### ORAU Team
**Dose Reconstruction Project for NIOSH**

Technical Information Bulletin – Bayesian Methods for Estimation of Unmonitored Y-12 External Penetrating Doses with a Time-Dependent Lognormal Model

<table>
<thead>
<tr>
<th>Document Number:</th>
<th>ORAUT-OTIB-0015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effective Date:</td>
<td>09/09/2004</td>
</tr>
<tr>
<td>Revision No.:</td>
<td>00</td>
</tr>
<tr>
<td>Controlled Copy No.:</td>
<td>______</td>
</tr>
<tr>
<td>Page 1 of 11</td>
<td></td>
</tr>
</tbody>
</table>

**Subject Expert:** Peter G. Groer

**Approval:** Signature on File Date: 09/09/2004
Judson L. Kenoyer, Task 3 Manager

**Concurrence:** Signature on File Date: 09/09/2004
Richard E. Toohey, Project Director

**Approval:** Signature on File Date: 09/09/2004
James W. Neton, Associate Director for Science

**Supersedes:** None
## RECORD OF ISSUE/REVISIONS

<table>
<thead>
<tr>
<th>ISSUE AUTHORIZATION DATE</th>
<th>EFFECTIVE DATE</th>
<th>REV. NO.</th>
<th>DESCRIPTION</th>
</tr>
</thead>
</table>
ACRONYMS AND ABBREVIATIONS

LOD  limit of detection
MLE  maximum likelihood estimate
mrem millirem
1.0 INTRODUCTION

This document describes Bayesian procedures in general, discusses a Bayesian regression and prediction analysis of external penetrating dose data for Y-12 with a time-dependent lognormal model, and an estimated a scale factor for imputation of unmonitored doses for an individual. The original data set (Watkins et al. 2004) for the regression and prediction analysis incorporated data from 1950 to 1970. This analysis only used the data from 1956 to 1961. For the regression analysis, all doses less than 15 mrem were replaced by 15 mrem. For the estimation of the scale factor, all observations less than or equal to 30 mrem [i.e., the reported limit of detection (LOD)] were treated as left-censored observations. This required some modification of the likelihood construction.

Bayesian methods work for all sizes of data sets, large or small. For the large data set (Watkins et al. 2004) considered here, maximum likelihood estimates (MLEs; Frome and Watkins 2004) are appropriate and produce numerically practically identical results. This is demonstrated by a comparison of Bayesian estimates and the MLEs for the regression analysis and the scaling procedure.

All analyses assumed that the doses in the original data set (Watkins et al. 2004) were precise with the exception of the doses that were treated as left censored.

2.0 REGRESSION AND PREDICTION ANALYSIS

2.1 APPROACH

The analysts estimated the parameters \( \log(\beta) \) and \( \mu \) for the following time-dependent lognormal model using Bayesian methods:

\[
f(y | \beta, \mu, \sigma) = \frac{1}{\sqrt{2\pi} \sigma y} e^{-\frac{(\log(y) - \log(\beta) - \mu)^2}{2\sigma^2}}
\]

The probability density shown above was used to construct the likelihood \( p(D | \log(\beta), \mu, \sigma) \) (\( D \) means data used). According to Bayes’ theorem the joint posterior density \( p(\log(\beta), \mu, \sigma | D) \) of the model parameters is proportional to the likelihood if uniform priors are used:

\[
p(\log(\beta), \mu, \sigma | D) \propto p(D | \log(\beta), \mu, \sigma)
\]

Because the parameters of greatest interest are \( \log(\beta) \) and \( \mu \), the joint density \( p(\log(\beta), \mu, \sigma | D) \) was first integrated over \( \sigma \). This integration can be done analytically if no left-censored observations are present, and it gives the following expression for the joint posterior density \( p(\log(\beta), \mu | D) \) for \( \log(\beta) \) and \( \mu \):

\[
p\left(\log(\beta), \mu | D\right) = \left(\sum_{i=1}^{l} \sum_{k=1}^{40} \log(y_{ik}) - \log(\beta) + \mu t_k \right)\left(\sum_{k=1}^{l} t_k - 1\right)^{-1/2}
\]

Where \( k \) labels the forty quarters from 1956 to 1965, and \( l_k \) is the number of observations \( y_{ik} \) in each quarter. The time \( t_k \) is measured in quarter years starting from 1956.
Prediction of unmonitored doses is accomplished with the predictive density. This is a strictly Bayesian concept that permits quantitative consideration of parameter uncertainty when making predictions. For the model under consideration the predictive density $p(y_f | D)$ for a unmonitored observation $y_f$ is given by:

$$p(y_f | D) = \int \int f(y_f | \beta, \mu, \sigma) p(\beta, \mu, \sigma | D) d\beta d\mu$$

The integration is over all model parameters $\theta = \{\beta, \mu, \sigma\}$. Section 2.2 shows sample plots of $p(y_f | D)$ for three different years. Point estimates based on maximum likelihood are compared with posterior expectations obtained by the Bayesian approach.

### 2.2 RESULTS

Figure 1 shows the joint posterior density $p(\log(\beta), \mu | D)$ after integration over $\sigma$, and Figures 2 and 3 show the marginal posterior densities of the model parameters $\log(\beta)$ and $\mu$, respectively.

![Figure 1. Joint posterior density $p(\log(\beta), \mu | D)$ (4.1 < $\log(\beta)$ < 4.35, 0.07 < $\mu$ < 0.12).](image)

The posterior expectations ($E$) and standard deviations ($SD$) of the parameters are:

$$E(\log(\beta)) = 4.2159, SD(\log(\beta)) = 0.02671$$
$$E(\mu) = 0.09281, SD(\mu) = 0.00453$$

This agrees with the estimates obtained with maximum likelihood methods in Frome and Watkins (2004), which are:

$$E(\log(\beta)) = 4.2159, SD(\log(\beta)) = 0.02668$$
$$E(\mu) = 0.09281, SD(\mu) = 0.00452$$
A detailed comparison for 1956 (i.e., \( t = 0 \)) gave the following results:

\[
E(\log(y_t) | D) = 4.2174 \text{ (which agrees with } E(\log(\beta)) \text{ from above because } t = 0.)
\]

\[
SD(\log(y_t) | D) = 0.9867
\]

\[
E(y_t | D) = 110.32
\]

\[
SD(y_t | D) = 142.37
\]

These results imply a geometric mean of \( GM(y_t | D) = 67.57 \) and a geometric standard deviation of \( GSD(y_t | D) = 2.69 \).

The corresponding estimates based on maximum likelihood techniques (Frome and Watkins 2004) are:

\[
E(\log(y_t) | D) = 4.2159 \text{ (which agrees with } E(\log(\beta)) \text{ from above because } t = 0)
\]

\[
SD(\log(y_t) | D) = 0.9916
\]
\[ E(y_f | D) = 110.78 \]
\[ SD(y_f | D) = 143.31 \]

Again, these results imply a geometric mean \( GM(y_f | D) = 67.75 \) [sic] and a geometric standard deviation \( GSD(y_f | D) = 2.69 \).

Figure 4 shows that the exact and the approximate lognormal predictive (dashed) densities are practically indistinguishable.

The right tail probability \( p(y_f > 346.2 | D) = 0.05 \) also agrees with the tail probability calculated with the approximate predictive density.

Figures 5 and 6 show two additional plots of \( p(y_f | D) \) for 1951 and 1957, respectively. Some point estimates for the same years are:

<table>
<thead>
<tr>
<th>Year</th>
<th>( E(\log(y_f)) )</th>
<th>( SD(\log(y_f)) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1951</td>
<td>4.6812</td>
<td>0.9928</td>
</tr>
<tr>
<td>1957</td>
<td>4.1243</td>
<td>0.9920</td>
</tr>
</tbody>
</table>

The corresponding MLEs are:

<table>
<thead>
<tr>
<th>Year</th>
<th>( MLE(\log(y_f)) )</th>
<th>( SD(\log(y_f)) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1951</td>
<td>4.6800</td>
<td>0.9914</td>
</tr>
<tr>
<td>1957</td>
<td>4.1231</td>
<td>0.9918</td>
</tr>
</tbody>
</table>
2.3 CONCLUSIONS

For this large data set, without censored observations, the difference between exact predictive densities and their lognormal approximations is negligible (see Figure 4). For the data set with left censoring at the reported LOD of 30 mrem, some difference between exact and approximate densities would be expected, dependent on the number of censored observations. The magnitude of the difference must be evaluated numerically because the likelihood terms for observations censored at, for example, $d$, shown below prevent symbolic integration over the parameter $\sigma$:

$$\int_{0}^{d} \frac{1}{\sqrt{2 \pi \sigma x}} e^{-\frac{(\text{Log}[x]-\mu)^2}{2 \sigma^2}} \, dx$$
3.0 BAYESIAN ESTIMATION OF A SCALE FACTOR FOR IMPUTING UNMONITORED DOSES

3.1 APPROACH

For a Y-12 worker, whose potential for exposure before 1961 is judged to be the same as after 1961, it is possible to estimate a scale factor $\varphi$ based on the worker’s post-1961 dose records. This parameter $\varphi$ measures the discrepancy between the worker’s dose and the population dose after 1961 (Frome and Groer 2004).

The Bayesian estimation of $\varphi$, defined above, starts again with a likelihood similar to $L(\varphi|d, \mu, \sigma)$ (Frome and Watkins 2004). The likelihood used for the Bayesian analysis is slightly different because $d_i = 0$ is treated as a left-censored observation whenever it occurs. Using a subscript $c$ (meaning censored) to differentiate it from $L(\varphi|d, \mu, \sigma)$ (Frome and Watkins 2004), the likelihood is now given by the following expression:

$$
L_c(\varphi|d, \mu, \sigma) \propto \prod_{i=1}^{17} \left( 1 - \frac{1}{\sigma_i} e^{-(\log(d_i) - \mu - \varphi)^2} \right) \times F_c(\varphi)
$$

Constants not depending on $\varphi$ have been dropped for brevity, and $i$ indicates all quarters for which $d_i > 0$. For the example presented below, $F_c(\varphi)$ is a product of three terms of the generic form:

$$
f_c(\varphi | \mu, \sigma) = \text{Erfc} \left( \frac{\mu + \varphi - \log(30)}{\sqrt{2}\sigma} \right)
$$

where $\mu_i$ and $\sigma_i$ are replaced by the corresponding numerical estimates for the quarters for which $d_i = 0$. The following integral defines the right-hand side of $f_c$ above:

$$
\int_0^{30} \frac{1}{\sigma_i d_i} e^{-(\log(d_i) - \mu - \varphi)^2} \, dd_i = \text{Erfc} \left( \frac{\mu + \varphi - \log(30)}{\sqrt{2}\sigma} \right)
$$

The upper limit of the integral (30 mrem) is the reported LOD mentioned above.

The posterior density of $\varphi$, $p(\varphi|D)$, is proportional to $L_c(\varphi|d, \mu, \sigma)$ given above ($D$ means the dose data of the individual in the example considered here).

3.2 RESULTS

Figure 7 shows a graph of the posterior density of $\varphi$. The expectation $[E(\varphi|D)]$ is 0.492, and the variance $[\text{Var}(\varphi|D)]$ is 0.0369.

A Bayesian analysis treating $d_i = 0$ as left-censored observations was applied to check the accuracy of the approximation of replacing non-detects by the conditional expectation of $y$, $\hat{y}_i$ (Frome and Groer 2004).

The slight difference between $E(\varphi|D)$ and the MLE result (Frome and Watkins 2004) is caused by the two different methods used to deal with non-detects.
Figure 7. Posterior density $p(\phi|D)$. 
REFERENCES

