

# Assessing Non-Inferiority Hypothesis in Two-Arm Trials with Log-Normal Data

Lahiru Wickramasinghe<sup>1</sup> and Saman Muthukumarana<sup>1\*</sup>

<sup>1</sup>Department of Statistics, University of Manitoba, Canada

## Article Information

Received date: Nov 06, 2015

Accepted date: Jan 08, 2016

Published date: Jan 10, 2016

### \*Corresponding author

Saman Muthukumarana, Department of Statistics, University of Manitoba, Canada, Email: Saman.Muthukumarana@umanitoba.ca

**Distributed under** Creative Commons CC-BY 4.0

**Keywords** Generalized p-value; Log-normal; Monte Carlo; Non-inferiority; Simulation

## Abstract

In health related studies, non-inferiority tests are used to demonstrate that a new treatment is not worse than a currently existing treatment by more than a pre-specified margin. In this paper we have proposed a Bayesian approach and compared it with two other methods available in the literature. We discuss three approaches; a Z-score approach, a generalized p-value approach and a Bayesian approach, to test the non-inferiority hypotheses in two-arm trials for ratio of log-normal means. The log-normal distribution is widely used to describe the positive random variables with positive skewness which is appealing for data arising from studies with small sample sizes. We demonstrate the approaches using data arising from an experimental aging study on cognitive penetrability of posture control. We also examine the suitability of three methods under various sample sizes via simulations. The results from the simulation studies indicate that the generalized p-value and the Bayesian approach reach an agreement approximately and the degree of the agreement increases when the sample sizes increase. However, the Z-score approach can produce unsatisfactory results even under large sample sizes.

## Introduction

Recently, in many clinical trials, the notion of equivalence and non-inferiority tests have become effective standard procedure. In classical or traditional hypothesis tests, well established theory and methods can be used to show whether two products or treatments are significantly differ from each other. The null hypothesis will be rejected if the test statistic is sufficiently large compared to the critical value or if the p-value of the test statistic is sufficiently small compared to the level of significance  $\alpha$ . More details and the methods for testing the hypotheses can be found in Lehmann et al. [1]. However, in pharmaceutical industry, when a new product or a treatment is introduced, the main goal is often to demonstrate that the new product or treatment is either equivalent or superior to the current or the old product or the treatment. In equivalence studies, the objective is to demonstrate that the effect of the new product or treatment is equivalent to the effect of the current or the old product or the treatment. Practically it is impossible to show the exact equivalence since an infinite sample size is often required. Then one can use an equivalence margin, denoted by  $\delta$ , which defines the range of values for which the effect of the new product or treatment is to be considered equivalent to the old product or the treatment. If the effects of the two treatments or the products differ by more than the equivalence margin in either direction, then equivalence does not hold. For example, we write the hypotheses for testing that the new drug Y is equivalent to the existing drug X as

$$H_0 : |\mu_x - \mu_y| \geq \delta \text{ vs. } H_1 : |\mu_x - \mu_y| < \delta \text{ where } \delta \rightarrow 0,$$

$\mu_x$  and  $\mu_y$  are the population means response of the existing drug and the new drug respectively.

In non-inferiority studies, the objective is to demonstrate the new treatment or the product is not worse than the existing treatment or the product by more than a pre-specified, small amount ( $\delta$ ). If you consider the above example, the null hypothesis is that the new drug (Y) is inferior to the existing drug (X) by at least a certain pre-specified amount ( $\delta$ ). The alternative is that the new drug (Y) is not inferior to the existing drug (X) by less than that pre-specified amount ( $\delta$ ). More specifically,

$$H_0 : \mu_x - \mu_y \geq \delta \text{ vs. } H_1 : \mu_x - \mu_y < \delta \text{ where } \delta > 0.$$

Here  $\delta$  is called the non-inferiority margin and it is defined to be strictly positive. A detailed discussion on equivalence and non-inferiority tests are given in Snapinn [2].

There is a well established literature for testing non-inferiority in two-arm trials with normal data. An overview of recent methods can be found in Snapinn [2], Walker et al. [3], Munk et al. [4] and Wellek [5]. Koti [6] also discuss the non-inferiority hypothesis testing in comparing the ratio of two normal means. In this paper, we consider the non-inferiority hypothesis testing with log-normal data. The log-normal distribution is appropriate for the random variables that are positive and also positively skewed (right skewed). The log-normal distributions are commonly used to model the data from many scientific disciplines: geology and mining, human medicine, environment,

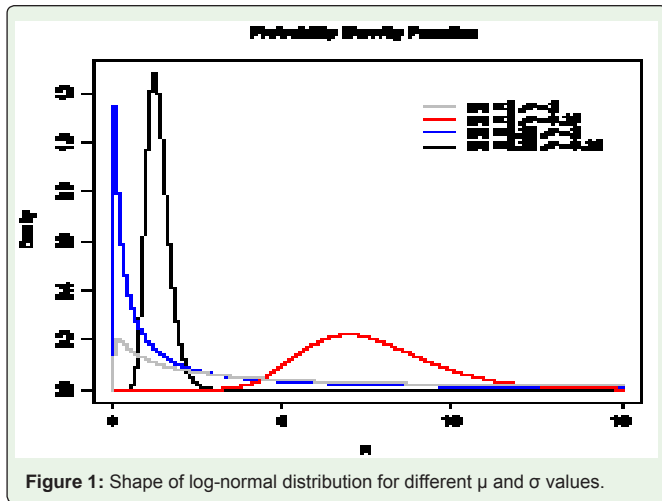


Figure 1: Shape of log-normal distribution for different  $\mu$  and  $\sigma$  values.

microbiology, plant physiology, ecology, food science, economics etc. We refer readers to Eckhard et al. [7] for various applications of log-normal distribution.

In Section 2, we discuss about the basic properties of log-normal distribution, Z-score test approach and a generalized p-value approach for non-inferiority hypothesis testing. We then propose a Bayesian approach for non-inferiority hypothesis testing in Section 3. In Section 4, we apply the methods proposed in this paper on a data set arising from an experimental aging study on cognitive penetrability of posture control. We also assess the suitability of the methods for various sample sizes via simulation studies in Section 5. We conclude with a short discussion in Section 6 based on results and methods discussed in the paper.

### Inference for Log-normal Data

Let  $X (0 < X < \infty)$  be a positive random variable from a log-normal distribution given by

$$f_X(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right]; & x > 0, -\infty < \mu < \infty, \sigma^2 > 0 \\ 0 & \text{Otherwise.} \end{cases}$$

Note that  $\mu$  is the location parameter or the log mean and  $\sigma$  is the scale parameter or the log standard deviation on the log-transformation. Figure 1 shows the density curves for different values of the location parameter ( $\mu$ ) and the shape parameter ( $\sigma$ ). The skewness decreases as the value of  $\mu$  increases and the value of  $\sigma$  decreases.

In a two samples problem, let  $X_1$  and  $X_2$  are two independent log-normal random variables distributed as

$$X_1 \sim \text{log-normal}(\mu_1, \sigma_1^2) \text{ and } X_2 \sim \text{log-normal}(\mu_2, \sigma_2^2)$$

Suppose that  $X_{1i}, i = 1, 2, \dots, n_1$  and  $X_{2j}, j = 1, 2, \dots, n_2$  denote random samples from  $X_1$  and  $X_2$ , respectively. Also, let  $Y_{1i} = \ln(X_{1i}), i = 1, 2, \dots, n_1$  and  $Y_{2j} = \ln(X_{2j}), j = 1, 2, \dots, n_2$ .

Let  $E(X_1) = \exp(\eta_1) = \mu_1 + \frac{\sigma_1^2}{2}$  and  $E(X_2) = \exp(\eta_2) = \mu_2 + \frac{\sigma_2^2}{2}$ . Then the problem of our interest is to compare the ratio of two log-normal means for non-inferiority. More specifically we test

$$H_0 : \frac{E(X_1)}{E(X_2)} \geq \delta \text{ vs. } H_1 : \frac{E(X_1)}{E(X_2)} < \delta \text{ where } \delta > 1$$

$$H_0 : \eta_1 - \eta_2 \geq \delta' \text{ vs. } H_1 : \eta_1 - \eta_2 < \delta' \text{ where } \delta' = \ln(\delta) > 0. \quad (1)$$

Note that the mean of log-normal distribution is a function of both  $\mu$  and  $\sigma^2$ , and it is difficult to obtain the exact or optimum tests for testing hypotheses in 1.

One can use a Z-score approach for the problem of testing two log-normal means as introduced by Zhou et al. [15]. A Z-score test statistic for testing  $H_0$  is given by

$$Z = \frac{\bar{Y}_1 - \bar{Y}_2 + \frac{1}{2}(S_1^2 - S_2^2) - \delta'}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} + \frac{1}{2}\left(\frac{S_1^4}{n_1 - 1} + \frac{S_2^4}{n_2 - 1}\right)}}. \quad (2)$$

When  $n_1$  and  $n_2$  are both large, the sampling distribution of Z is approximately standard normal under  $H_0$ . The Z-score approach is recommended for large samples but for small samples power and type I error are too conservative or too liberal.

On the other hand, one can also take a generalized p-value approach introduced by Krishnamoorthy et al. [8]. The generalized p-value was first introduced by Weerahandi and Tsui [9] as an extended version of the classical p-value. In classical p-value method, there are a few challenges: difficult to find the suitable test statistic, difficult to find the sampling distribution of the test statistic and involves many nuisance parameters. The nuisance parameters are unknown parameters which are required to construct a realistic model but there is no interest in making inferences about them. As a result, the exact solution may not exist, hence the approximate solution with restrictions may suggest. A comprehensive discussion of the generalized p-value is given in the book by Weerahandi [10].

Krishnamoorthy et al. [8] proposed an algorithm for testing hypotheses on log-normal means. We extend the approach and implement the algorithm in R for testing the non-inferiority hypotheses in 1.

Let  $X_{1i}, i = 1, 2, \dots, n_1$ , and  $X_{2j}, j = 1, 2, \dots, n_2$  denote random samples from the log-normal distributions of  $X_1$  and  $X_2$  respectively. Also, let  $Y_{1i} = \ln(X_{1i}), i = 1, 2, \dots, n_1$  and  $Y_{2j} = \ln(X_{2j}), j = 1, 2, \dots, n_2$ . Then define

$$\bar{Y}_i = \frac{1}{n} \sum_{j=1}^{n_i} Y_{ij} \text{ and } S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2, i=1, 2.$$

Let  $\bar{y}_1, \bar{y}_2, s_1^2$  and  $s_2^2$  denote the observed values of  $\bar{Y}_1, \bar{Y}_2, S_1^2$  and  $S_2^2$  respectively.

Let

$$T_{3i} = \bar{y}_i - \frac{\bar{y}_i - \mu_i}{S_i / \sqrt{n_i}} \times \frac{s_i}{\sqrt{n_i}} + \frac{\sigma_i^2}{2S_i^2} s_i^2$$

$$= \bar{y}_i - \frac{Z_i}{U_i / \sqrt{n_i - 1}} \frac{s_i}{\sqrt{n_i}} + \frac{s_i^2}{2U_i^2 / (n_i - 1)} \quad i=1, 2, \quad (3)$$

where  $Z_i = \frac{\bar{y}_i - \mu_i}{\sigma_i / \sqrt{n_i}} \sim N(0, 1)$  and  $U_i^2 = (n_i - 1) \frac{S_i^2}{\sigma_i^2} \sim \chi_{n_i - 1}^2$  for  $i=1, 2$  and these random variables are independent. Define the generalized test variable

$$T_3 = T_{31} - T_{32} - (\eta_1 - \eta_2)$$

and let

$$T_4 = T_{31} - T_{32}$$

So that  $T_3 = T_4 - (\eta_1 - \eta_2)$ . Thus the generalized p-value for testing the hypotheses in 1 is given by

$$P(T_3 \geq 0 | \eta_1 - \eta_2 = \delta') = P(T_4 \geq \delta'). \quad (4)$$

we develop a Bayesian approach non-inferiority hypothesis testing in Section 3.

### Bayesian Inference for Log-Normal Data

The main idea behind Bayesian hypothesis testing is to use the Bayesian approach for determining the strength of evidence. In the frequentist approach for hypothesis testing, inferences are based on the probability of data conditioning on the null hypothesis, whereas the Bayesian approach is based on the probability of hypotheses conditioning on the data. Also, the frequentist approach tests one hypothesis, the null hypothesis ( $H_0$ ) against its alternative hypothesis ( $H_1$ ). But, the Bayesian approach can test multiple hypotheses and takes advantage of prior information.

For the Bayesian hypothesis testing, first we need to find the posterior distribution of the unknown parameter  $\theta$  either analytically or by simulation. Given the posterior distribution and the data ( $\mathbf{x}$ ), we simply calculate  $P(\theta \in \Theta_0 | \mathbf{x})$  and decide in favor of the null hypothesis ( $H_0$ ) if  $P(\theta \in \Theta_0 | \mathbf{x})$  is sufficiently large. In order to make a better comparison of the hypotheses, the prior and posterior information can also be combined in a ratio (called Bayes factor) unless the posterior probability,  $(P(\theta \in \Theta_0 | \mathbf{x}))$  is close to 0 or 1.

Let  $X_1$  and  $X_2$  are two independent log-normal random variables distributed as

$$X_1 \sim \text{log-normal}(\mu_1, \sigma_1^2) \text{ and } X_2 \sim \text{log-normal}(\mu_2, \sigma_2^2).$$

Suppose that  $X_{1i}, i = 1, 2, \dots, n_1$  and  $X_{2j}, j = 1, 2, \dots, n_2$  denote random samples from  $X_1$  and  $X_2$ , respectively. Then the likelihood function is given by

$$L(\mathbf{x} | \mu_1, \sigma_1^2, \mu_2, \sigma_2^2) = \prod_{i=1}^{n_1} f(x_{1i} | \mu_1, \sigma_1^2) \prod_{j=1}^{n_2} f(x_{2j} | \mu_2, \sigma_2^2) \\ = \left(\frac{1}{\sigma_1^n \sigma_2^{n_2}}\right) \left(\frac{1}{\sqrt{2\pi}}\right)^{n+n_2} \left(\frac{1}{\prod_{i=1}^{n_1} \prod_{j=1}^{n_2} x_{ij}}\right) \times \exp \left[ - \left( \frac{\sum_{i=1}^n (\ln(x_{1i}) - \mu_1)^2}{2\sigma_1^2} + \frac{\sum_{j=1}^{n_2} (\ln(x_{2j}) - \mu_2)^2}{2\sigma_2^2} \right) \right].$$

Where  $f(\cdot)$  is the probability density function.

The choice for a prior distribution on the location parameter ( $\mu_i$ ) is the normal distribution. The obvious choice for a prior on the scale parameter ( $\sigma_i^2$ ) is the inverse gamma distribution, since not only it is conjugate, but also the support of inverse gamma distribution is restricted to positive real numbers. In notation, we write the prior distributions of unknown parameters  $\mu$  and  $\sigma^2$  as follows

$$\mu_i \sim N(\mu_{0i}, \sigma_{0i}^2) \text{ and } \sigma_i^2 \sim \text{Inv-Gamma}(\alpha_i, \beta_i).$$

Then, 
$$\pi(\mu_i) = \frac{1}{\sqrt{2\pi}\sigma_{0i}} \exp \left[ -\frac{(\mu_i - \mu_{0i})^2}{2\sigma_{0i}^2} \right] \text{ and}$$

$$\pi(\sigma_i^2) = \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} \sigma_i^{-2(\alpha_i+1)} \exp \left( -\frac{\beta_i}{\sigma_i^2} \right) \text{ for } i = 1, 2.$$

Here, the parameters  $\mu_{0i}, \sigma_{0i}^2, \alpha_i$  and  $\beta_i$  are assigned fixed values, so they are called hyper parameters. One can also add another level to the hierarchy as hyper priors. The hyper priors provide more information but also they add complexity to the model.

Assuming prior independence, the joint prior distribution  $\theta =$

$$(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \text{ is}$$

$$\pi(\theta) = \pi(\mu_1) \times \pi(\sigma_1^2) \times \pi(\mu_2) \times \pi(\sigma_2^2) \\ = \frac{\beta_1^{\alpha_1} \beta_2^{\alpha_2} \sigma_1^{-2(\alpha_1+1)} \sigma_2^{-2(\alpha_2+1)}}{2\pi\sigma_{01}\sigma_{02}\Gamma(\alpha_1)\Gamma(\alpha_2)} \exp \left[ - \left( \frac{(\mu_1 - \mu_{01})^2}{2\pi\sigma_{01}^2} + \frac{(\mu_2 - \mu_{02})^2}{2\pi\sigma_{02}^2} + \frac{\beta_1}{\sigma_1^2} + \frac{\beta_2}{\sigma_2^2} \right) \right]. \quad (5)$$

Then the posterior distribution is

$$\pi(\theta | \mathbf{x}) \propto L(\mathbf{x} | \theta) \times \pi(\theta) = \frac{\beta_1^{\alpha_1} \beta_2^{\alpha_2} \sigma_1^{-2(\alpha_1+n_1+2)} \sigma_2^{-2(\alpha_2+n_2+2)}}{2\pi^{(n_1+n_2+2)} \sigma_{01} \sigma_{02} \Gamma(\alpha_1) \Gamma(\alpha_2) \prod_{i=1}^{n_1} x_{1i} \prod_{j=1}^{n_2} x_{2j}} \exp \left[ - \frac{(\mu_1 - \mu_{01})^2}{2\sigma_{01}^2} - \frac{(\mu_2 - \mu_{02})^2}{2\sigma_{02}^2} \right] \\ \times \exp \left[ - \frac{1}{2} \left( \frac{2\beta_1 + \sum_{i=1}^{n_1} (\ln(x_{1i}) - \mu_1)^2}{\sigma_1^2} + \frac{2\beta_2 + \sum_{j=1}^{n_2} (\ln(x_{2j}) - \mu_2)^2}{\sigma_2^2} \right) \right]. \quad (6)$$

It is difficult to identify the distribution in 6 using distribution theory since it is not in the closed form. This leads to obtain the distribution in 6 using Markov Chain Monte Carlo (MCMC) simulations. To simulate the values from the posterior distribution in 6, first we need to find the full conditional distributions of  $\mu_i$  and  $\sigma_i^2$ .

The full conditional distribution of  $\mu_1$  is

$$\pi(\mu_1 | \mathbf{x}, \mu_2, \sigma_1^2, \sigma_2^2) \sim N \left( \frac{\frac{\mu_{01} + \sum_{i=1}^{n_1} \ln(x_{1i})}{\sigma_{01}^2 + \sigma_1^2}}{\left(\frac{1}{\sigma_{01}^2} + \frac{n_1}{\sigma_1^2}\right)}, \frac{1}{\left(\frac{1}{\sigma_{01}^2} + \frac{n_1}{\sigma_1^2}\right)} \right).$$

And the full conditional distribution of  $\mu_2$  is

$$\pi(\mu_2 | \mathbf{x}, \mu_1, \sigma_1^2, \sigma_2^2) \sim N \left( \frac{\frac{\mu_{02} + \sum_{j=1}^{n_2} \ln(x_{2j})}{\sigma_{02}^2 + \sigma_2^2}}{\left(\frac{1}{\sigma_{02}^2} + \frac{n_2}{\sigma_2^2}\right)}, \frac{1}{\left(\frac{1}{\sigma_{02}^2} + \frac{n_2}{\sigma_2^2}\right)} \right).$$

The full conditional distribution of  $\sigma_1^2$  is

$$\pi(\sigma_1^2 | \mathbf{x}, \mu_1, \mu_2, \sigma_2^2) \sim \text{Inv-Gamma} \left( \alpha_1 + \frac{n_1}{2}, \frac{2\beta_1 + \sum_{i=1}^{n_1} (\ln(x_{1i}) - \mu_1)^2}{2} \right).$$

The full conditional distribution of  $\sigma_2^2$  is

$$\pi(\sigma_2^2 | \mathbf{x}, \mu_1, \mu_2, \sigma_1^2) \sim \text{Inv-Gamma} \left( \alpha_2 + \frac{n_2}{2}, \frac{2\beta_2 + \sum_{j=1}^{n_2} (\ln(x_{2j}) - \mu_2)^2}{2} \right).$$

The derivations of the likelihood function, the joint prior distribution, the posterior distribution and the full conditional distributions can be found in Wickramasinghe [11]. We proposed a Gibbs sampling algorithm using these full conditionals as follows.

1. Give  $\mu_1, \mu_2, \sigma_1^2$  and  $\sigma_2^2$  starting values, say  $\mu_1^{[0]}, \mu_2^{[0]}, \sigma_1^{2[0]}$  and  $\sigma_2^{2[0]}$ .
2. Generate  $\mu_1^{[1]}, \mu_2^{[1]}, \sigma_1^{2[1]}$  and  $\sigma_2^{2[1]}$  from the full conditionals as describe earlier

$$\mu_1^{[1]} \sim \pi(\mu_1 | x, \mu_2^{[0]}, \sigma_1^{2[0]}, \sigma_2^{2[0]}), \mu_2^{[1]} \sim \pi(\mu_2 | x, \mu_1^{[0]}, \sigma_1^{2[0]}, \sigma_2^{2[0]}),$$

$$\sigma_1^{2[1]} \sim \pi(\sigma_1^2 | x, \mu_1^{[1]}, \mu_2^{[1]}, \sigma_2^{2[0]}) \text{ and } \sigma_2^{2[1]} \sim \pi(\sigma_2^2 | x, \mu_1^{[1]}, \mu_2^{[1]}, \sigma_1^{2[0]}).$$

3. Then repeat the previous step recursively using a loop replacing the starting values.
4. The chain of values produced by this procedure is a Markov chain, and it converges to its equilibrium distribution which is

$$\pi(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2 | \mathbf{x}).$$

We now derive the prior and posterior predictive distribution of a new value  $y$  using the joint prior distribution in 5 and the posterior distribution in 6.

The prior predictive probability of a new value  $y$  is given by

$$f(y) = \int_{-\infty}^{\infty} \int_0^{\infty} f(y | \theta) \pi(\theta) d\sigma^2 d\mu$$

$$\propto \frac{1}{y} \int_{-\infty}^{\infty} \exp\left[-\left(\frac{\mu - \mu_0}{2\sigma_0^2}\right)^2\right] \frac{1}{\left[(\ln y - \mu)^2 + 2\beta\right]^{\left(\frac{\alpha+1}{2}\right)}} d\mu. \quad (7)$$

It is difficult to calculate the integral in 7. So  $f(y)$  cannot be derived directly and we use a Monte Carlo method to sample from  $f(y)$ . To sample from  $f(y)$ , follow the following steps.

For  $j = 1$  to  $m$ ,

1. Generate  $\mu_j$  from  $p(\mu)$ .
2. Generate  $\sigma_j^2$  from  $p(\sigma^2)$ .
3. Generate  $y_j^*$  from  $p(x | \mu_j, \sigma_j^2)$ .

End of  $j$  loop.

Note that  $y_1^*, y_2^*, \dots, y_m^*$  are an independent and identically (iid) sample from  $f(y)$ .

The posterior predictive probability of a future observation  $y$ , given the observed data  $x$  is

$$f(y | x) = \int_{-\infty}^{\infty} \int_0^{\infty} f(y | \theta) \pi(\theta | x) d\sigma^2 d\mu$$

$$\propto \frac{1}{y} \int_{-\infty}^{\infty} \exp\left[-\left(\frac{\mu - \mu_0}{2\sigma_0^2}\right)^2\right] \frac{1}{\left[(\ln y - \mu)^2 + 2\beta + \sum_{i=1}^n (\ln(x_i) - \mu)^2\right]^{\left(\frac{\alpha+n+1}{2}\right)}} d\mu. \quad (8)$$

Again, the distribution in 8 is not in the closed form. So  $f(y|x)$  cannot be derived directly and we use a Monte Carlo method to sample from  $f(y|x)$ . To sample from  $f(y|x)$ , follow the following steps.

For  $j = 1$  to  $m$ ,

1. Generate  $\mu_j$  from  $p(\mu, \sigma^2 | x)$ .
2. Generate  $\sigma_j^2$  from  $p(\mu, \sigma^2 | x)$ .
3. Generate  $y_j^*$  from  $p(x | \mu_j, \sigma_j^2)$ .

End of  $j$  loop.

Note that  $y_1^*, y_2^*, \dots, y_m^*$  are an independent and identically (iid) sample from  $f(y|x)$ .

The descriptive statistics of the sample from the posterior predictive distribution can be compared to the observed data to assess the model fit. If the model fits well, the predictive distribution should be relatively likely to the original data. On the other hand, if there is a large variation between observed data and the data from the predictive distribution, it indicates that the model performs poorly. We perform an analysis using predictive simulations in Section 4.

### Data Analysis

We illustrate the methods discussed in Section 2 and 3 using a published data arising from an experimental aging study from Teasdale et al. [12]. The data set consists of the mean sway range in the forward-backward plane and the mean sway range in the side-to-side plane. Note that forward-backward plane and side-to-side plane are two methods of physical treatments on penetrability of posture control. Nine elderly and eight young adults were participated in the experiment. All the participants stood bare-foot on the force platform with feet together and arms along the body. They were asked to maintain an upright stable posture on the force platform and the sway range in the forward-backward plane and the sway range in the side-to-side plane were calculated. For each participant, 24 trials were given and for each trial, the participants need to maintain an upright stable posture on the force platform for 20s. The mean sway ranges in the forward-backward plane and in the side-to-side plane were calculated and the data are given below with  $n_1 = 17$  and  $n_2 = 17$ .

The mean sway range (in millimeters) in the forward-backward plane ( $X_1$ ): 19, 30, 20, 19, 29, 25, 21, 24, 50, 25, 21, 17, 15, 14, 14, 22, 17.

The mean sway range (in millimeters) in the side-to-side plane ( $X_2$ ): 14, 41, 18, 11, 16, 24, 18, 21, 37, 17, 10, 16, 22, 12, 14, 12, 18.

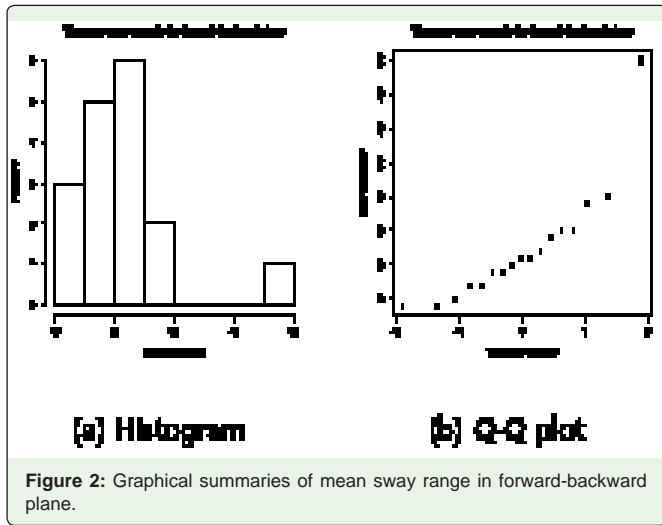
Figure 2 and Figure 3 show that the histogram and Q-Q plot of the mean sway range in the forward-backward plane and the mean sway range in the side-to-side plane respectively. Both histogram and Q-Q plot show that the distributions of the mean sway range in the forward-backward plane and the side-to-side plane are positively skewed (right skewed).

Table 1 also indicates that the distributions of the mean sway range in the forward-backward plane and the side-to-side plane positively skewed (Mean > Median)

Figure 4 shows the Q-Q plots for log transformed mean sway range in the forward-backward plane and the side-to-side plane

**Table 1:** Descriptive statistics.

	Forward-backward	Side-to-side
Mean	22.47	18.88
Median	21	17
Standard Deviation	8.54	8.53



and the distributions of the transformed data are more nearly normal. This suggests that the distributions of mean sway range in the forward-backward plane and the side-to-side plane might follow log-normal distributions, but without more information, we cannot predict accurately.

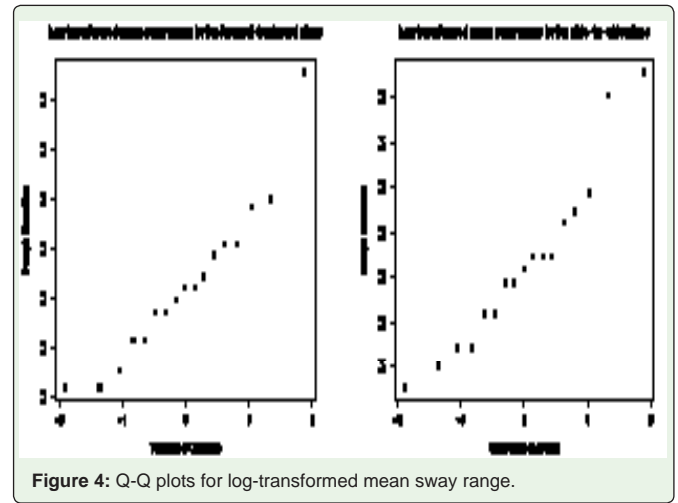
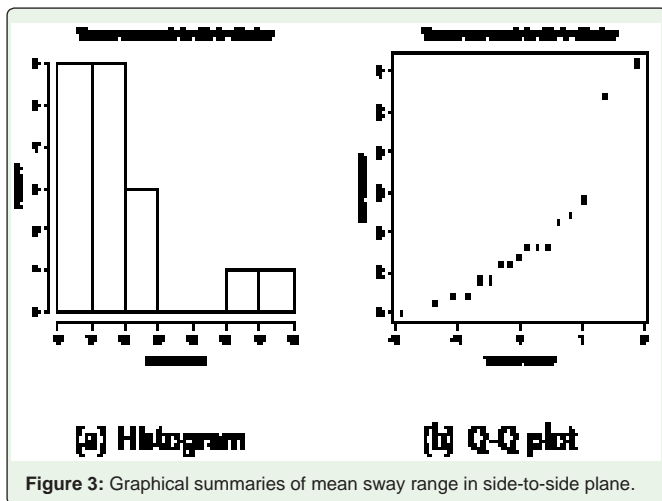
We use Kolmogorov-Smirnov (KS) test to assess the appropriateness of the log-normal distribution for the data as follows,

$H_0$ : The mean sway range follows a log-normal distribution

$H_1$ : The mean sway range does not follow a log-normal distribution.

Table 2 shows that for both planes, we don't have sufficient evidence to reject the log-normal fit. So we can assume that the log-normal model adequately describes the mean sway ranges in the forward-backward plane and the side-to-side plane.

Table 3 gives the method of moment estimates (MME) and the maximum likelihood estimates (MLE) of log-normal parameters for the mean sway ranges in the forward-backward plane and the side-to-side plane. The MME's and the MLE's of  $\mu$  and  $\sigma^2$  are very close for both planes.



For illustration purpose, we assume that a practically meaningful test is provided by  $\delta' = 0.01$  millimetres. The null hypothesis is that the mean sway range in the side-to-side plane ( $X_2$ ) is inferior to the mean sway range in the forward-backward plane ( $X_1$ ) by at least 1.01 millimetres. The alternative is that the mean sway range in the side-to-side plane ( $X_2$ ) is not inferior to the mean sway range in the forward-backward plane ( $X_1$ ) by less than 1.01 millimetres. Accordingly the hypotheses to be tested are given below

$$H_0 : \frac{E(X_1)}{E(X_2)} \geq 1.01 \text{ vs } H_1 : \frac{E(X_1)}{E(X_2)} < 1.01$$

$$H_0 : \eta_1 - \eta_2 \geq 0.01 \text{ vs } H_1 : \eta_1 - \eta_2 < 0.01. \quad (9)$$

We obtain the generalized p-value of 0.86 using the algorithm describe in Section 2. On the other hand, the Z-score approach gives a p-value of 0.92. All these results lead to the conclusion that the data do not provide sufficient evidence to indicate that the ratio of the means of the mean sway ranges in the forward-backward plane and the side-to-side plane is less than 1.01.

We now illustrate the Bayesian approach discussed in Section 3. Note that the prior distribution of location parameter ( $\mu$ ) is normal distribution. When no information is available on  $\mu$ , a usual choice for the prior mean of location parameter ( $\mu_0$ ) is the zero. Note that this prior choice centers the prior belief of  $\mu$  around zero. To make

**Table 2:** KS test statistic and p-value.

	Forward-backward	Side-to-side
KS test statistic	0.1262	0.1747
p-value	0.9494	0.6773

**Table 3:** The MME's and the MLE's of parameters.

	Forward-backward	Side-to-side
MME of $\mu$	3.05	2.85
MME of $\sigma^2$	0.13	0.18
MLE of $\mu$	3.04	2.86
MLE of $\sigma^2$	0.11	0.15



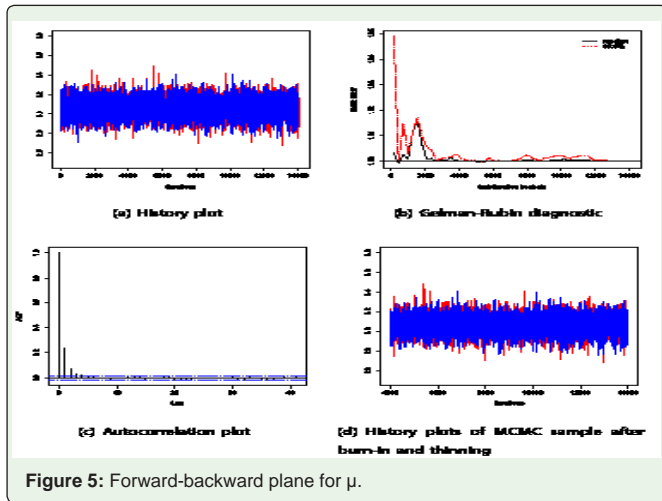


Figure 5: Forward-backward plane for  $\mu$ .

this prior diffuse, we can set the prior variance of location parameter ( $\sigma_0^2$ ) to a large value, (for example 10000). The prior distribution of scale parameter ( $\sigma^2$ ) is inverse-gamma distribution. We set  $\alpha$  and  $\beta$  to a small value,  $\alpha = \beta = 0.01$ .

One can also take an empirical Bayes approach in selecting prior parameters using historical data [16]. For example one can set the prior variance to a very small value (0.01). The mean ( $\mu$ ) and the variance ( $\sigma^2$ ) of the mean sway range in the forward-backward plane are 3.06 and 0.10 respectively. The prior distribution of location parameter ( $\mu$ ) is normal distribution. We have set prior mean of location parameter ( $\mu_0$ ) to 3.06 and the prior variance of location parameter ( $\sigma_0^2$ ) to 0.01. The prior distribution of scale parameter ( $\sigma^2$ ) is inverse-gamma distribution. We have set prior mean of scale parameter to 0.10 and the prior variance of scale parameter to 0.01.

$$E(\sigma^2) = \frac{\beta}{\alpha - 1} = 0.10 \text{ and } Var(\sigma^2) = \frac{\beta^2}{(\alpha - 1)^2 (\alpha - 2)} = 0.01.$$

This will give  $\alpha = 3$  and  $\beta = 0.2$ . Similarly, for the mean sway range in the side-to-side plane, the values of hyper parameters are  $\mu_0 = 3.04$ ,  $\sigma_0^2 = 0.01$ ,  $\alpha = 4.3$  and  $\beta = 0.5$ . We perform the Bayesian analysis under these prior settings. We obtain two random samples (two chains) with 10,000 observations from the posterior distributions for  $\mu$  and  $\sigma^2$  of the mean sway range in the forward-backward plane and

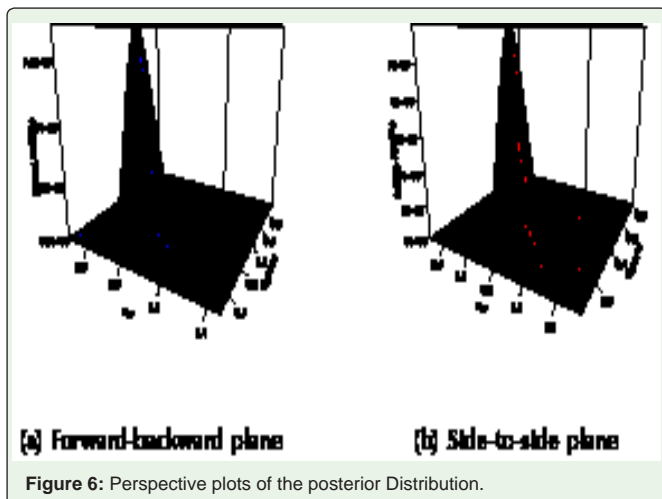


Figure 6: Perspective plots of the posterior Distribution.

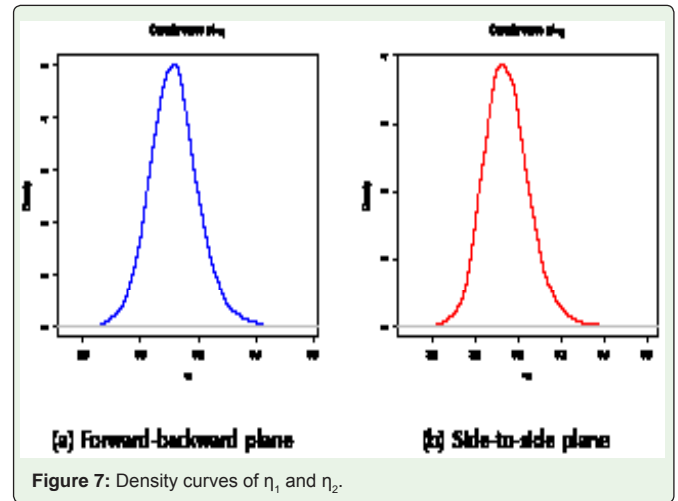


Figure 7: Density curves of  $\eta_1$  and  $\eta_2$ .

the side-to-side plane. So we have run two chains for 14,000 iterations with a burn-in of 4,000.

The first and easiest way to check the convergence of the MCMC algorithm is to visually inspect the trace plots or history plots of  $\mu$  and  $\sigma^2$  of the mean sway range in the forward-backward plane and the side-to-side plane. Figure 5 (a) shows the history plots of  $\mu$  of the mean sway range in the forward-backward plane. Here we have run two chains simultaneously and each chain shows in different colour (red or blue). The plots do not show any particular patterns and all the chains appear to be overlapping one another and mixing well. Similarly, we examined the other parameters of the mean sway range in the forward-backward plane and the side-to-side plane and chains are mixing well.

To have a more precise view on convergence, we use the Brooks-Gelman diagnostics which is introduced by Brooks et al. [13] to calculate the Brooks-Gelman convergence statistics ( $R_c$ ). Figure 5 (b) plots the Brooks-Gelman diagnostic for  $\mu$  of the mean sway range in the forward-backward plane for two chains of 14000 iterations. For the convergence,

$R_c$  should be close to 1 and  $R_c$  is indeed close to 1, roughly after 3000 iterations. So we have considered a burn-in period of 4000 iterations in obtaining posterior summaries.

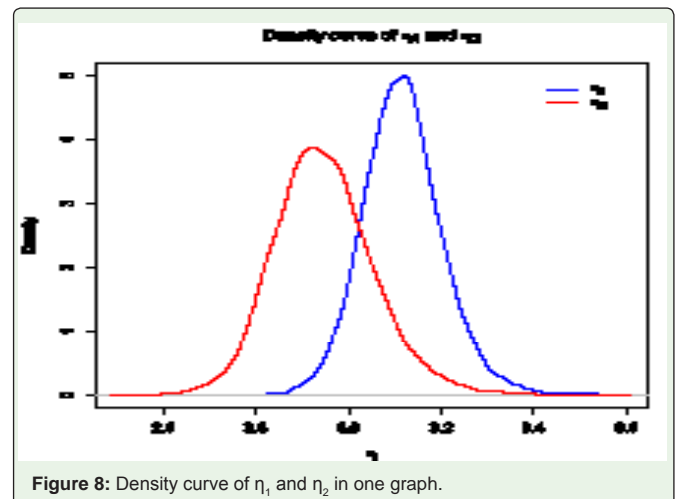


Figure 8: Density curve of  $\eta_1$  and  $\eta_2$  in one graph.

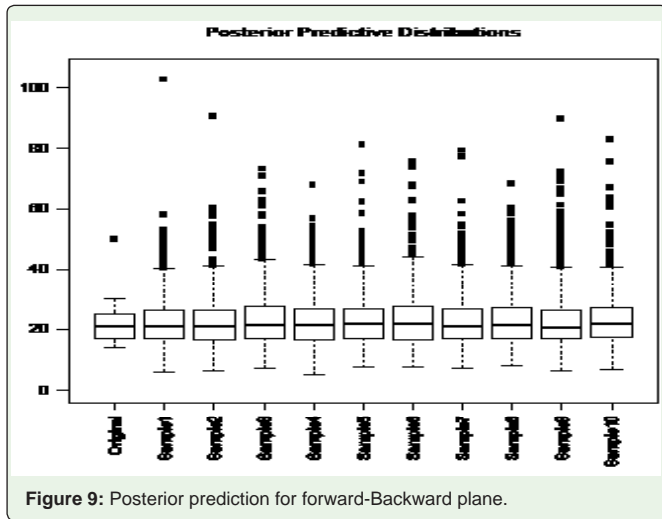


Figure 9: Posterior prediction for forward-Backward plane.

The next step is to obtain random samples after the chains are settled down to the posterior distribution. Figure 5 (c) shows the autocorrelations by lag for  $\mu$  of the mean sway ranges in the forward-backward plane for two chains of 6000 iterations. The correlation at lag 0 is 1 and the rest of the correlations are very close to zero. It indicates that we have a random sample from the posterior distribution and the thinning is not needed in this case.

Table 4 reports the posterior summaries, Monte Carlo Error and sample standard deviation for  $\mu$  and  $\sigma^2$  of the mean sway range in the forward-backward plane and the side-to-side plane. The MC Errors of each parameters of both planes less than 5% of the sample standard deviations. We have obtained the random samples with 10,000 independent and identically distributed observations from the posterior distributions.

Figure 5 (d) shows the history plots of  $\mu$  of the mean sway range in the forward-backward plane after burn-in. Figure 6 shows the perspective plots for the joint posterior of the mean sway range in the forward-backward plane and the side-to-side plane. These perspective plots make 3D plots of Posterior distributions of a surface over  $\mu$ - $\sigma^2$  plane. The plots clearly show that the joint posterior is unimodal

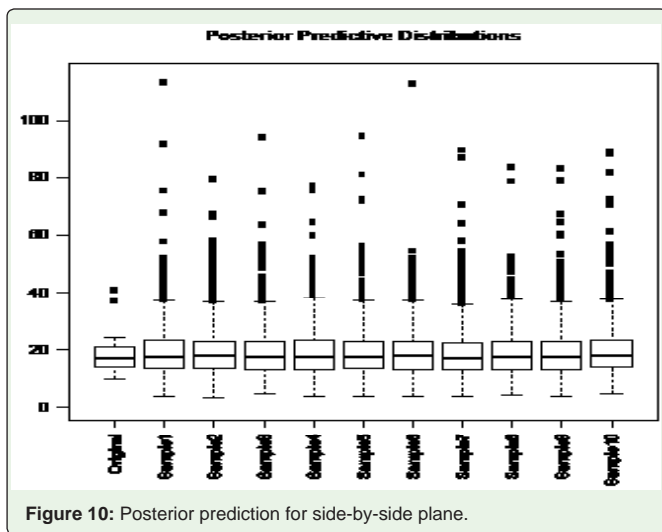


Figure 10: Posterior prediction for side-by-side plane.

Table 4: Posterior estimates of the parameters.

	Parameter	Estimate	MC Error	SD	5%*SD
Forward-Backward	$\mu$	3.05	0.0006	0.082	0.0041
	$\sigma^2$	0.12	0.0004	0.046	0.0023
Side-to-Side	$\mu$	2.84	0.0007	0.101	0.0051
	$\sigma^2$	0.16	0.0005	0.07	0.0035

Table 5: Descriptive statistics of posterior prediction for forward-backward plane.

	Mean	Median	Inter-quartile range (IQR)
Original	22.47	21	8
Sample1	22.52	21.17	9.38
Sample2	22.36	20.95	9.81
Sample3	23.02	21.32	10.48
Sample4	22.84	21.51	10.03
Sample5	22.87	21.65	9.64
Sample6	23.1	21.91	11.03
Sample7	22.6	21.13	9.9
Sample8	23.05	21.42	9.86
Sample9	22.6	20.75	9.51
Sample10	23.07	21.71	9.7

indicating that MCMC analysis is more meaningful. Therefore we can now make inferences on  $\eta_1$  and  $\eta_2$ .

Note that our interest is on  $\eta_1$  and  $\eta_2$ . Figure 7 shows the density curves of  $\eta_1$  and  $\eta_2$ . Figure 8 combines both density curves of  $\eta_1$  and  $\eta_2$  into one graph.

We computed Bayesian probability under  $H_0 (P(H_0|\mathbf{x}))$  for testing hypotheses in 9 and the value is 0.88. Since  $P(H_0|\mathbf{x})$  is sufficiently large, we decide in favor of Null Hypothesis ( $H_0$ ). This leads to the conclusion that the data do not provide the sufficient evidence to indicated that the ratio of the means of the mean sway ranges in the forward-backward plane and the side-to-side plane is less than 1.01. We also examine the prediction ability of the model using posterior predictive distribution described in Section 3.

Figure 9 shows the posterior predictions of 10 samples with 10,000 observations for the mean sway ranges in the forward-backward plane. The first boxplot represents the original sample and the next

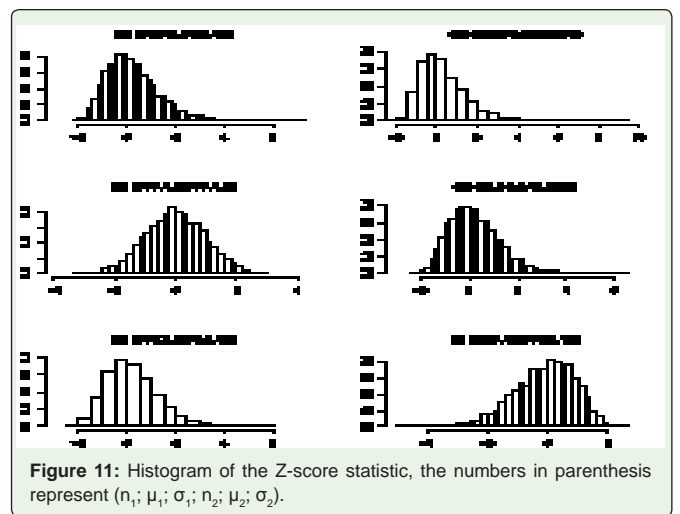


Figure 11: Histogram of the Z-score statistic, the numbers in parenthesis represent  $(n_1; \mu_1; \sigma_1; n_2; \mu_2; \sigma_2)$ .

**Table 6:** Descriptive statistics of posterior prediction for side-by-side plane.

	Mean	Median	Inter-quartile range (IQR)
Original	18.88	17	8.53
Sample1	19.28	17.46	9.51
Sample2	19.26	17.77	9.43
Sample3	18.88	17.31	9.49
Sample4	18.99	17.44	9.94
Sample5	19.04	17.45	9.52
Sample6	19.06	17.73	9.73
Sample7	18.88	17.03	9.17
Sample8	18.79	17.32	10
Sample9	19.17	17.55	9.45
Sample10	19.61	17.68	9.53

10 boxplots represent the predicted samples. We compare the last 10 boxplots (predicted samples) with the first boxplot (original sample) and the observed dataset appears to be consistent with the generated datasets.

Furthermore, Table 5 compares the descriptive statistics of the predicted samples with the original sample of the mean sway range in the forward-backward plane. The mean and median of each predicted samples are approximately equal to the mean and median of the original sample. But the inter-quartile range of predicted samples are higher than the inter-quartile range of the original sample due to the heterogeneity of data. Figure 9 and Table 5 lead to the conclusion that there is no difference between predicted samples and the original sample of the mean sway range in the forward-backward plane.

Similarly, Figure 10 shows the posterior predictions of 10 samples with 10,000 observations for the mean sway range in the side-to-side plane. The first boxplot represents the original sample and the next 10 boxplots represent the predicted samples. We have compared the last 10 boxplots (predicted samples) with the first boxplot (original sample) and it seems that no difference can be claimed.

Table 6 compares the descriptive statistics of the predicted samples with the original sample of the mean sway range in the side-to-side plane. The mean and median of each predicted samples are approximately equal to the mean and median of the original sample. But the inter-quartile range of predicted samples are higher than the inter-quartile range of the original sample due to the heterogeneity of data. Figure 9 and Table 5 lead to the conclusion that there is no difference between predicted samples and the original sample of the mean sway range in the side-to-side plane.

**Simulation Studies**

We also conducted various simulation studies to compare the performance of three methods; Z-score, generalized p-value and Bayesian approaches. For Z-score and generalized p-value methods, the type I error probabilities for various combinations of  $n_1, n_2, \mu_1, \sigma_1^2$  and  $\sigma_2^2$  are reported in Table 7. For the simulations, we have taken  $\mu_2 = 0$ , without loss of generality. The type I error probability corresponds to the parameter combinations satisfying  $\mu_1 + \frac{1}{2}\sigma_1^2 - \mu_2 + \frac{1}{2}\sigma_2^2 = 0.01$  with 10,000 runs.

**Table 7:** Size of the generalized p-value test and the Z-score test at 5% significance level when  $\mu_2 = 0$  and  $H_0: \eta_1 - \eta_2 \geq 0.01$  vs.  $H_1: \eta_1 - \eta_2 < 0.01$ .

$\eta_1$	$\eta_2$	$\mu_1$	$\sigma_1^2$	$\sigma_2^2$	Size	
					Generalized p-value	Z-Score
4	4	1.01	2	4	0.0338	0.0098
		0.01	3	3	0.0374	0.0375
		5.01	2	12	0.0344	0.0001
		0.01	12	12	0.0398	0.0166
10	10	1.01	2	4	0.0427	0.0133
		0.01	3	3	0.0464	0.0388
		5.01	2	12	0.0424	0.0004
		0.01	12	12	0.0489	0.026
25	25	0.01	1	1	0.0454	0.0436
		0.01	5	5	0.051	0.0447
		0.01	10	10	0.0488	0.0382
		0.01	100	100	0.0505	0.0342
		2.01	4	8	0.0463	0.0199
		4.01	8	16	0.0473	0.0137
40	25	0.01	1	1	0.0516	0.0357
		0.01	5	5	0.053	0.034
		0.01	10	10	0.0533	0.0257
		5.01	2	12	0.0483	0.0079
25	40	0.01	1	1	0.049	0.0542
		0.01	5	5	0.0471	0.0558
		0.01	10	10	0.0502	0.0583
		5.01	2	12	0.0489	0.0154
40	40	8.01	4	20	0.0469	0.0148
		14.01	4	32	0.0457	0.0121
100	25	0.01	1	1	0.053	0.0293
		0.01	5	5	0.0514	0.016
		0.01	10	10	0.0482	0.0118
		13.01	4	30	0.0494	0.0058
25	100	0.01	1	1	0.0425	0.0666
		0.01	5	5	0.0478	0.0819
		0.01	10	10	0.0486	0.0847
		13.01	4	30	0.0445	0.0262
100	100	13.01	4	30	0.0514	0.0225
200	50	0.01	1	1	0.0534	0.0356
		0.01	5	5	0.0502	0.0259
		0.01	10	10	0.0498	0.0228



50	200	0.01	1	1	0.0439	0.0535
		0.01	5	5	0.0532	0.0711
		0.01	10	10	0.0477	0.0739
200	200	24.01	12	60	0.05	0.0319
		40.01	2	82	0.0491	0.0272

For smaller values of  $n_1$  and  $n_2$  ( $n_1 < 25$  and  $n_2 < 25$ ), Z-score test has type I error probabilities which are too small, whereas the test based on generalized p-value is extremely satisfactory for controlling type I error probability. For  $H_1: \eta_1 - \eta_2 < 0.01$ , the Z-score test type I error probabilities which are too small when  $n_1 > n_2$ , and they are too large when  $n_1 < n_2$ . The histogram in Figure 11 (d), is skewed to the right and the corresponding type I error probability to the Z-score test is too small. On the other hand, the histogram of Z-score statistics in Figure 11 (f), is skewed to the left and the corresponding type I error probability to the Z-score test is too large. But, the test based on the generalized p-value controls type I error quite satisfactorily, regardless of  $n_1$  and  $n_2$ .

The power for various combinations of  $n_1, n_2, \mu_1, \sigma_1^2$  and  $\sigma_2^2$  are reported in Table 8. For  $H_1: \eta_1 - \eta_2 < 0.01$ , when  $n_1 > n_2$ , the Z-score test has a smaller power compared to the generalized p-value test. On the other hand when  $n_1 < n_2$ , the Z-score test has a larger power compared to the generalized p-value test. Also, there are situations where the power of the Z-score test is smaller than the significance level of 0.05 (last row in Table 8), indicating that Z-score test is biased.

Zhou et al. [15] claimed that, when both  $n_1$  and  $n_2$  are large, the distribution of the Z-score statistic is approximately standard normal under  $H_0$ . The histogram of the Z-score statistic in Figure 11 indicate that the distribution of Z-score statistic is skewed when both sample sizes are large as 100 but  $(n_1, \mu_1, \sigma_1^2)$  is not approximately equal to  $(n_2, \mu_2, \sigma_2^2)$ . So the normal approximation is appropriate only when both samples are very large and  $(n_1, \mu_1, \sigma_1^2)$  is approximately equal to  $(n_2, \mu_2, \sigma_2^2)$ .

Table 9 shows the generalized p-value and the Bayesian probabilities for testing hypothesis 9. The generalized p-value has performed well for both large and small samples. The Bayesian probabilities also perform very well and indicate a very good agreement in conclusion based on the Bayes factor. The Bayes factor is not calibrated, so one of the future researches is to develop a robust method to compare the Bayesian approach and the generalized p-value approach using the relative belief ratio as discussed in Muthukumarana and Evans [14].

**Discussion**

In this article, we have derived a Bayesian Inference procedure for testing non-inferiority hypothesis testing for the ratio of two independent log-normal means. In non-inferiority studies, the objective is to demonstrate the new treatment or the product is not worse than the existing treatment or the product by more than a pre-specified, small amount ( $\delta$ ).

The mean of log-normal distribution is a function of both  $\mu$  and  $\sigma^2$  and it is difficult to obtain exact or optimum tests for testing non-inferiority hypotheses. Zhou et al. [15] have introduced the Z-score approach for the problem of testing the equality of two log-normal

**Table 8:** Power of the generalized p-value test and the Z-score test at 5% significance level when  $\mu_2 = 0$  and  $H_0: \eta_1 - \eta_2 \geq 0.01$  vs.  $H_1: \eta_1 - \eta_2 < 0.01$ .

$\eta_1$	$\eta_2$	$\mu_1$	$\sigma_1^2$	$\sigma_2^2$	Power	
					Generalized p-value	Z-Score
4	4	0	4	12	0.1592	0.0347
		0	4	20	0.265	0.0358
		3	4	12	0.0525	0.0018
		4	1	11	0.0693	0.0002
10	10	0	4	12	0.3961	0.229
		0	4	20	0.6753	0.4149
		3	4	12	0.0849	0.0056
		4	1	11	0.1113	0.0041
25	25	1	1	5	0.376	0.2302
		1	5	9	0.1536	0.0925
		1	10	14	0.1045	0.0658
		0	2	4	0.3582	0.3136
		0	7	9	0.1369	0.1066
40	25	1	1	5	0.4051	0.2365
		1	5	9	0.1766	0.0821
		1	4	9	0.305	0.1716
		1	9	14	0.1598	0.0831
25	40	1	1	5	0.4656	0.378
		1	5	9	0.16	0.1566
		1	4	9	0.3005	0.2746
		1	9	14	0.1527	0.1489
100	25	1	1	5	0.4187	0.2344
		1	5	9	0.2003	0.0777
		0	1	2	0.4192	0.274
		0	1	3	0.705	0.5452
		1	4	7	0.1294	0.0434
		1	9	12	0.0871	0.0262
25	100	1	1	5	0.6665	0.6602
		1	5	9	0.1895	0.2485
		0	1	2	0.4006	0.4751
		0	1	3	0.8155	0.8511
		1	4	7	0.1235	0.1608
		1	9	12	0.0783	0.1246

100	100	0	1	2	0.6717	0.5147
		0	1	3	0.8314	0.7157
		1	1	5	0.3157	0.2914
		1	5	9	0.2714	0.2213
50	200	1	1	4	0.5116	0.5043
		1	5	8	0.1495	0.186
		0	1	2	0.6831	0.7067
		0	1	3	0.9801	0.9847
		0	1	2	0.6831	0.7067
		0	1	3	0.9801	0.9847
200	50	1	1	4	0.3175	0.213
		1	5	8	0.1469	0.0765
		0	1	2	0.6126	0.515
		1	1	4	0.3175	0.213
		1	5	8	0.1469	0.0765
		0	1	2	0.6126	0.515
200	200	0	1	2	0.7156	0.6012
		0	1	3	0.9701	0.9574
		1	1	4	0.3027	0.2644
		1	5	8	0.2223	0.2057
300	25	0.8	12	18	0.0589	0.0228

means and we have extended this idea for testing non-inferiority hypotheses. This approach is recommended for large samples but for small samples the power and the type I error are too conservative or too liberal. Although, the samples are large, still there are situations where the Z-score test is unsatisfactory. But, the Z-score test performs well in terms of the power and the type I error when the  $\mu$ 's and  $\sigma_2$ 's close to each other under large sample sizes.

There are also few challenges in classical p-value method in hypothesis testing such as difficult to find the suitable test statistics, difficult to find the sampling distribution of the test statistic and involves many nuisance parameters. To overcome these challenges, Tsui and Weerahandi [9] introduced the concept of generalized p-value approach.

We utilized the generalized p-value approach for testing non-inferiority hypotheses. The generalized p-value approach controls type I error quite satisfactorily, regardless of  $n_1$  and  $n_2$ . The generalized p-value approach has performed well for both large and small samples but the Bayesian approach also showed a good agreement with conclusion in non-inferiority hypothesis testing.

**Acknowledgement**

Lahiru Wickramasinghe (email: wickraml@cc.umanitoba.ca) is Ph.D student and Saman Muthukumarana (email: Saman.Muthukumarana@ad.umanitoba.ca) is Assistant Professor in the Department of Statistics, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada. Muthukumarana has been partially supported by grant from the Natural Sciences and Engineering Research Council of Canada. The authors thank two anonymous reviewers whose comments helped to improve the manuscript.

**References**

1. Lehmann EL, Romano JS. Testing statistical hypotheses. *Metrika*.2006; 64: 255-256.

**Table 9:** The generalized p-value and the Bayesian probabilities under  $H_0$  and  $H_1$  when  $\mu_2 = 0$  and  $H_0 : \eta_1 - \eta_2 \geq 0.01$  vs  $H_1 : \eta_1 - \eta_2 < 0.01$ .

$\eta_1$	$\eta_2$	$\mu_1$	$\sigma_1^2$	$\sigma_2^2$	Generalized p-value	$P(H_0   \mathbf{x})$	$P(H_1   \mathbf{x})$	Bayes Factor
					0.7334	0.9561	0.0439	23.6883
4	4	0	3	14	0.0121	0.3709	0.6291	0.6413
4	4	0.01	3	3	0.616	0.7212	0.2788	2.8136
25	25	0	3	0	0.9994	1	0	$\infty$
25	25	0	3	14	0.0001	0	1	0
25	25	0.01	3	3	0.3425	0.9154	0.0846	11.7689
25	4	0	3	0	0.9998	0.9977	0.0023	47.1811
25	4	0	3	14	0.0001	0.2085	0.7915	0.2865
25	4	0.01	3	3	0.3305	0.8804	0.1196	8.0065
4	25	0	3	0	0.9916	0.7483	0.2517	3.2336
4	25	0	3	14	0.0249	0.0324	0.9676	0.0364
4	25	0.01	3	3	0.8359	0.9298	0.0702	14.4061
100	100	0	3	0	1	1	0	$\infty$
100	100	0	3	14	0	0	1	0
100	100	0.01	3	3	0.1671	0.8388	0.1612	5.6596

2. Snapinn S. Noninferiority trials. *Curr Control Trials Cardiovasc Med*. 2000; 1: 19-21.
3. Walker E, Nowacki AS. Understanding Equivalence and Non-inferiority Testing. *J Gen Intern Med*. 2011; 26: 192-196.
4. Munk A, Trampisch H. Therapeutic Equivalence-Clinical Issues and Statistical Methodology in Noninferiority Trials. *Biometrical Journal*. 2005; 47: 7-9.
5. Wellek S. Testing Statistical Hypotheses of Equivalence, vol. 22. Chapman and Hall/CRC. 2003.
6. Koti KM. New Tests for Assessing Non-Inferiority and Equivalence from Survival Data. *Open J Statistics*. 2013; 3: 55-64.
7. Eckhard L, Werner AS, Markus A. Log-normal Distributions across the Sciences: Keys and Clues. American Institute of Biological Sciences. 2001; 51: 341-352.
8. Krishnamoorthy K, Mathews T. Inferences on the means of lognormal distributions using generalized p-values and generalized confidence intervals. *Journal of Statistical Planning and Inference*. 2003; 115: 103-121.
9. Weerahandi S, Tsui K. Generalized p-Values in Significance Testing of Hypotheses in the Presence of Nuisance Parameters. *Journal of the American Statistical Association*. 1989; 84: 602-607.
10. Weerahandi S. Exact Statistical Methods for Data Analysis. New York: Springer-Verlag, 1995.
11. Wickramasinghe L. Non-inferiority hypothesis testing in two-arm trials with log-normal data. Master's thesis, University of Manitoba. 2015.
12. Teasdale N, LaRue J, Bard C, Fleury M. On the cognitive penetrability of posture control. *Exp Aging Res*. 1993; 19: 1-13.
13. Brooks SP, Gelman A. General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*. 1998; 7: 434-455.
14. Muthukumarana S, Evans M. Bayesian inference in two-arm trials using relative belief ratios. *Pharmaceut Statist*. 2015; 14: 471-478.
15. Zhou XH, Gao S, Hui SL. Methods for comparing the means of two independent log-normal samples. *Biometrics*. 1997; 53: 1129-1135.
16. Bayes T, Price R. An Essay towards Solving a Problem in the Doctrine of Chances. *Phil Trans*. 1763; 53: 370-418.