ANALYSIS OF UNBALANCED OCCUPATIONAL EXPOSURE DATA USING A BAYESIAN RANDOM EFFECTS MODEL

Justin Harvey [1]
Department of Statistics and Actuarial Science, University of Stellenbosch, Private Bag X1, Matieland, 7602
e-mail: *jharvey@sun.ac.za*
and
Abrie van der Merwe
Department of Mathematical Statistics and Actuarial Science, University of the Free State, Box 339, Bloemfontein, 9300
e-mail: *matheeme@ufs.ac.za*

Key words: Bayes, non-informative prior, occupational exposure limit (OEL), one-way random effects model.

Summary: Krishnamoorthy and Mathew (2002) made use of a one-way random effects model to analyse lognormally distributed data. They proposed the use of generalized confidence intervals and generalized p-values (frequentist methods) when a balanced data set (an equal number of measurements for each observational unit) was available. Their method was demonstrated on occupational exposure data. Harvey and van der Merwe (2014) developed a Bayesian approach, using objective prior distributions, to analyse the same situation as analysed by Krishnamoorthy and Mathew (2002). Krishnamoorthy and Guo (2005) subsequently extended the generalized confidence intervals and generalized p-value approaches to "unbalanced" data sets (unequal number of measurements for each observational unit). Similarly, in this article the purpose is to extend the Bayesian approach in Harvey and van der Merwe (2014) to unbalanced data. Several non-informative priors are evaluated. Occupational exposure data is used for comparison. The Bayesian approach developed is applicable to any one-way random effects model with lognormally distributed unbalanced data. Results from this article indicate that the Bayesian approach is comparable to the frequentist approaches and indeed offers additional modelling ability.

1. Introduction

Lognormally distributed data occur frequently in practice. For example, in studies of the treatment of the human immunodeficiency virus (HIV) important variables, such as the viral load, are often found to be normally distributed after taking an appropriate log-transformation (Chu et al., 2010). In

[1] Corresponding author.

AMS: 62C10

addition to the transformed data more closely resembling a normal distribution such data is assumed to be lognormally distributed since the log-transformation is often used to stabilize variances when modelling longitudinal data (see Li, Chu and Gallant (2006), Chu et al. (2005) and Lyles, Williams and Chuachoowong (2001) for further discussions). Lognormally distributed data is not limited to only HIV-related data. Another example of its occurrence is within occupational exposure research (Finkelstein and Verma, 2001).

Many different methods have been proposed for the analysis of lognormally distributed data from a one-way random effects model. For example, Krishnamoorthy and Mathew (2002) proposed the one-way random effects model for the analysis of balanced and lognormally distributed occupational exposure data (styrene exposures more specifically) measured on factory workers. Fifteen random workers each had styrene exposures (in mg/m^3) measured at 10 different time points. A comprehensive discussion of occupational exposure data is beyond the scope of this article. The interested reader is referred to the original articles by Krishnamoorthy and Mathew (2002) for a more complete discussion as well as texts by Rappaport, Kromhout and Symanski (1993), Heerderik and Hurley (1994), Lyles, Kupper and Rappaport (1997a) and Lyles, Kupper and Rappaport (1997b).

Krishnamoorthy and Mathew (2002) modelled the mean styrene exposure from a frequentist point of view by means of generalized confidence intervals. Furthermore, they developed a generalized p-value approach that allowed them to test specific hypotheses regarding the occupational exposure limit (OEL), specifically the probability of mean exposure exceeding the OEL. The OEL is a pre-determined exposure threshold that has clinical significance (in some sense). This threshold may differ depending on the clinical setting and usually represents the maximum level of exposure that is deemed (according to clinical judgement) safe for any particular worker. It is not determined in any way by the statistician or from the data itself. Refer to Lyles et al. (1997a) and Lyles et al. (1997b) for a further discussion of the OEL within occupational exposure research. The concept of an OEL is not restricted to occupational exposure data and similar thresholds exist within other research fields. For example, in Chu et al. (2010) a similar threshold was applied that defined the level at which an HIV positive patient became eligible for highly active antiretroviral therapy (HAART). In insurance applications we may be interested in the probability of an insurance claim exceeding a pre-specified boundary. It is, therefore, evident that the methods developed by Krishnamoorthy and Mathew (2002) are not only applicable to occupational exposure research, but to any one-way random effects model where the data have a lognormal distribution.

Harvey and van der Merwe (2014) proposed a Bayesian approach to analysing lognormally distributed data from a one-way random effects model. In their article they presented an objective Bayesian approach for modelling the mean styrene exposure previously described and compared the effect of several non-informative priors. It was shown that the Bayesian approach has several distinct advantages over the generalized confidence intervals and generalized p-value approaches. The most evident advantage is the flexibility of the Bayesian approach that allows for the modelling of mean exposure for individual workers.

The previous examples and articles discussed considered the case of balanced data from a one-way random effects model. The data were balanced in that there were an equal number of observations for each observational unit, i.e. for each worker. Unfortunately, the case of balanced data is overly simplistic and unbalanced data (where every observational unit does not have the same number of measurements) is often encountered in practice. According to Lewsey, Gardiner and Gettinby

(1997) unbalanced data is common and can arise due to any number of factors, such as equipment failure, patients being lost to follow-up (in clinical trials) or even problems with data collection. For example, in Littell, Stoup and Fruend (2002) in the chapter on Mixed Models unbalanced data from a clinical trial performed in multiple trial centres (the observational units) was used and unbalancing occurred simply due to the inability of every trial centre to recruit the same number of patients. For a more detailed discussion on unbalanced data refer to Schafer and Graham (2002) and for more examples of the occurrence of unbalanced data refer to the illustrative examples in Liao, Lin and Iyer (2005). The analysis of unbalanced data will in turn require an analysis framework that specifically accounts for the unequal number of observations (Lewsey et al., 1997).

It was for this reason that Krishnamoorthy and Guo (2005) extended the generalized confidence intervals and generalized p-values approaches to account for unbalanced data. However, in all other respects the problem-statement was the same as in Krishnamoorthy and Mathew (2002).

The purpose of this article is to extend the methods presented in Harvey and van der Merwe (2014) to the situation where we have unbalanced data. In other words, we will develop an objective Bayesian framework to model unbalanced lognormally distributed data from a one-way random effects model. In order to complete the Bayesian specification of the model suitable prior distributions have to be derived and their performance compared. This forms a large part of this article. The selection and determination of non-informative priors in multi-parameter settings is not an easy task and it has been observed that the selection of a specific prior distribution could have unexpectedly dramatic effects on the posterior distribution. We will consider the Reference prior (Berger and Bernardo, 1992) and the Probability-Matching prior (Datta and Ghosh, 1995). The effectiveness of the proposed prior distributions will be compared.

Lastly, the Bayesian methods developed here are not unique to occupational exposure research. However, in order to aid comparison with the generalized confidence intervals and generalized p-value approaches in Krishnamoorthy and Guo (2005) all derivations will be made assuming the same occupational exposure context.

2. Description of the setting

The setting for this article is similar to the setting described in Harvey and van der Merwe (2014). The point of departure, conceptually, is that the data now is unbalanced. In Harvey and van der Merwe (2014) the data represented the amount of exposure to a particular agent and for each worker there were exactly the same number of observations. In the unbalanced case we have an unequal number of observations for each worker (the mechanism by which this "missing" data is generated is not of interest in this article). Even though this is a minor conceptual change all derivations of priors and posterior distributions would necessarily change.

Therefore, in the unbalanced case we have the diagrammatic representation of the data shown in Table 1.

Let X_{ij} represent the j-th shift-long exposure measurement for the i-th worker, where $j = 1, \ldots, n_i$ and $i = 1, \ldots, k$. Therefore, there are n_i measurements for the i-th worker, which results in the "unbalanced" nature of the data. The X_{ij} are lognormally distributed and therefore $Y_{ij} = ln(X_{ij})$

Table 1: Representation of shift exposure data.

	Shift-long exposure measurements			
Worker	**1**	**2**	$\dots$	$\mathbf{n}_i$
1	x_{11}	x_{12}	$\dots$	x_{1n_1}
2	x_{21}	x_{22}	$\dots$	x_{2n_2}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
k	x_{k1}	x_{k2}	$\dots$	x_{kn_k}

are normally distributed. The situation can be represented by the following one-way random effects model:

$$Y_{ij} = \mu + \tau_i + e_{ij}, \qquad i = 1, \dots, k; j = 1, \dots, n_i,$$

where μ is the general mean, $\tau_i \sim N\left(0, \sigma_\tau^2\right)$ and $e_{ij} \sim N\left(0, \sigma_e^2\right)$. All the random variables are independent of each other and here τ_i represents the random effect due to the i-th worker.

Given the lognormal distribution of the X_{ij},

$$\mu_{x_i} = E\left(X_{ij}|\tau_i\right) = E\left(\exp\left[Y_{ij}\right]|\tau_i\right) = \exp\left(\mu + \tau_i + {}^{\sigma_e^2}/_2\right)$$

and μ_{x_i} is the mean exposure for the i-th worker. Note that $ln(E(X_{ij}|\tau_i)) = \mu + \tau_i + \sigma_e^2/2$. Furthermore, τ_i is unknown but we do know that $\tau_i \sim N\left(0, \sigma_\tau^2\right)$ and therefore, $ln(E(X_{ij})) \sim N\left(\mu + \sigma_e^2/2, \sigma_\tau^2\right)$. Let θ denote the probability that μ_{x_i} exceeds the OEL. Thus, $\theta = P(\mu_{x_i} > OEL) = P(ln(\mu_{x_i}) > ln(OEL)) = 1 - \Phi\left(\left\{ln(OEL) - \mu - \sigma_e^2/2\right\}/\sigma_\tau\right)$, where $\Phi(\cdot)$ denotes the c.d.f. of the standard normal distribution. The hypothesis of interest is:

$$H_0 : \theta \geq A \quad \text{vs.} \quad H_1 : \theta < A,$$

where A is a specific quantity that is usually small (usually between 0.1 and 0.01), according to Krishnamoorthy and Guo (2005). Based on our earlier definition of θ it is evident that the above hypothesis (for a given value of A) is equivalent to the following:

$$H_0 : \ \mu + Z_{1-A}\sigma_\tau + \tfrac{1}{2}\sigma_e^2 \geq ln(OEL) \quad \text{vs.} \quad H_1 : \ \mu + Z_{1-A}\sigma_\tau + \tfrac{1}{2}\sigma_e^2 < ln(OEL)$$

where Z_{1-A} denotes the $100(1-A)$-th percentile of the standard normal distribution. For a given significance level α, the null hypothesis will be rejected if a $100(1-\alpha)\%$ upper confidence limit for $\mu + Z_{1-A}\sigma_\tau + \frac{1}{2}\sigma_e^2$ is less than $ln(OEL)$.

Table 2 is an example of a data set of simulated "styrene exposures." This data set is the same as the data set used in Harvey and van der Merwe (2014), except that observations have been removed at random, resulting in an "unbalanced"design. Removing observations at random to create an unbalanced design is an accepted practice in literature (refer to the examples given in Liao et al. (2005)). This simulated data set was used instead of real data to enable simple comparison to the results obtained for the balanced design in Harvey and van der Merwe (2014). It was simulated to represent the data used in Krishnamoorthy and Mathew (2002), given that only certain sufficient statistics were presented in their article and not the complete data set. This data will serve as a basis

Table 2: Simulated styrene exposures.

	Shift-long exposures									
Worker	1	2	3	4	5	6	7	8	9	10
1	95.6	64.7	50.9	87.4	82.3	149.9	33.4	77.5	70.8	60.9
2	57.4	82.3	174.2	107.8	98.5	129	121.5	95.6	92.8	133
3	84.8	214.9	79.8	169	149.9	164	84.8	84.8	114.4	
4	68.7	77.5	54.1	41.3	64.7	46.5	59.1	45.2	54.1	
5	114.4	101.5	49.4	101.5	90	52.5	114.4	79.8	68.7	87.4
6	87.4	242.3	145.5	133	174.2	214.9	137	129	169	179.5
7	54.1	75.2	84.8	55.7	90	70.8	60.9	101.5	64.7	95.6
8	64.7	95.6	57.4	95.6	82.3	101.5	92.8	60.9	101.5	98.5
9	137	208.5	92.8	159.2	92.8	82.3	90			
10	125.2	87.4	121.5	90	154.5	107.8	117.9	179.5	129	129
11	42.5	73	50.9	59.1	49.4	66.7				
12	57.4	68.7	59.1	64.7	55.7	92.8	42.5			
13	101.5	149.9	111.1	77.5	111.1	84.8	64.7	62.8		
14	68.7	101.5	111.1	179.5	82.3	174.2	174.2	87.4	145.5	114.4
15	121.5	77.5	145.5	174.2	77.5	92.8	159.2	129	104.6	77.5

for discussion in this article and will help us define and illustrate the objectives of the article. An example based on real data is given later in this article. Table 2 represents the X_{ij} data points, from which the $Y_{ij} = ln(X_{ij})$ can easily be obtained.

From these data we have the following definitions and associated results (these results will be used in all subsequent applications and examples):

$$k = 15$$

$$v_1 = \sum_{i=1}^{k} (n_i - 1);\ v_2 = k - 1$$

$$\bar{Y}_{i\bullet} = \tfrac{1}{n_i} Y_{i\bullet} = \tfrac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} = [4.3\ \ 4.7\ \ 4.8\ \ 4.0\ \ 4.4\ \ 5.0\ \ 4.3\ \ 4.4\ \ 4.8\ \ 4.9\ \ 4.0\ \ 4.1\ \ 4.5\ \ 4.8\ \ 4.7]'$$

$$n_i = [10\ \ 10\ \ 9\ \ 9\ \ 10\ \ 10\ \ 10\ \ 10\ \ 7\ \ 10\ \ 10\ \ 6\ \ 7\ \ 8\ \ 10\ \ 10]'$$

$$\bar{Y}_{\bullet\bullet} = \tfrac{1}{k} \sum_{i=1}^{k} \bar{Y}_{i\bullet} = 4.508$$

$$SS_e = v_1 m_1 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\bullet})^2 = 10.492 = \text{“within workers sum of squares”}$$

$$SS_\tau = v_2 m_2 = \sum_{i=1}^{k} n_i (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 = 11.885 = \text{“between workers sum of squares.”}$$

3. Bayesian methodology

The basis for analysing any situation from a Bayesian perspective is the following relationship, a well-known result of Bayes' theorem: $Posterior \propto Likelihood \times Prior$. The likelihood function (in matrix form) is given by:

$$L\left(\mu,\boldsymbol{\tau},\sigma_e^2,\sigma_\tau^2|\mathbf{Y}\right) = \left(2\pi\sigma_e^2\right)^{-\frac{1}{2}\tilde{n}}\exp\left\{-\frac{1}{2\sigma_e prior^2}\left(\mathbf{Y}-\mu\mathbf{1}-Z\boldsymbol{\tau}\right)'\left(\mathbf{Y}-\mu\mathbf{1}-Z\boldsymbol{\tau}\right)\right\}\times \left(2\pi\sigma_\tau^2\right)^{-\frac{1}{2}k}\exp\left\{-\frac{1}{2\sigma_\tau^2}\boldsymbol{\tau}'\boldsymbol{\tau}\right\}, \quad (1)$$

where $\tilde{n} = \sum_{i=1}^{k} n_i$, $\boldsymbol{\tau}' = \begin{bmatrix}\tau_1 & \tau_2 & \ldots & \tau_k\end{bmatrix}$, $\mu\mathbf{1} = [\mu\ \ \mu\ \cdots\ \mu]'$,

$$Z'_{\tilde{n}\times k} = \begin{bmatrix} 1_1 & 1_2 & \cdots & 1_{n_1} & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1_1 & 1_2 & \cdots & 1_{n_2} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 1_1 & 1_2 & \cdots & 1_{n_k} \end{bmatrix}$$

and $\mathbf{Y} = \begin{bmatrix} y_{11}\ y_{12}\ \cdots\ y_{1n_1}\ \cdots\ y_{k1}\ y_{k2}\ \cdots\ y_{kn_k}\end{bmatrix}'$.

Now, we already know, from the specification of the random effects model, that $\tau_i \sim N\left(0,\sigma_\tau^2\right)$ with $i = 1,\ 2,\ \ldots,\ k$. Since this is the case we would, therefore, like to define prior distributions for μ, σ_e^2 and σ_τ^2. To simplify the mathematical derivations (since the posterior can then be expressed in hierarchical form) we will define the quantity

$$\tilde{r} = \frac{\sigma_\tau^2}{\sigma_e^2}$$

and then define prior distributions in terms of μ, σ_e^2 and $\tilde{r}$ instead. In order to derive prior distributions we first need to derive the integrated likelihood function, $L\left(\mu,\sigma_e^2,\sigma_\tau^2|\mathbf{Y}\right)$ for the following model:

$$\mathbf{Y} = \mu\mathbf{1} + Z\boldsymbol{\tau} + \mathbf{e},$$

where $\mathbf{e} \sim N\left(\mathbf{0},\sigma_e^2 I_{\tilde{n}}\right)$ and $\boldsymbol{\tau} \sim N\left(\mathbf{0},\sigma_\tau^2 I_k\right)$. This is given by the following:

$$L\left(\mu,\sigma_e^2,\sigma_\tau^2|\mathbf{Y}\right) \propto \left(\sigma_e^2\right)^{-\frac{1}{2}(\tilde{n}-k)}\prod_{i=1}^{k}\left(\frac{1}{n_i\sigma_\tau^2+\sigma_e^2}\right)^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\left[\frac{v_1 m_1}{\sigma_e^2}+\sum_{i=1}^{k}\frac{n_i\left(\bar{y}_{i\bullet}-\mu\right)^2}{n_i\sigma_\tau^2+\sigma_e^2}\right]\right\}. \quad (2)$$

For sake of brevity only the necessary results are shown here. Full derivations with intermediate results can be found in Harvey (2012). Now, if $\tilde{r} = \frac{\sigma_\tau^2}{\sigma_e^2}$ then it follows that:

$$L\left(\mu,\sigma_e^2,\tilde{r}|\mathbf{Y}\right) \propto \left(\sigma_e^2\right)^{-\frac{1}{2}\tilde{n}}\prod_{i=1}^{k}\left(\frac{1}{n_i\tilde{r}+1}\right)^{\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma_e^2}\left[v_1 m_1+\sum_{i=1}^{k}\frac{n_i\left(\bar{y}_{i\bullet}-\mu\right)^2}{n_i\tilde{r}+1}\right]\right\}. \quad (3)$$

In this article, two non-informative priors are compared, namely the Probability-matching and Reference priors. These priors often lead to procedures with good frequentist properties while retaining the Bayesian flavour. 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. An in-depth discussion of the nature and merits of the Reference and Probability-matching priors lies outside the scope of this article, but the interested reader should consult Berger and Bernardo (1992) as well as Datta and Ghosh (1995).

The Probability-Matching prior for the parameters $\left(\mu,\tilde{r},\sigma_e^2\right)$ is given by:

$$P\left(\mu,\tilde{r},\sigma_e^2\right) \propto \frac{1}{\sigma_e^2}\left\{\sum_{i=1}^{k}\frac{n_i^2}{(1+\tilde{r}n_i)^2}-\frac{1}{n}\left(\sum_{i=1}^{k}\frac{n_i}{1+\tilde{r}n_i}\right)^2\right\}^{\frac{1}{2}}. \tag{4}$$

The Reference prior for the parameter groupings $\left(\mu,\tilde{r},\sigma_e^2\right)$, $\left(\tilde{r},\ \mu,\sigma_e^2\right)$ and $\left(\tilde{r},\sigma_e^2,\ \mu\right)$ is given by:

$$P_{R_1}\left(\mu,\tilde{r},\sigma_e^2\right) \propto \frac{1}{\sigma_e^2}\left\{\sum_{i=1}^{k}\frac{n_i^2}{(1+\tilde{r}n_i)^2}-\frac{1}{\tilde{n}}\left(\sum_{i=1}^{k}\frac{n_i}{1+\tilde{r}n_i}\right)^2\right\}^{\frac{1}{2}}. \tag{5}$$

This is coincidentally the same as the Probability-Matching prior and therefore, the Probability-Matching prior is also the Reference prior for these specific parameter orderings.

The Reference prior for the parameter groupings $\left(\mu,\sigma_e^2,\tilde{r}\right)$, $\left(\sigma_e^2,\mu,\tilde{r}\right)$ and $\left(\sigma_e^2,\tilde{r},\mu\right)$ is given by:

$$P_{R_2}\left(\mu,\ \sigma_e^2\ ,\ \tilde{r}\right) \propto \frac{1}{\sigma_e^2}\left\{\sum_{i=1}^{k}\frac{n_i^2}{(1+\tilde{r}n_i)^2}\right\}^{\frac{1}{2}}. \tag{6}$$

It should be evident that if we substitute $n_1 = n_2 = \ldots = n_k = n$ and $\tilde{n} = kn$ in Equations (5) and (6), i.e. assume we have the balanced case and if we transform $\tilde{r}$ back to σ_τ^2, then Equations (5) and (6) become the Jeffreys Independence priors as used in Harvey and van der Merwe (2014).

3.1. Joint posterior distribution for μ , σ_e^2 and $\tilde{r}$

We are now able to examine the distribution of the posterior distribution of μ ,σ_e^2 and $\tilde{r}$. This is based on the previous derivations. From the formulation of the Bayesian model we know the following:

$$p\left(\mu,\ \sigma_e^2\ ,\ \tilde{r}|\ \mathbf{Y}\right) \propto L\left(\mu,\ \sigma_e^2\ ,\ \tilde{r}\mid\mathbf{Y}\right)p\left(\mu,\ \sigma_e^2\ ,\ \tilde{r}\right),$$

where

$$L\left(\mu,\sigma_e^2,\tilde{r}|\mathbf{Y}\right) \propto \left(\sigma_e^2\right)^{-\frac{1}{2}\tilde{n}}\prod_{i=1}^{k}\left(\tfrac{1}{n_i\tilde{r}+1}\right)^{\frac{1}{2}}\exp\left\{-\tfrac{1}{2\sigma_e^2}\left[v_1m_1+\sum_{i=1}^{k}\tfrac{n_i(\bar{y}_{i\bullet}-\mu)^2}{n_i\tilde{r}+1}\right]\right\}.$$

If we use the Probability-Matching prior as defined by Equation (4), which is the same as the Reference prior for the first ordering of parameters as described in Equation (5), then the joint posterior distribution is given by:

$$P_{R_1}\left(\mu,\tilde{r},\sigma_e^2|\mathbf{Y}\right) \propto \left(\frac{1}{\sigma_e^2}\right)^{\frac{1}{2}(\tilde{n}+2)}\prod_{i=1}^{k}\left(\frac{1}{n_i\tilde{r}+1}\right)^{\frac{1}{2}}\times\exp\left\{-\frac{1}{2\sigma_e^2}\left[v_1m_1+\sum_{i=1}^{k}\frac{n_i(\bar{y}_{i\bullet}-\mu)^2}{n_i\tilde{r}+1}\right]\right\}$$
$$\times\left\{\sum_{i=1}^{k}\frac{n_i^2}{(1+\tilde{r}n_i)^2}-\frac{1}{n}\left(\sum_{i=1}^{k}\frac{n_i}{1+\tilde{r}n_i}\right)^2\right\}^{\frac{1}{2}}, \tag{7}$$

where $\nu_1 m_1 = SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\bullet})^2$.

From Equation (7) it follows that the joint posterior can be expressed hierarchically as

$$P_{R_1}\left(\mu, \tilde{r}, \sigma_e^2 | \mathbf{Y}\right) = p\left(\mu | \mathbf{Y}, \tilde{r}, \sigma_e^2\right) \times p\left(\sigma_e^2 | \tilde{r}, \mathbf{Y}\right) \times p(\tilde{r} | \mathbf{Y}),$$

where

$$\mu | \mathbf{Y}, \tilde{r}, \sigma_e^2 \sim N\left(\hat{\mu}, \; \sigma_e^2 \left(\sum_{i=1}^{k} \frac{n_i}{1+\tilde{r}n_i}\right)^{-1}\right) \tag{8}$$

and

$$\hat{\mu} = \frac{\sum_{i=1}^{k} \bar{y}_{i\bullet} \frac{n_i}{1+\tilde{r}n_i}}{\sum_{i=1}^{k} \frac{n_i}{1+\tilde{r}n_i}}.$$

In addition,

$$P_{R_1}\left(\sigma_e^2 | \tilde{r}, \mathbf{Y}\right) = K_1 \left(\frac{1}{\sigma_e^2}\right)^{\frac{1}{2}(\tilde{n}+1)} \exp\left\{-\frac{1}{2\sigma_e^2}\left[\nu_1 m_1 + \sum_{i=1}^{k} \frac{n_i(\bar{y}_{i\bullet} - \hat{\mu})^2}{n_i \tilde{r} + 1}\right]\right\}, \tag{9}$$

which is an inverse Gamma distribution. Furthermore, we know that

$$K_1 = \left\{\frac{1}{2}\left[\nu_1 m_1 + \sum_{i=1}^{k} \frac{n_i(\bar{y}_{i\bullet} - \hat{\mu})^2}{n_i \tilde{r} + 1}\right]\right\}^{\frac{1}{2}(\tilde{n}-1)}$$

and

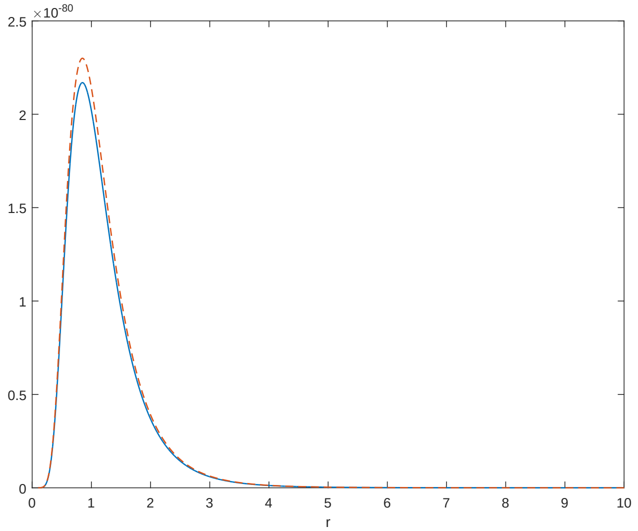
$$P_{R_1}(\tilde{r}|\mathbf{Y}) \propto \prod_{i=1}^{k} \left(\frac{1}{n_i \tilde{r}+1}\right)^{\frac{1}{2}} \times \left(\sum_{i=1}^{k} \frac{n_i}{1+\tilde{r}n_i}\right)^{-\frac{1}{2}} \times \left\{\sum_{i=1}^{k} \frac{n_i^2}{(1+\tilde{r}n_i)^2} - \frac{1}{n}\left(\sum_{i=1}^{k} \frac{n_i}{1+\tilde{r}n_i}\right)^2\right\}^{\frac{1}{2}} \times \left[\nu_1 m_1 + \sum_{i=1}^{k} \frac{n_i(\bar{y}_i - \hat{\mu})^2}{n_i \tilde{r}+1}\right]^{-\frac{1}{2}(n-1)}. \tag{10}$$

If we use the alternate ordering for parameters as described in Equation (6) we find that the joint posterior distribution has the same hierarchical structure, except that

$$P_{R_2}(\tilde{r}|\mathbf{Y}) \propto \prod_{i=1}^{k} \left(\frac{1}{n_i \tilde{r}+1}\right)^{\frac{1}{2}} \times \left(\sum_{i=1}^{k} \frac{n_i}{1+\tilde{r}n_i}\right)^{-\frac{1}{2}} \times \left(\sum_{i=1}^{k} \frac{n_i^2}{(1+\tilde{r}n_i)^2}\right)^{\frac{1}{2}} \times \left[\nu_1 m_1 + \sum_{i=1}^{k} \frac{n_i(\bar{y}_{i\bullet} - \hat{\mu})^2}{n_i \tilde{r}+1}\right]^{-\frac{1}{2}(n-1)}, \tag{11}$$

where $0 < \tilde{r} < \infty$.

Figure 1 depicts these two posterior distributions. Furthermore, the posterior distribution of $\mu + \tau_i$ (the mean of the i-th worker in the data set) given σ_e^2 and $\tilde{r}$ is normal with the following mean and variance:



*For further details see Harvey (2012) and van der Merwe, Pretorius and Meyer (2006).

Figure 1: Two Reference priors. Reference prior 1 (solid) and Reference prior 2 (dashed).

$$E\left\{(\mu+\ \tau_i)\mid \mathbf{Y}\ ,\ \sigma_e^2\ ,\ \tilde{r}\right\}=\ \tfrac{\tilde{r}\ n_i}{1+\tilde{r}n_i}\bar{y}_{i\bullet}+\ \tfrac{1}{1+\tilde{r}n_i}\hat{\mu}$$

and

$$Var\left\{(\mu+\ \tau_i)\mid \mathbf{Y}\ ,\ \sigma_e^2\ ,\ \tilde{r}\right\}=\ \tfrac{\sigma_e^2}{1+\tilde{r}n_i}\ \left\{\tilde{r}+\ \tfrac{1}{1+\tilde{r}n_i}\ \left(\sum_{i=1}^{k}\tfrac{n_i}{1+\tilde{r}n_i}\right)^{-1}\right\}.$$

From the above it follows that

$$\mu+\ \tau_i+\ \tfrac{1}{2}\sigma_e^2\mid \mathbf{Y}\ ,\ \sigma_e^2\ ,\ \tilde{r}$$

is normally distributed with mean

$$E\left\{\left(\mu+\ \tau_i\ +\ \frac{1}{2}\sigma_e^2\right)\mid \mathbf{Y}\ ,\ \sigma_e^2\ ,\ \tilde{r}\right\}=\ \frac{\tilde{r}\ n_i}{1+\tilde{r}n_i}\bar{y}_{i\bullet}+\ \frac{1}{1+\tilde{r}n_i}\hat{\mu}+\ \frac{1}{2}\sigma_e^2 \tag{12}$$

and variance

$$Var\left\{\left(\mu+\ \tau_i+\ \frac{1}{2}\sigma_e^2\right)\mid \mathbf{Y}\ ,\ \sigma_e^2\ ,\ \tilde{r}\right\}=\frac{\sigma_e^2}{1+\tilde{r}n_i}\ \left\{\tilde{r}+\ \frac{1}{1+\tilde{r}n_i}\ \left(\sum_{i=1}^{k}\frac{n_i}{1+\tilde{r}n_i}\right)^{-1}\right\}. \tag{13}$$

Now, we are interested in the posterior distribution of

$$\exp\left(\mu+\ \tau_i+\ {}^{\sigma_e^2}/_2\right), \tag{14}$$

for $i=1,\ 2,\ \ldots,\ k$, in other words, for each worker.

Given σ_e^2 and $\tilde{r}$ we can now simulate from Equation (14) by simulating from a normal distribution with mean and variance specified by Equations (12) and (13) respectively. Using these results we are able to simulate and test hypotheses for individuals (e.g. individual workers). The results will be presented in later sections.

3.2. Procedure for simulation study

The purpose of this article is to describe the behaviour of the various prior distributions in the setting described earlier. Although detailed descriptions will be given in relevant sections, here we offer a broad description of the simulation of σ_e^2 and $\tilde{r}$ values from the distributions obtained in previous sections, including the final simulation of μ, which will ultimately enable the simulation of quantities such as those defined by Equation (14). The simulation procedure can be described broadly as follows:

1. Simulate a value of $\tilde{r}$ using either Equation (10) or (11), based on the choice of prior distribution. Since neither Equation (10) nor Equation (11) is a known distribution and cannot be solved in closed form the use of the Rejection method as described in Rice (1995) will be used.

2. Each value of $\tilde{r}$ simulated in the previous step will then be substituted into Equation (9) to simulate a value of σ_e^2. In this case the distribution is of a known form, i.e. an Inverse Gamma distribution, and therefore we can simulate σ_e^2 by making use of the fact that:

$$\left\{\frac{1}{\sigma_e^2}\left[v_1 m_1 + \sum_{i=1}^{k} \frac{n_i(\bar{y}_{i\bullet} - \hat{\mu})^2}{n_i\tilde{r} + 1}\right]\right\} \sim \chi^2_{n-1}.$$

It follows that a simulated value of σ_e^2 can be obtained from the equation

$$\frac{1}{\chi^2_{n-1}}\left\{\left[v_1 m_1 + \sum_{i=1}^{k} \frac{n_i(\bar{y}_{i\bullet} - \hat{\mu})^2}{n_i\tilde{r} + 1}\right]\right\} = \sigma_e^2 .$$

Using the values of σ_e^2 and $\tilde{r}$ simulated in the previous steps we can simulate values of μ (if desired) from Equation (8). All the desired quantities are based on these variables in some manner.

4. An upper confidence bound and test for the overall mean exposure

In Krishnamoorthy and Guo (2005) one of the primary interests is testing the hypothesis of whether the occupational exposure in an individual (discussed previously in Equations (12) and (13)) or group of workers exceeds a pre-specified threshold. If we consider making inferences about the total group, we are interested in the distribution of the overall mean exposure, which for this unbalanced case can be represented as:

$$\mu_x = \exp\left\{\mu + \frac{\sigma_e^2}{2}(\tilde{r} + 1)\right\} = e^{\theta}. \tag{15}$$

Now we know from Equation (8) that

$$\mu|\mathbf{Y}, \tilde{r}, \sigma_e^2 \sim N\left(\hat{\mu}\,,\ \sigma_e^2\left(\sum_{i=1}^{k} \frac{n_i}{1+\tilde{r}n_i}\right)^{-1}\right)$$

and therefore,

$$\theta = \mu + \tfrac{\sigma_e^2}{2}(\tilde{r} + 1)\,|\mathbf{Y},\, \tilde{r}, \sigma_e^2 \sim N(E(\theta)\,,\, Var(\theta))$$

is normally distributed with the following mean and variance:

$$E(\theta) = E\left\{\mu + \frac{\sigma_e^2}{2}(\tilde{r} + 1)\,|\mathbf{Y},\, \tilde{r}, \sigma_e^2\right\} = \hat{\mu} + \frac{\sigma_e^2}{2}(\tilde{r} + 1), \tag{16}$$

$$Var(\theta) = Var\left\{\mu + \frac{\sigma_e^2}{2}(\tilde{r} + 1)\,|\mathbf{Y},\, \tilde{r}, \sigma_e^2\right\} = \sigma_e^2\left(\sum_{i=1}^{k} \frac{n_i}{1 + \tilde{r}n_i}\right)^{-1}. \tag{17}$$

Thus, given $\tilde{r}$ and σ_e^2 we simulate θ from a normal distribution with mean and variance defined by Equations (16) and (17) and substitute this into Equation (15). We then repeat this process l $(= 100\,000)$ times.

Additionally Krishnamoorthy and Guo (2005) also simulated the following parameter (and inference regarding this variable will be made using the Bayesian methodology developed previously):

$$\xi = \mu + Z_{1-A}\sigma_\tau + \tfrac{1}{2}\sigma_e^2,$$

where A is a suitably chosen parameter between 0 and 1 and Z_{1-A} denotes the $100(1-A)$-th percentile of the standard normal distribution. Using a specific value of OEL the following hypothesis can be tested:

$$H_0:\ \xi\ \geq ln(OEL),$$

against the alternative hypothesis

$$H_1:\ \xi\ < ln(OEL).$$

For example, if our choice of A is 0.05 then essentially we are testing (one-sided) whether at least 5% of the workers have mean exposure levels in excess of the chosen OEL. In practice the OEL is chosen to be a clinically relevant value. The specific choice of OEL is not the primary concern of this research and has previously been discussed.

In order to replicate the methodology of Krishnamoorthy and Guo (2005) from a Bayesian perspective the simulation study was performed for a range of values of OEL and A.

Let $\xi = \mu + \sigma_e(Z_{1-A}\tilde{r}^{1/2} + \sigma_e/2)$. This representation is the same as the earlier definition of ξ, re-written for convenience in terms of σ_e and $\tilde{r}$ instead of σ_e^2 and σ_τ. We know from Equation (8) that $\xi \mid \mathbf{Y},\, \tilde{r}, \sigma_e^2$ is normally distributed with:

$$E\left\{\xi \mid \mathbf{Y},\, \tilde{r}, \sigma_e^2\right\} \sim \hat{\mu} + \sigma_e(Z_{1-A}\tilde{r}^{1/2} + \tfrac{\sigma_e}{2})$$

and

$$Var\left\{\xi|\mathbf{Y}, \tilde{r}, \sigma_e^2\right\} \sim \sigma_e^2\left(\sum_{i=1}^{k} \tfrac{n_i}{1+\tilde{r}n_i}\right)^{-1}.$$

This result is used for simulating the necessary observations for ξ necessary for hypothesis testing. A value of $\tilde{r}$ is simulated from either Equation (10) or Equation (11). This is then used to simulate a value of σ_e^2 from Equation (9) (as previously described). Using these simulated observations we can take the necessary square roots and simulate from the above-mentioned normal distribution. The simulations were performed for several choices of OEL (= [130; 140; 150; 160; 170; 180]) and for several choices of A (= [0.1; 0.05; 0.025; 0.001]), as was done in Harvey and van der Merwe (2014).

5. Results from the simulation study

Using the methodology derived previously a simulation study was conducted to simulate 100 000 observations for each particular type of analysis. Using the unbalanced data provided in Table 2 we were able to simulate observations relating to occupational exposure in the workplace. Since two Reference priors were derived the simulations were repeated for each of these Reference priors. The results are presented in the following sections.

5.1. Results: Individual worker means

As mentioned it was possible to simulate observations from the posterior distribution for each of the 15 workers, using both Reference priors. Selected results will be shown here for the purposes of illustration. Simulation results for all 15 workers can be found in Harvey (2012).

From Tables 3 and 4 we can see that the results from the first and second Reference priors are comparable, with no large differences between the various Reference priors. Reference prior 2 results in slightly wider confidence intervals, which is easily attributed to the shape of the distribution, as seen in Figure 1. The effect of "unbalancing" however, is observed. For example, workers 4 and 11 both had comparable mean exposure levels (arithmetic means of 55.98 and 55.7 respectively from the data set), but were at the two extremes (in this hypothetical data set) with regards to unbalancing (worker 4 had 10 exposure observations, while worker 11 only had 6 observations). The results for worker 15 agree with those found in the balanced case, as found in Harvey (2012) and Harvey and van der Merwe (2014). It is interesting to note that in both cases the probability of exceeding an OEL of 130 was less than 0.001 (based on 100 000 simulated observations). It thus appears that the Bayesian methodology is able to model the effect of unbalancing and account for additional uncertainty introduced by the unbalancing, particularly at a worker-specific level.

Table 3: Simulation summary results: Reference prior 1.

Worker	$P(\mu_{x_i} > 130)$	90% CI		95% CI		Mean	Median	Mode
		Low	High	Low	High			
Worker 4	<0.001	52.25	71.81	50.67	74.07	61.45	61.14	59.75
Worker 11	<0.001	51.78	75.94	49.90	78.86	63.10	62.68	61.25
Worker 15	0.0715	98.32	132.21	95.56	136.12	114.41	113.94	115.25

Table 4: Simulation summary results: Reference prior 2.

Worker	$P(\mu_{x_i} > 130)$	90% CI		95% CI		Mean	Median	Mode
		Low	High	Low	High			
Worker 4	<0.001	52.26	71.81	50.68	74.13	61.47	61.13	59.75
Worker 11	<0.001	51.78	75.99	49.98	78.91	63.14	62.70	61.75
Worker 15	0.0721	98.30	132.21	95.63	136.08	114.40	113.94	113.25

5.2. Results: Overall mean exposure

The next results relate to the overall mean exposure, that is the exposure of the group of 15 workers as a whole. The results in Table 5 and Figures 2 and 3 were obtained for the two Reference prior distributions (the relevant information for each histogram is displayed in the Table 5). It is evident that for overall mean exposure the results from Reference prior 1 and Reference prior 2 are similar.

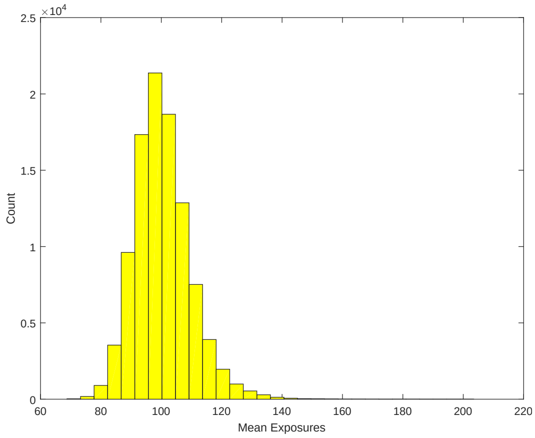


Figure 2: Overall mean exposure: Reference prior 1.

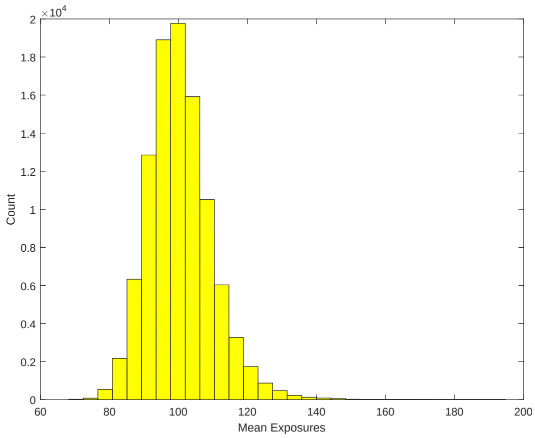


Figure 3: Overall mean exposure: Reference prior 2.

The results from Table 5 as well as Figures 2 and 3 are based on 100 000 simulations. We see very little difference between the two Reference prior distributions. For comparison, refer to the results in Harvey and van der Merwe (2014) and Harvey (2012).

Table 5: Simulation summary results of overall mean exposure.

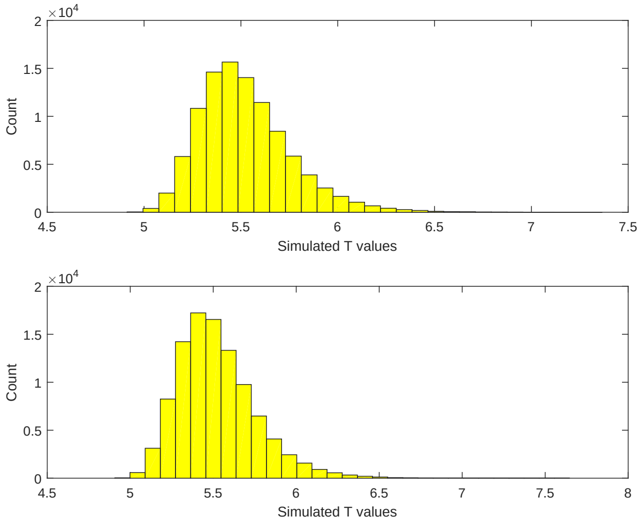
All Workers	$P(\mu_{overall} > 130)$	90% CI		95% CI		Mean	Median	Mode
		Low	High	Low	High			
Reference prior 1	0.0064	87.07	116.70	84.62	120.99	100.45	99.68	98.75
Reference prior 2	0.0064	87.02	116.81	84.77	121.21	100.46	99.65	98.75

5.3. Results: Hypothesis testing

Lastly, Krishnamoorthy and Guo (2005) tested hypotheses regarding the group of workers using the following parameter:

$$\xi = \mu + Z_{1-A}\sigma_\tau + \tfrac{1}{2}\sigma_e^2,$$

where A is a suitably chosen parameter between 0 and 1 and Z_{1-A} denotes the $100(1-A)$-th percentile of the standard normal distribution. Recall from previous derivations that this is equivalent to $\xi = \mu + \sigma_e(Z_{1-A}\tilde{r}^{1/2} + \sigma_e/2)$. Several different values of A were chosen in addition to several different OEL limits. The results in Figure 4 are once again produced for both Reference prior distributions. From Figure 4 we can see that only 0.1% or more of workers had occupational exposure



$100(1-\alpha)$*-th percentile = 5.9577 (Reference prior 1)*
$100(1-\alpha)$*-th percentile = 5.9547 (Reference prior 2)*

Figure 4: Hypothesis testing for Reference priors 1 (above) 2 (below): $A = 0.001$.

levels (in log units) in excess of 5.9577 and 5.9547 (99.9th percentile) respectively for Reference priors 1 and 2.

5.4. Application to real data

The previous results were based on data that was simulated, according to the data used in Krishnamoorthy and Mathew (2002) for the balanced case and again in Krishnamoorthy and Guo (2005) for the unbalanced case. Data had to be simulated since the original data was not available, but only certain descriptive results. The use of this simulated data still allowed sufficient comparison between the Bayesian and frequentist approaches in previous sections.

In Krishnamoorthy and Mathew (2009) a data set containing styrene exposures (measured in mg/m^3) from laminators at a boat manufacturing plant was used. Krishnamoorthy and Mathew (2009) stated that the interest was in estimating the proportion of workers for whom the exposure exceeds a chosen OEL. The chosen OEL in this case was 213 mg/m^3, i.e. 5.3613 units on the log scale (justification for this choice of OEL was given in Lyles et al. (1997a) and Lyles et al. (1997b)). The unbalanced data given in Table 6 was obtained from the original balanced data set. It is lognormally distributed and can be analysed using a one-way random effects model.

Table 6: Styrene exposures on laminators at a boat manufacturing plant (log scale).

	Observations		
Worker	**1**	**2**	**3**
1	3.071	3.871	
2	4.319	4.396	5.045
3	5.221	4.876	5.058
4	4.572	5.116	5.578
5	5.351	3.925	4.217
6	5.889	4.893	4.775
7	5.192	4.457	5.097
8		4.807	5.345
9		5.271	5.454
10	5.188	4.499	5.340
11	5.970	5.660	5.175
12	5.619	1.843	5.545
13	4.200	5.294	4.945

To obtain this unbalanced data set observations were randomly removed from the original balanced data set. As previously mentioned, this practice is accepted in literature (refer to Liao et al. (2005) for examples where a similar method was used to obtain unbalanced data). The third observation from worker 1 (2.965) and the first observation from workers 8 (4.477) and 9 (5.060) were removed. The data was then analysed according to the Bayesian methodology developed previously, using both Reference prior distributions.

The Bayesian methodology allowed modelling of exposures for specific workers (as opposed to unknown future workers), as in Krishnamoorthy and Guo (2005). For example, for worker 1 the probability that his/her mean exposure exceeded the OEL of 213 was 0.05568 and 0.05576 for Reference prior 1 and 2 respectively.

The overall mean exposure was also modelled similarly to previous sections and the probability that the mean exposure of the entire group of workers exceeded the OEL of 213 was 0.22552 for Reference prior 1 and 0.22542 for Reference prior 2.

With regards to hypothesis testing, it was found that the probability that an exposure measurement exceeds 7.1827 (on the log scale) is less than 0.1%, with both prior distributions yielding similar results.

6. Conclusion

In this article the usefulness of the Bayesian methodology to the proposed setting of occupational exposure data was examined, specifically for the case where there are an unequal number of observations for each observational unit (worker). The one-way random effects model was adapted to account for unbalanced data using the chosen prior distributions. One of the advantages of the Bayesian model is that it is able to model results for individual workers and not simply for an unknown future worker. Only a few non-informative prior distributions have been derived in this article, but they do by no means represent an exhaustive list. The derivation and comparison of all possible prior distributions was not an objective of this research. However, the derivation and application of other non-informative priors could be used to refine the analysis and improve performance. Ultimately, if subjective prior information is available this could lead to significant improvements in prediction of future exposure for individual workers as well as for groups of workers.

References

BERGER, J. O. AND BERNARDO, J. M. (1992). *On the Development of the Reference Prior Method*. Claredon Press: Oxford, pp. 35–60.

CHU, H., GANGE, S., LI, X., HOOVER, D., LIU, C., CHMIEL, J., AND JACOBSON, L. (2010). The effect of HAART on HIV RNA trajectory among treatment naïve men and women: a segmental Bernoulli/lognormal random effects model with left censoring. *Epidemiology*, **21** (4), 25–34.

CHU, H., MOULTON, L., MACK, W., PASSARO, D., BARROSO, P., AND MUÑOZ, A. (2005). Correlating two continuous variables subject to detection limits in the context of mixture distributions. *Journal of the Royal Statistical Society, Series C*, **54**, 831–845.

DATTA, G. S. AND GHOSH, J. K. (1995). On priors providing frequentist validity of Bayesian inference. *Biometrika*, **82**, 37–45.

FINKELSTEIN, M. AND VERMA, D. (2001). Exposure estimation in the presence of nondetectable values: another look. *American Industrial Hygiene Association Journal*, **62**, 195–198.

HARVEY, J. (2012). *Bayesian Inference for the Lognormal Distribution*. Unpublished Ph.D. dissertation, Bloemfontein, South Africa.

HARVEY, J. AND VAN DER MERWE, A. (2014). Modelling occupational exposure using a random effects model: a Bayesian approach. *South African Statistical Journal*, **48**, 61–71.

HEERDERIK, D. AND HURLEY, F. (1994). Occupational exposure assessment: Investigating why exposure measurements vary. *Applied Occupational and Environmental Hygiene*, **9**, 71.

KRISHNAMOORTHY, K. AND GUO, H. (2005). Assessing occupational exposure via the one-way random effects model with unbalanced data. *Journal of Statistical Planning and Inference, Elsevier*, **128**, 219–229.

KRISHNAMOORTHY, K. AND MATHEW, T. (2002). Assessing occupational exposure via the one-way random effects model with balanced data. *Journal of Agricultural, Biological and Environmental Statistics*, **3**, 440–451.

KRISHNAMOORTHY, K. AND MATHEW, T. (2009). *Statistical Tolerance Regions: Theory, Applications and Computation.* John Wiley and Sons: New York.

LEWSEY, J., GARDINER, W., AND GETTINBY, G. (1997). A study of simple unbalanced factorial designs that use type II and type III sums of squares. *Communications in Statistics - Simulation and Computation*, **26** (4), 1315–1328.

LI, X., CHU, H., AND GALLANT, J. (2006). Bimodal virologic response to antiretroviral therapy for HIV infection: an application using a mixture model with left censoring. *Journal of Epidemiology and Community Health*, **60**, 811–818.

LIAO, C., LIN, T., AND IYER, H. (2005). One- and two-sided tolerance intervals for general balanced mixed models and unbalanced one-way random models. *Technometrics*, **47** (3), 323–335.

LITTELL, R., STOUP, W. W., AND FRUEND, R. J. (2002). *SAS for Linear Models. 4th edition.* SAS Institute Inc.: Cary, North Caroline, USA.

LYLES, R. H., KUPPER, L. L., AND RAPPAPORT, S. M. (1997a). Assessing regulatory compliance of occupational exposures via the balanced one-way random effects ANOVA model. *Journal of Agricultural, Biological, and Environmental Statistics*, **2**, 64–86.

LYLES, R. H., KUPPER, L. L., AND RAPPAPORT, S. M. (1997b). Lognormal distribution based exposure assessment method for unbalanced data. *Annals of Occupational Hygiene*, **41**, 63–76.

LYLES, R. H., WILLIAMS, J., AND CHUACHOOWONG, R. (2001). Correlating two viral load assays with known detection limits. *Biometrics*, **57**, 1238–1244.

RAPPAPORT, S. M., KROMHOUT, H., AND SYMANSKI, E. (1993). Variation of exposure between workers in homogeneous exposure groups. *American Industrial Hygiene Association Journal*, **54**, 654–662.

RICE, A. (1995). *Mathematical Statistics and Data Analysis*. Duxbury Press: Belmont, California.

SCHAFER, J. L. AND GRAHAM, J. W. (2002). Missing data: our view of the state of the art. *Psychological Methods*, **7** (2), 147–177.

VAN DER MERWE, A. J., PRETORIUS, A., AND MEYER, J. (2006). Bayesian tolerance intervals for the unbalanced one-way random effects model. *Journal of Quality Technology*, **38**, 280–293.

Manuscript received, 2014-03-18, revised, 2015-06-30, accepted, 2015-07-26.