Bayesian Confidence Intervals for the Mean of a Lognormal Distribution: A Comparison with the MOVER and Generalized Confidence Interval Procedures

J. Harvey[a,b],
P.C.N. Groenewald[b]
&
A.J. van der Merwe[b]

[a]Centre for Statistical Consultation
Department of Statistics and Actuarial Science
University of Stellenbosch
Private Bag X1
Stellenbosch
South Africa
jharvey@sun.ac.za (corresponding author)

[b]Department of Mathematical Statistics and Actuarial Science
University of the Free State
P.O. Box 339
Bloemfontein
9300
South Africa
Matheeme.sci@ufs.ac.za

Abstract

The lognormal distribution is currently used extensively to describe the distribution of positive random variables. This is especially the case with data pertaining to occupational health and other biological data. One particular application of the data is statistical inference with regards to the mean of the data. Other authors, namely Zou, Taleban and Huo (2009), have proposed procedures involving the so-called “method of variance estimates recovery” (MOVER), while an alternative approach based on simulation is the so-called generalized confidence interval, discussed by Krishnamoorthy and Mathew (2003). In this paper we compare the performance of the MOVER-based confidence interval estimates and the generalized confidence interval procedure to coverage of credibility intervals obtained using Bayesian methodology using a variety of different prior distributions to estimate the appropriateness of each. An extensive simulation study is conducted to evaluate the coverage accuracy and interval width of the proposed methods. For the Bayesian approach both the equal-tail and highest posterior density (HPD) credibility intervals are presented. Various prior distributions (independence Jeffreys' prior, the Jeffreys-rule prior, namely, the square root of the determinant of the Fisher Information matrix and reference and probability-matching priors) are evaluated and compared to determine which give the best coverage with the most efficient interval width. The simulation studies show that the constructed Bayesian confidence intervals have satisfying coverage probabilities and in some cases outperform the MOVER and generalized confidence interval results.

Keywords: Bayesian procedure; Lognormal; Highest Posterior Density; MOVER; Credibility intervals; Coverage probabilities.

1 Introduction

Lognormally distributed data presents itself in a number of scientific fields. According to Limpert et al. (2001), the distribution may be used to approximate right skewed data that arises in a wide variety of scientific settings. Particularly in the area of health costs the log distribution has been extensively used by other authors and numerous statistical methods have been developed.

For testing the equality of means from two skewed populations Zhou, Gao and Hui (1997) and Zhou, Melfi and Hui (1997) first proposed a Z-score method for populations that have lognormal distributions. Zhou and Tu (1999) then extended the scenario by proposing a likelihood ratio test for instance when the populations contain both zero and non-zero observations. Furthermore, Zhou and Gao (1997) did propose confidence intervals for the one-sample lognormal mean, but until Zhou and Tu (2000) no confidence intervals had been proposed for methods that compare the means of two populations. Even though Zhou & Tu (1999) provided a test of whether the means were the same, if they were indeed found to be different the method they proposed did not add any additional information on the relative differences and magnitudes of the two population means, as a confidence interval would indeed.

As mentioned, Zhou and Tu (2000) considered the problem of constructing confidence intervals for the ratio of two means of independent populations that contain both lognormal and zero observations. The context of the proposed techniques is as described earlier concerning the excess charges of diagnostic testing. For the purposes of the analysis a maximum likelihood method and a two-stage bootstrap method was used. An extensive simulation study was conducted to ascertain the coverage accuracy, interval width and relative bias of the proposed methods. The focus was also on inferences about the overall population means, including zero costs.

According to Hannig, Lidong, Abdel-Karim and Iyer (2006) simultaneous confidence intervals for certain lognormal parameters are useful in pharmaceutical studies. In bioequivalence studies comparing a test drug to a reference drug it is of interest to compare the mean response of the two drugs to ensure that they are (more or less) equally effective. An important variable in these studies is the area under the curve relating the plasma drug concentration in a patient to the elapsed time after the drug is administered. It is usually assumed that the area under the curve is lognormally distributed. So the parameter of interest is the ratio of means of two lognormal distributions.

Also, as mentioned in Zhou, Taliban and Huo (2009) since many health cost data may be positively skewed the literature dealing with the analysis of the means of lognormal data has increased substantially. This includes procedures for a single sample mean, a difference between two sample means and additional zero values for each of these cases.

Many authors have proposed methods based on the simulation of pivotal statistics, commonly referred to as generalized confidence intervals. These methods have generated a series of papers on the means of lognormal data (see for example Krishnamoorthy and Mathew (2003), Tian (2005), Chen and Zhou (2006), Krishnamoorthy, Mathew and Romachandren (2006), Tian and Wu (2007 [a],[b]) and Bebu and Mathew (2008)). For further details refer to Zou et al (2009). The simulation of pivotal statistics is a frequentist method that is in effect rather similar to some of the Bayesian procedures that will be proposed in the next sections.

Instead of adapting a simulation approach for making inferences on the lognormal mean, Zou, Taliban and Huo (2009) proposed procedures involving the so-called "method of variance estimates recovery" (MOVER). The MOVER method was designed in order to apply to a general scenario and also to provide adequate coverage rates in estimation procedures relating to lognormally distributed data. The issue of zero values is also potentially problematic where in practical situations (many arising from the medical field) non-zero data is right-skewed, but zero-valued observations are present in the data. The MOVER method can also be used to handle zero values quite easily. The advantage of the MOVER is therefore that it is easily applicable to many different settings with little more than a basic knowledge of introductory statistical text.

In this paper an extensive simulation study will be conducted to ascertain the coverage accuracy of the proposed methods. These proposed methods are all based on a Bayesian framework, where the choice of prior distribution is the factor of interest. Specifically, the choice of different prior distributions in different parameters settings and the appropriateness of each is of primary importance. This will be compared to the MOVER algorithm developed by Zou et al (2009) and the generalized confidence interval procedure proposed by Krishnamoorthy and Mathew (2003) by means of simulation studies. Another setting that is also addressed in this paper is that of zero values, particularly zero costs in medical health costs. The focus was also on inferences about the overall population means, including zero costs.

As mentioned a Bayesian approach to the problem will be taken. Depending on the choice of prior distribution it will be shown that the Bayesian procedure has equal or better coverage accuracy than both the MOVER method and generalized confidence interval procedure. In the next section we begin with a formulation of the model and a specification of all parameters and distributions of interest. In further sections we compare the performance of the methods for different prior distributions by conducting a simulation study to assess specific quantities of the proposed credibility intervals in pre-defined finite sample sizes (the same as those used by Zhou and Tu [2000]).

2 Description of the Setting

The first setting is the straightforward lognormal distribution, i.e. with no zero valued observations/costs. Bayesian methods will be compared to the MOVER method and generalized confidence interval methods.

The second setting allows for the possibility of zero-valued observations. The lognormal distribution in itself does not allow for zero values to be included in the data. This suggests an interesting setting, namely the analysis of data that contains both zero and non-zero values, with the non-zero values being lognormally distributed.

3 The Case of No Zero-Valued Observations

3.1 Model Formulation

From the specification of the problem in the Introduction we can assume that the population of interest contains only non-zero (positive observations) Furthermore, it is assumed that the observations are distributed lognormally with parameters μ and σ^2. Now, let $X_1, X_2, \dots, X_n$ be a random sample and let $M = E(X_i)$. From this preliminary setting specification we wish to construct credibility intervals for mean, M. As mentioned the observations are lognormally distributed: $\ln(X_i) \sim N(\mu, \sigma^2)$, for $i = 1, \dots, n$. From this it follows that the mean is given by:

$$M = exp\left(\mu + \frac{1}{2}\sigma^2\right).$$

3.2 Intervals Based on a Bayesian Procedure

Denote $y_i = ln(X_i)$ then the likelihood function is given by:

$$L(\mu, \sigma^2 | data) \propto \prod_{i=1}^{n} \left(\frac{1}{\sigma^2}\right)^{1/2} exp\left[-\frac{(y_i - \mu)^2}{2\sigma^2}\right] \tag{1}$$

The choice of prior to be used in this setting will be discussed in further sections. Given the previous specification of the likelihood, the Fisher Information Matrix can be written as:

$$I(\mu, \sigma^2) = -E\left\{\frac{\partial^2}{\partial^2(\mu, \sigma^2)} lnL(\mu, \sigma^2 | data)\right\}$$

Therefore,

$$I(\mu, \sigma^2) = diag\left[\frac{n}{\sigma^2} \quad \frac{n}{2\sigma^4}\right] \tag{2}$$

3.2.1 Independence Jeffreys Prior:

Since $\boldsymbol{\theta} = (\mu, \sigma^2)$ is unknown the prior

$$p(\boldsymbol{\theta}) \propto \frac{1}{\sigma^2} \tag{3}$$

will be specified for the unknown parameters. This is known as the independence Jeffreys prior. In (3) we have assumed μ and $\sigma^2$to be independently distributed, *a priori*, with μ and $log\sigma^2$ each uniformly distributed. See Zellner (1971) and Box and Tiao (1973) for further discussion. Combining the likelihood function (1) and the prior density function (3) the joint posterior density function can be written as:

$$P(\boldsymbol{\theta}|data) = \left(\frac{2\pi\sigma^2}{n}\right)^{-1/2} exp\left(-\frac{n}{2\sigma^2}(\mu - \hat{\mu})^2\right)\left(\frac{\nu\hat{\sigma}^2}{2}\right)^{1/2\nu}\left(\frac{(\sigma^2)^{-1/2(\nu+2)} exp\left[-\frac{\nu\hat{\sigma}^2}{2\sigma^2}\right]}{\Gamma\left(\frac{\nu}{2}\right)}\right) \tag{4}$$

where $\hat{\mu} = \frac{1}{n}\sum_1^n y_i$, $\nu = n - 1$ and

$\hat{\sigma}^2 = \frac{1}{n-1}\sum_1^n(y_i - \hat{\mu})^2$.

From (4) it follows that the posterior distribution can be defined as the conditional posterior distribution of μ, which is normal:

$$\mu|\sigma^2, data \sim N\left(\hat{\mu}, \frac{\sigma^2}{n}\right) \tag{5}$$

and for σ^2, the posterior density function is an Inverted Gamma density, specifically:

$$P(\sigma^2|data) = \left(\frac{\nu\hat{\sigma}^2}{2}\right)^{1/2\nu}\left(\frac{(\sigma^2)^{-1/2(\nu+2)} exp\left[-\frac{\nu\hat{\sigma}^2}{2\sigma^2}\right]}{\Gamma\left(\frac{\nu}{2}\right)}\right). \tag{6}$$

From equation (6) it follows that $\tau^* = \frac{\nu\ \hat{\sigma}^2}{\sigma^2}$ has a chi-square distribution with ν degrees of freedom. From classical statistics (if $\hat{\sigma}^2$ is considered to be random) it is well known that τ^* is also distributed chi-square with ν degrees of freedom. This agreement between classical and Bayesian statistics is only true if the prior $p(\sigma^2) \propto \sigma^{-2}$ is used. If some other prior distributions are used, for example $p(\sigma^2) \propto \sigma^{-3}$ or $p(\sigma^2) \propto constant$, then the posterior of τ^* will still be a chi-square distribution but the degrees of freedom will be different.

The method proposed here to find the Bayesian credibility intervals for lnM , the log of the mean, is through Monte Carlo simulation. Since $lnM = \mu + \frac{1}{2}\sigma^2$ standard routines can be used in the simulation procedure.

3.2.2 Simulation Procedure:

The following simulation was obtained from the preceding theory using the MATLAB® package:

1. For given values of μ, σ^2 and n, sample values $\hat{\mu}$ and $\hat{\sigma}^2$ were drawn from $\hat{\mu} \sim N\left(\mu , \frac{\sigma^2}{n}\right)$ and $\hat{\sigma}^2 = \frac{u\sigma^2}{\nu}$, where $u \sim \chi^2_\nu$. This was repeated 10000 times. It is only necessary to simulate the sufficient statistics $\hat{\mu}$ and $\hat{\sigma}^2$ for the sample data and not the complete sample.
2. For every pair of $\hat{\mu}$ and $\hat{\sigma}^2$ values, μ and σ^2 can be simulated from their respective posterior distributions, as given in (5) and (6).

3. For the 10000 generated values, $lnM = \mu + \frac{1}{2}\sigma^2$ can be calculated and ordered. The equal-tailed and shortest 95% Highest Posterior Density Intervals can then be obtained.
4. The coverage and average widths of the intervals from the 10000 samples can be calculated.

3.2.3 Alternative Prior Distributions

3.2.3.1 Jeffreys-Rule Prior:

As mentioned in the Abstract and Introduction to this document, one of the objectives is to compare the Bayesian procedure for different choices of prior distributions for $\boldsymbol{\theta} = [\mu\, \sigma^2]'$, the unknown parameters. In the previous two sections we discussed the analysis methods using Jeffreys' non-informative prior and the resulting simulation technique respectively.

In this and subsequent sections different choices of prior distributions will be discussed in an effort to eventually compare the results. The choice of prior applied in this section is the square root of the determinant of the Fisher Information Matrix, which is an adaptation of the Jeffreys' rule used in the previous section.

The Jeffreys'-Rule prior becomes

$$p(\boldsymbol{\theta}) \propto \sigma^{-3} \qquad (7)$$

This was derived from $|I(\boldsymbol{\theta})|^{1/2}$, which was defined in (2). In (3) we have assumed μ and σ^2 to be independently distributed, *a priori*, with μ and $log\sigma^2$ each uniformly distributed. Combing the likelihood function (1) and the prior density function (7) it follows that τ^* has a chi-square distribution with $\nu + 1$ degrees of freedom and the posterior distribution of σ^2 is as defined in (6) with $\nu + 1$ instead of ν.

A similar simulation procedure to the one previously described can be used only with $\sigma^2 = \frac{\nu\hat{\sigma}^2}{\chi^2_{\nu+1}}$.

3.2.3.2 Reference and Probability Matching Priors

In addition to the two previously mentioned Jeffreys' Prior distributions the following prior distributions were all tested:

$$p(\boldsymbol{\theta}) \propto \frac{1}{\sigma}\sqrt{1 + \frac{2}{\sigma^2}} \qquad (8)$$

$$p(\boldsymbol{\theta}) \propto \frac{1}{\sigma^2}\sqrt{1 + \frac{2}{\sigma^2}} \qquad (9)$$

Prior distributions (8) and (9) are the reference and probability-matching priors, the derivation of which will be discussed and provided in later sections of this paper.

In each case, 10000 samples were simulated and for each sample 10000 Bayesian simulations were made to obtain credibility intervals.

3.3 A Comparison of the Bayesian Methods with the Generalized Confidence Interval Procedure and the MOVER

The object of the paper by Zou et al (2009, page 3760) was to demonstrate the MOVER in different scenarios, i.e. for a few different combinations of n and σ^2, where $\mu = -\frac{1}{2}\sigma^2$. The same scenarios will be presented here for the Bayesian confidence intervals and these will be compared to the results from the MOVER method and generalized confidence interval procedure to evaluate the performance of the Bayesian confidence intervals for both the equal-tailed and HPD intervals. The following characteristics are reported:

- coverage probabilities
- average interval lengths

A nominal significance level of $\alpha = 0.05$ will be used for each parameter setting.

The confidence limits for the MOVER, as given by Zou et al (2009) on page 3758, are:

$$L = \hat{\mu} + \frac{\hat{\sigma}^2}{2} - \sqrt{Z^2_{\alpha/2}\frac{\hat{\sigma}^2}{n}\left\{\frac{\hat{\sigma}^2}{2}\left(1 - \frac{\nu}{\chi^2_{1-\alpha/2,\nu}}\right)\right\}^2}$$

$$U = \hat{\mu} + \frac{\hat{\sigma}^2}{2} + \sqrt{Z^2_{\alpha/2}\frac{\hat{\sigma}^2}{n}\left\{\frac{\hat{\sigma}^2}{2}\left(\frac{\nu}{\chi^2_{\alpha/2,\nu-1}}\right)\right\}^2}$$

The simulation procedure as described for lnM in Section 3.2.2 can be summarized as follows:

$$lnM = \hat{\mu} + \frac{Z}{\sqrt{\frac{\tau^*}{\nu}}}\frac{\hat{\sigma}}{n} + \frac{\hat{\sigma}^2}{\frac{2\tau^*}{\nu}}$$

where
$Z \sim N(0,1)$ and $\tau^* \sim \chi^2_\nu$.

The generalized confidence interval procedure discussed by Krishnamoorthy and Mathew (2003) involves simulating the pivotal statistic:

$$lnM = \hat{\mu} - \frac{Z}{\sqrt{\frac{\tau^*}{\nu}}}\frac{\hat{\sigma}}{n} + \frac{\hat{\sigma}^2}{\frac{2\tau^*}{\nu}}$$

The only difference is the change of the sign. The reason for this is that from a frequentist point of view $\hat{\mu}|\mu, \sigma^2 \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, while Bayesian statisticians would say that $\mu|\hat{\mu}, \sigma^2 \sim N\left(\hat{\mu}, \frac{\sigma^2}{n}\right)$. Since $Z \sim N(0,1)$ the two equations for lnM above are the same. As mentioned, this will be the case only if the prior $p(\mu, \sigma^2) \propto \sigma^{-2}$ is used.

3.4 Discussion of Results for the Simulation Studies

The objective of this report was to compare these results against results obtained from a Bayesian-based simulation study using a specifically chosen set of prior distributions and to evaluate the performance of each prior distribution against both the other distributions and the results obtained by Zou et al (2009), overall. The generalized confidence intervals will also be discussed.

The following table presents the summary statistics of the results in both the Zou et al (2009) simulation study and the Bayesian simulation study using Jeffreys' Independence prior.

The same designs are used as considered by Zou et al (2009), where (< , >)% refers to the proportion of cases where the interval is below or above the true value respectively:

Table 1
Comparison of the MOVER and Independence Jeffreys' Prior for Constructing Two-sided 95% Confidence Intervals for $\mu + \frac{1}{2}\sigma^2$

		MOVER		GCI		Jeffreys' (Equal-tail)		Jeffreys' (HPD)	
n	σ^2	Cover (< , >)%	Width	Cover (< , >)%	Width	Cover (< , >)%	Width	Cover (< , >)%	Width
5	0.5	93.47 (3.19, 3.34)	2.54	93.99 (2.00, 4.01)	2.69	94.31 (1.87, 3.82)	2.68	95.67 (3.01, 1.32)	2.27
	1.0	94.65 (2.75, 2.60)	4.66	93.98 (1.92, 4.10)	4.78	93.81 (2.16, 4.03)	4.73	95.69 (3.27, 1.04)	3.80
	1.5	95.03 (2.92, 2.05)	6.67	94.15 (2.07, 3.78)	6.76	93.55 (2.09, 4.36)	6.75	95.83 (3.48, 0.69)	5.26
	2.0	95.10 (2.99, 1.91)	8.76	93.77 (2.36, 3.87)	8.82	94.13 (2.11, 3.76)	8.68	95.70 (3.66, 0.64)	6.64
	2.5	95.30 (2.89, 1.81)	10.55	94.08 (2.38, 3.54)	10.59	94.21 (2.14, 3.65)	10.82	95.84 (3.80, 0.36)	8.16
	3.0	95.35 (2.60, 2.05)	12.72	93.90 (2.09, 4.01)	12.74	94.35 (2.07, 3.58)	12.71	95.81 (3.68, 0.51)	9.50
20	0.5	94.24 (3.37, 2.39)	0.74	94.56 (2.35, 3.09)	0.76	94.56 (2.29, 3.15)	0.77	95.06 (3.00, 1.94)	0.75
	1.0	95.19 (2.87, 1.94)	1.20	94.90 (2.17, 2.93)	1.22	95.09 (1.88, 3.03)	1.22	95.76 (2.76, 1.48)	1.19
	1.5	94.76 (3.04, 2.20)	1.63	94.36 (2.40, 3.24)	1.64	94.77 (2.26, 2.97)	1.65	95.32 (3.28, 1.40)	1.59
	2.0	94.94 (2.89, 2.17)	2.04	94.39 (2.39, 3.22)	2.06	95.16 (2.30, 2.54)	2.05	95.77 (3.30, 0.93)	1.96
	2.5	95.24 (2.61, 2.15)	2.43	94.89 (2.14, 2.97)	2.44	94.98 (2.27, 2.75)	2.54	95.40 (3.48, 1.12)	2.39
	3.0	95.41 (2.82, 1.77)	2.86	94.97 (2.37, 2.66)	2.87	95.59 (2.36, 3.05)	2.86	95.06 (3.62, 1.32)	2.71

The equal-tailed results from the Jeffreys' Independence prior should match the generalized confidence interval (GCI) results as found in Zou et al (2009). This is due to the formulation of the model as described in previous sections. From the above simulations it is apparent that the equal-tail intervals above do indeed match the GCI results, for all practical purposes.

In comparison to the MOVER confidence intervals, the equal-tail intervals also seem to compare reasonably well, with insignificant differences in both the proportion of confidence intervals above and below the true parameter, but the width of the intervals are larger than those of the MOVER. Naturally, as the sample size increases the width of the interval tends to decrease. On the other hand, if the variance increases the width of the intervals will also increase.

Thus, the equal-tail intervals do not offer an improvement on the MOVER method. However, when considering the HPD intervals, which are only possible through the Bayesian framework in this setting, a large improvement on the MOVER can be gained, particularly when n is small. Of particular interest to note is that these HPD intervals result in considerable reductions in interval width. Also of note is that the proportion of intervals above the true parameter is considerably less than both the MOVER and the equal-tail intervals.

So, the performance of the Jeffreys' Independence prior is comparable (or improved for HPD intervals) to the MOVER. However, in terms of the literature, Box and Tiao (1973), this would be the natural choice of prior distribution in this setting and thus, its accuracy is an expected result. As mentioned previously though, other prior distributions were also chosen to examine their effectiveness in this situation. These other prior distributions were mentioned in section 3.2.4. The table below represents the results of these distributions, once again compared to the MOVER. However, these were only performed for the extreme values of σ^2 in Table 1. Also, only the coverage and interval widths are presented (the derivation and potential justification of the Reference and Probability-Matching prior distributions are given in later sections of this paper).

Table 2
Comparison of the MOVER and Other Prior Distributions for Constructing Two-sided 95% Confidence Intervals for $\mu + \frac{1}{2}\sigma^2$

			Equal-Tail / MOVER		HPD Intervals	
n	σ^2	Prior / Method	Cover %	Width	Cover %	Width
5	0.5	MOVER	93.47	2.54	N/A	N/A
		$p(\boldsymbol{\theta}) \propto \sigma^{-3}$	91.36	1.82	91.68	1.66
		Reference Prior	94.79	3.36	96.76	2.79
		Probability-Matching Prior	92.27	2.09	93.30	1.89
5	3	MOVER	95.35	12.72	N/A	N/A
		$p(\boldsymbol{\theta}) \propto \sigma^{-3}$	91.54	7.68	91.23	6.27
		Reference Prior	93.57	11.89	96.41	10.09
		Probability-Matching Prior	92.76	8.68	93.35	7.37
20	0.5	MOVER	94.24	0.74	N/A	N/A
		$p(\boldsymbol{\theta}) \propto \sigma^{-3}$	94.39	0.74	94.14	0.72
		Reference Prior	94.88	0.77	95.26	0.76
		Probability-Matching Prior	94.74	0.74	94.81	0.73
20	3	MOVER	95.41	2.86	N/A	N/A
		$p(\boldsymbol{\theta}) \propto \sigma^{-3}$	94.59	2.66	94.21	2.54
		Reference Prior	94.71	2.99	95.66	2.84
		Probability-Matching Prior	94.78	2.78	94.95	2.65

It appears as though the coverage of the other prior is not as good as the Independence Jeffreys' prior, particularly for small sample sizes. However, as the sample size increases the effect of the prior distribution seems to decrease and the results are comparable.

From Table 2 it is also clear that the reference prior is comparable to the MOVER and seems to have better coverage than both the Jeffreys' rule ($p(\boldsymbol{\theta}) \propto \sigma^{-3}$) and probability-matching priors.

<u>Example 1:</u> Two-sided 95% Bayesian Confidence Intervals for Lognormal Means and the Difference between Two Lognormal Means

On page 3761 of Zou et al (2009) the following example was given (as referenced in Zhou et al (1997)): the effect of race on the cost of medical care for type I diabetes was investigated using MOVER. Log transformed cost data for 119 black patients and 106 white patients. For the black patients the following log-transformed data was available: $\hat{\mu}_1 = \bar{y}_1 = 9.06694$ and $\hat{\sigma}_1^2 = s_1^2 = 1.82426$. For the white patients this was $\hat{\mu}_2 = \bar{y}_2 = 8.69306$ and $\hat{\sigma}_2^2 = s_2^2 = 2.69186$. For the MOVER the following results were obtained and this is compared to results obtained from the Bayes methods:

	MOVER	Jeffreys-Rule	Independence Jeffreys	Reference Prior	Probability Matching Prior
<u>Black Patients</u>					
Lower Limit (equal-tail)	15806.00	15970.19	15882.58	16021.01	15906.92
Upper Limit (equal-tail)	31388.77	31759.23	31752.94	32025.42	31557.68
Lower Limit (HPD)		15708.48	15626.14	15753.04	15671.29
Upper Limit (HPD)		31063.21	31121.86	31351.21	30982.03
<u>White Patients</u>					
Lower Limit (equal-tail)	14842.03	15097.15	15039.52	15081.46	14962.31
Upper Limit (equal-tail)	39722.09	40079.56	40156.06	40483.85	40019.03
Lower Limit (HPD)		14744.94	14668.91	14765.88	14461.04
Upper Limit (HPD)		38799.26	38847.41	39404.13	38289.58
<u>Difference</u>					
Lower Limit (equal-tail)	-19112.18	-19156.98	-19241.18	-19812.57	-19169.17
Upper Limit (equal-tail)	11371.14	11229.07	11381.24	11457.97	11361.90
Lower Limit (HPD)		-19161.99	-19254.62	-19820.21	-19170.15
Upper Limit (HPD)		11225.65	11367.43	11446.41	11360.83

Since the sample sizes are large the interval lengths for the different procedures are more or less the same. It is however, clear that the HPD intervals for the Independence Jeffreys' prior are somewhat shorter for black and white patients than those of the MOVER.

4 The Case of Zero-Valued Observations

4.1 Model Formulation

A further problem suggested by Zou et al (2009) and other authors (e.g. Zhou and Tu (2000)) is that of the inclusion of zero-valued observations. This setting we assume that the non-zero-valued observations are lognormally distributed, but zero valued observations are also present in the data.

We assume that the probability of obtaining a zero observation from the population is δ where $0 < \delta < 1$. This can be specified in the likelihood function by means of a binomial distribution. Furthermore, we assume that the non-zero observations are distributed lognormally in the same way as described in section 3.1. From this preliminary setting specification we wish to construct credibility intervals for $\widetilde{M}$, where:

$$\widetilde{M} = (1 - \delta)exp\left(\mu + \frac{1}{2}\sigma^2\right).$$

The only adjustment to the setting mentioned in section 3.1 is that $\hat{\delta} = \frac{n_0}{n}$, where n_0 is the number of zero-valued observations, n_1 is the number of non-zero observations and $n = n_0 + n_1$.

4.2 Intervals Based on a Bayesian Procedure

The unknown parameters can now be defined by: $\widetilde{\boldsymbol{\theta}}' = [\delta \quad \mu \quad \sigma^2]$ and then the likelihood function is given by:

$$L(\widetilde{\boldsymbol{\theta}}|data) \propto \delta^{n_0}(1 - \delta)^{n_1} \prod_{i=1}^{n_1} \left(\frac{1}{\sigma^2}\right)^{1/2} exp\left[-\frac{(y_i - \mu)^2}{2\sigma^2}\right] \tag{10}$$

From equation (10) the Fisher Information matrix is obtained as:

$$I(\widetilde{\boldsymbol{\theta}}) = diag\left[\frac{n}{\delta\ (1-\delta\ )} \quad \frac{n\ (1-\delta\ )}{\sigma^2} \quad \frac{n\ (1-\delta\ )}{2\sigma^4}\right] \tag{11}$$

4.2.1 Prior Distributions:

The following prior distributions were then chosen:

Independence Jeffreys' Prior: $p(\widetilde{\boldsymbol{\theta}}) \propto \sigma^{-2}\delta^{-1/2}(1 - \delta)^{-1/2}$ (12)

Jeffreys' Rule Prior: $p(\widetilde{\boldsymbol{\theta}}) \propto \sigma^{-3}\delta^{-1/2}(1 - \delta)^{1/2}$ (13)

which is obtained from $|I(\widetilde{\boldsymbol{\theta}})|^{1/2}$. The following: $p(\delta) \propto \delta^{-1/2}(1 - \delta)^{-1/2}$ is the prior proposed by Jeffreys' (1967) for the binomial parameter.

In addition to these prior distributions, the Reference and Probability-Matching prior will also be evaluated.

It is easy to show that the posterior distribution of δ in the case of Independence Jeffreys' prior is a Beta distribution, specifically $B\left(n_0 + \frac{1}{2};\ n_1 + \frac{1}{2}\right)$, while the posterior distribution of δ for the Jeffreys' Rule prior is $B\left(n_0 + \frac{1}{2};\ n_1 + \frac{3}{2}\right)$. The posteriors $p(\delta|data)$ and $p(\mu\,, \sigma^2|data)$ will be independently distributed.

4.2.2 Simulation Procedure:

The simulation procedure is similar to that previously mentioned except for the simulation from the Beta distribution to obtain δ.

4.3 Discussion of Results from Simulation Study

Zou et al(2009) used the MOVER for what they termed the one sample Δ-distribution, which is the same situation as the previously described setting. The following results were obtained:

Table 3

Comparison of the MOVER and GCI against Independence Jeffreys' Prior for Zero Values Included for Constructing Two-sided 95% Confidence Intervals for $ln(1 - \delta) + \left(\mu + \frac{1}{2}\sigma^2\right)$

			MOVER		GCI		Equal-Tail		HPD Intervals	
δ	n	σ^2	Cover %	Width	Cover %	Width	Cover %	Width	Cover %	Width
0.1	15	1	95.03 (3.60, 1.37)	1.65	95.53 (2.34, 2.13)	1.72	95.03 (2.59, 2.38)	1.66	95.36 (1.26, 3.38)	1.59
		2	95.22 (3.13, 1.65)	2.78	95.50 (2.26, 2.24)	2.85	95.14 (2.62, 2.24)	2.79	95.56 (0.91, 3.53)	2.62
		3	94.87 (2.90, 2.23)	3.88	94.94 (2.35, 2.71)	3.94	94.86 (2.73, 2.41)	3.89	95.42 (0.80, 3.78)	3.59
	25	1	95.21 (3.09, 1.70)	1.17	95.95 (2.03, 2.02)	1.22	95.75 (2.30, 2.45)	1.18	95.34 (1.34, 3.32)	1.15
		2	94.94 (2.87, 2.19)	1.93	95.21 (2.19, 2.60)	1.97	95.00 (2.52, 2.48)	1.93	95.43 (1.22, 3.35)	1.86
		3	95.09 (2.80, 2.11)	2.67	95.31 (2.26, 2.43)	2.71	94.85 (2.60, 2.55)	2.66	95.22 (1.04, 3.74)	2.54
	50	1	95.10 (3.02, 1.88)	0.78	95.79 (2.26, 1.95)	0.80	94.93 (2.47, 2.60)	0.78	94.98 (1.78, 3.24)	0.77
		2	95.16 (2.87, 1.97)	1.26	95.41 (2.37, 2.22)	1.29	95.28 (2.47, 2.25)	1.27	95.56 (1.47, 2.97)	1.24
		3	94.86 (2.69, 2.45)	1.73	94.87 (2.43, 2.70)	1.76	94.93 (2.68, 2.39)	1.73	95.13 (1.53, 3.34)	1.69
0.2	15	1	95.17 (3.30, 1.53)	1.87	95.99 (2.14, 1.87)	1.98	95.48 (2.44, 2.08)	1.89	96.18 (1.08, 2.74)	1.80
		2	95.41 (2.78, 1.81)	3.13	95.70 (2.02, 2.28)	3.23	94.98 (2.76, 2.26)	3.13	95.78 (0.88, 3.34)	2.91
		3	94.97 (3.11, 1.92)	4.38	94.93 (2.54, 2.53)	4.47	94.60 (2.83, 2.57)	4.37	95.10 (0.81, 4.09)	3.99
	25	1	95.17 (3.20, 1.63)	1.30	96.01 (2.14, 1.85)	1.36	95.32 (2.49, 2.19)	1.30	95.81 (1.45, 2.74)	1.27
		2	95.34 (2.79, 1.87)	2.10	95.70 (2.11, 2.19)	2.16	95.34 (2.47, 2.19)	2.11	95.83 (1.08, 3.09)	2.03
		3	95.27 (2.74, 1.99)	2.91	95.57 (2.17, 2.26)	2.96	94.92 (2.48, 2.60)	2.90	95.26 (0.89, 3.85)	2.76
	50	1	95.00 (2.99, 2.01)	0.85	95.56 (2.26, 2.18)	0.88	94.90 (2.62, 2.48)	0.86	95.20 (1.84, 2.96)	0.85
		2	94.95 (2.92, 2.13)	1.37	95.28 (2.36, 2.36)	1.39	95.09 (2.74, 2.17)	1.37	95.20 (1.72, 3.08)	1.34
		3	95.30 (2.58, 2.12)	1.87	95.39 (2.24, 2.37)	1.90	95.13 (2.77, 2.10)	1.87	95.46 (1.45, 3.09)	1.83

From the above it is evident that the interval lengths and coverage of the equal-tail intervals are very similar to the MOVER, with the lengths being almost identical. It is interesting to note however, that proportion of intervals above and below the true value differ substantially. The Bayesian HPD intervals are therefore a large improvement on the MOVER and generalized confidence intervals.

In Table 1 the equal tail Bayesian intervals using the Jeffreys' prior $p(\mu, \sigma^2) \propto \sigma^{-2}$ are identical to the generalized confidence intervals. This will not be the case in Table 3. The reason for this is the simulation of δ. In the Bayesian case (using the Independence

Jeffreys prior) the posterior distribution of δ is the Beta, $B\left(n_0 + \frac{1}{2}, n_1 + \frac{1}{2}\right)$, distribution while Zou et al (2009) (see also Tian (2005)) used two pivotal quantities, $B(n_0 + 1, n_1)$ and $B(n_0, n_1 + 1)$ for δ, which are combined with the pivotal quantity of the lognormal mean to simulate $ln\widetilde{M} = ln(1 - \delta) + \left(\mu + \frac{1}{2}\sigma^2\right)$. From Table 3 it is also clear that the Bayesian equal-tail intervals are shorter than those of the generalized confidence procedure with just as good or better coverage probabilities.

Tian (2006) also considered an approach based on the adjusted log-likelihood ratio statistics for constructing a confidence interval for the mean of lognormally distributed data with excess zeros. Because of different parameter values only a few results could be compared. It does seem, however, that the procedures described in Table 3 result in better results than the adjusted log likelihood method.

Once again, other prior distributions were also evaluated, but only for the case of $\delta = 0.1$. In addition, the proportion of intervals above and below the true parameter values were also not recorded.

Table 4
Comparison of the MOVER and Other Prior Distributions for Constructing Two-sided 95% Confidence Intervals for $ln(1-\delta)+\left(\mu+\frac{1}{2}\sigma^2\right)$

n	σ^2	Prior / Method	Equal-Tail / MOVER		HPD Intervals	
			Cover %	Width	Cover %	Width
15	1	MOVER	95.03	1.65		
		$p(\tilde{\boldsymbol{\theta}}) \propto \sigma^{-3}\delta^{-1/2}(1-\delta)^{1/2}$	94.35	1.53	94.10	1.48
		Reference Prior	95.45	1.75	96.35	1.67
		Probability-Matching Prior	96.25	1.67	94.41	1.59
	2	MOVER	95.22	2.78		
		$p(\tilde{\boldsymbol{\theta}}) \propto \sigma^{-3}\delta^{-1/2}(1-\delta)^{1/2}$	94.06	2.52	93.93	2.39
		Reference Prior	95.01	3.02	96.10	2.81
		Probability-Matching Prior	94.76	2.68	94.79	1.59
	3	MOVER	94.87	3.88		
		$p(\tilde{\boldsymbol{\theta}}) \propto \sigma^{-3}\delta^{-1/2}(1-\delta)^{1/2}$	94.47	3.45	94.02	3.23
		Reference Prior	95.26	4.27	96.48	3.90
		Probability-Matching Prior	94.22	3.76	94.45	3.49
25	1	MOVER	95.21	1.17		
		$p(\tilde{\boldsymbol{\theta}}) \propto \sigma^{-3}\delta^{-1/2}(1-\delta)^{1/2}$	94.12	1.12	94.18	1.10
		Reference Prior	95.03	1.19	95.46	1.17
		Probability-Matching Prior	94.82	1.15	94.81	1.12
	2	MOVER	94.94	1.93		
		$p(\tilde{\boldsymbol{\theta}}) \propto \sigma^{-3}\delta^{-1/2}(1-\delta)^{1/2}$	94.65	1.82	94.47	1.76
		Reference Prior	94.96	1.99	95.89	1.92
		Probability-Matching Prior	95.14	1.88	95.20	1.82
	3	MOVER	95.09	2.67		
		$p(\tilde{\boldsymbol{\theta}}) \propto \sigma^{-3}\delta^{-1/2}(1-\delta)^{1/2}$	94.82	2.51	94.33	2.41
		Reference Prior	95.01	2.78	95.82	2.64
		Probability-Matching Prior	94.44	2.60	94.35	2.49
50	1	MOVER	95.10	0.78		
		$p(\tilde{\boldsymbol{\theta}}) \propto \sigma^{-3}\delta^{-1/2}(1-\delta)^{1/2}$	94.79	0.76	94.72	0.76
		Reference Prior	94.93	0.79	94.95	0.79
		Probability-Matching Prior	94.80	0.77	94.85	0.76
	2	MOVER	95.16	1.26		
		$p(\tilde{\boldsymbol{\theta}}) \propto \sigma^{-3}\delta^{-1/2}(1-\delta)^{1/2}$	94.37	1.23	94.25	1.21
		Reference Prior	95.34	1.28	95.78	1.26
		Probability-Matching Prior	94.79	1.25	95.00	1.23
	3	MOVER	94.86	1.73		
		$p(\tilde{\boldsymbol{\theta}}) \propto \sigma^{-3}\delta^{-1/2}(1-\delta)^{1/2}$	95.03	1.68	94.51	1.65
		Reference Prior	95.11	1.77	95.53	1.72
		Probability-Matching Prior	94.65	1.72	94.66	1.68

In this instance it appears as though the probability matching prior, $p(\tilde{\boldsymbol{\theta}}) \propto \delta^{-1/2}(1-\delta)^{-1/2}\sigma^{-2}\left(1+\frac{2}{\sigma^2}\right)^{1/2}$, and the Jeffreys' Rule prior ($p(\tilde{\boldsymbol{\theta}}) \propto \sigma^{-3}\delta^{-1/2}(1-\delta)^{1/2}$) tend to give coverage probabilities less than 0.95. The reference prior, $p(\tilde{\boldsymbol{\theta}}) \propto \delta^{-1/2}(1-\delta)^{-1/2}\sigma^{-1}\left(1+\frac{2}{\sigma^2}\right)^{1/2}$ on the other hand gives the correct coverage probabilities, but the intervals are wider than that of the MOVER.

Example 2: *Two-sided 95% Confidence Intervals for* $(1-\ \delta)exp\left(\mu+\frac{1}{2}\sigma^2\right)$

On page 3761 of Zou et al (2009) the following example was given (as referenced in Zhou and Tu (2000)): diagnostic test charges on 40 patients were investigated. Among them, 10 patients had no diagnostic test charges and the charges for the remaining patients were approximated using a lognormal distribution. On the log scale the following values were observed: $\bar{y}=6.8535$ and $s^2=1.8696$.

For the MOVER the following interval was obtained and this is compared to results obtained from the Bayes and GCI methods:

	MOVER	GCI	Jeffreys-Rule	Independence Jeffreys	Reference Prior	Probability Matching Prior
Lower Limit (equal-tail)	955.50	970.81	1002.18	975.03	996.10	982.41
Upper Limit (equal-tail)	4491.55	4687.37	4690.34	4310.66	4802.44	4519.89
Lower Limit (HPD)			958.05	906.87	931.60	926.10
Upper Limit (HPD)			4345.73	3932.04	4356.93	4111.21

From the table it is clear that the intervals do not differ much. The intervals for the MOVER and the probability-matching prior are for all practical purposes the same. The shortest intervals are obtained from the Independence Jeffreys' and Probability-Matching prior. As mentioned before, the equal-tail Independence Jeffreys interval and the GCI interval will not be the same because of the difference in the simulation of δ.

5 Ratio between Two Samples

The last situation that will be investigated in this paper is similar to the rest of section 4 only with the added complexity of analyzing the results when not only a single sample is taken, but when the ratio of two populations containing zero values is considered.

This situation was not specifically addressed in Zou et al (2009), but it has been presented, particularly with regards to diagnostic charges in Zhou and Tu (2000). Once again, the method will be compared to the results obtained by the MOVER methodology, maximum likelihood and bootstrap estimates (as evaluated by Zhou and Tu (2000)) and within the Bayesian framework itself a variety of different priors will be tested to evaluate performance.

The formulation of the model is very similar to the framework presented in sections 4.1 and 4.2. This can basically be described as follows:

From the specification of the problem in the Introduction we can assume that the populations of interest contain both zero and non-zero (positive observations) and we furthermore assume that the probability of obtaining a zero observation from the *j*-th population (*j* = 1,2) is δ_j where $0 \leq \delta_j \leq 1$. Furthermore, we assume that the non-zero observations are distributed lognormally with parameters μ_j and σ_j^2. Now, let $X_{1j}, X_{2j}, ..., X_{nj}$ be a random sample from the j^{th} population and let $M_j = E(X_{ij})$. From this preliminary setting specification we wish to construct credibility intervals for the ratio of the means, M_1 and M_2, of the two populations. As in Zhou and Tu (2000) we assume that in the j^{th} sample the non-zero observations come first: $X_{ij} > 0$, and $\ln(X_{ij}) \mid n_{j1} \sim N(\mu_j, \sigma_j)$, for $i = 1, ..., n_{j1}$. In addition, $X_{ij} = 0$, for $i = n_{j1} + 1, ..., n_j$ and $n_{j0} = n_j - n_{j1} \sim Bin(n_j, \delta_j)$. From this it follows that the mean of the j^{th} population, which is a function of μ_j, σ_j^2 and δ_j, is given by:

$$M_j = (1-\delta_j)\exp(\mu_j + \frac{1}{2}\sigma_j^2).$$

To compare the two population means we will construct credibility intervals (Bayesian confidence intervals) for the ratio of the means:

$$\frac{M_1}{M_2} = \frac{(1-\delta_1)\exp(\mu_1 + \frac{1}{2}\sigma_1^2)}{(1-\delta_2)\exp(\mu_2 + \frac{1}{2}\sigma_2^2)}.$$

Denote $y_{ij} = \ln X_{ij}$ and $\boldsymbol{\theta} = [\delta_1 \quad \mu_1 \quad \sigma_1^2\ \delta_2 \quad \mu_2 \quad \sigma_2^2]'$ then the likelihood function is given by:

$$L(\boldsymbol{\theta} \mid data) \propto \prod_{j=1}^{2} \{\delta_j^{n_{j0}} (1-\delta_j)^{n_{j1}} \prod_{i=1}^{n_{j1}} (\frac{1}{\sigma_j^2})^{\frac{1}{2}} \exp[-\frac{(y_{ij}-\mu_j)^2}{2\sigma_j^2}]\} \tag{14}$$

Therefore,

$$I(\boldsymbol{\theta}) = diag\left[\frac{n_1}{\delta_1(1-\delta_1)} \quad \frac{n_1(1-\delta_1)}{\sigma_1^2} \quad \frac{n_1(1-\delta_1)}{2\sigma_1^4} \quad \frac{n_2}{\delta_2(1-\delta_2)} \quad \frac{n_2(1-\delta_2)}{\sigma_2^2} \quad \frac{n_2(1-\delta_2)}{2\sigma_2^4}\right] \tag{15}$$

The Independence Jeffreys' prior is then given by:

$$p(\boldsymbol{\theta}) \propto \prod_{j=1}^{2} \sigma_j^{-2} \delta_j^{-\frac{1}{2}} (1-\delta_j)^{-\frac{1}{2}} \tag{16}$$

The Jeffreys' Rule prior then becomes

$$p(\boldsymbol{\theta}) \propto \prod_{j=1}^{2} \sigma_j^{-3} \delta_j^{-\frac{1}{2}} (1-\delta_j)^{\frac{1}{2}} \tag{17}$$

As was done in Zhou and Tu (2000) we will use computer simulations to study the operating characteristics of the proposed Bayesian confidence interval procedure in finite sample sizes. Random sample sizes containing both zero and lognormal observations are generated using the following different sample sizes:

Table 5
Sample Sizes Analysed by Monte Carlo Simulation Techniques

n_1	n_2
10	10
25	25
50	50
100	100
10	25
25	10
25	50

Zero proportions with different skewness coefficients are also considered. Based on these generated samples the credibility intervals are constructed.

In the following table the parameter settings used in the simulation study are presented (the skewness coefficients for samples 1 and 2 are reported under headings γ_1 and γ_2):

Table 6
Parameter Settings used in the Simulation Study

Design	σ_1^2	σ_2^2	δ_1	δ_2	γ_1	γ_2
1	3.0	1.0	0.0	0.0	96.4851	6.1849
2	4.0	4.0	0.0	0.0	414.3593	414.3593
3	3.0	1.0	0.1	0.1	100.9809	6.1763
4	2.0	0.5	0.0	0.1	23.7323	2.6848
5	2.0	0.5	0.1	0.2	24.5572	2.5806

The results from the simulation study performed by Zhou and Tu (2000) for the Maximum Likelihood and Bootstrap methods have been supplied as well for the purposes of comparison. In addition, the MOVER and Bayesian frameworks for the two different Jeffreys' prior distributions are compared (due to space only average results for the different designs are given). The coverage probabilities and interval lengths are for $lnM_1 - lnM_2 = ln(1-\delta_1) + \left(\mu_1 + \frac{1}{2}\sigma_1^2\right) - ln(1-\delta_2) - \left(\mu_2 + \frac{1}{2}\sigma_2^2\right)$.

Table 7
Comparison of Results for the Ratio of Two Populations – Summary Results for Simulation Studies – 95% Two-sided Confidence Intervals for $lnM_1 - lnM_2$

		Equal – Tail*		HPD Intervals	
Design	Method	Coverage	Width	Coverage	Width
1	ML	92.85 (6.36, 0.80)	2.36		
	Bootstrap	92.66 (3.93, 3.41)	2.63		
	Independence Jeffreys'	95.20 (2.42, 2.38)	3.22	96.01 (1.18, 2.81)	3.08
	Jeffreys' Rule	94.82 (2.11, 3.06)	2.89	95.19 (1.12, 3.68)	2.78
	MOVER	95.68 (1.98, 2.34)	3.23		
2	ML	94.76 (2.78, 2.46)	3.85		
	Bootstrap	93.69 (3.24, 3.07)	3.95		
	Independence Jeffreys'	95.23 (2.38, 2.39)	5.52	96.52 (1.59, 1.89)	5.34
	Jeffreys' Rule	94.73 (2.52, 2.74)	4.90	95.95 (1.88, 2.17)	4.78
	MOVER	95.71 (2.05, 2.24)	5.61		
3	ML	92.37 (6.75, 0.88)	2.51		
	Bootstrap	92.98 (3.93, 3.08)	2.70		
	Independence Jeffreys'	95.28 (2.32, 2.40)	3.59	96.02 (1.22, 2.76)	3.40
	Jeffreys' Rule	94.68 (2.24, 3.08)	3.15	95.15 (1.20, 3.66)	3.02
	MOVER	95.62 (2.11, 2.27)	3.62		
4	ML	92.74 (6.24, 1.02)	1.73		
	Bootstrap	92.94 (3.80, 3.26)	1.86		
	Independence Jeffreys'	95.46 (2.35, 2.18)	2.29	95.97 (1.35, 2.69)	2.20
	Jeffreys' Rule	94.76 (2.28, 2.96)	2.07	95.00 (1.33, 3.66)	2.01
	MOVER	95.52 (2.13, 2.35)	2.30		
5	ML	92.74 (6.19, 1.08)	1.87		
	Bootstrap	93.46 (3.49, 3.05)	1.98		
	Independence Jeffreys'	95.41 (2.35, 2.24)	2.58	95.95 (1.33, 2.72)	2.46
	Jeffreys' Rule	94.98 (2.12, 2.90)	2.28	95.28 (1.23, 3.49)	2.20
	MOVER	95.68 (2.06, 2.26)	2.57		
Overall	ML	93.09 (5.66, 1.25)	2.46		
	Bootstrap	93.15 (3.68, 3.17)	2.62		
	Independence Jeffreys'	95.32 (2.36, 2.32)	3.44	96.09 (1.33, 2.57)	3.30
	Jeffreys' Rule	94.79 (2.25, 2.95)	3.06	95.31 (1.35, 3,33)	2.96
	MOVER	95.64 (2.07, 2.29)	3.46		

*Refers to Equal-tail Bayesian intervals, Maximum Likelihood Methods, Bootstrap and MOVER estimates.

In this case it is once again evident that the Bayesian methods are substantially better than both the ML and bootstrap methods. Both the ML and Bootstrap methods results in substantial undercoverage. The MOVER is a vast improvement on these methods and results in adequate coverage. However, the Bayesian methods result in adequate coverage and particularly the HPD intervals result in more efficient (narrower) intervals than even the MOVER. According to Table 7, the prior that gives the shortest HPD intervals with correct coverage is the Jeffreys' Rule prior. The Independence Jeffreys' prior on the other hand gives the best coverage for equal-tail intervals.

5.1 Example: Rainfall Data

For the purposes of comparison of the different methods, an example was chosen using raw data obtained from the South African Weather Service. The data consisted of the monthly rainfall totals for the cities of Bloemfontein and Kimberley, two South African cities, over a period of 69 to 70 years of measurement. However, these two cities are both located in relatively arid regions and are characterised by mainly summer rainfall. For that reason, the winter months of June do contain some rainfall data, but also contain many years where the total monthly rainfall data was zero. Probability plots as well as the Shapiro-Wilk (1965) test indicate that the lognormal distribution is a better fit than the normal distribution.

The aim is to produce two-sided 95% confidence intervals for the ratio of the two means, $\frac{M_1}{M_2} = \frac{(1-\delta_1)exp\left(\mu_1+\frac{1}{2}\sigma_1^2\right)}{(1-\delta_2)exp\left(\mu_2+\frac{1}{2}\sigma_2^2\right)}$. If show that the confidence interval includes 1 then the conclusion is that there is no difference between the mean rainfall.

The data can be summarised as follows:

Table 8
Summary of the Rainfall Data

City	Parameter	Value
Bloemfontein	Number of Years of Available Data	70
	Number of Zero Valued Observations	18
	Mean of Log-Transformed Data	1.9578
	Variance of Log-Transformed Data	2.1265
Kimberley	Number of Years of Available Data	69
	Number of Zero Valued Observations	10
	Mean of Log-Transformed Data	1.0526
	Variance of Log-Transformed Data	3.1589

In order to compare the results, both the MOVER was applied to the data as well as the Bayesian methods described in the preceding text for the following priors: Independence Jeffery's Prior (Prior 1 in the table), the Jeffery's Rule Prior (Prior 2), the Reference Prior and the Probability Matching Prior. The maximum likelihood and bootstrap procedures derived by Zhou and Tu (2000) are also given. The results are presented in the following table:

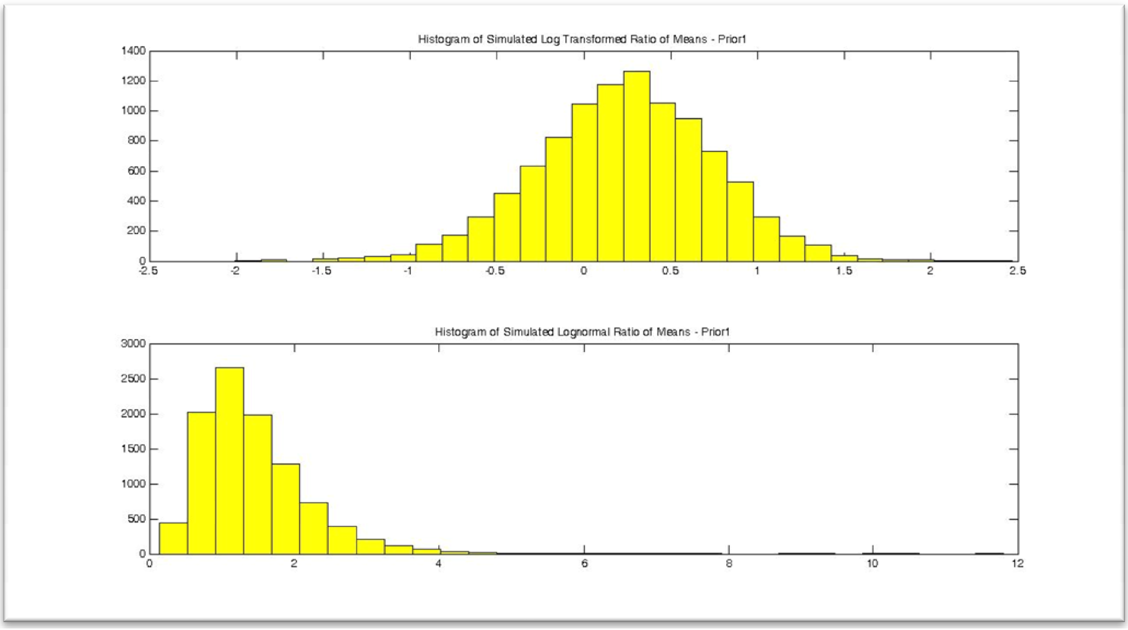
Table 9
Summary of Results for the Rainfall Data – Two-sided 95% Confidence Intervals for $\frac{M_1}{M_2} = \frac{(1-\delta_1)exp\left(\mu_1+\frac{1}{2}\sigma_1^2\right)}{(1-\delta_2)exp\left(\mu_2+\frac{1}{2}\sigma_2^2\right)}$

	Maximum Likelihood	Bootstrap	MOVER	Prior 1	Prior 2	Reference Prior	Probability Matching Prior
Lower Limit	0.5009	0.5271	0.4486	0.4485	0.4520	0.4410	0.4587
Upper Limit	3.2805	3.4710	3.3428	3.3245	3.3048	3.3548	3.3141

From Table 9 it is clear that the Bayesian and MOVER intervals are for practical purposes the same. The maximum likelihood and bootstrap intervals on the other hand differ somewhat from these intervals. It can furthermore be seen that 1 is included in all the 95% confidence intervals and thus there is no difference between the mean rainfall for the two cities.

The following graphs also illustrate these results:

<u>Prior 1</u>



6 Probability-matching and Reference Priors for $M = e^{\mu+\frac{1}{2}\sigma^2}$, the Mean of the Lognormal Distribution

Probability-matching and reference priors often lead to procedures with good frequency properties while retaining the Bayesian flavor. The fact that the resulting posterior intervals of level $1-\alpha$ are also good frequentist intervals at the same level is a very desirable situation.

6.1 The Probability-matching prior for $M = e^{\mu+\frac{1}{2}\sigma^2}$

Datta and Ghosh (1995) derived the differential equation that a prior must satisfy if the posterior probability of a one-sided credibility interval (Bayesian confidence interval) for a parametric function and its frequentist probability agree up to $0\left(n^{-1}\right)$, where n is the sample size. They proved that the agreement between the posterior probability and the frequentist probability holds if and only if the differential equation

$$\sum_{\alpha=1}^{m}\frac{\partial}{\partial\theta_\alpha}\{\eta_\alpha(\boldsymbol{\theta})p(\boldsymbol{\theta})\}=0$$

is satisfied, where $p(\boldsymbol{\theta})$ is the probability-matching prior distribution for $\boldsymbol{\theta}$, the vector of unknown parameters.

Also,

$$\nabla_t=\left[\frac{\partial}{\partial\theta_1}t(\boldsymbol{\theta}),\ldots,\frac{\partial}{\partial\theta_m}t(\boldsymbol{\theta})\right]'$$

and

$$\eta(\boldsymbol{\theta})=\frac{F^{-1}(\boldsymbol{\theta})\nabla_t(\boldsymbol{\theta})}{\sqrt{\nabla_t'(\boldsymbol{\theta})F^{-1}(\boldsymbol{\theta})\nabla_t(\boldsymbol{\theta})}}=\left[\eta_1(\boldsymbol{\theta}),\ldots,\eta_m(\boldsymbol{\theta})\right]'.$$

It is clear that $\eta'(\boldsymbol{\theta})F(\boldsymbol{\theta})\eta(\boldsymbol{\theta})=1$ for all $\boldsymbol{\theta}$ where $F^{-1}(\boldsymbol{\theta})$ is the inverse of $F(\boldsymbol{\theta})$. $F(\boldsymbol{\theta})$ is the Fisher information matrix of $\boldsymbol{\theta}$ and $t(\boldsymbol{\theta})$ is the parameter of interest.

The results of Datta and Ghosh (1995) will be briefly reviewed in the following theorem:

6.1.1 Theorem 1

For the mean, $M = e^{\mu+\frac{1}{2}\sigma^2}$, of the lognormal distribution, the probability matching prior is given by:

$$p_P\left(\mu,\sigma^2\right)\propto\frac{1}{\sigma^2}\sqrt{1+\frac{2}{\sigma^2}}. \tag{18}$$

<u>Proof:</u> The proof is given in the appendix.
Multiplying (18) by the likelihood function in equation (1) it follows that:

$$p_P(\sigma_j^2|data) \propto (\sigma_j^2)^{-\frac{1}{2}(\nu_{j1}+2)}\left(1+\frac{2}{\sigma_j^2}\right)^{\frac{1}{2}} exp\left[-\frac{\nu_{j1}\hat{\sigma}_j^2}{2\sigma_j^2}\right]$$

for $j = 1,2$

(19)

Simulation from (19) can be obtained using the rejection method. Simulation of μ_j and δ_j are as before.

6.2 The Reference prior for $M = e^{\mu+\frac{1}{2}\sigma^2}$

The determination of reasonable, non-informative priors in multiparameter problems is not easy; common non-informative priors, such as Jeffreys' prior, can have features that have an unexpectedly dramatic effect on the posterior distribution. In recognition of this problem Berger and Bernardo (1992) proposed the *reference prior* approach to the development of non-informative priors. As in the case of the Jeffreys and probability-matching priors, the reference prior method is derived from the Fisher information matrix. Reference priors depend on the group ordering of the parameters. Berger and Bernardo (1992) suggested that multiple groups, ordered in terms of inferential importance, are allowed, with the reference prior being determined through a succession of analyses for the implied conditional problems. They particularly recommend the reference prior based on having each parameter in its own group, i.e. having each conditional reference prior be only one dimensional.

As mentioned by Pearn and Wu (2005) the reference prior maximises the difference in information (entropy) about the parameter provided by the prior and posterior distributions. In other words, the reference prior is derived in such a way that it provides as little as possible information about the parameter.

The following theorem can now be stated.

6.2.1 Theorem 2

For the mean, $M = e^{\mu+\frac{1}{2}\sigma^2}$, of the lognormal distribution, the reference prior relative to the ordered parameterisation (μ,σ^2) is given by:

$$p_R(\mu,\sigma^2) \propto \frac{1}{\sigma}\sqrt{1+\frac{2}{\sigma^2}}. \qquad (20)$$

Proof: The proof is given in the appendix.
The posterior distribution of σ_j^2 is now

$$p_P(\sigma_j^2|data) \propto (\sigma_j^2)^{-\frac{1}{2}(\nu_{j1}+1)}\left(1+\frac{2}{\sigma_j^2}\right)^{\frac{1}{2}} exp\left[-\frac{\nu_{j1}\hat{\sigma}_j^2}{2\sigma_j^2}\right]$$

for $j = 1,2$

(22)

7 Appendix

7.1 Proof of Theorem 1

The probability matching prior is derived from the inverse of the Fisher information matrix. Now

$$F^{-1}(\boldsymbol{\theta}) = F^{-1}(\mu, \sigma^2) = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix}.$$

and

$$t(\boldsymbol{\theta}) = e^{\mu + \frac{1}{2}\sigma^2} = M$$

from which it follows that

$$\frac{\partial t(\boldsymbol{\theta})}{\partial \mu} = e^{\mu + \frac{1}{2}\sigma^2} \text{ and } \frac{\partial t(\boldsymbol{\theta})}{\partial \sigma^2} = \frac{1}{2} e^{\mu + \frac{1}{2}\sigma^2}.$$

Also

$$\nabla_t'(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial t(\boldsymbol{\theta})}{\partial \mu} & \frac{\partial t(\boldsymbol{\theta})}{\partial \sigma^2} \end{bmatrix} = e^{\mu + \frac{1}{2}\sigma^2} \begin{bmatrix} 1 & \frac{1}{2} \end{bmatrix}$$

and

$$\nabla_t'(\boldsymbol{\theta}) F^{-1}(\boldsymbol{\theta}) = e^{\mu + \frac{1}{2}\sigma^2} \begin{bmatrix} \sigma^2 & \sigma^4 \end{bmatrix}$$

$$\nabla_t'(\boldsymbol{\theta}) F^{-1}(\boldsymbol{\theta}) \nabla_t(\boldsymbol{\theta}) = e^{2\mu + \sigma^2} \left(\sigma^2 + \frac{1}{2}\sigma^4 \right)$$

and

$$\sqrt{\nabla_t'(\boldsymbol{\theta}) F^{-1}(\boldsymbol{\theta}) \nabla_t(\boldsymbol{\theta})} = e^{\mu + \frac{1}{2}\sigma^2} \sqrt{\sigma^2 + \frac{1}{2}\sigma^4}.$$

Therefore

$$\eta'(\boldsymbol{\theta}) = \frac{\nabla_t'(\boldsymbol{\theta}) F^{-1}(\boldsymbol{\theta})}{\sqrt{\nabla_t'(\boldsymbol{\theta}) F^{-1}(\boldsymbol{\theta}) \nabla_t(\boldsymbol{\theta})}} = [\eta_1(\boldsymbol{\theta}) \ \ \eta_2(\boldsymbol{\theta})] = \frac{1}{\sqrt{\sigma^2 + \frac{1}{2}\sigma^4}} [\sigma^2 \quad \sigma^4]$$

For a prior $p_P(\boldsymbol{\theta}) = p_P(\mu, \sigma^2)$ to be a probability matching prior, the differential equation

$$\frac{\partial}{\partial \mu}[\eta_1(\boldsymbol{\theta}) \ p_P(\boldsymbol{\theta})] + \frac{\partial}{\partial \sigma^2}[\eta_2(\boldsymbol{\theta}) \ p_P(\boldsymbol{\theta})] = 0$$

must be satisfied. If we take

$$p_P(\mu, \sigma^2) \propto \frac{1}{\sigma^2} \sqrt{1 + \frac{2}{\sigma^2}}$$

then the differential equation will be satisfied.

7.2 **Proof of Theorem 2**

The Fisher information matrix of $\boldsymbol{\theta} = \left[\mu, \sigma^2\right]$ per unit observation is given by:

$$F(\boldsymbol{\theta}) = F\left(\mu, \sigma^2\right) = \begin{bmatrix} 1/\sigma^2 & 0 \\ 0 & 1/2\sigma^4 \end{bmatrix}.$$

The parameter of interest is the mean of the lognormal distribution

$$t(\boldsymbol{\theta}) = e^{\mu + \frac{1}{2}\sigma^2} = M$$

Define $A = \dfrac{\partial\left(\mu, \sigma^2\right)}{\partial\left(t(\boldsymbol{\theta}), \sigma^2\right)} = \begin{bmatrix} 1/t(\boldsymbol{\theta}) & -1/2 \\ 0 & 1 \end{bmatrix}.$

Hence, the Fisher information matrix under the reparameterisation $\left(t(\boldsymbol{\theta}), \sigma^2\right)$ is given by

$$F\left(t(\boldsymbol{\theta}), \sigma^2\right) = A'F\left(\mu, \sigma^2\right)A = \begin{bmatrix} \dfrac{1}{t^2(\boldsymbol{\theta})\sigma^2} & \dfrac{-1}{2t(\boldsymbol{\theta})\sigma^2} \\ \dfrac{-1}{2t(\boldsymbol{\theta})\sigma^2} & \dfrac{1}{4\sigma^2} + \dfrac{1}{2\sigma^4} \end{bmatrix}.$$

Following the notation of Berger and Bernardo (1992), the functions $h_j, (j = 1, 2)$, which are needed to calculate the reference prior for the group ordering $\left(t(\boldsymbol{\theta}), \sigma^2\right)$, can be obtained from $F\left(t(\boldsymbol{\theta}), \sigma^2\right)$ as follows:

$$h_1^{\frac{1}{2}} = \left| \frac{1}{t^2(\boldsymbol{\theta})\sigma^2} - \left(\frac{-1}{2t(\boldsymbol{\theta})\sigma^2}\right)^2 \left(\frac{1}{4\sigma^2} + \frac{1}{2\sigma^4}\right)^{-1} \right|^{\frac{1}{2}} = \frac{1}{t(\boldsymbol{\theta})}\left(\frac{1}{\sigma^2} - \frac{1}{2+\sigma^2}\right)^{\frac{1}{2}}$$

and

$$h_2^{\frac{1}{2}} = \left[\frac{1}{2\sigma^2}\left(\frac{1}{2} + \frac{1}{\sigma^2}\right)\right]^{\frac{1}{2}}.$$

Therefore, the reference prior relative to the ordered parameterisation $\left(t(\boldsymbol{\theta}), \sigma^2\right)$ is given by

$$p_R\left(t(\boldsymbol{\theta}), \sigma^2\right) \propto \frac{1}{t(\boldsymbol{\theta})} \frac{1}{\sigma} \sqrt{1 + \frac{2}{\sigma^2}}.$$

In the $\left(\mu, \sigma^2\right)$ parameterisation this corresponds to

$$p_R\left(\mu, \sigma^2\right) \propto \frac{1}{t(\boldsymbol{\theta})} \frac{1}{\sigma} \sqrt{1 + \frac{2}{\sigma^2}} \left(t(\boldsymbol{\theta})\right) \propto \frac{1}{\sigma} \sqrt{1 + \frac{2}{\sigma^2}}.$$

This is the same result derived by Roman (2008).

8 BIBLIOGRAPHY

Bebu, I. and Mathew, T. (2008). "Comparing the Means and Variances of a Bivariate Lognormal Distribution," *Statistics in Medicine*, 27, pp 2684 – 2696.

Berger, J.O., Bernardo, J.M. (1992). "On the development of the Reference Prior Method," *Bayesian Statistics 4: Proceedings of the Fourth Valencia Meeting.* Eds. Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. Oxford, U.K., Clarendon Press, pp. 35 – 60.

Box, G.E.P., Tiao, G.C. (1973). "Bayesian Inference in Statistical Analysis," *Reading, MA*, Addison-Wesley.

Chen, Y.H., Zhou, X.H. (2006). "Interval estimates for the ratio and difference of two lognormal means," *Statistics in Medicine*, 25, pp 4099 – 4113.

Datta, G.S., Ghosh, J.K. (1995). "On priors providing Frequentist Validity of Bayesian Inference," *Biometrika*, 82, pp. 37 – 45.

Hannig, J., Lidong, E., Abdel-Karim, A. and Iyer, H. (2006). "Simultaneous Fiducial Generalized Confidence Intervals for Ratios of Means of Lognormal Distributions," *Australian Journal of Statistics*, 35, pp 261-269.

Krishnamoorthy, K., 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, pp. 103 - 121.

Krishnamoorthy, K., Mathew, T., Ramachandran, G. (2006). "Generalized p-values and confidence intervals: A novel approach for analyzing lognormally distributed exposure data," *Journal of Occupational and Environmental Hygiene*, 3 pp 642 – 650.

Jeffreys, H. (1967). "Theory of Probability," *Oxford University Press*, London.

Pearn, W.L., Wu, C.W. (2005). "Process Capability Assessment for Index C_{pk} based on the Bayesian Approach," *Metrika*, 61, p. 211

Rice, A.J. (1995). "Mathematical Statistics and Data Analysis – 2nd ed.," *Duxbury Press*, Belmont, California.

Roman, L.P. (2008). "Reference prior distributions for prediction problems," *Current technical report*.

Shapiro, S.S., Wilk, M.B. (1965). "An Analysis of Variance Test for Normality (Complete Samples)." *Biometrika*, 52: pp. 591 - 611.

Schenker, N., (1985). "Qualms about bootstrap confidence intervals," *Journal of the*

American Statistical Association, 80, pp. 360 – 361.

Starfield, B., Weiner, J., Mumford, L., Steinwachs, D. (1991). "Ambulatory Care Groups: A Categorization of Diagnoses for Research and Management." *Health Serv. Res*. 26: pp. 53 - 74.

Tian, L., Wu, J. (2007a). "Inferences on the common mean of several lognormal populations: The generalized variable approach," *Biometrical Journal*, 49, pp 944 – 951.

Tian, L., Wu, J. (2007b). "Inferences on the mean response in a log-regression model: The generalized variable approach," *Statistics in Medicine*, 25, pp 5180 - 5188.

Tian, L., (2005). "Inferences on the Mean of Zero-inflated Lognormal Data: the Generalized Variable Approach." *Statistics in Medicine*, pp. 3223 - 3232.

Tian, L., Wu, J., (2006). "Confidence Intervals for the Mean of Lognormal Data with Excess Zeros." *Biometrical Journal*, 48 no. 1: pp. 149 - 156.

Weiner, J., Starfield, B., Mumford, L. (1991) "Development and Application of a Population-Oriented Measure of Ambulatory Care Casemix." *Medical Care*. 29: pp. 452 - 472

Zellner, A. (1971). "An Introduction to Bayesian Inference in Econometrics," *Wiley*, New York, NY.

Zhou, X.H., Gao, S. (1997). "Confidence intervals for the lognormal mean," *Statistical Medicine*, 16, 783 – 790.

Zhou, X.H., Gao, S., Hui, S.L. (1997). "Methods for comparing the means of two independent lognormal samples," *Biometrics*, 53, pp. 1129 – 1135.

Zhou, X.H., Melfi, C.A., Hui, S.L. (1997b). "Methods for Comparison of Cost Data." *Annal of Internal Medicine*, 127: pp. 752 - 756.

Zhou, X.H., Tu, W. (1999). "Comparison of Several Independent Population Means when their Samples Contain Lognormal and Possibly Zero Observations" *Biometrics*, 55, pp. 645 - 651.

Zhou, X.H., 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, pp. 201 – 210.

Zou, Y.Z., Taleban, J., Huo, C.Y. (2009). "Confidence interval estimation for lognormal data with application to health economics," *Computational Statistics and Data Analysis*, 53, pp. 3755 – 3764.

Zou, G.Y., Taleban, J., Huo, C.Y. (2009). “Simple confidence intervals for lognormal means and their differences with environmental applications,” *Econometrics*, 20, pp. 172–180.