_1)(\widehat{\theta}_2 - l_2)]^2 - u_1l_2(2\widehat{\theta}_1 - u_1)(2\widehat{\theta}_2 - l_2)}}{l_2(2\widehat{\theta}_2 - l_2)}. \quad (8)$$

Note that nowhere in the derivation of R_l^M and R_u^M did we enforce symmetry on the sampling distributions of $\widehat{\theta}_i$ $(i = 1, 2)$. Instead, we used two separate normal distributions to recover variance estimates. By so doing, we deliberately accommodated the skewness of the sampling distributions in constructing confidence limits. However if the sampling distributions are indeed symmetric, as in the case of a normal distribution, we can show R_l

and R_u yield limits identical to those obtained using Fieller's Theorem. This can be seen by substituting zs_i, where s_i is the standard error for $\widehat{\theta}_i$, for $\widehat{\theta}_i - l_i$ and $u_i - \widehat{\theta}_i$ in the numerators, and $\widehat{\theta}_i - zs_i$ for l_i and $\widehat{\theta}_i + zs_i$ for u_i in denominators of R_l and R_u. Specifically we then have

$$R_l = \frac{\widehat{\theta}_1\widehat{\theta}_2 - rz^2 s_1 s_2 - \sqrt{[\widehat{\theta}_1\widehat{\theta}_2 - rz^2 s_1 s_2]^2 - (\widehat{\theta}_1 - zs_1)(\widehat{\theta}_2 - zs_2)(2\widehat{\theta}_1 - \widehat{\theta}_1 + zs_1)(2\widehat{\theta}_2 - \widehat{\theta}_2 + zs_2)}}{(\widehat{\theta}_2 - zs_2)(2\widehat{\theta}_2 - \widehat{\theta}_2 + zs_2)}$$
$$= \frac{\widehat{\theta}_1\widehat{\theta}_2 - rz^2 s_1 s_2 - \sqrt{(\widehat{\theta}_1\widehat{\theta}_2 - rz^2 s_1 s_2)^2 - (\widehat{\theta}_1^2 - z^2 s_1^2)(\widehat{\theta}_2^2 - z^2 s_2^2)}}{\widehat{\theta}_2^2 - z^2 s_2^2}$$

and

$$R_u = \frac{\widehat{\theta}_1\widehat{\theta}_2 - rz^2 s_1 s_2 + \sqrt{(\widehat{\theta}_1\widehat{\theta}_2 - rz^2 s_1 s_2)^2 - (\widehat{\theta}_1^2 - z^2 s_1^2)(\widehat{\theta}_2^2 - z^2 s_2^2)}}{\widehat{\theta}_2^2 - z^2 s_2^2}.$$

Since the MOVER approach begins with the construction of confidence intervals for the numerator and denominator, separately, it promotes interpretation of an estimated ratio in that specific context. The importance of this point has been pointed out previously in the health economics literature (Laska et al., 1997).

4. Applications

4.1. Incidence Estimates of Human Immunodeficiency Virus

We now apply the MOVER to obtain a confidence interval about the incidence of human immunodeficiency virus 1 (HIV-1) as estimated using a two-stage design (Janssen et al., 1998).

In monitoring and controlling the spread of human immunodeficiency virus 1 (HIV-1), differentiating individuals with early infection from those infected for a longer period is difficult. A two-stage testing strategy has been proposed to reduce this difficulty (Janssen et al., 1998). First, a highly sensitive enzyme immunoassay for HIV is applied to each specimen. Those found to be positive in the first stage are tested again by a 'detuned' enzyme immunoassay. Persons who are tested positive again are considered to be prevalent cases while those tested negative at the second stage are considered to be incident cases. This method is referred to as the serologic testing algorithm for recent HIV seroconversion (STARHS). The incidence is estimated as the number of incident cases divided by the time between serocoversion on the two-stage test, which may be estimated using laboratory incubation. It is reasonable to assume that incident cases follow a Poisson distribution and time may follow a normal distribution.

A distinct feature of our approach is that confidence intervals for the numerator and denominator are obtained from two independent sources (Janssen et al., 1998; Cole et al., 2007). A virtue of the MOVER is that it not only extracts statistical information from the confidence limits, but also reflects the sampling distributions.

As an illustration, consider the data presented by Janssen et al. (1998) which involved over 2.7 million first-time American Red Cross blood donor in 32 collection regions in the United States between May 1993 and December 1996. The summary data and the

calculations are presented in Table 1. We considered several scenarios, as in Cole et al. (2007) who adopted a simulation approach. For comparison we also considered Fieller's theorem, and two versions of the MOVER approach, one with numerator limits obtained using the exact method and the other obtained using the score method. The results show that the limits obtained using simulation and those obtained by Fieller's theorem tend to be similar, and both tend to provide smaller limits than those obtained by the MOVER.

To evaluate the performance of the confidence interval procedures, we conducted a simulation study. For each of the 21 parameter combinations (3 σ for $T \times 7$ for Poisson mean λ), we generated 10000 datasets and applied the four confidence interval procedures to each of them. For the simulation approach, we used 10000 draws for each data set. Each procedure was evaluated in terms of the proportion that the 10000 intervals that contained the true incidence, and the proportion of runs in which the intervals lay completely on the left (missing left) and on the right (missing right) of the true incidence. Mean interval width was also computed.

Results in Table 2 show that Fieller's theorem provides undercoverage when the value of λ is not large. The coverage increases toward the nominal level with increasing λ, although with unbalanced tail errors. The computationally intensive simulation method by Cole et al. (2007) performs similarly to Fieller's theorem. However we also point out that the method used by Cole et al. (2007) has been characterized as 'naive' and should not be trusted (Greenland, 2004, p. 1390). In contrast, the intervals provided by the MOVER approach perform well, with the interval procedure based on the score method for the Poisson mean performing particularly well in terms of coverage and tail errors. There are no substantial differences with respect to interval width.

4.2. Cost-Effectiveness in Health Services Research

We now evaluate the MOVER approach in the context of health economic evaluation where the skewness of cost data is a common feature. Patient costs are weighted sums of counts of resources used, where the unit costs serve as weights. Thus, 'total cost is really a weighted mixture of other distributions' (Briggs et al., 2002, p. 394). Typically, cost will be highly skewed with a few patients incurring highly expensive costs and many patients having little or no cost. Of course, the investigator may log-transform the data before conducting inferences. However, this would ultimately produce a geometric mean, which is less than the average cost, hence leading to an underestimated budget (Drummond et al., 2005, Ch 8).

The lognormal distribution has often been adopted in modeling health costs (Briggs et al., 1999; Zhou, 2002; Briggs et al., 2005). Yet statistical inferences involving a lognormal mean has been perceived as a challenging problem in the literature. For example, it has been stated that 'obtaining the confidence interval for the lognormal estimator is a non-trivial problem since it is a function of two transformed sample estimates' (Briggs et al., 2005, p. 422). However, confidence interval estimation for single lognormal means and their difference has now been obtained easily using the MOVER (Zou and Donner, 2008; Zou et al., 2009b).

The existing literature provides seemingly conflicting results on confidence interval construction for the incremental cost-effectiveness ratio (ICER). After evaluating 3840 pa-

rameter combinations, Briggs et al. (1999) suggested that Fieller's theorem is satisfactory for the ICER, even when cost data are generated from a lognormal distribution. On the other hand, Dinh and Zhou (2006) showed that Fieller's theorem as well as three versions of the bootstrap all performed poorly when cost data are assumed to follow a lognormal distribution. The confusing results by Briggs et al. (1999) may have arisen from using unrealistic parameter values for the lognormal distribution. Specifically, the used for the coefficient of variation (CV) were 10% to 40%, which are too low compared with those that arise from real data (Zhou, 2002; Briggs et al., 2005). This is because for the lognormal distribution,

$$CV = \sqrt{\exp(\sigma^2) - 1}.$$

Thus,

$$\sigma^2 = \ln(CV^2 + 1),$$

which gives $\sigma^2 = 0.15$ for a CV of 40%. With such a low value, the lognormal distribution is very close to a normal distribution, which explains the conclusions by Briggs et al. (1999).

Our purpose here is to evaluate the performance of the MOVER in setting confidence limits for the ICER, even when the data follow a lognormal distribution. Letting μ_i^c represent mean costs and μ_i^e represent mean effects, and denoting the ICER by R gives

$$R = \frac{\mu_1^c - \mu_2^c}{\mu_1^e - \mu_2^e},$$

which is equivalent to

$$\mu_1^c - \mu_2^c - R(\mu_1^e - \mu_2^e) = 0$$

or

$$(\mu_1^c - R\mu_1^e) - (\mu_2^c - R\mu_2^e) = 0. \tag{9}$$

Our strategy is to obtain the values of R_l and R_u that satisfy this equation. We use three steps to reach this goal: 1) we obtain L_i and U_i for $\mu_i^c - R\mu_i^e$ $(i = 1, 2)$, using the MOVER. 2) we apply the MOVER again to get L and U for $(\mu_1^c - R\mu_1^e) - (\mu_2^c - R\mu_2^e)$, and 3) we solve the the quadratic equations for R to obtain R_l and R_u. The components required for the first step are confidence limits and point estimates for mean costs $\bar{c}_i(l_i^c, u_i^c)$ and mean effects $\bar{e}_i(l_i^e, u_i^e)$, as well as correlation estimates r_i for each group.

Applying the MOVER results in confidence limits for $\mu_i^c - R\mu_i^e$ given by

$$L_i = \bar{c}_i - R\bar{e}_i - \left[(\bar{c}_i - l_i^c)^2 + R^2(u_i^e - \bar{e}_i)^2 - 2Rr_i(\bar{c}_i - l_i^c)(u_i^e - \bar{e}_i)\right]^{1/2}$$

and

$$U_i = \bar{c}_i - R\bar{e}_i + \left[(u_i^c - \bar{c}_i)^2 + R^2(\bar{e}_i - l_i^e)^2 - 2Rr_i(u_i^c - \bar{c}_i)(\bar{e}_i - l_i^e)\right]^{1/2}.$$

Applying the MOVER again, we have the confidence limits for $(\mu_1^c - R\mu_1^e) - (\mu_2^c - R\mu_2^e)$ given by

$$L = \bar{c}_1 - \bar{c}_2 - R(\bar{e}_1 - \bar{e}_2) - \left[(\bar{c}_1 - R\bar{e}_1 - L_1)^2 + (U_2 - (\bar{c}_2 - R\bar{e}_2))^2\right]^{1/2}$$

and

$$U = \bar{c}_1 - \bar{c}_2 - R(\bar{e}_1 - \bar{e}_2) + \left[(U_1 - (\bar{c}_1 - R\bar{e}_1))^2 + ((\bar{c}_2 - R\bar{e}_2) - L_2)^2\right]^{1/2}.$$

Setting $L = 0$ yields

$$R_l = \frac{B_1 - \sqrt{B_1^2 - A_1 C_1}}{A_1}$$

where

$$\begin{cases} A_1 = (\bar{e}_1 - \bar{e}_2)^2 - (u_1^e - \bar{e}_1)^2 - (\bar{e}_2 - l_2^e)^2 \\ B_1 = (\bar{e}_1 - \bar{e}_2)(\bar{c}_1 - \bar{c}_2) - r_1(\bar{c}_1 - l_1^c)(u_1^e - \bar{e}_1) - r_2(u_2^c - \bar{c}_2)(\bar{e}_2 - l_2^e) \\ C_1 = (\bar{c}_1 - \bar{c}_2)^2 - (\bar{c}_1 - l_1^c)^2 - (u_2^c - \bar{c}_2)^2 \end{cases}.$$

and setting $U = 0$ yields

$$R_u = \frac{B_2 + \sqrt{B_2^2 - A_2 C_2}}{A_2}$$

where

$$\begin{cases} A_2 = (\bar{e}_1 - \bar{e}_2)^2 - (\bar{e}_1 - l_1^e)^2 - (u_2^e - \bar{e}_2)^2 \\ B_2 = (\bar{e}_1 - \bar{e}_2)(\bar{c}_1 - \bar{c}_2) - r_1(u_1^c - \bar{c}_1)(\bar{e}_1 - l_1^e) - r_2(\bar{c}_2 - l_2^c)(u_2^e - \bar{e}_2) \\ C_2 = (\bar{c}_1 - \bar{c}_2)^2 - (u_1^c - \bar{c}_1)^2 - (\bar{c}_2 - l_2^c)^2 \end{cases}.$$

We note that regarding R as the reciprocal of willingness-to-pay (usually a known constant) in equation (9), the left side becomes

$$\mu_1^c - \mu_2^c - R(\mu_1^e - \mu_2^e),$$

which is identical to the net health benefit, a quantity proposed to replace the ICER due to its perceived unpleasant statistical properties (Stinnett and Mullahy, 1998). This observation suggests immediately a new confidence interval procedure for the net health benefit using the first two steps listed above. Thus, in contrast to what has been suggested in the health economics literature (Stinnett and Mullahy, 1998; Drummond et al., 2005), there is no need to avoid the ICER purely on statistical grounds. Moreover, switching to net health benefit cannot alleviate the skewness problem.

We now employ simulation to evaluate the confidence interval for ICER based on the MOVER, as compared to method using the delta method and Fieller's theorem. In this evaluation, we generated data according to the bivariate distributions used by Dinh and Zhou (2006). Specifically we have

1) Bivariate normal

$$\begin{pmatrix} C_{1i} \\ E_{1i} \end{pmatrix} \sim N\left[\begin{pmatrix} 40000 \\ 60 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 = 200000 & \rho_1\sigma_1\sigma_2 \\ \rho_1\sigma_1\sigma_2 & \sigma_2^2 = 5 \end{pmatrix}\right]$$

$$\begin{pmatrix} C_{2i} \\ E_{2i} \end{pmatrix} \sim N\left[\begin{pmatrix} 30000 \\ 50 \end{pmatrix}, \begin{pmatrix} \sigma_3^2 = 100000 & \rho_2\sigma_3\sigma_4 \\ \rho_2\sigma_3\sigma_4 & \sigma_4^2 = 10 \end{pmatrix}\right]$$

2) Semi-lognormal

$$\begin{pmatrix} \ln C_{1i} \\ E_{1i} \end{pmatrix} \sim N\left[\begin{pmatrix} 8 \\ 4 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 = 2 & \rho_1\sigma_1\sigma_2 \\ \rho_1\sigma_1\sigma_2 & \sigma_2^2 = 0.5 \end{pmatrix}\right]$$

$$\begin{pmatrix} \ln C_{2i} \\ E_{2i} \end{pmatrix} \sim N\left[\begin{pmatrix} 6 \\ 3 \end{pmatrix}, \begin{pmatrix} \sigma_3^2 = 2 & \rho_2\sigma_3\sigma_4 \\ \rho_2\sigma_3\sigma_4 & \sigma_4^2 = 0.5 \end{pmatrix}\right]$$

3) Bivariate lognormal

$$\begin{pmatrix} \ln C_{1i} \\ \ln E_{1i} \end{pmatrix} \sim N\left[\begin{pmatrix} 8 \\ 4 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 = 2 & \rho_1\sigma_1\sigma_2 \\ \rho_1\sigma_1\sigma_2 & \sigma_2^2 = 0.5 \end{pmatrix}\right]$$

$$\begin{pmatrix} \ln C_{2i} \\ \ln E_{2i} \end{pmatrix} \sim N\left[\begin{pmatrix} 6 \\ 3 \end{pmatrix}, \begin{pmatrix} \sigma_3^2 = 2 & \rho_2\sigma_3\sigma_4 \\ \rho_2\sigma_3\sigma_4 & \sigma_4^2 = 0.5 \end{pmatrix}\right]$$

Using the results in Shimizu (1983), we can show that the correlation estimates for semi-lognormal and bivariate lognormal data are given by

$$r = \frac{\rho\sigma_1}{\sqrt{\exp(\sigma_1^2) - 1}}$$

and

$$r = \frac{\exp(\rho\sigma_1\sigma_2) - 1}{\sqrt{[\exp(\sigma_1^2) - 1][\exp(\sigma_2^2) - 1]}}$$

respectively. Reliable confidence intervals for a lognormal mean and a difference between two lognormal means have previously been developed using the MOVER (Zou and Donner, 2008; Zou et al., 2009b). In the simulation study, we used the confidence interval for a single lognormal mean given by $[\exp(L), \exp(U)]$, where

$$\begin{cases} L = \bar{x} + s^2/2 - \sqrt{z^2 s^2/n + \{s^2/2[1 - (n-1)/\chi^2_{1-\alpha/2}]\}^2} \\ U = \bar{x} + s^2/2 + \sqrt{z^2 s^2/n + \{s^2/2[1 - (n-1)/\chi^2_{\alpha/2}]\}^2} \end{cases}$$

with $\bar{x}$ and s^2 denoting respectively the estimates of mean and variance based on the log-transformed data, and χ^2_q denoting qth percentile of the chi-square distribution with $n-1$ degrees of freedom.

When both cost and effect data are assumed to follow a normal distribution, the results in Table 3 indicate that Fieller's theorem is preferred even in fairly large sample sizes, because the delta method tends to provided lopsided confidence intervals. For lognormally distributed data, Tables 4 and 5 show clearly that there is no reason to use the delta method for estimating the ICER, even for sample sizes as large as 1000 subjects per group. Compared to the delta method, Fieller's theorem improves overall coverage slightly in these cases, but still in a lopsided fashion. In contrast, the MOVER performs very well in both the semi-lognormal and bivariate lognormal cases. It is also of interest to note that in large sample sizes, the intervals based on the MOVER are actually narrower than those provided by the delta method and Fieller's theorem.

5. Conclusion

We have presented an extension of Fieller's Theorem for constructing confidence intervals for ratios. Our approach is more general than Fieller's theorem because it does not assume the numerator and denominator are normally distributed. Furthermore, when the normality assumption is satisfied, our procedure reduces to Fieller's theorem. This is in contrast to the existing large literature (Zerbe, 1978; Guiard, 1989; Young et al., 1990, 1997; Hirschberg et al., 2008; Walter et al., 2008) in which normality is invariably assumed. Our general procedure should prove useful in a wide range of settings, as illustrated by the consideration of two interval estimation problems for which satisfactory procedures are either currently unavailable or unduly complicated.

Inspired by the comment that enforced symmetry is the source of deficiency in conventional confidence interval methods (Efron and Tibshirani, 1993, p.180), we developed our approach in three steps. First, we re-parameterized a ratio to a difference. Second, we applied the MOVER to obtain confidence limits for the difference. Finally, we exploited the constraint that the difference must be zero to obtain confidence limits for the ratio. Since it begins with confidence limits for the numerator and denominator separately, our procedure for setting confidence limits for a ratio automatically avoids the problem of interpreting inferences for a ratio without specifically taking into account the numerator and denominator (Laska et al., 1997). A further advantage of our approach is that it lends itself naturally to constructing simultaneous confidence intervals for the numerator and denominator as well as their ratio (Piepho and Emrich, 2005), again without the normality assumption.

The validity of the proposed confidence interval approach relies heavily on the validity of the confidence intervals for the numerator and denominator. One may alternatively be tempted to consider the bootstrap because of its nonparametric nature. We would like to remind the reader that '[t]he bootstrap is a fairly crude form of inference, that can be used when the data analyst is either unable or unwilling to carry out more extensive modeling. Nonparametric bootstrap inferences are asymptotically efficient. $\cdots$ [T]hey do not enjoy exact finite sample nonparametric properties' (Efron and Tibshirani, 1993, p. 395). Thus, although both the bootstrap and the MOVER are intended to provide large-sample confidence intervals that accommodate asymmetry, the former discards available knowledge, while the latter deliberately utilizes it.

Acknowledgments

The authors' work was partially supported by an Individual Discovery Grant from Natural Sciences and Engineering Research Council (NSERC) of Canada. Guang Yong Zou is a recipient of Early Researcher Award, Ontario Ministry of Research and Innovation, Canada. Helpful comments from Julia Taleban and Stephanie N Dixon are gratefully acknowledged.

Table 1. Confidence intervals about HIV incidence estimates using the serologic testing algorithm for recent HIV seroconversion (STARHS) among persons donating blood for the first time at American Red Cross sites, May 1993 to December 1996 (Data from Janssen et al, 1998, JAMA 280; 42–48)

		Calendar year				
		1993	1994	1995	1996	Total
Total samples tested: N		460,507	763,141	754,551	739,711	2,717,910
Established infections: B		107	177	141	122	547
Recent infections: A		15	24	12	18	69
Incidence per 10^5 per year $\frac{A}{N-B} \times \frac{365.25}{T} \times 10^5$, $T = 129$		9.22	8.91	4.50	6.89	7.19
95% CI for incidence						
1) $T = 129$ fixed						
	Fieller	4.56, 13.89	5.34, 12.47	1.96, 7.05	3.71 10.07	5.49, 8.88
	Simulation †	4.92, 14.14	5.57, 12.62	2.25, 7.13	3.83, 10.34	5.52, 8.96
	Exact‡	5.16, 15.21	5.71, 13.25	2.33, 7.87	4.08, 10.89	5.59, 9.10
	Approximate§	5.15, 14.87	5.70, 13.05	2.32, 7.65	4.08, 10.68	5.59, 9.04
2) $T \sim Normal(129, 10.2^2)$						
	Fieller	4.50, 14.40	5.25, 13.00	1.94, 7.29	3.66, 10.46	5.30, 9.43
	Simulation	4.78, 14.63	5.42, 13.16	2.13, 7.44	3.84, 10.61	5.33, 9.46
	MOVER-e¶	5.09, 15.69	5.59, 13.75	2.30, 8.09	4.02, 11.25	5.39, 9.61
	MOVER-a*	5.07, 15.35	5.58, 13.55	2.29, 7.88	4.01, 11.05	5.39, 9.56
3) $T \sim Normal(129, 30^2)$						
	Fieller	4.18, 19.11	4.74, 17.74	1.82, 9.55	3.36, 14.04	4.52, 13.63
	Simulation	4.37, 19.16	4.85, 17.69	1.97, 9.63	3.49, 14.07	4.52, 13.68
	MOVER-e	4.64, 20.21	4.98, 18.32	2.12, 10.27	3.63, 14.69	4.57, 13.73
	MOVER-a	4.63, 19.91	4.98, 18.16	2.11, 10.07	3.63, 14.51	4.57, 13.71
4) $T \sim Normal(129, 50^2)$						
	Fieller	3.75, 39.87	4.15, 37.97	1.66, 19.64	2.98, 29.61	3.82, 30.17
	Simulation	3.75, 35.84	4.09, 33.72	1.69, 17.78	2.93, 26.46	3.72, 26.86
	MOVER-e	4.10, 40.77	4.32, 38.40	1.89, 20.25	3.19, 30.12	3.85, 30.25
	MOVER-a	4.10, 40.52	4.32, 38.28	1.89, 20.08	3.18, 29.97	3.85, 30.23

† Simulation limits are based the 2.5 percentile and 97.5 percentile of 100,000 replications; ‡ Exact limits about A are given by $0.5\chi^2_{0.025,2A}$, $0.5\chi^2_{0.975,2A+2}$; § Approximate limits about A are $(\sqrt{A} - 1.96/2)^2$, $(\sqrt{A+0.5} + 1.96/2)^2$; ¶ The MOVER method as applied to exact limits for A; * The MOVER method as applied to approximate limits for A.

Table 2. Performance of procedures for constructing a two-sided 95% confidence interval for incidence using a two-stage testing algorithm based on 10,000 runs. Within each run, the simulation-based interval was based 10,000 Monte Carlo draws. Duration (T) of seroconversion was generated using Normal distribution with mean 129 and variance σ^2, while number of recent infections (A) from Poisson distribution with mean λ.

		Fieller Theorem	Simulation	MOVER-e	Mover-a
σ	λ	CV (ML, MR)% † Width‡	CV (ML, MR)% Width	CV (ML, MR)% Width	CV (ML, MR)% Width
10	5	87.47 (11.90, 0.63) 0.07	93.03 (5.95, 1.02) 0.07	97.81 (0.60, 1.59) 0.08	94.44 (3.97, 1.59) 0.07
	10	92.77 (6.54, 0.69) 0.09	93.70 (5.17, 1.13) 0.10	97.50 (1.00, 1.50) 0.10	95.78 (2.76, 1.46) 0.10
	15	92.26 (6.75, 0.99) 0.12	94.15 (4.48, 1.37) 0.12	96.15 (1.98, 1.87) 0.13	96.05 (2.09, 1.86) 0.12
	20	94.46 (4.43, 1.11) 0.14	94.82 (3.76, 1.42) 0.13	96.16 (1.80, 2.04) 0.14	96.10 (1.86, 2.04) 0.14
	30	94.20 (4.44, 1.36) 0.17	94.36 (3.92, 1.72) 0.16	95.67 (2.17, 2.16) 0.17	95.43 (2.41, 2.16) 0.17
	50	94.13 (4.14, 1.73) 0.21	94.32 (3.62, 2.06) 0.21	95.21 (2.21, 2.58) 0.22	94.93 (2.49, 2.58) 0.22
	100	94.77 (3.93, 1.30) 0.30	95.12 (3.59, 1.29) 0.30	95.63 (2.56, 1.81) 0.31	95.32 (2.87, 1.81) 0.31
30	5	87.31 (12.15, 0.54) 0.07	92.96 (6.09, 0.95) 0.07	97.82 (0.73, 1.45) 0.08	94.89 (3.69, 1.42) 0.07
	10	92.55 (6.61, 0.84) 0.10	93.44 (5.18, 1.38) 0.10	96.64 (1.41, 1.95) 0.10	95.29 (2.83, 1.88) 0.10
	15	93.13 (5.76, 1.11) 0.12	94.00 (4.44, 1.56) 0.12	95.98 (1.86, 2.16) 0.13	95.41 (2.46, 2.13) 0.12
	20	93.71 (5.30, 0.99) 0.14	94.12 (4.42, 1.46) 0.14	95.93 (2.08, 1.99) 0.14	95.41 (2.60, 1.99) 0.14
	30	94.08 (4.73, 1.19) 0.17	94.45 (4.11, 1.44) 0.17	96.15 (2.04, 1.81) 0.18	95.65 (2.56, 1.79) 0.17
	50	94.17 (4.16, 1.67) 0.22	94.43 (3.65, 1.92) 0.22	95.52 (2.28, 2.20) 0.23	95.16 (2.64, 2.20) 0.22
	100	94.54 (3.75, 1.71) 0.31	94.75 (3.56, 1.69) 0.31	95.33 (2.62, 2.05) 0.32	95.08 (2.87, 2.05) 0.32
50	5	87.98 (11.61, 0.41) 0.07	92.58 (6.52, 0.90) 0.07	97.86 (0.67, 1.47) 0.08	95.34 (3.28, 1.38) 0.07
	10	92.91 (6.42, 0.67) 0.10	93.81 (5.21, 0.98) 0.10	96.79 (1.51, 1.70) 0.11	95.78 (2.58, 1.64) 0.10
	15	93.78 (5.23, 0.99) 0.12	94.30 (4.35, 1.35) 0.12	96.20 (1.77, 2.03) 0.13	95.44 (2.54, 2.02) 0.12
	20	94.26 (4.44, 1.30) 0.14	94.69 (3.58, 1.73) 0.14	96.00 (1.82, 2.18) 0.15	95.58 (2.30, 2.12) 0.14
	30	94.14 (4.21, 1.65) 0.17	94.35 (3.61, 2.04) 0.17	95.88 (1.79, 2.33) 0.18	95.36 (2.32, 2.32) 0.17
	50	94.68 (3.82, 1.50) 0.22	94.74 (3.51, 1.75) 0.22	95.91 (1.97, 2.12) 0.23	95.62 (2.29, 2.09) 0.23
	100	94.98 (3.40, 1.62) 0.33	95.18 (3.13, 1.69) 0.33	95.67 (2.39, 1.94) 0.33	95.38 (2.68, 1.94) 0.33

† CV: coverage percentage; ML: missing the parameter value from its left; MR: missing the parameter value from its right. ‡ Width: average width of 10,000 intervals.

Table 3. Performance of procedures for constructing a two-sided 95% confidence interval for ICER when cost and effect data follow a bivariate normal distribution.

		Delta Method	Fieller Theorem
(n_1, n_2)	(ρ_1, ρ_2)	CV (ML, MR)% † Width‡	CV (ML, MR)% Width
(100, 100)	(-.30,0.30)	94.93 (3.41,1.66) 152.5	95.15 (2.39,2.46) 155.3
(100, 200)	(-.30,0.30)	94.67 (3.42,1.91) 126.7	94.70 (2.73,2.57) 128.0
(500, 500)	(-.30,0.30)	95.04 (3.06,1.90) 68.50	95.01 (2.60,2.39) 68.74
(500,1000)	(-.30,0.30)	94.81 (2.89,2.30) 56.88	94.70 (2.62,2.68) 56.99
(1000,1000)	(-.30,0.30)	94.84 (3.05,2.11) 48.43	94.85 (2.76,2.39) 48.51
(1000,2000)	(-.30,0.30)	95.09 (2.65,2.26) 40.24	95.16 (2.31,2.53) 40.28
(100, 100)	(-.30,0.00)	94.95 (3.49,1.56) 155.1	95.16 (2.54,2.30) 157.9
(100, 200)	(-.30,0.00)	94.62 (3.29,2.09) 128.4	94.70 (2.49,2.81) 129.7
(500, 500)	(-.30,0.00)	94.94 (2.80,2.26) 69.81	94.98 (2.34,2.68) 70.06
(500,1000)	(-.30,0.00)	94.90 (2.87,2.23) 57.70	94.72 (2.62,2.66) 57.82
(1000,1000)	(-.30,0.00)	95.16 (2.75,2.09) 49.37	95.23 (2.48,2.29) 49.46
(1000,2000)	(-.30,0.00)	94.97 (2.71,2.32) 40.82	94.99 (2.50,2.51) 40.86
(100, 100)	(0.00,0.00)	94.71 (3.42,1.87) 152.5	94.93 (2.45,2.62) 155.3
(100, 200)	(0.00,0.00)	94.70 (3.33,1.97) 124.9	94.67 (2.69,2.64) 126.2
(500, 500)	(0.00,0.00)	95.07 (3.01,1.92) 68.45	95.28 (2.51,2.21) 68.70
(500,1000)	(0.00,0.00)	94.98 (2.96,2.06) 56.05	95.09 (2.61,2.30) 56.16
(1000,1000)	(0.00,0.00)	94.82 (2.92,2.26) 48.47	94.89 (2.63,2.48) 48.55
(1000,2000)	(0.00,0.00)	94.87 (2.82,2.31) 39.64	94.89 (2.66,2.45) 39.68
(100, 100)	(0.00,0.00)	94.16 (3.97,1.87) 152.3	94.56 (2.82,2.62) 155.1
(100, 200)	(0.00,0.00)	94.43 (3.72,1.85) 124.7	94.68 (2.85,2.47) 125.9
(500, 500)	(0.00,0.00)	95.06 (3.03,1.91) 68.44	95.20 (2.54,2.26) 68.69
(500,1000)	(0.00,0.00)	94.92 (2.79,2.29) 56.10	94.87 (2.54,2.59) 56.22
(1000,1000)	(0.00,0.00)	94.96 (2.85,2.19) 48.49	95.01 (2.54,2.45) 48.57
(1000,2000)	(0.00,0.00)	95.08 (2.68,2.24) 39.64	95.17 (2.44,2.39) 39.68
(100, 100)	(0.30,-.30)	94.32 (3.80,1.88) 152.2	94.42 (2.80,2.78) 155.0
(100, 200)	(0.30,-.30)	94.56 (3.44,2.00) 123.1	94.87 (2.49,2.64) 124.3
(500, 500)	(0.30,-.30)	94.98 (3.03,1.99) 68.51	95.05 (2.61,2.34) 68.76
(500,1000)	(0.30,-.30)	95.00 (2.85,2.15) 55.23	94.94 (2.51,2.55) 55.33
(1000,1000)	(0.30,-.30)	94.86 (2.90,2.24) 48.43	95.02 (2.51,2.47) 48.52
(1000,2000)	(0.30,-.30)	95.21 (2.57,2.22) 39.09	95.20 (2.33,2.47) 39.13
(100, 100)	(0.30,0.00)	94.81 (3.41,1.78) 149.4	95.08 (2.43,2.49) 152.1
(100, 200)	(0.30,0.00)	94.86 (3.24,1.90) 121.3	94.82 (2.48,2.70) 122.5
(500, 500)	(0.30,0.00)	94.82 (2.79,2.39) 67.12	94.88 (2.40,2.72) 67.36
(500,1000)	(0.30,0.00)	94.85 (2.92,2.23) 54.41	94.95 (2.50,2.55) 54.52
(1000,1000)	(0.30,0.00)	95.08 (2.66,2.26) 47.51	95.02 (2.39,2.59) 47.59
(1000,2000)	(0.30,0.00)	95.07 (2.63,2.30) 38.49	95.20 (2.36,2.44) 38.53

† CV: coverage percentage; ML: missing the parameter value from its left; MR: missing the parameter value from its right. ‡ Width: median width of 10,000 intervals.

Table 4. Performance of procedures (based on 10000 runs) for constructing a two-sided 95% confidence interval for ICER when cost and effect data follow a semi-lognormal distribution.

		Delta Method	Fieller Theorem	MOVER
(n_1, n_2)	(ρ_1, ρ_2)	CV (ML, MR)% † WD ‡	CV (ML, MR)% WD	CV (ML, MR)% WD
(50, 50)	(-.60,-.30)	84.40 (15.5,0.06) 3946	86.83 (13.0,0.21) 4102	94.41 (3.14,2.45) 5758
(100, 100)	(-.60,-.30)	87.67 (12.1,0.20) 3126	89.15 (10.6,0.29) 3185	94.48 (3.07,2.45) 3749
(500, 500)	(-.60,-.30)	91.77 (7.73,0.50) 1629	92.22 (7.21,0.57) 1635	94.65 (2.88,2.47) 1561
(1000,1000)	(-.60,-.30)	92.95 (6.41,0.64) 1200	93.05 (6.20,0.75) 1202	94.44 (3.01,2.55) 1098
(50, 50)	(-.60,0.00)	84.87 (15.1,0.08) 3925	86.88 (13.0,0.13) 4076	94.69 (3.14,2.17) 5802
(100, 100)	(-.60,0.00)	87.76 (12.1,0.13) 3110	89.12 (10.7,0.21) 3167	94.78 (2.90,2.32) 3738
(500, 500)	(-.60,0.00)	92.58 (6.90,0.52) 1620	92.99 (6.44,0.57) 1626	94.90 (2.60,2.50) 1563
(1000,1000)	(-.60,0.00)	92.67 (6.61,0.72) 1193	92.94 (6.29,0.77) 1195	94.62 (3.03,2.35) 1095
(50, 50)	(-.60,0.30)	84.89 (15.0,0.07) 3905	86.72 (13.2,0.11) 4053	94.72 (2.97,2.31) 5749
(100, 100)	(-.60,0.30)	87.53 (12.4,0.09) 3084	88.81 (11.0,0.18) 3142	95.12 (2.82,2.06) 3729
(500, 500)	(-.60,0.30)	91.79 (7.79,0.42) 1611	92.15 (7.32,0.53) 1617	94.93 (2.74,2.33) 1551
(1000,1000)	(-.60,0.30)	92.72 (6.65,0.63) 1190	92.87 (6.43,0.70) 1192	94.93 (2.80,2.27) 1091
(50, 50)	(0.00,-.30)	84.04 (15.8,0.14) 3710	85.93 (13.9,0.14) 3846	95.05 (2.85,2.10) 5482
(100, 100)	(0.00,-.30)	87.18 (12.6,0.19) 2928	87.97 (11.8,0.21) 2980	95.04 (3.00,1.96) 3558
(500, 500)	(0.00,-.30)	91.74 (7.68,0.58) 1548	91.90 (7.49,0.61) 1553	95.04 (2.62,2.34) 1489
(1000,1000)	(0.00,-.30)	92.50 (6.82,0.68) 1143	92.63 (6.66,0.71) 1145	94.98 (2.57,2.45) 1048
(50, 50)	(0.00,0.00)	84.16 (15.7,0.13) 3675	85.66 (14.2,0.14) 3811	94.84 (3.06,2.10) 5463
(100, 100)	(0.00,0.00)	86.86 (13.0,0.16) 2894	87.70 (12.1,0.18) 2946	94.92 (3.08,2.00) 3539
(500, 500)	(0.00,0.00)	91.71 (7.77,0.52) 1533	91.97 (7.49,0.54) 1539	94.92 (2.82,2.26) 1489
(1000,1000)	(0.00,0.00)	92.67 (6.59,0.74) 1142	92.76 (6.49,0.75) 1143	94.74 (2.76,2.50) 1043
(50, 50)	(0.00,0.30)	83.24 (16.6,0.14) 3640	84.90 (15.0,0.15) 3772	94.62 (3.51,1.87) 5461
(100, 100)	(0.00,0.30)	87.91 (11.9,0.17) 2914	88.61 (11.2,0.18) 2964	95.09 (2.72,2.19) 3532
(500, 500)	(0.00,0.30)	92.23 (7.18,0.59) 1535	92.39 (6.96,0.65) 1540	95.34 (2.28,2.38) 1483
(1000,1000)	(0.00,0.30)	93.07 (6.28,0.65) 1136	93.23 (6.11,0.66) 1138	94.99 (2.54,2.47) 1041
(50, 50)	(0.60,-.30)	83.31 (16.5,0.19) 3376	84.13 (15.8,0.12) 3498	95.21 (2.98,1.81) 5178
(100, 100)	(0.60,-.30)	86.71 (13.1,0.19) 2703	87.10 (12.7,0.16) 2750	95.27 (2.84,1.89) 3357
(500, 500)	(0.60,-.30)	90.95 (8.50,0.55) 1458	91.00 (8.47,0.53) 1463	95.37 (2.42,2.21) 1417
(1000,1000)	(0.60,-.30)	92.59 (6.85,0.56) 1084	92.63 (6.83,0.54) 1086	95.33 (2.83,1.84) 995.4
(50, 50)	(0.60,0.00)	83.44 (16.4,0.14) 3378	84.26 (15.7,0.09) 3502	95.12 (3.18,1.70) 5204
(100, 100)	(0.60,0.00)	86.29 (13.6,0.14) 2696	86.66 (13.2,0.11) 2743	95.34 (2.80,1.86) 3356
(500, 500)	(0.60,0.00)	91.93 (7.63,0.44) 1461	91.96 (7.62,0.42) 1466	95.22 (2.49,2.29) 1417
(1000,1000)	(0.60,0.00)	93.03 (6.40,0.57) 1083	93.05 (6.40,0.55) 1085	95.21 (2.62,2.17) 992.8
(50, 50)	(0.60,0.30)	83.08 (16.7,0.18) 3375	83.80 (16.1,0.08) 3496	95.13 (3.20,1.67) 5167
(100, 100)	(0.60,0.30)	86.17 (13.6,0.27) 2689	86.51 (13.3,0.19) 2736	95.34 (2.86,1.80) 3358
(500, 500)	(0.60,0.30)	91.48 (8.08,0.44) 1452	91.51 (8.10,0.39) 1457	95.60 (2.51,1.89) 1407
(1000,1000)	(0.60,0.30)	92.27 (7.04,0.69) 1080	92.27 (7.07,0.66) 1082	95.26 (2.57,2.17) 990.0

† CV: coverage percentage; ML: missing the parameter value from its left; MR: missing the parameter value from its right. ‡ WD: median width of 10,000 intervals.

Table 5. Performance of procedures (based on on 10000 runs) for constructing a two-sided 95% confidence interval for ICER when cost and effect data follow a bivariate-lognormal distribution.

		Delta Method	Fieller Theorem	MOVER
(n_1, n_2)	(ρ_1, ρ_2)	CV (ML, MR)% † WD ‡	CV (ML, MR)% WD	CV (ML, MR)% WD
(50, 50)	(-.40,-.60)	86.50 (13.5,0.03) 218.6	92.21 (7.17,0.62) 251.9	94.76 (2.60,2.64) 315.7
(100, 100)	(-.40,-.60)	89.21 (10.7,0.13) 169.1	92.60 (6.85,0.55) 181.0	94.88 (2.58,2.54) 199.5
(500, 500)	(-.40,-.60)	92.54 (6.94,0.52) 84.06	93.71 (5.45,0.84) 85.16	94.66 (2.89,2.45) 81.18
(1000,1000)	(-.40,-.60)	93.52 (5.73,0.75) 61.34	94.10 (4.92,0.98) 61.70	94.58 (2.54,2.88) 56.87
(50, 50)	(-.40,0.00)	86.81 (13.2,-.00) 216.9	92.44 (7.10,0.46) 250.6	94.88 (2.53,2.59) 317.8
(100, 100)	(-.40,0.00)	89.24 (10.7,0.11) 166.0	92.69 (6.73,0.58) 177.7	94.86 (2.56,2.58) 197.6
(500, 500)	(-.40,0.00)	93.21 (6.21,0.58) 83.97	94.26 (4.88,0.86) 85.01	94.85 (2.60,2.55) 81.07
(1000,1000)	(-.40,0.00)	93.22 (6.01,0.77) 61.01	93.74 (5.16,1.10) 61.38	94.40 (2.90,2.70) 56.69
(50, 50)	(-.40,0.60)	86.58 (13.4,0.01) 212.5	92.11 (7.49,0.40) 244.8	94.94 (2.63,2.43) 313.8
(100, 100)	(-.40,0.60)	89.07 (10.8,0.09) 165.2	92.80 (6.77,0.43) 176.6	95.08 (2.46,2.46) 196.9
(500, 500)	(-.40,0.60)	92.27 (7.25,0.48) 82.80	93.50 (5.72,0.78) 83.81	94.77 (2.65,2.58) 80.14
(1000,1000)	(-.40,0.60)	93.09 (6.18,0.73) 60.59	93.78 (5.29,0.93) 60.95	94.96 (2.65,2.39) 56.15
(50, 50)	(0.00,-.60)	86.11 (13.8,0.11) 200.7	91.73 (7.84,0.43) 229.5	95.12 (2.38,2.50) 295.9
(100, 100)	(0.00,-.60)	89.06 (10.8,0.17) 153.6	92.05 (7.40,0.55) 163.4	94.92 (2.73,2.35) 185.9
(500, 500)	(0.00,-.60)	92.26 (7.11,0.63) 78.21	93.22 (5.83,0.95) 79.14	95.11 (2.49,2.40) 76.24
(1000,1000)	(0.00,-.60)	92.95 (6.27,0.78) 57.41	93.53 (5.42,1.05) 57.75	94.74 (2.69,2.57) 53.43
(50, 50)	(0.00,0.00)	86.22 (13.7,0.08) 199.1	91.52 (8.08,0.40) 226.8	95.28 (2.56,2.16) 292.7
(100, 100)	(0.00,0.00)	88.44 (11.4,0.12) 153.3	91.99 (7.64,0.37) 163.2	95.19 (2.48,2.33) 185.3
(500, 500)	(0.00,0.00)	92.33 (7.14,0.53) 77.69	93.41 (5.80,0.79) 78.60	95.11 (2.56,2.33) 76.07
(1000,1000)	(0.00,0.00)	93.32 (5.92,0.76) 57.31	93.79 (5.17,1.04) 57.63	94.82 (2.57,2.61) 53.18
(50, 50)	(0.00,0.60)	85.76 (14.1,0.14) 194.3	91.08 (8.41,0.51) 221.5	95.18 (2.71,2.11) 290.4
(100, 100)	(0.00,0.60)	89.25 (10.6,0.19) 151.5	92.36 (7.23,0.41) 161.0	95.18 (2.42,2.40) 182.6
(500, 500)	(0.00,0.60)	93.03 (6.37,0.60) 77.33	93.97 (5.23,0.80) 78.22	95.30 (2.18,2.52) 75.46
(1000,1000)	(0.00,0.60)	93.40 (5.86,0.74) 56.76	93.94 (5.14,0.92) 57.09	94.89 (2.47,2.64) 52.84
(50, 50)	(0.40,-.60)	85.70 (14.1,0.24) 170.1	91.01 (8.60,0.39) 192.5	95.88 (2.27,1.85) 261.1
(100, 100)	(0.40,-.60)	88.19 (11.6,0.26) 133.7	91.32 (8.30,0.38) 141.5	95.86 (2.04,2.10) 165.3
(500, 500)	(0.40,-.60)	91.76 (7.60,0.64) 69.17	92.57 (6.69,0.74) 69.88	95.60 (2.24,2.16) 68.53
(1000,1000)	(0.40,-.60)	92.95 (6.37,0.68) 51.30	93.43 (5.82,0.75) 51.56	95.85 (2.09,2.06) 47.98
(50, 50)	(0.40,0.00)	85.73 (14.0,0.24) 169.8	90.84 (8.88,0.28) 192.0	95.66 (2.51,1.83) 264.0
(100, 100)	(0.40,0.00)	88.06 (11.6,0.33) 131.9	91.21 (8.44,0.35) 139.2	95.89 (2.15,1.96) 164.2
(500, 500)	(0.40,0.00)	92.43 (6.97,0.60) 69.19	93.12 (6.13,0.75) 69.91	95.78 (2.12,2.10) 68.29
(1000,1000)	(0.40,0.00)	93.16 (6.11,0.73) 50.96	93.49 (5.71,0.80) 51.20	95.65 (2.26,2.09) 47.78
(50, 50)	(0.40,0.60)	85.14 (14.7,0.21) 167.1	90.14 (9.58,0.28) 187.9	96.12 (2.19,1.69) 258.5
(100, 100)	(0.40,0.60)	88.39 (11.3,0.27) 130.7	91.02 (8.66,0.32) 138.1	95.81 (2.30,1.89) 163.9
(500, 500)	(0.40,0.60)	92.06 (7.43,0.51) 68.26	92.95 (6.46,0.59) 68.95	95.71 (2.28,2.01) 67.46
(1000,1000)	(0.40,0.60)	92.75 (6.44,0.81) 50.57	93.16 (5.99,0.85) 50.82	95.72 (2.25,2.03) 47.39

† CV: coverage percentage; ML: missing the parameter value from its left; MR: missing the parameter value from its right. ‡ WD: median width of 10,000 intervals.

References

Agresti, A., Coull, B., 1998. Approximate is better than "exact" for interval estimation of binomial proportions. *American Statistician* **52**, 119–126.

Bartlett, M. S., 1953. Approximate confidence intervals. 2. more than one unknown parameter. *Biometrika* **40**, 306–317.

Briggs, A., Nixon, R., Dixon, S., Thompson, S., 2005. Parametric modelling of cost data: some simulation evidence. *Health Economics* **14**, 421–428.

Briggs, A. H., Mooney, C. Z., Wonderling, D. E., 1999. Constructing confidence intervals for cost-effectiveness ratios: An evaluation of parametric and non-parametric techniques using Monte Carlo simulation. *Statistics in Medicine* **18**, 3245–3262.

Briggs, A. H., O'Brien, B. J., Blackhouse, G., 2002. Thinking outside the box: Recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies. *Annual Review of Public Health* **23**, 377–401.

Cole, S. R., Chu, H., Brookmeyer, R., 2007. Confidence intervals for biomarker-based human immunodeficiency virus incidence estimates and differences using prevalent data. *American Journal of Epidemiology* **165**, 94–100.

Des Jarlais, D. C., Perlis, T., Arasteh, K., et al., 2005. Hiv incidence among injection drug users in New York City, 1990 to 2002: use of serologic test algorithm to assess expansion of HIV prevention services. *American Journal of Public Health* **95**, 1439–1444.

Dinh, P., Zhou, X. H., 2006. Nonparametric statistical methods for cost-effectiveness analysis. *Biometrics* **62**, 576–586.

Donner, A., Zou, G. Y., 2002. Interval estimation for a difference between intraclass kappa statistic. *Biometrics* **58**, 209–215.

Drummond, M. F., Sculpher, M. J., Torrance, G. W., O'Brien, B. J., Stoddar, G. L., 2005. *Methods for the Economic Evaluation of Health Care Programmes* (3rd ed). New York: Oxford University Press.

Efron, B., 1987. Better bootstrap confidence intervals (with discussion). *Journal of the American Statistical Association* **82**, 171–200.

Efron, B., Tibshirani, R. J., 1993. *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Fieller, E. C., 1940. The biological standardization of insulin (with discussion). *Supplement to the Journal of the Royal Statistical Society* **7**, 1–64.

Fieller, E. C., 1944. A fundamental formula in the statistics of biological assay, and some applications. Quarterly Journal of Pharmacy and Pharmacology 17, 117–123.

Fieller, E. C., 1947. Some remarks on the statistical background in bio-assay. *Analyst* **72**, 37–43.

Fieller, E. C., 1954. Some problems in interval estimation. *Journal of the Royal Statistical Society, Ser B* **16**, 175–185.

Goldenberg, D. P., Frieden, R. W., Haack, J. A., Morrison, T. B., 1989. Mutational analysis of a protein-folding pathway. *Nature* **338**, 127–132.

Greenland, S., 2004. Interval estimation by simulation as an alternative to and extension of confidence intervals. *International Journal of Epidemiology* **33**, 1389–1397.

Guiard, V., 1989. Some remarks on the estimation of the ratio of the expectation values of a twodimensional normal random variable (correction of the theorem of Milliken). *Biometrical Journal* **31**, 681–697.

Hauschke, D., Pigeot, I., 2005. Establishing efficacy of a new experimental treatment in the 'gold standard' design. *Biometrical Journal* **47**, 782–786.

Hirschberg, J. G., Lye, J. N., Slottje, D. J., 2008. Inferential methods for elasticit exhaustivey estimates. *Journal of Econometrics* **147**, 299–315.

Howe, W. G., 1974. Approximate confidence limits on the mean of $X + Y$ where X and Y are two tabled independent random variable. *Journal of the American Statistical Association* **69**, 789–794.

Janssen, R. S., Satten, G. A., Stramer, S. L., et al., 1998. Strategy to detect early HIV-1 infection for use in incidence estimates and for clinical and prevention purposes. *Journal of the American Medical Association* **280**, 42–48.

Laska, E. M., Meisner, M., Siegel, C., 1997. Statistical inference for cost-effectiveness ratio. *Health Economics* **6**, 229–242.

Miller, S. E., Capps Jr, O., Wells, G. J., 1984. Confidence intervals for elasticities and flexibilities from linear equations. *American Journal of Agricultural Economics* **66**, 392–396.

Newcombe, R. G., 1998. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine* **17**, 873–890.

O'Brien, B. J., Drummond, M. F., Labelle, R. J., 1994. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Medical Care* **32**, 150–160.

Piepho, H. P., Emrich, K., 2005. Simultaneous confidence intervals for two estimable functions and their ratio under a linear model. *American Statistician* **59**, 292–300.

Ruczinski, I., Sosnick, T. R., Plaxco, K. W., 2006. Methods for the accurate estimation of confidence intervals on protein folding ϕ value. *Protein Science* **15**, 2257–2264.

Schenker, N., Gentleman, J. F., 2001. On judging the significance of differences by examining the overlap between confidence intervals. *American Statistician* **55**, 182–186.

Severens, J. L., De Boo, T. M., Konst, E. M., 1999. Uncertainty of incremental cost-effectiveness ratios - A comparison of Fieller and bootstrap confidence intervals. *International Journal of Technology Assessment in Health Care* **15**, 608–614.

Shimizu, K., 1983. UMVU estimation for covariance and first product moments of transformed variables. *Communications in Statistics–Theory and Methods* **12**, 1661–1674.

Sitter, R. R., Wu, C. F. J., 1954. Construction of confidence limits about effect measures: A general approach. *Journal of the Royal Statistical Society, Series B* **16**, 175–185.

Stinnett, A., Mullahy, J., 1998. Net health benefit: A new framework for the analysis of uncertainty in cost-effectiveness alaysis. *Medical Decision Making* **18**, S68–S80.

Suissa, S., 1999. Relative excess risk: an alternative measure of comparative risk. *American Journal of Epidemiology* **150**, 279–282.

Wakker, P., Klaassen, M. P., 1995. Confidence intervals for cost/effectiveness ratios. *Health Economics* **3**, 373–381.

Walter, S. D., Gafni, A., Birch, S., 2008. A geometric confidence ellipse approach to the estimation of the ratio of two variables. *Statistics in Medicine* **27**, 5956–5974.

Wilson, E. B., 1927. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association* **22**, 209–212.

Wong, J. B., Mulrow, C., Sox, H. C., 2009. Health policy and cost-effectiveness analysis: yes we can. yes we must (editorial). *Annals of Internal Medicine* **150**, 274–275.

Young, D. A., Zerbe, G. O., Hay, W. W., 1990. Fieller's theorem, the likelihood and the delta-method. *Biometrics* **46**, 709–718.

Young, D. A., Zerbe, G. O., Hay, W. W., 1997. Fieller's theorem, Scheffe simultaneous confidence intervals, and ratios of parameters of linear and nonlinear mixed-effects models. *Biometrics* **53**, 838–847.

Zerbe, G. O., 1978. Fieller theorem and general linear-model. *American Statistician* **32**, 103–105.

Zhou, X. H., 2002. Inferences about population means of health care costs. *Statistical Methods in Medical Research* **11**, 327–339.

Zou, G. Y., 2008. On the estimation of additive interaction by use of the four-by-two table and beyond. *American Journal of Epidemiology* **168**, 212–224.

Zou, G. Y., Donner, A., 2008. Construction of confidence limits about effect measures: A general approach. *Statistics in Medicine* **27**, 1693–1702.

Zou, G. Y., Huang, W., Zhang, X., 2009a. A note on confidence interval estimation for a linear function of binomial proportions. *Computational Statistics and Data Analysis* **53**, 1080–1085.

Zou, G. Y., Huo, Y., Taleban, J., 2009b. Simple confidence intervals for lognormal means and their differences with environmental applications. *Environmetrics* **20**, 172–180.