UW Biostatistics Working Paper Series

9-7-2006

Generalized confidence intervals for the ratio or difference of two means for lognormal populations with zeros

Yea-Hung Chen
University of Washington, yeahung@u.washington.edu

Xiao-Hua Zhou
University of Washington, azhou@u.washington.edu

Generalized confidence intervals for the ratio or difference of two means for lognormal populations with zeros

Yea-Hung Chen[1,*] and Xiao-Hua Zhou[1,†]

[1] Department of Biostatistics, University of Washington

Box 357232, Seattle WA 98195, U.S.A

[*] *email*: yeahung@u.washington.edu

[†] *email*: azhou@u.washington.edu

Abstract

We discuss in this article methods for analyzing lognormal data that may include zeros. Specifically, we are interested in interval estimation for the ratio or difference of the population means. We propose here two generalized pivotal (GP) approaches: a "true" GP method and an "approximate" GP method. Additionally, we propose two likelihood-based approaches: a signed log-likelihood ratio (SLLR) method and a modified SLLR method. Our simulation studies suggest that the approximate generalized pivotal approach outperforms all other known methods; it results in highly accurate coverage frequencies and fairly low bias, even in small sample settings.

Key words: Skewed; Zeros; Lognormal; Confidence intervals; Generalized pivotal; Log-likelihood ratio.

1 Introduction

In many naturally occurring situations, the populations of interest may be reasonably described by lognormal distributions. Examples from the health sciences include measurements of hazardous materials or substances (Lyles and Kupper, 1996), health costs (Nixon and Thompson, 2004), and microarray spot intensities (Hoyle et al., 2002). The scientific question in these scenarios may often involve inference on the population means. Note that while the median is commonly regarded as a desirable summary for skewed data, it is not always a quantity of scientific interest.

In a recent article (Chen and Zhou, 2006) we discussed methods for estimating the ratio or difference of two lognormal means. These methods implicitly assume that the data and the populations are positive. In many situations however, it may be most sensible to assume that the populations may in fact include zeros. This assumption should be guided by scientific notions, not the data themselves. Note that a positive sample does not necessarily suggest that the corresponding population is positive as well: a population with zeros can certainly produce an entirely positive sample.

In many situations, there may be a good reason to believe not only that the population includes zeros but also that the positive subset of the population is lognormal. The second assumption should be supported by the data (model checking techniques are discussed in the next section). We discuss in this article methods for analyzing data that are believed to have arisen from such distributions. Specifically, we are interested in methods for estimating the ratio or difference of two population means. Appropriate interval estimates or hypothesis tests for these quantities are lacking. Zhou and Tu (2000) have proposed maximum likelihood and bootstrap confidence interval approaches

for the ratio of the means, but neither method appears to perform very well. The methods perform particularly poorly in small sample situations: the confidence intervals in these settings often have poor coverage probabilities (Zhou and Tu, 2000).

We propose in the present article generalized pivotal approaches for the ratio or difference of the population means. Additionally, we propose two likelihood-based approaches. Our focus is on confidence intervals. Note however that each of the methods could be modified to obtain a hypothesis test, if desired.

The structure of the paper is as follows. In Section 2, we present the model and the parameters of interest. We also discuss some model checking techniques. In Sections 3, 4 and 5, we present the methods of estimation (longer formulas and arguments are provided in the Appendix). In Section 6, we provide the results of some simulation studies. We close the paper with an example and a discussion (Sections 7 and 8).

2 The model

As indicated in the introduction, we are interested in analyzing lognormal populations that include zeros. Zhou and Tu (2000) have shown that these populations can be represented by a mixture model. Suppose we have two populations of interest. Let $\boldsymbol{W}_1 = (W_{11}, \ldots, W_{1n_1})$ denote a random sample from the first population and let $\boldsymbol{W}_2 = (W_{21}, \ldots, W_{2n_2})$ denote a random sample from the second population. Assume that W_{ij} takes on a value of zero with probability π_i. Else, let W_{ij} take on a positive value with a probability density function (pdf) given by a lognormal function. The pdf of W_{ij} is:

$$f(w_{ij}) = 1_{[w_{ij}=0]}\pi_i + 1_{[w_{ij}>0]}\frac{1}{\sqrt{2\pi}w_{ij}\sigma_i}\exp\left\{-\frac{1}{2\sigma_i^2}(\log w_{ij} - \mu_i)^2\right\}, \quad (1)$$

where $1_{[w_{ij} \in S]}$ denotes an indicator function, such that $1_{[w_{ij} \in S]} = 1$ if $w_{ij} \in S$ and $1_{[w_{ij} \in S]} = 0$ otherwise.

Assume that the W_{ij} are independent, and index the data such that the first n_{i1} observations of W_{ij} are positive and the remaining $n_{i0} = n_i - n_{i1}$ observations of W_{ij} are equal to 0. Also, let $\boldsymbol{Y}_i$ denote the population-specific random vector of the log-transformed positive variables, such that $\boldsymbol{Y}_i = (\log W_{i1}, \ldots, \log W_{in_{i1}})$. The joint pdf of $\boldsymbol{W} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2, n_{i0}, n_{20})$ can then be given by:

$$\begin{aligned} f(\boldsymbol{y}_1, \boldsymbol{y}_2, n_{10}, n_{20}) &= \binom{n_1}{n_{10}} \pi_1^{n_{10}} (1-\pi_1)^{n_{11}} \binom{n_2}{n_{20}} \pi_2^{n_{20}} (1-\pi_0)^{n_{21}} \\ &\quad \times \left(\frac{1}{\sqrt{2\pi}}\right)^{n_{11}+n_{21}} \left(\frac{1}{\sigma_1}\right)^{n_{11}} \left(\frac{1}{\sigma_2}\right)^{n_{21}} \\ &\quad \times \exp\left\{ -\frac{1}{2\sigma_1^2} \sum_{j=1}^{n_{11}} (y_{1j} - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{j=1}^{n_{21}} (y_{2j} - \mu_2)^2 \right\}. \end{aligned} \tag{2}$$

Note that the function reduces when $n_{10} = 0$ or $n_{20} = 0$.

We are interested in the ratio and difference of the population means. The mean, m_i, of population i can be derived from (1) and is given by:

$$m_i = E(W_{ij}) = (1 - \pi_i) \exp\left(\mu_i + \frac{1}{2}\sigma_i^2\right). \tag{3}$$

The ratio of the means, m_1/m_2, is simply:

$$m_1/m_2 = \left(\frac{1-\pi_1}{1-\pi_2}\right) \exp\left(\mu_1 + \frac{1}{2}\sigma_1^2 - \mu_2 - \frac{1}{2}\sigma_2^2\right). \tag{4}$$

We will often refer to the natural logarithm of this ratio, which we will denote ψ:

$$\psi = \log\left(m_1/m_2\right) = \log(1-\pi_1) + \mu_1 + \frac{1}{2}\sigma_1^2 - \log(1-\pi_2) - \mu_2 - \frac{1}{2}\sigma_2^2. \tag{5}$$

The difference of the means $m_1 - m_2$, which we will denote δ, is:

$$\delta = m_1 - m_2 = (1 - \pi_1) \exp\left(\mu_1 + \frac{1}{2}\sigma_1^2\right) - (1 - \pi_2) \exp\left(\mu_2 + \frac{1}{2}\sigma_2^2\right). \quad (6)$$

The lognormal assumption can and should be checked. Recall that a population is lognormal (by definition) if and only if the log-transformed values follow a normal distribution. Thus, we can check to see whether a sample is lognormal by checking whether the natural logarithm of each sample's positive subset follows a normal distribution. We recommend that this be done via the use of normal quantile plots and the Shapiro-Wilk test for normality. These should be applied to each of the transformed samples.

3 A generalized pivotal approach

We introduce here our generalized pivotal approaches for $\exp\psi = m_1/m_2$ and $\delta = m_1 - m_2$. Weerahandi (1993) defines a generalized pivotal (GP) as a statistic that has (a) a distribution free of unknown parameters, and (b) an observed value that does not depend on nuisance parameters. Note that the generalized pivotal is allowed to be a function of nuisance parameters, whereas conventional pivotal quantities can only be a function of the sample and the parameter of interest.

Generalized pivotal quantities can be used to obtain interval estimates for the parameter of interest; Weerahandi calls such intervals generalized confidence intervals. To find a 100γ percent generalized confidence interval, it is necessary to find a region C_γ of the pivotal space such that the probability that the pivotal quantity is in C_γ is equal to the confidence coefficient, γ. The generalized confidence interval is simply the region of the parameter space corresponding to C_γ.

3.1 GP approach for the ratio of means

Krishnamoorthy and Mathew (2003) have shown that Weerahandi's methods can be used to obtain generalized confidence intervals for the ratio of two lognormal means. We show here that the approach can be generalized to lognormal populations that also include zeros.

Let $\hat{\mu}_i$ denote the maximum likelihood (ML) estimate for μ_i, and let s_i^2 be the bias-corrected ML estimate for σ_i^2:

$$\hat{\mu}_i = \frac{1}{n_{i1}} \sum_{j=1}^{n_{i1}} y_{ij} \tag{7}$$

$$s_i^2 = \frac{1}{n_{i1}-1} \sum_{j=1}^{n_{i1}} (y_{ij} - \hat{\mu}_i)^2. \tag{8}$$

Define the statistics T_1 and T_2 as:

$$T_i = \hat{\mu}_i - \frac{Z_i}{U_i/\sqrt{n_{i1}-1}} \frac{s_i}{\sqrt{n_{i1}}} + \frac{1}{2} \frac{s_i^2}{U_i^2/(n_{i1}-1)}, \tag{9}$$

where $Z_i \sim \mathrm{N}(0,1)$ and $U_i^2 \sim \chi^2_{n_{i1}-1}$. Additionally, let $\tilde{\pi}_i$ denote the Agresti-Coull estimate for π_i (Agresti and Coull, 1998):

$$\tilde{\pi}_i = \frac{n_{i0} + \frac{1}{2} z_\alpha^2}{n_i + z_\alpha^2},$$

where z_α is the quantile of the standard normal curve corresponding to α, and α is one minus the confidence coefficient of the desired confidence interval.[1] Finally, define f_i as the function:

$$f_i(n_i, n_{i0}) = \tilde{\pi}_i - Z_{i+2} \sqrt{\tilde{\pi}_i(1-\tilde{\pi}_i)/(n_i + z_\alpha^2)},$$

[1] Our GP approach utilizes the Agresti-Coull (A-C) normal approximation. Other approximations to the binomial distribution are possible. These include the traditional Wald approximation and the Wilson approximation. Our studies suggest however that a GP method with an A-C approximation slightly outperforms GP methods with Wald or Wilson approximations (it seems to provide more accurate coverage frequencies).

where $Z_{i+2} \sim \mathrm{N}(0,1)$ and define S_i as:

$$S_i = \begin{cases} 0 & \text{if } f_i(n_i, n_{i0}) < 0 \\ \text{Uniform}(99/100, 1) & \text{if } f_i(n_i, n_{i0}) > 1 \\ f_i(n_i, n_{i0}) & \text{otherwise} \end{cases} \tag{10}$$

where $\text{Uniform}(a, b)$ denotes a uniformly distributed continuous random variable on the interval $[a, b]$.[2] The statisitc T_R:

$$T_R = T_1 + \log(1 - S_1) - T_2 - \log(1 - S_2) \tag{11}$$

is a generalized pivotal statistic for ψ. A proof is provided in Appendix A.

We show in the Appendix that the observed value of T_R – the value of T_R given the sample – is precisely equal to ψ. Thus, a $100(1-\alpha)\%$ two-sided generalized confidence interval for ψ is simply the $100(\alpha/2)$ and $100(1-\alpha/2)$ percentiles of T_R. We can find these percentiles by generating k values of T_R, where k denotes some "large" positive integer. The algorithm is as follows:

(For $j = 1$ to k)

Generate values for $Z_1, Z_2, Z_3, Z_4, U_1^2$, and U_2^2

Calculate T_R

(End loop)

Order the k values of T_R; find the 100α and $100(1-\alpha)$ percentiles; denote these $T_{R(l)}$ and $T_{R(u)}$, respectively.

[2]Note that we have provided a conditional definition of S_i because – as should be apparent from our definition of T_R – S_i is emblematic of π_i. Thus, it is conceptually desirable to restrict the function to ensure that it is not greater than 1 or less than 0. In fact, the presence of the logarithmic functions requires that S_i be less than 1, so some correction is strictly necessary since f_i can indeed be greater than 1 (f_i is likely to be greater than 1 when π_i is large, and is likely to be less than 0 when π_i is small). The definition of S_i for $f_i > 1$ is admittedly somewhat arbitrary.

A $100(1-\alpha)\%$ generalized confidence interval for $\exp\psi = m_2/m_2$ is:

$$(\exp T_{R(l)}, \exp T_{R(u)}). \tag{12}$$

3.2 GP for the difference of means

We also propose a generalized confidence interval for the difference of the population means, $\delta = m_1 - m_2$. Let T_1, T_2, S_1, and S_2 be the statistics given by (9) and (10). The statistic:

$$T_D = \exp\{T_1 + \log(1-S_1)\} - \exp\{T_2 + \log(1-S_2)\} \tag{13}$$

is a generalized pivotal statistic for δ. This may be shown via arguments similar to those used in the proof for T_R, provided in Appendix A. It should be apparent from that section that the observed value of T_D is equal to δ. Thus, a $100(1-\alpha)$ two-sided generalized confidence interval for δ is given by the $100(\alpha/2)$ and $100(1-\alpha/2)$ percentiles of T_D. They may be found via the following algorithm:

(For $j = 1$ to k)

Generate values for $Z_1, Z_2, Z_3, Z_4, U_1^2$, and U_2^2

Calculate T_D

(End loop)

Order the k values of T_D; find the 100α and $100(1-\alpha)$ percentiles; denote these $T_{D(l)}$ and $T_{D(u)}$, respectively.

A $100(1-\alpha)\%$ generalized confidence interval for $\delta = m_2 - m_2$ is:

$$(T_{D(l)}, T_{D(u)}). \tag{14}$$

4 An approximate GP approach

Recall that T_i (9) utilizes the bias-corrected ML estimate for σ_i^2. Alternatively, we may have defined the statistic using the non-corrected ML estimate, $\hat{\sigma}_i^2$:

$$\hat{\sigma}_i^2 = \frac{1}{n_{i1}} \sum_{j=1}^{n_{i1}} (y_{ij} - \hat{\mu}_i)^2 \tag{15}$$

in place of s_i^2. We can use this modified form of T_i, which we will call T_i^*, to define modified forms of T_R (11) and T_D (13). That is, we can define T_R^* as T_R, with T_i replaced with T_i^*. Likewise, we can define T_D^* as T_D, with T_i replaced with T_i^*.

The statistic T_R^* is not in fact a generalized pivotal statistic for ψ, nor is T_D^* a generalized pivotal statistic for δ. As should be apparent from the arguments provided in Appendix A, the observed values of T_R^* and T_D^* will actually depend on nuisance parameters. Nevertheless, T_R^* and T_D^* are "approximate" GP statistics: their observed values each approach the parameter of interest when the samples are large. Of course, the idea of using approximate generalized pivotal (AGP) statistics may seem strange. Our own experience has revealed however that AGP approaches can actually have desirable properties. In fact, the "generalized pivotal statistics" described in our earlier article (Chen and Zhou, 2006) were, contrary to our claims, actually AGP statistics. A "true" GP method, which uses the bias-corrected ML estimate for σ_i^2, is presented in Krishnamoorthy and Mathew's 2003 article. Our analysis has indicated however, that the method discussed in our article (an AGP approach) outperforms the true GP method in small sample settings; the former approach typically results in more accurate coverage frequencies.

Given this knowledge, we felt it would be reasonable to propose AGP approaches here. We define to the AGP methods to be entirely analogous to the GP methods, with T_R^* or T_D^* used in place of T_R or T_D. That is, the AGP methods use the non-corrected ML estimate for σ_i^2 as opposed

to the bias-corrected ML estimate s_i^2.

5 Likelihood based approaches

We propose here two likelihood approaches for estimating m_1/m_2 or $m_1 - m_2$. These are: a signed log-likelihood ratio (SLLR) approach and a modified signed log-likelihood approach (MSLLR).

5.1 A signed log-likelihood ratio approach

Wu et al. (2002) have shown that the signed log-likelihood ratio (SLLR) statistic can be used to obtain confidence intervals for the ratio of two lognormal means. We propose here an approach for situations in which the populations may also include zeros.

Define $\boldsymbol{\lambda}$ as the vector of nuisance parameters $\boldsymbol{\lambda} = (\mu_2, \sigma_1, \sigma_2, \pi_1, \pi_2)$ and define $\boldsymbol{t}$ as the vector of statistics $\boldsymbol{t} = (t_1, t_2, t_3, t_4, t_5, t_6) = (\sum y_{1j}, \sum y_{2j}, \sum y_{1j}^2, \sum y_{2j}^2, n_{11}, n_{21})$. The log-likelihood function is simply the natural logarithm of (2), and can be expressed as a function of $\boldsymbol{t}$:

$$\begin{aligned}
\ell(\psi; \boldsymbol{\lambda}, \boldsymbol{t}) &= \left\{\psi - \log(1-\pi_1) - \frac{1}{2}\sigma_2^2 + \log(1-\pi_2) + \mu_2 + \frac{1}{2}\sigma_2^2\right\}\frac{1}{\sigma_1^2}t_1 + \frac{\mu_2}{\sigma_2^2}t_2 \\
&\quad - \frac{1}{2\sigma_1^2}t_3 - \frac{1}{2\sigma_2^2}t_4 - t_5\log\sigma_1 + (n_1 - t_5)\log\pi_1 + t_5\log(1-\pi_1) \\
&\quad - \left\{\psi - \log(1-\pi_1) - \frac{1}{2}\sigma_2^2 + \log(1-\pi_2) + \mu_2 + \frac{1}{2}\sigma_2^2\right\}^2\frac{1}{2\sigma_1^2}t_5 \\
&\quad - t_6\log\sigma_2 + (n_2 - t_6)\log\pi_2 + t_6\log(1-\pi_2) - \frac{\mu_2^2}{2\sigma_2^2}t_6. \qquad (16)
\end{aligned}$$

Note that the function reduces when one or more of the samples does not contain zeros. This because $t_5 = n_1$ when $n_{10} = 0$, and $t_6 = n_2$ when $n_{20} = 0$.

The signed log-likelihood ratio statistic, which we will denote r, is given by:

$$r(\psi) = \mathrm{sgn}(\hat{\psi} - \psi)\{2[\ell(\hat{\psi}, \hat{\boldsymbol{\lambda}}) - \ell(\psi, \hat{\boldsymbol{\lambda}}_\psi)]\}^{1/2}, \qquad (17)$$

where $\hat{\psi}$ and $\hat{\boldsymbol{\lambda}}$ denote the maximum likelihood estimates of ψ and $\boldsymbol{\lambda}$. These can be derived from (16) and are as follows:

$$\begin{aligned}
\hat{\mu}_i &= \frac{1}{n_{i1}} \sum_{j=1}^{n_{i1}} y_{ij} \\
\hat{\sigma}_i^2 &= \frac{1}{n_{i1}} \sum_{j=1}^{n_{i1}} (y_{ij} - \hat{\mu}_i)^2 \\
\hat{\pi}_i &= n_{i0}/n_i \qquad (18) \\
\hat{\psi} &= \log(1-\hat{\pi}_1) + \hat{\mu}_1 + \frac{1}{2}\hat{\sigma}_1^2 - \log(1-\hat{\pi}_2) - \hat{\mu}_2 - \frac{1}{2}\hat{\sigma}_2^2. \qquad (19)
\end{aligned}$$

The expression $\hat{\boldsymbol{\lambda}}_\psi$ denotes the constrained maximum likelihood estimate of $\boldsymbol{\lambda}$ at ψ. That is, it is the maximum likelihood estimate of $\boldsymbol{\lambda}$ at some fixed value of ψ. This may be obtained computationally via the use of a maximization algorithm. We used the `optim` function in R (R Development Core Team, 2006).

The distribution of r is said to approximate the standard normal to the first order (see for instance: Barndorff-Nielsen, 1986). Thus, a $100(1-\alpha)$ percent confidence interval for ψ is simply given by the region:

$$\{\psi; -z_{\alpha/2} \leq r(\psi) \leq z_{\alpha/2}\}. \qquad (20)$$

Finding the lower and upper bounds is a non-trivial task, as the task requires some sort of computational strategy. We used `uniroot`, a root-finding function included in R (R Development Core Team, 2006). This approach worked quite well.

Signed log-likelihood ratio confidence intervals for $\delta = m_1 - m_2$ are also possible. The approach is entirely analogous to that described above: one simply needs to rewrite ℓ as a function of δ.

5.2 A modified signed log-likelihood ratio approach

Barndorff-Nielsen (1986; 1991) has developed a modified form of the signed log-likelihood ratio (MSLLR) statistic. It is advantageous because the distribution of the statistic is said to approximate the standard normal to the third order (Barndorff-Nielsen, 1986; 1991). Wu et al. (2002) have shown that the MSLLR can be used to estimate the ratio of two lognormal means. We propose here an approach for situations in which the populations may include zeros.

Let $\boldsymbol{\lambda}, \boldsymbol{t}$, and ℓ be as defined in Section 5.1. Also, define $\boldsymbol{\theta}$ as the vector of parameters $(\psi, \boldsymbol{\lambda})$. The modified signed log-likelihood ratio statistic (Barndorff-Nielsen, 1986; Barndorff-Nielsen, 1991), which we will denote r^*, is:

$$r^*(\psi) = r(\psi) + \frac{1}{r(\psi)} \log \left\{ \frac{u(\psi)}{r(\psi)} \right\}. \tag{21}$$

The statistic u is:

$$u(\psi) = \frac{|\ell_{\boldsymbol{t}}(\hat{\psi}, \hat{\boldsymbol{\lambda}}) - \ell_{\boldsymbol{t}}(\psi, \hat{\boldsymbol{\lambda}}_\psi), \ell_{\boldsymbol{\lambda t}}(\psi, \hat{\boldsymbol{\lambda}}_\psi)|}{|\ell_{\boldsymbol{\theta t}}(\hat{\psi}, \hat{\boldsymbol{\lambda}})|} \times \left\{ \frac{|\ell_{\boldsymbol{\theta\theta}}(\hat{\psi}, \hat{\boldsymbol{\lambda}})|}{|\ell_{\boldsymbol{\lambda\lambda}}(\psi, \hat{\boldsymbol{\lambda}}_\psi)|} \right\}^{1/2}, \tag{22}$$

where $\ell_{\boldsymbol{t}} = \partial\ell/\partial\boldsymbol{t}$, $\ell_{\boldsymbol{\lambda t}} = \partial^2\ell/\partial\boldsymbol{\lambda}\partial\boldsymbol{t}$, $\ell_{\boldsymbol{\theta t}} = \partial^2\ell/\partial\boldsymbol{\theta}\partial\boldsymbol{t}$, $\ell_{\boldsymbol{\theta\theta}} = -\partial^2\ell/\partial\boldsymbol{\theta}^2$, and $\ell_{\boldsymbol{\lambda\lambda}} = -\partial^2\ell/\partial\boldsymbol{\lambda}^2$. Precise expressions for these functions are provided in the appendix. Note that the comma in the numerator of the first fraction is meant to indicate that the matrix is formed by adjoining the columns of $\ell_{\boldsymbol{t}}(\hat{\psi}, \hat{\boldsymbol{\lambda}}) - \ell_{\boldsymbol{t}}(\psi, \hat{\boldsymbol{\lambda}}_\psi)$ and $\ell_{\boldsymbol{\lambda t}}(\psi, \hat{\boldsymbol{\lambda}}_\psi)$.

As indicated above, the distribution of r^* is said to approximate the standard normal to the third order. Thus, a $100(1-\alpha)$ percent confidence interval for ψ can be given by the region:

$$\{\psi; -z_{\alpha/2} \leq r^*(\psi) \leq z_{\alpha/2}\}. \tag{23}$$

The upper and lower bounds may be found computationally via, for instance, the `uniroot` function

in R (R Development Core Team, 2006).

A MSLLR approach for δ is also possible, and is entirely analogous to the approach described above. Our exploration suggests however, that derivation of the method would require fairly tedious mathematical computation, involving a large amount of irreducible terms.

6 Simulation studies

We report and discuss here simulations for the methods for estimating $\exp\psi = m_1/m_2$. Note that we have also conducted simulations for the methods for estimating $\delta = m_1 - m_2$, but these studies revealed patterns fairly consistent with those observed in the simulations for m_1/m_2 (and are thus not reported here).

Six methods were considered in our study. These include the four methods we have proposed above: the GP and the AGP approaches, the SLLR approach, and the MSLLR approach. Additionally, we conducted simulations for the two existing approaches: the maximum likelihood (ML) and bootstrap methods proposed by Zhou and Tu (2000).

A complete list of our sample sizes and parameter values is provided in Table 1. Note that we have chosen to focus on small sample settings because, as discussed in our introduction, the existing methods (the ML and bootstrap methods) have been reported to perform fairly poorly in small sample scenarios.

For each of the designs in Table 1, we generated 10,000 sets of samples. Each set of samples was used to construct a 95 percent confidence interval for the ratio of means, $\exp\psi = m_1/m_2$, using each of the six methods referred to above. We used $k = 500$ pivotal quantities (per simulation) for each generalized confidence interval, and $k = 500$ bootstrap samples for each bootstrap interval. All simulations were carried out in the R programming environment (R Development Core Team, 2006).

Our complete results are presented in Table 2. Coverage frequencies and bias measurements are summarized in Figure 1. Coverage is the percent of confidence intervals that included the true value, and median width is the median width for each set of intervals (10,000 per simulation). We also provide in this table summaries of the error frequencies. Left error is the percent of the intervals that were to the left of the true value; right error is the percent that were to the right of the parameter. Relative bias is a relative comparison of the two frequencies:

$$\text{relative bias} = \frac{(\text{right error}) - (\text{left error})}{(\text{right error}) + (\text{left error})}.$$

Note that the quantity takes on a positive value when right error exceeds left error.

We are primarily interested in the coverage frequencies and the relative bias measurements. The AGP method appears to be the clear winner. It resulted in highly accurate coverage frequencies and did not reveal a strong amount of bias. The GP and MSLLR methods also resulted in fairly good coverage and low bias, but the coverage frequencies were not as accurate as those provided by the AGP method. The GP method tended to result in over coverage and the MSLLR method tended to result in under coverage. All other techniques – the ML, bootstrap, and SLLR approaches – revealed poor coverage probabilities. The ML and bootstrap approaches appear to be particularly undesirable as they can result in a large amount of bias.

7 Example

We consider here an example from the health sciences. Callahan et al. (1997) were involved in a cohort study of older adults at a large primary care group practice. The cohort consisted of 3767 elderly adults who had completed testing on the Centers for Epidemiologic Studies Depression Scale (CES-D). Clinical information for these individuals was collected over a two year period. These

records were used to classify the patients into various subgroups, defined by the Ambulatory Care Group system (abbreviated ACG, and now known as the Adjusted Clinical Group system). The researchers were interested in estimating the mean test cost in each of the patient subgroups.

For our discussion here, we will compare the Group 8 to Group 9. The mean cost in Group 8 is \$1788.67 and the mean for Group 9 is \$498.32. Data from both groups are highly skewed and include zeros: 10 of the 40 observations in Group 8 and 22 of 125 observations in Group 9 are equal to zero. Thus, we propose checking to see whether the data might be reasonably described by our mixture model. Normal quantile plots for the log-transformed positive data are provided in Figure 2. These provide no evidence against lognormality: the points all fall fairly close to the quantile lines. Likewise, the Shapiro-Wilk tests do not suggest that the data are not lognormal: the p-value for each group, 0.202 for Group 8 and 0.316 for Group 9, exceeds $\alpha = 0.05$ (the null hypothesis is that the transformed sample is normal). It appears then that the lognormal model is appropriate. Additionally, it seems fairly reasonable to assume that the data are independent.

Having assessed the model assumptions, we proceed to analyze the data. We assume here that we are only interested in relative differences between the two means. Thus, we will use the AGP method to obtain a confidence interval for the ratio of the population means. We are using the AGP method because, as emphasized in Section 6, the AGP method provides very accurate coverage frequencies and fairly low bias. A 95% AGP two-sided confidence interval for the ratio of the group means, comparing Group 8 to Group 8, is $(2.24, 11.68)$. Thus, it appears that the mean cost for Group 8 does in fact exceed the mean cost of Group 9. Moreover, because the interval does not include 2, it appears that the mean for Group 8 is more than twice the mean for Group 9.

8 Discussion

Our simulation studies clearly suggest that the AGP method is the preferred method of estimation, particularly in small sample settings. Of course, the studies reported here examined only a finite set of parameter values and sample sizes. Additional studies (not reported in detail here) indicate that the AGP approach continues to perform extremely well for other values of n_i, δ_i, and σ_i. It should be noted that while the AGP method outperforms the GP method in small sample settings, further exploration (not reported in detail) indicates that the two methods perform equally well when the n_i are relatively large.

Parametric approaches, such as those examined in this article, are often criticized because they may perform poorly when the distributional assumptions are violated. Our own studies, not reported here, suggest that the AGP method (as well as the other approaches) are fairly robust to the lognormal assumption. That is, they can continue to provide fairly good coverage even when the lognormal assumption is violated. Further comments regarding this matter are provided in our previous article (Chen and Zhou, 2006).

The fact that the AGP method outperforms the GP method in small sample settings has, in our eyes, very interesting theoretical implications. It seems possible for instance, that GP methods for other scenarios may be "improved" via the use of an approximate GP. We feel that it would be informative to explore the extent to which we can, or should, relax Weerahandi's requirements.

Acknowledgements

This work was supported in part by U.S. Department of Veterans Affairs, Veterans Affairs Health Administration, HSR&D grant ECI-03-206. This paper presents the findings and conclusions of the authors. It does not necessarily represent those of VA HSR&D Service.

References

[1] Agresti A. and Coull B.A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *American Statistician* **52**, 119-126.

[2] Barndorff-Nielsen O.E. (1986). Inference on full and partial parameters, based on the standarized signed log-likelihood ratio. *Biometrika* **73**, 307-322.

[3] Barndorff-Nielsen O.E. (1991). Modified signed log likelihood ratio. *Biometrika* **78**, 557-563.

[4] Callahan C.M., Kesterson J.G., and Tierney W.M. (1997). Association of symptoms of depression with diagnostic test charges. *Annals of Internal Medicine* **126**, 426-432.

[5] Chen Y.-H. and Zhou X.-H. (2006). Interval estimates for the ratio and difference of two lognormal means *Statistics in Medicine*. (In Press).

[6] Hoyle D.C., Rattray M., Jupp R., and Brass A. (2002). Making sense of microarray data distributions. *Bioinformatics* **18**, 576-584.

[7] Krishnamoorthy K. and Mathew T. (2003). Inferences on the means of lognormal distributions using generalized p-values and generalized confidence intervals. *Journal of Statistical Planning and Inference* **115**, 103-120.

[8] Lyles R.H. and Kupper L.L. (1996). On strategies for comparing occupational exposure data to limits. *American Industrial Hygiene Association Journal* **57**, 6-15.

[9] Nixon R.M. and Thompson S.G. (2004). Parametric modelling of cost data in medical studies. *Statistics in Medicine* **23**, 1311-1331.

[10] R Development Core Team (2006). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

[11] Weerahandi S. (1993). Generalized confidence intervals. *Journal of the American Statistical Association* **88**, 899-905.

[12] Wu J., Jiang G., Wong A.C.M., and Sun X. (2002). Likelihood analysis for the ratio of means of two independent log-normal distributions. *Biometrics* **58**, 463-469.

[13] Zhou X.-H. and Tu W. (2000). Interval estimation for the ratio in means of log-normally distributed medical costs with zero values. *Computational Statistics and Data Analysis* **35**, 201-210.

A Argument that T_R is a GP statistic

To show that T_R is a generalized pivotal statistic for $\psi = \log m_1/m_2$ it is necessary to show (a) that its distribution does not depend on any unknown parameters, and (b) that the observed value is free of nuisance parameters. It is apparent from (9) and (10) that the first of the two conditions is satisfied. Note that the distributions of T_i and S_i do not depend on any unknown parameters. Consequently, the distribution of T_R itself does not depend on unknown parameters.

To show that the second condition is satisfied, it is useful to rewrite our statistics as functions of the parameters. First, we rewrite the T_i. Define the random variables $\bar{Y}_i$ and S_i^2 as:

$$\bar{Y}_i = \frac{1}{n_{i1}} \sum_{j=1}^{n_{i1}} Y_{ij} \qquad S_i^2 = \frac{1}{n_{i1}-1} \sum_{j=1}^{n_{i1}} (Y_{ij} - \bar{Y}_i)^2.$$

The Y_{ij} are assumed to be normally distributed with mean μ_i and variance σ_i^2, so $\sqrt{n_{i1}}(\bar{Y}_i - \mu_i)/\sigma_i \sim \mathrm{N}(0,1)$ and $(n_{i1}-1)S_i^2/\sigma_i^2 \sim \chi^2_{n_{i1}-1}$. We can say then that $Z_i = \sqrt{n_{i1}}(\bar{Y}_i - \mu_i)/\sigma_i$ and $U_i^2 = (n_{i1}-1)S_i^2/\sigma_i^2$, and rewrite T_i as:

$$\begin{aligned} T_i &= \hat{\mu}_i - \frac{Z_i}{U_i/\sqrt{n_{i1}-1}} \frac{s_i}{\sqrt{n_{i1}}} + \frac{1}{2} \frac{s_i^2}{U_i^2/(n_{i1}-1)} \\ &= \hat{\mu}_i - \frac{\bar{Y}_i - \mu_i}{S_i/\sqrt{n_{i1}}} s_i/\sqrt{n_{i1}} + \frac{1}{2} \frac{\sigma_i^2}{S_i^2} s_i^2. \end{aligned}$$

It is also useful to rewrite the S_i. Let N_{i0} denote a Binomial(n_i, π_i) random variable:

$$N_{i0} \sim \text{Binomial}(n_i, \pi_i),$$

and define P_i as:

$$P_i = \frac{N_{i0} + \frac{1}{2}z_\alpha^2}{n_i + z_\alpha^2}.$$

Agresti-Coull confidence intervals assume that $(P_i - \pi_i)/\sqrt{P_i(1-P_i)/(n_i + z_\alpha^2)} \sim \text{N}(0,1)$. We can say then that $Z_{i+2} = (P_i - \pi_i)/\sqrt{P_i(1-P_i)/(n_i + z_\alpha^2)}$. The function f_i can then be rewritten as:

$$\begin{aligned} f_i(n_i, n_{i0}) &= \tilde{\pi}_i - Z_{i+2}\sqrt{\tilde{\pi}_i(1-\tilde{\pi}_i)/(n_i + z_\alpha^2)} \\ &= \tilde{\pi}_i - \frac{P_i - \pi_i}{\sqrt{P_i(1-P_i)/(n_i + z_\alpha^2)}}\sqrt{\tilde{\pi}_i(1-\tilde{\pi}_i)/(n_i + z_\alpha^2)}. \end{aligned}$$

The observed value of T_i, the value of the statistic given the sample of interest, is:

$$\begin{aligned} \text{observed value of } T_i &= \hat{\mu}_i - \frac{\hat{\mu}_i - \mu_i}{s_i/\sqrt{n_{i1}}} s_i/\sqrt{n_{i1}} + \frac{1}{2}\frac{\sigma_i^2}{s_i^2}s_i^2 \\ &= \mu_i + \frac{1}{2}\sigma_i^2. \end{aligned}$$

The observed value of S_i when $0 \le f_i \le 1$ is, by definition of S_i, the observed value of f_i:

$$\begin{aligned} \text{observed value of } f_i &= \tilde{\pi}_i - \frac{\tilde{\pi}_i - \pi_i}{\sqrt{\tilde{\pi}_i(1-\tilde{\pi}_i)/(n_i + z_\alpha^2)}}\sqrt{\tilde{\pi}_i(1-\tilde{\pi}_i)/(n_i + z_\alpha^2)} \\ &= \pi_i. \end{aligned}$$

From the above, it is apparent that the observed value of T_R is simply the parameter of interest, ψ. We have shown now that T_R is indeed a generalized pivotal statistic for ψ: the distribution of the statistic does not depend on unknown parameters, and the observed value is free of nuisance

parameters.

Similar arguments to those provided above can be used to show that T_D is a GP for δ. Additionally, it should be apparent from the above that T_R^* and T_D^* are not GP statistics. Note specifically that the observed value of T_i^* is not equal to $\mu_i + \frac{1}{2}\sigma_i^2$.

B Components of the u statistic

First, let $\boldsymbol{\lambda}$, $\boldsymbol{t}$, and $\boldsymbol{\theta}$ be as defined in Section 3.1. Also, define μ_1 as:

$$\mu_1(\psi, \boldsymbol{\lambda}) = \psi - \log(1-\pi_1) - \frac{1}{2}\sigma_1^2 + \log(1-\pi_2) + \mu_2 + \frac{1}{2}\sigma_2^2.$$

Recall that the log-likelihood function, ℓ, reduces when one or more of the samples does not contain zeros. Thus, some of the components of u depend on n_{10} or n_{20}.

The partial derivative of ℓ with respect to $\boldsymbol{t}$ is:

$$\begin{aligned}
\ell_{\boldsymbol{t}}(\psi, \boldsymbol{\lambda}) &= \Big(\frac{\mu_1}{\sigma_1^2}, \frac{\mu_2}{\sigma_2^2}, -\frac{1}{2\sigma_1^2}, -\frac{1}{2\sigma_2^2}, \\
&\quad -\log\sigma_1 - \frac{\mu_1^2}{2\sigma_1^2} + \{\log(1-\pi_1) - \log\pi_1\}\, 1_{[n_{10}\neq 0]}, \\
&\quad -\log\sigma_2 - \frac{\mu_2^2}{2\sigma_2^2} + \{\log(1-\pi_2) - \log\pi_2\}\, 1_{[n_{20}\neq 0]}\Big).
\end{aligned}$$

The partial derivative of ℓ with respect to $\boldsymbol{\lambda}$ and $\boldsymbol{t}$ is $\ell_{\boldsymbol{\lambda t}}(\psi, \boldsymbol{\lambda}) = [a_{ij}]_{6\times 5}$, where:

$$\begin{aligned}
(a_{11}, a_{12}, a_{13}, a_{14}, a_{15}) &= \Big(\frac{1}{\sigma_1^2}, -\frac{1}{\sigma_1} - \frac{2\mu_1}{\sigma_1^3}, \frac{\sigma_2}{\sigma_1^2}, \frac{1}{\sigma_1^2(1-\pi_1)}, -\frac{1}{\sigma_1^2(1-\pi_2)}\Big) \\
(a_{21}, a_{22}, a_{23}, a_{24}, a_{25}) &= \Big(\frac{1}{\sigma_2^2}, 0, -\frac{2\mu_2}{\sigma_2^3}, 0, 0\Big) \\
(a_{31}, a_{32}, a_{33}, a_{34}, a_{35}) &= \Big(0, \frac{1}{\sigma_1^3}, 0, 0, 0\Big) \\
(a_{41}, a_{42}, a_{43}, a_{44}, a_{45}) &= \Big(0, 0, \frac{1}{\sigma_2^3}, 0, 0\Big)
\end{aligned}$$

$$
\begin{aligned}
(a_{51}, a_{52}, a_{53}, a_{54}, a_{55}) &= \Big(-\frac{\mu_1}{\sigma_1^2}, -\frac{1}{\sigma_1} + \frac{\mu_1}{\sigma_1} + \frac{\mu_1^2}{\sigma_1^3}, -\frac{\mu_1\sigma_2}{\sigma_1^2}, \\
&\quad -\frac{1}{1-\pi_1} - \frac{\mu_1}{\sigma_1^2(1-\pi_1)} - \frac{1}{\pi_1} 1_{[n_{10}\neq 0]}, \frac{\mu_1}{\sigma_1^2(1-\pi_2)} \Big) \\
(a_{61}, a_{62}, a_{63}, a_{64}, a_{65}) &= \Big(-\frac{\mu_2}{\sigma_2^2}, 0, -\frac{1}{\sigma_2} + \frac{\mu_2^2}{\sigma_2^3}, 0, -\frac{1}{1-\pi_2} - \frac{1}{\pi_2} 1_{[n_{20}\neq 0]} \Big).
\end{aligned}
$$

The determinant of $\ell_{\boldsymbol{\theta t}}$ is:

$$
|\ell_{\boldsymbol{\theta t}}(\psi, \boldsymbol{\lambda})| = \begin{cases}
\{\sigma_1^5\sigma_2^5\pi_1(1-\pi_1)\pi_2(1-\pi_2)\}^{-1} & \text{if } n_{10} \neq 0 \text{ and } n_{20} \neq 0 \\
\{\sigma_1^5\sigma_2^5\pi_1(1-\pi_1)\}^{-1} & \text{if } n_{10} \neq 0 \text{ and } n_{20} = 0 \\
\{\sigma_1^5\sigma_2^5\pi_2(1-\pi_2)\}^{-1} & \text{if } n_{10} = 0 \text{ and } n_{20} \neq 0 \\
(\sigma_1^5\sigma_2^5)^{-1} & \text{if } n_{10} = 0 \text{ and } n_{20} = 0
\end{cases}
$$

and the determinant of $\ell_{\boldsymbol{\theta\theta}}$ is:

$$
|\ell_{\boldsymbol{\theta\theta}}(\psi, \boldsymbol{\lambda})| = \begin{cases}
4n_1^3n_2^3(1-\pi_1)(1-\pi_2)(\sigma_1^4\sigma_2^4\pi_1\pi_2)^{-1} & \text{if } n_{10} \neq 0 \text{ and } n_{20} \neq 0 \\
4n_1^3n_2^3(1-\pi_1)(\sigma_1^4\sigma_2^4\pi_1)^{-1} & \text{if } n_{10} \neq 0 \text{ and } n_{20} = 0 \\
4n_1^3n_2^3(1-\pi_2)(\sigma_1^4\sigma_2^4\pi_2)^{-1} & \text{if } n_{10} = 0 \text{ and } n_{20} \neq 0 \\
4n_1^3n_2^3(\sigma_1^4\sigma_2^4)^{-1} & \text{if } n_{10} = 0 \text{ and } n_{20} = 0
\end{cases}
$$

Finally, the matrix $\ell_{\boldsymbol{\lambda\lambda}}$ is given by $\ell_{\boldsymbol{\lambda\lambda}}(\psi, \boldsymbol{\lambda}) = [b_{ij}]_{5\times 5}$ where:

$$
\begin{aligned}
(b_{11}, b_{12}, b_{13}, b_{14}, b_{15}) &= \Big(\frac{t_5}{\sigma_1^2} + \frac{t_6}{\sigma_2^2}, \frac{2t_1}{\sigma_1^3} - \frac{t_5}{\sigma_1} - \frac{2\mu_1 t_5}{\sigma_1^3}, \frac{2t_2}{\sigma_2^3} + \frac{\sigma_2 t_5}{\sigma_1^2} - \frac{2\mu_2 t_6}{\sigma_2^3}, \\
&\quad \frac{t_5}{\sigma_1^2(1-\pi_1)}, -\frac{t_5}{\sigma_1^2(1-\pi_2)} \Big) \\
(b_{21}, b_{22}, b_{23}, b_{24}, b_{25}) &= \Big(b_{12}, -\frac{3t_1}{\sigma_1^2} - \frac{6\mu_1 t_1}{\sigma_1^4} + \frac{3t_3}{\sigma_1^4} + t_5 - \frac{t_5}{\sigma_1^2} + \frac{3\mu_1 t_5}{\sigma_1^2} + \frac{3\mu_1^2 t_5}{\sigma_1^4}, \\
&\quad \frac{2\sigma_2 t_1}{\sigma_1^3} - \frac{\sigma_2 t_5}{\sigma_1} - \frac{2\mu_1\sigma_2 t_5}{\sigma_1^3}, \frac{2t_1}{\sigma_1^3(1-\pi_1)} - \frac{t_5}{\sigma_1(1-\pi_1)} - \frac{2\mu_1 t_5}{\sigma_1^3(1-\pi_1)}, \\
&\quad -\frac{2t_1}{\sigma_1^3(1-\pi_2)} + \frac{t_5}{\sigma_1(1-\pi_2)} + \frac{2\mu_1 t_5}{\sigma_1^3(1-\pi_2)} \Big)
\end{aligned}
$$

$$
\begin{aligned}
(b_{31}, b_{32}, b_{33}, b_{34}, b_{35}) &= \Big(b_{13}, b_{23}, -\frac{t_1}{\sigma_1^2} - \frac{6\mu_2 t_2}{\sigma_2^4} + \frac{3t_4}{\sigma_2^4} + \frac{\sigma_2^2 t_5}{\sigma_1^2} + \frac{\mu_1 t_5}{\sigma_1^2} - \frac{t_6}{\sigma_2^2} + \frac{3\mu_2^2 t_6}{\sigma_2^4}, \\
&\quad \frac{\sigma_2 t_5}{\sigma_1^2(1-\pi_1)}, -\frac{\sigma_2 t_5}{\sigma_1^2(1-\pi_2)}\Big) \\
(b_{41}, b_{42}, b_{43}, b_{44}, b_{45}) &= \Big(b_{14}, b_{24}, b_{34}, -\frac{t_1}{\sigma_1^2(1-\pi_1)^2} + \frac{t_5}{(1-\pi_1)^2} + \frac{t_5}{\sigma_1^2(1-\pi_1)^2} \\
&\quad + \frac{\mu_1 t_5}{\sigma_1^2(1-\pi_1)^2} + \frac{n_1 - t_5}{\pi_1^2} 1_{[n_{10}\neq 0]}, -\frac{t_5}{\sigma_1^2(1-\pi_1)(1-\pi_2)}\Big) \\
(b_{51}, b_{52}, b_{53}, b_{54}, b_{55}) &= \Big(b_{15}, b_{25}, b_{35}, b_{45}, \frac{t_1}{\sigma_1^2(1-\pi_2)^2} + \frac{t_5}{\sigma_1^2(1-\pi_2)^2} - \frac{\mu_1 t_5}{\sigma_1^2(1-\pi_2)^2} \\
&\quad + \frac{t_6}{(1-\pi_2)^2} + \frac{n_2 - t_6}{\pi_2^2} 1_{[n_{20}\neq 0]}\Big).
\end{aligned}
$$

Note that the components described above should be evaluated at the maximum likelihood or constrained maximum likelihood estimates, as indicated in (22).

Table 1: Parameters and sample sizes used in the simulation studies.

Design	n_1	n_2	δ_1	δ_2	μ_1	σ_1^2	σ_2^2
1a	10	10	0.1	0.1	0	1	1
1b	25	25					
1c	10	25					
2a	10	10	0.1	0.1	0.75	0.5	2
2b	25	25					
2c	10	25					
3a	10	10	0.1	0.2	0	1	1
3b	25	25					
3c	10	25					
3d	25	10					
4a	10	10	0.1	0.2	0.75	0.5	2
4b	25	25					
4c	10	25					
4d	25	10					
5a	10	10	0.1	0.2	−0.75	2	0.5
5b	25	25					
5c	10	25					
5d	25	10					

Table 2: Results from the simulation studies.

Design	Method	Coverage	Med. width	Left error	Right error	Rel. bias
1a	GP	96.19	4.80	1.96	1.85	−0.03
	AGP	95.66	4.18	2.29	2.05	−0.06
	SLLR	93.68	3.28	3.23	3.09	−0.02
	MSLLR	94.05	4.20	3.06	2.89	−0.03
	ML	92.88	2.61	3.63	3.49	−0.02
	Bootstrap	92.44	2.60	3.75	3.81	0.01
1b	GP	95.99	1.85	2.00	2.01	0.00
	AGP	95.82	1.78	2.06	2.12	0.01
	SLLR	94.10	1.69	2.89	3.01	0.02
	MSLLR	94.67	1.80	2.59	2.74	0.03
	ML	94.14	1.57	2.85	3.01	0.03
	Bootstrap	93.68	1.55	3.06	3.26	0.03
1c	GP	95.55	3.95	2.88	1.57	−0.29
	AGP	94.85	3.31	3.65	1.50	−0.42
	SLLR	93.10	2.69	4.13	2.77	−0.20
	MSLLR	94.11	3.78	3.26	2.63	−0.11
	ML	92.50	1.95	5.54	1.96	−0.48
	Bootstrap	91.84	1.87	6.25	1.91	−0.53

Design	Method	Coverage	Med. width	Left error	Right error	Rel. bias
2a	GP	95.81	3.83	2.22	1.97	−0.06
	AGP	95.20	3.75	1.94	2.86	0.19
	SLLR	93.28	3.34	2.38	4.34	0.29
	MSLLR	93.94	3.52	2.66	3.40	0.12
	ML	90.28	3.77	0.71	9.01	0.85
	Bootstrap	89.52	3.74	0.88	9.60	0.83
2b	GP	95.30	1.94	2.53	2.17	−0.08
	AGP	95.08	1.95	2.27	2.65	0.08
	SLLR	93.91	1.92	2.27	3.82	0.25
	MSLLR	94.62	1.96	2.52	2.86	0.06
	ML	92.70	2.06	0.91	6.39	0.75
	Bootstrap	92.07	2.04	1.05	6.88	0.74
2c	GP	95.58	2.60	2.93	1.49	−0.33
	AGP	95.21	2.45	3.07	1.72	−0.28
	SLLR	94.29	2.29	2.52	3.19	0.12
	MSLLR	94.71	2.51	2.48	2.81	0.06
	ML	93.60	2.26	1.54	4.86	0.52
	Bootstrap	93.25	2.20	1.87	4.88	0.45

Design	Method	Coverage	Med. width	Left error	Right error	Rel. bias
3a	GP	96.36	5.55	1.32	2.32	0.27
	AGP	95.73	4.86	1.57	2.70	0.26
	SLLR	92.95	3.95	1.90	5.15	0.46
	MSLLR	93.86	4.98	1.76	4.38	0.43
	ML	92.65	2.83	3.31	4.04	0.10
	Bootstrap	92.08	2.92	3.18	4.74	0.20
3b	GP	95.28	2.16	1.01	3.71	0.57
	AGP	94.85	2.09	1.09	4.06	0.58
	SLLR	92.96	2.01	1.30	5.74	0.63
	MSLLR	93.74	2.13	1.21	5.05	0.61
	ML	94.18	1.66	2.59	3.23	0.11
	Bootstrap	93.65	1.68	2.59	3.76	0.18
3c	GP	95.58	4.48	1.90	2.52	0.14
	AGP	95.15	3.79	2.39	2.46	0.01
	SLLR	92.74	3.13	2.61	4.65	0.28
	MSLLR	93.53	4.32	1.97	4.50	0.39
	ML	92.85	2.05	5.17	1.98	−0.45
	Bootstrap	92.26	2.00	5.58	2.16	−0.44
3d	GP	96.08	3.08	0.87	3.05	0.56
	AGP	95.42	3.02	0.82	3.76	0.64
	SLLR	92.54	2.74	1.45	6.01	0.61
	MSLLR	94.65	2.88	1.41	3.94	0.47
	ML	92.33	2.36	1.91	5.76	0.50
	Bootstrap	91.11	2.53	1.65	7.24	0.63

Design	Method	Coverage	Med. width	Left error	Right error	Rel. bias
4a	GP	95.98	4.48	1.60	2.42	0.20
	AGP	95.32	4.40	1.40	3.28	0.40
	SLLR	92.34	4.06	1.44	6.22	0.62
	MSLLR	94.06	4.25	1.79	4.15	0.40
	ML	89.13	4.15	0.80	10.07	0.85
	Bootstrap	87.92	4.35	0.79	11.29	0.87
4b	GP	95.35	2.31	1.48	3.17	0.36
	AGP	94.87	2.31	1.23	3.90	0.52
	SLLR	93.08	2.32	1.05	5.87	0.70
	MSLLR	94.47	2.34	1.19	4.34	0.57
	ML	92.17	2.22	0.83	7.00	0.79
	Bootstrap	91.30	2.24	0.86	7.84	0.80
4c	GP	95.50	3.02	1.95	2.55	0.13
	AGP	95.11	2.87	1.96	2.93	0.20
	SLLR	93.39	2.74	1.45	5.16	0.56
	MSLLR	94.12	2.96	1.41	4.47	0.52
	ML	93.18	2.43	1.51	5.31	0.56
	Bootstrap	92.73	2.42	1.67	5.60	0.54
4d	GP	95.60	3.66	1.39	3.01	0.37
	AGP	94.97	3.77	1.05	3.98	0.58
	SLLR	92.18	3.58	1.25	6.57	0.68
	MSLLR	94.76	3.61	1.63	3.61	0.38
	ML	88.20	3.82	0.61	11.19	0.90
	Bootstrap	86.51	4.11	0.57	12.92	0.92

Design	Method	Coverage	Med. width	Left error	Right error	Rel. bias
5a	GP	95.65	16.12	1.56	2.79	0.28
	AGP	95.27	11.54	2.15	2.58	0.09
	SLLR	93.12	7.23	3.14	3.74	0.09
	MSLLR	93.29	14.57	2.49	4.22	0.26
	ML	90.81	3.14	8.42	0.77	−0.83
	Bootstrap	90.52	3.12	8.50	0.98	−0.79
5b	GP	94.53	3.36	1.25	4.22	0.54
	AGP	94.64	3.09	1.47	3.89	0.45
	SLLR	93.66	2.85	1.99	4.35	0.37
	MSLLR	93.71	3.31	1.58	4.71	0.50
	ML	93.23	1.96	5.72	1.05	−0.69
	Bootstrap	92.70	1.93	5.97	1.33	−0.64
5c	GP	94.80	14.84	2.06	3.14	0.21
	AGP	94.52	10.31	2.91	2.57	−0.06
	SLLR	92.66	6.52	3.78	3.56	−0.03
	MSLLR	93.66	14.04	2.24	4.10	0.29
	ML	89.29	2.69	10.32	0.39	−0.93
	Bootstrap	88.19	2.53	11.29	0.52	−0.91
5d	GP	95.82	4.09	0.85	3.33	0.59
	AGP	95.49	3.81	0.94	3.57	0.58
	SLLR	93.51	3.27	1.97	4.52	0.39
	MSLLR	94.25	3.68	1.53	4.22	0.47
	ML	93.51	2.34	4.69	1.80	−0.45
	Bootstrap	93.25	2.44	4.30	2.45	−0.27

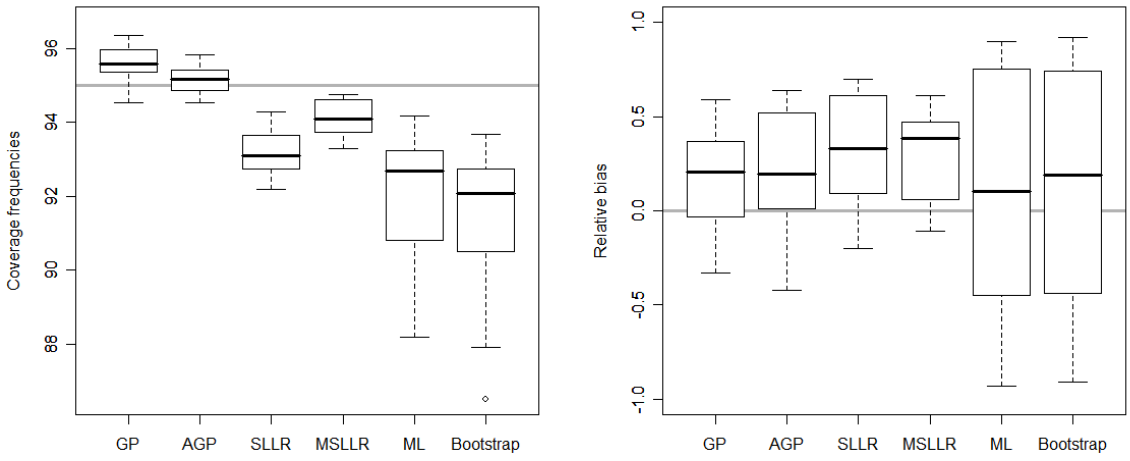

Figure 1: Coverage and bias results from the simulation studies, by method. All 108 studies (18 designs $\times$ 6 methods) are represented.

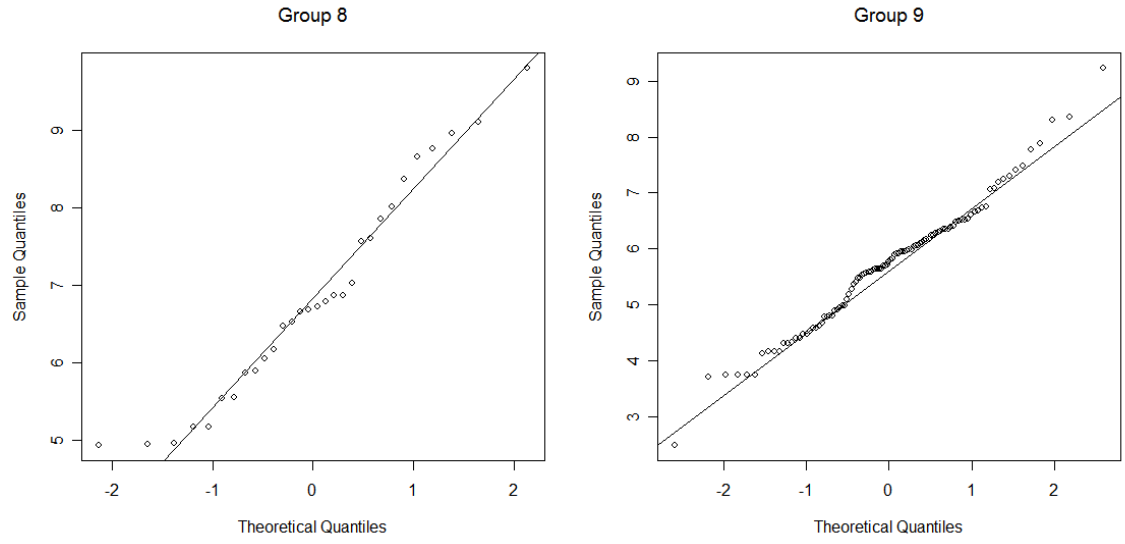

Figure 2: Normal quantile plots for the log-transformed positive costs.