Bayesian and classical solutions for binomial cytogenetic dosimetry problem

Márcia D. Branco[1], Heleno Bolfarine
Departamento de Estatística, Universidade de São Paulo
Caixa Postal 66281, CEP 05315-970, São Paulo, Brasil

Pilar Iglesias and Reinaldo Boris Arellano-Valle
Departamento de Estadística, Pontificia Universidad Católica de Chile
Casilla 306, Santiago 22, Chile

SUMMARY

The main interest of the cytogenetic dosimetry is the prevision of an unknown radiation dosage based in cytogenetic analysis. In this paper the dosimetry problem is formulated as a linear calibration problem for binary response data. Two approaches are considered for inference on the quantity of interest, which is expressed as a calibration parameter in a discrete response variable situation. One is based on the maximum likelihood approach, which depends on large sample results and the second one is based on a Markov chain Monte Carlo (MCMC) simulation approach using BUGS. Application to a data set obtained from blood cultures exposed in vitro to Cobalt 60 (.^{60}Co) at the Energetic Nuclear Research Center (IPEN - Brasil) is considered.

1 Introduction

The calibration problem can be briefly described as follows. There are two related responses x and y, where x represents the true value of the characteristic of interest and y a variable related to it. The statistical controlled calibration problem considers that the variables x and y are related through a fuctional form as in the usual regression problem. However, the main interest center on estimation of an unknown value of x. At a first stage a collection of values $x_1, x_2, \ldots, x_n$ are fixed and corresponding values $y_1, y_2, \ldots, y_n$ are observed. At a second stage, k replications of the response are observed, $y_{01}, y_{02}, \ldots, y_{0k}$, are observed, corresponding at an unknown value x_0. The interest is on estimating x_0 given $(x_i, y_i), i = 1, 2, \ldots, n$ and $y_{0j}, j = 1, 2, \ldots, k$. The special case where y and x are linearly related and y is normally distributed has been extensively considered. A good exposition of this area is

[1]Correspondence to: Márcia D. Branco, e-mail:mbranco@ime.usp.br

presented in Osborne[25] and Brown [7]. Extensions for Student-t models and more generally, elliptical linear models are presented in Branco et al. [5,6].

In this paper, it is considered that the response variable y is binomial, with parameter p. In this case, a function of p is modelled as a linear function of the independent variable x.

The cytogenetic dosimetry is concerned with the relationship between dosage as some form of exposure to radiation and response, as some measure of genetic aberration. The exposure is typically very dificult to measure and the cytogenetic analysis can be helpful to make inference about the unknown exposure dosage. The cytogenetic dosimetry experiment *in vitro* considers samples of cell cultures exposure to a range of doses levels of a given agent. At each dose level some measure of cell disability is recorded. In the application developed at section 3, the agent is gamma radiation and cell disability is the presence of micronuclei (MN) at the cells. We recall that Cobalt 60 ($.^{60}Co$) is a heavy radioactive isope of the mass number 60 produced in nuclear reactors and used as a source of gamma rays (as for radiotherapy). Therefore, $k_1, k_2, \ldots, k_n$ cells are exposed to fixed dose levels $d_1, d_2, \ldots, d_n$, respectively. The response $y_1, y_2, \ldots, y_n$ are the number of cells with MN, for each dose level. The interest centers on estimating the unknown dose, d_0, to which k_0 new cells were exposed, based in the number of cells with MN, y_0, and a dose-response model. See Madruga et al. [22] for details on the data set.

The model used for describing the dose-response relationship considered in the paper is described in Finney [11], where it is suggested the use of the logit and probit models to study the problem. Two components are considered: the stimuli (radiation, for example) and the response observed in a subject (blood cells, for example). The denomination dose is used to describe the intensity of the stimuli at which the subject is submitted. Tolerance, denoted by T, is the value used to specify the limit of the stimuli, after which a response is expected (cell deformation, for example). Moreover, tolerance is a population characteristic varying with the population units. Given a dose level d, a response is expected in subjects with $T \leq d$. Thus, the expected proportion of subjects with positive response is $p = P[T \leq d] = \int_0^d g(t)dt$, where $g(t)$ is the probability density function associated with T. Since T is a positive random variable, the transformation $X = \log T$ may be considered, taking values in $\Re$ and for which we consider $p = P[X \leq x] = \int_{-\infty}^x f(t)dt$, $x \in \Re$. If f is the normal density then the probit model follows. To establish the calibration problem, let y be the positive response among n subjects submitted to a value x of the independent variable. Considering $y|x \sim Bin(k, p)$ in the probit model, x and y are related through the nonlinear model

$$p = P[X \leq x] = \Phi(\frac{x-\mu}{\sigma}),$$

where $\Phi(z) = P[Z \leq z]$, with $Z \sim N(0, 1)$. Then, $\Phi^{-1}(p) = \beta_1 + \beta_2 x$, with $\beta_1 = -\mu/\sigma$ and $\beta_2 = 1/\sigma$, $\beta_1 \in \Re$ and $\beta_2 \in \Re^+$. Thus, a linear transformation is obtained relating x and a function of p, $\Phi^{-1}(p)$. In the logistic case, the transformation obtained is

$$\log \frac{p}{1-p} = \beta_1 + \beta_2 x.$$

Estimates obtained by using the logistic or the probit model are similar, except for small (close to zero) or large (close to one) values of p, as considered, for example, in Lloyd [21]. Estimates for β_1 and β_2 can be obtained by using the maximum likelihood approach, which are computed by using numerical techniques. It is available in any statistical software. The Bayesian methodology for analyzing logistic regression models abound in the literature. See, for example, Zellner and Rossi [30], Albert and Chib [1] and Bedrick et al. [3]. The above references mainly address the issue of Bayesian inference on the regression coefficients. In this paper, the main interest is focused on the calibration problem which seems not to have been considered in the literature using either classical or Bayesian approaches. As it happens in the case where interest centers on the regression coefficients, there is no analytical or closed form posteriors for the calibration problem.

Section 2 presents classical (based on the maximum likelihood approach) and Bayesian (based on the MCMC methodology) solutions to the calibration problem under the binomial model for logit and probit link functions. The problem of model comparisons is also investigated. An asymptotic approximation is considered for the posterior distribution for estimating x. Finally in Section 3 we present an application to a data set reported in Madruga [22] on the number of blood cells affected by $.^{60}Co$ radiation.

2 The binomial calibration model

In this section, we consider the binomial calibration model,

$$y_i | x_i, \beta_1, \beta_2 \overset{ind}{\sim} Bin(k_i, F(\beta_1 + \beta_2 x_i)), \tag{2.1}$$

$i = 0, 1, \ldots, n$, where β_1, β_2 and x_0 are unknown parameters and F is a (known) continue distribution function, which has a continue density function f. Note that if F is the distribution function of the standard normal distribution, then the probit model follows and if F is the distribution function of a logist distribution, then the logit model follows. It follows from (2.1) that the likelihood function can be written as

$$L(\beta_1, \beta_2, x_0) = \prod_{i=0}^{n} \binom{k_i}{y_i} [F(\beta_1 + \beta_2 x_i)]^{y_i} [1 - F(\beta_1 + \beta_2 x_i)]^{k_i - y_i}. \tag{2.2}$$

Thus, it is not simple to deal with the likelihood (2.2) in the sense of obtaining explicit expressions for the maximum likelihood estimator (MLE) and for the posterior distribution of x_0. To overcome this difficulty, two different approximations are considered. One is based on the asymptotic distribution of the MLE and the other approximation is based on the Markov chain Monte Carlo approach to posterior approximation, by using BUGS (Spiegelhater et al. [28]).

2.1 The maximum likelihood approach

It is well known that under certain regularity conditions the distribution of the MLE of (β_1, β_2, x_0) can be approximated (see Lehmann [20]) by a normal distribution with mean (β_1, β_2, x_0) and the covariance matrix as the inverse of the Fisher information matrix evaluated at the MLE. In the following we discuss the derivation of the maximum likelihood estimators for the binomial calibration problem discussed above. As such, considering the reparametrization

$$(\beta_1, \beta_2, x_0) \to (\beta_1, \beta_2, p_0),$$

where

$$p_0 = F(\beta_1 + \beta_2 x_0),$$

and taking the logarithm of the likelihood function (2.2), we obtain the log-likelihood given by

$$l(\beta_1, \beta_2, p_0) \propto y_0 \log p_0 + (k_0 - y_0)\log(1 - p_0) + \sum_{i=1}^{n} y_i \log[F(\beta_1 + \beta_2 x_i)] + \sum_{i=1}^{n}(k_i - y_i)\log[1 - F(\beta_1 + \beta_2 x_i)]. \tag{2.3}$$

Let $\hat{p}_0$ and $(\hat{\beta}_1, \hat{\beta}_2)$ be the MLE of p_0 and (β_1, β_2), respectively. Thus, from (2.3) it follows that $\hat{p}_0 = y_0/k_0$ and $(\hat{\beta}_1, \hat{\beta}_2)$ is a function of the calibration data (k_i, x_i, y_i), $i = 1, \ldots, n$. Note that $\hat{p}_0$ and $(\hat{\beta}_1, \hat{\beta}_2)$ are independent. To obtain the MLE $\hat{x}_0$ of x_0, we note that

$$x_0 = \frac{F^{-1}(p_0) - \beta_1}{\beta_2}, \tag{2.4}$$

so that by using the invariance property of the MLE, it follows that

$$\hat{x}_0 = \frac{F^{-1}(y_0/k_0) - \hat{\beta}_1}{\hat{\beta}_2}. \tag{2.5}$$

In particular,

i) $\hat{x}_0 = (\Phi(y_0/k_0) - \hat{\beta}_1)/\hat{\beta}_2$, for the probit model, and

ii) $\hat{x}_0 = (\log\{y_0/(k_0 - y_0)\} - \hat{\beta}_1)/\hat{\beta}_2$, for the logit model.

As mentioned previously, it is known from the likelihood theory for generalized linear models (see Lloyd [21]) that the MLE of (β_1, β_2) can not be obtained explicitly, and numerical algorithms such as the Newton-Raphson must be used to compute them. Thus, from the MLE of (β_1, β_2), the MLE of x_0 can be computed by using (2.5) and S-Plus subroutines for logit and probit link functions, for example.

The asymptotic variance of the MLE $\hat{x}_0$ is considered next. Let

$$\boldsymbol{I}_N(\boldsymbol{\theta}) = \Big(\Big(E\Big\{\frac{\partial l}{\partial \theta_i}\frac{\partial l}{\partial \theta_j}\Big\}\Big)\Big),$$

where $\boldsymbol{\theta} = (\beta_1, \beta_2, p_0)$ and $N = \sum_{i=0}^{n} k_i$, be the Fisher information matrix corresponding to the log-likelihood function (2.3). Thus, after some algebraic manipulations, it can be shown that

$$\mathbf{I}_N(\boldsymbol{\theta}) = \begin{pmatrix} \sum\limits_{i=1}^{n} k_i w_i & \sum\limits_{i=1}^{n} k_i w_i x_i & 0 \\ \sum\limits_{i=1}^{n} k_i w_i x_i & \sum\limits_{i=1}^{n} k_i w_i x_i^2 & 0 \\ 0 & 0 & \frac{k_0}{p_0(1-p_0)} \end{pmatrix}, \tag{2.6}$$

where

$$w_i = \frac{f^2(\beta_1 + \beta_2 x_i)}{F(\beta_1 + \beta_2 x_i)[1 - F(\beta_1 + \beta_2 x_i)]},$$

$i = 0, 1, \ldots, n$ and f is the density function corresponding to the distribution function F. Note that $w(t) = F(t)(1 - F(t))$, for the logistic model. Assuming that $k_i/N \to \lambda_i > 0$, as $N \to \infty$, with $\sum_{i=0}^{n} \lambda_i = 1$, if follows that

$$\frac{1}{N}\mathbf{I}_N(\boldsymbol{\theta}) \to \mathbf{I}(\boldsymbol{\theta}),$$

as $N \to \infty$, where $\mathbf{I}(\boldsymbol{\theta})$ is as in (2.6) with k_i replaced by λ_i, $i = 0, 1, \ldots, n$. Thus, letting $\hat{\boldsymbol{\theta}}$ be the MLE of $\boldsymbol{\theta}$, it follows for large N that $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is approximately normally distributed with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}^{-1}(\boldsymbol{\theta})$ (see Lehmann [20]), that is,

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim AN(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\theta})).$$

Consequently, since $x_0 = x_0(\boldsymbol{\theta})$ (see (2.4)), we have that

$$\sqrt{N}(\hat{x}_0 - x_0) \sim AN(0, \Delta(\boldsymbol{\theta})),$$

where

$$\Delta(\boldsymbol{\theta}) = (\frac{\partial x_0}{\partial \boldsymbol{\theta}})\mathbf{I}^{-1}(\boldsymbol{\theta})(\frac{\partial x_0}{\partial \boldsymbol{\theta}})',$$

and from (2.4),

$$(\frac{\partial x_0}{\partial \boldsymbol{\theta}}) = \left(-\frac{1}{\beta_2}, -\frac{g(p_0)-\beta_1}{\beta_2^2}, \frac{g'(p_0)}{\beta_2}\right),$$

with $g(u) = F^{-1}(u)$, the link function, so that

$$g'(u) = \frac{dg(u)}{du} = \frac{1}{f(g(u))}.$$

Thus, after some algebraic manipulations we obtain that

$$\Delta(\boldsymbol{\theta}) = \frac{\beta_2^2 \sum_{i=1}^n \lambda_i w_i x_i^2 - 2\beta_2(g(p_0)-\beta_1)\sum_{i=1}^n \lambda_i w_i x_i + (g(p_0)-\beta_1)^2 \sum_{i=1}^n \lambda_i w_i}{\beta_2^4\{(\sum_{i=1}^n \lambda_i w_i)(\sum_{i=1}^n \lambda_i w_i x_i^2) - (\sum_{i=1}^n \lambda_i w_i x_i)^2\}}$$

$$+\frac{p_0(1-p_0)[g'(p_0)]^2}{\lambda_0 \beta_2^2}.$$

By the assumption that $p_0 = F(\beta_1 + \beta_2 x_0)$, it follows that $g(p_0) = F^{-1}(p_0) = \beta_1 + \beta_2 x_0$ and $g'(p_0) = 1/f(g(p_0)) = 1/f(\beta_1 + \beta_2 x_0)$. Thus, in terms of $\boldsymbol{\theta} = (\beta_1, \beta_2, x_0)$ the asymptotic variance of $\sqrt{N}(\hat{x}_0 - x_0)$ is given by

$$\Delta(\boldsymbol{\theta}) = \frac{1}{\beta_2^2}\Big\{\frac{\sum_{i=1}^n \lambda_i w_i x_i^2 - 2x_0 \sum_{i=1}^n \lambda_i w_i x_i + x_0^2 \sum_{i=1}^n \lambda_i w_i}{(\sum_{i=1}^n \lambda_i w_i)(\sum_{i=1}^n \lambda_i w_i x_i^2) - (\sum_{i=1}^n \lambda_i w_i x_i)^2} + \frac{1}{\lambda_0 w_0}\Big\}.$$

Notice that the above asymptotic variance require f to be nonnull on $\Re$. Further, for large N, $\lambda_i \approx k_i/N$, $i = 0, \ldots, n$, so that $\Delta(\boldsymbol{\theta})$ can be estimated consistently by $\Delta(\hat{\boldsymbol{\theta}})$.

2.2 The Bayesian approach

As mentioned before, it is not possible to obtain explicit expressions for the posterior distribution of x_0. In fact, from (2.1), it follows that

$$\pi(x_0|y_0, y_1, \ldots, y_n) \propto \int\int \prod_{i=0}^n [F(\beta_1 + \beta_2 x_i)]^{y_i}[1 - F(\beta_1 + \beta_2 x_i)]^{k_i - y_i}\pi(\beta_1, \beta_2|x_0)d\beta_1 d\beta_2.$$

The last integral is intractable even for logit and probit models or for the case where non-informative or reference priors are considered. So, to overcome such difficulties we consider the MCMC methodology for approximating to the posterior distribution. As is well known, the main idea behind MCMC is to build up a Markovian process whose stationary distribution (with density f) is the one of interest. Among the MCMC methods, the most popular approach is the Gibbs sampler, introduced in Bayesian inference by Gemman and Gemman [16] while studying problems related to image processing. The books by Robert and Casella [26] and Chen et al. [10] contain a comprehensive review of these methods with applications for logistic regression models.

In the case of the binomial calibration model with probit or logit links, the likelihoods are logconcave (Wedderburn [29]). So the adaptive rejection algorithm (Gilks and Wild [18]) can be used and implemented by using the software BUGS developed by Spiegelhater et al. [28]. It is a free software and can be obtained from the world wide web page $http : ||www.mrc_bsu.com.ac.uk|bugs$. We specify normal priors for β_1 and β_2 with large variances (flat prior) and $x_0 \sim N(m_0, v_0)$. This will guaranted a proper posterior distribution. For a more recent discussion about this see, for example, Gelfand and Sahu [15].

Remark 2.1. Another alternative to approximating the posterior distribution is to consider the normal approximation (see Section 2.1). Under general regularity conditions (Chen [10]), the posterior distribution of x_0, can be approximated for large N by the normal distribution

$$N\Big(\hat{x}_0, \frac{\Delta(\hat{\boldsymbol{\theta}})}{N}\Big),$$

where $\hat{\boldsymbol{\theta}} = (\hat{\beta}_1, \hat{\beta}_2, \hat{x}_0)$ is the MLE of $\boldsymbol{\theta} = (\beta_1, \beta_2, x_0)$ and $\Delta(\boldsymbol{\theta})$ is the asymptotic variance of $\sqrt{N}(\hat{x}_0 - x_0)$ (see Section 2.1). Thus, the credibility interval for x_0 coincides with the classical interval that follows by using the normal approximation to the distribution of the MLE $\hat{x}_0$.

Another aspect of interest is to decide which of the two link functions is more appropriate for a particular data set. The binomial calibration model with the logistic (probit) link function is denoted by $M_1(M_2)$. The Bayes factor can be computed with the aim of deciding for one of the two models. Let $p_i(\mathbf{y}|\boldsymbol{\theta}_i)$ and $\pi_i(\boldsymbol{\theta}_i)$, respectively, the distribution of the data $\mathbf{y} = (y_1, \ldots, y_n)'$ and the prior distribution for the parameter vector $\boldsymbol{\theta}_i$ under model M_i, $i = 1, 2$. Thus, the Bayes factor for model M_2 against model M_1 is given by

$$B_{21}(\mathbf{y}) = \frac{m_2(\mathbf{y})}{m_1(\mathbf{y})},$$

where $m_i(\mathbf{y})$ is the marginal (predictive) distribution of $\mathbf{y}$ under M_i, $i = 1, 2$. The predictive distribution can be approximated by using Monte Carlo methods (see, for example, Bedrick et al. [2] and Carlin and Chib [8]). Because the Bayes factor can be extremely sensitive to the specified prior $\pi(\boldsymbol{\theta}_i)$ (see, for example, O'Hagan [24] and de Santi and Spezaferri [27]), several authors have proposed the use of robust Bayes factors and Partial Bayes factors. One of them is the pseudo Bayes factor which is easy to compute and is implemented in the program BUGS. It was introduced in Geisser and Eddy [12] (see also Gelfand et al. [13] and Gelfand and Dey [14]) and it is based on the conditional predictive densities $p(y_r|\mathbf{y}_{(r)})$, where $\mathbf{y}_{(r)} = (y_1, \ldots, y_{r-1}, y_{r+1}, \ldots, y_n)$.

The pseudo-Bayes factor for model M_1 against model M_2 is

$$PSFB_{12} = \frac{\prod_{r=1}^{n} p_1(y_r|\mathbf{y}_{(r)})}{\prod_{r=1}^{n} p_2(y_r|\mathbf{y}_{(r)})}.$$

Using Monte Carlo methods and the fact that

$$p(y_r|\mathbf{y}_{(r)}) = \int p(y_r|\boldsymbol{\theta}, \mathbf{y}_{(r)})\pi(\boldsymbol{\theta}|\mathbf{y}_{(r)})d\boldsymbol{\theta},$$

we can write (see Gelfand and Dey [14])

$$p(y_r|\mathbf{y}_{(r)}) = \left(\int \frac{1}{p(y_r|\mathbf{y}_{(r)}, \boldsymbol{\theta})}\pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}\right)^{-1},$$

which can be estimated by

$$\hat{p}(y_r|\mathbf{y}_{(r)}) = s\left(\sum_{i=1}^{s} \frac{1}{p(y_r|\mathbf{y}_{(r)}, \boldsymbol{\theta}^{(i)})}\right)^{-1},$$

where s is the size of the sample generated by using BUGS from the posterior of $\boldsymbol{\theta}$.

In our case, y_r is independent of $\mathbf{y}_{(r)}$ given $\boldsymbol{\theta}$, so that

$$\hat{p}(y_r|\mathbf{y}_{(r)}) = s\left(\sum_{i=1}^{s} \frac{1}{p(y_r|\boldsymbol{\theta}^{(i)})}\right)^{-1}$$

with

$$p(y_r|\boldsymbol{\theta}^{(i)}) = \binom{k_r}{y_r}\left[F(\alpha^{(i)} + \beta^{(i)}x_r)\right]^{y_r}\left[1 - F(\alpha^{(i)} + \beta^{(i)}x_r)\right]^{k_r - y_r},$$

$i = 1, \ldots, s$ and $r = 1, \ldots, n$. The estimates $c_r(l) = \hat{p}_l(y_r|\mathbf{y}_{(r)})$ can be plotted against r for $l = 1, 2$, which together with $c(l) = \prod_{r=1}^{n} c_r(l)$ will give indication of which model to select.

3 Analysis of citogenetic data

The data considered in the following is part of the data analyzed in Madruga et al. [22]. The experiment was conducted at the São Paulo Nuclear Institute. Presence of micronucley (MN) indicates cell aberration. We consider here only the presence (or absence) of the MN. Table 3.1 presents the frequency of cells with MN in blood samples from two healthy older subjects, which were exposed to gamma radiation ($.^{60}Co$) with doses $20, 50, 100, 200, 300, 400$ and 500.

We consider the transformation $x_i = \log(d_i)$, $i = 1, \ldots, 8$, where d_i represents the i-th dose value, which is previously fixed. For each one of the groups a model is specified by considering

$$y_i|x_i \sim B(k_i, p_i), \quad \text{with} \quad p_i = F(\beta_1 + \beta_2 x_i),$$

$i = 1, \ldots, 8$, where y_i is the frequency of MN cells associated with the i-th dose value, k_i is the number of cells exposed to dose d_i, p_i is the probability of a cell exposed to the i-th dose value will present micronucley and F is a distribution function.

Using the maximum likelihood approach, three models are considered: the probit, the logit and the Student-t with $\nu = 8$ degree of freedom. The three models are compared by using the mean squared error (MSE) computed using cross validation. The results are presented in the Table 3.2.. Note, from the table, that the results for the logit and Student-t with 8 degrees of freedom links are very close which is not unexpected as the logistic distribution is well approximated by a Student-t distribution with 8 degrees of freedom (Mudholkar and George [23]). We can see that the probit model performs worst according to the MSE criterion.

Table 3.3 presents maximum likelihood estimators and large sample confidence intervals (C.I.) for d_0 using the Student-t model.

The graphical results presented in Figure 3.1 relate the value of d_i (horizontal axis) with the value of p_i (vertical axis). As we can see the linear fitting for the logit, probit or Student-t links are quite reasoable. Madruga et al. [22] considered a nonlinear fit at logit. Though, they obtain a little better fitting, they have difficulty to obtain the real credibility interval for the unknown dose d_0.

We recall that the main interest here is not fitting the model but the prevision of a new dose associated with a new individual. Madruga et al. [22] consider a new observed value, that is $y_0 = 1117$ cells with MN in a total of 2427 evaluated cells. In this case we do not know the value of the dose the cells are exposed to. Relative to Table 3.1, a dose value larger than 500 is suggested, yielding an extrapolation problem.

Table 3.4 presents classical and Bayesian point and interval estimates based on the probit and logit link functions. The Bayesian computation is based on normal prior specification for x_0, with mean $m_0 = \bar{x}$ and variance $v_0 = 10$, where $\bar{x}$ is the average of the fixed value $x_i's$.

Also normal prior specifications for β_1 and β_2 are considered with a large variances (10^3) (flat prioris). The Gibbs samples were generated by using a program implemented in the software WinBUGS (see Spiegehalter et al. [28]) with an average time of 46 seconds used to generate a sample of size 90,000 disregarding the 10,000 initial iterations. Convergence was verifyed by using the Geweke's statistics (Geweke [17]) and also by looking at the graphics of the generated values.

The classical and Bayesian confidence intervals are somewhat close in length for probit and logit links. However, the estimates obtained with the probit link is larger than the estimates obtained with the logit link. The Student-t link estimates follow the logit link estimates.

Using BUGS, we also computed the conditional predictive densities. As we can see in the Figure 3.2., the logit link performs better than the probit link for the most part of the time. However there is not a uniform best model.

4 Conclusion

The present paper considers Bayesian and classical approaches for the calibration problem with binomial response under logit and probit links. A new kind of link is proposed, the Student-t link. The three models are considered with classical, or asymptotic Bayesian solutions. The application in Section 3 suggests us that, results using Student-t link ($\nu = 8$) are close to the results by using logit link. In the MCMC Bayesian solution, it is not straightforward to implement the Student-t model. However, that can be done by introducing latent variables as consided in Branco [4]. The Bayesian approach is very helpful for model comparison as we can see from Figure 3.2. As remarked before, here we consider the response variable as binomial, but in the original data set the response is multinomial. It is under current investigation using the multinomial calibration model and will be reported in future work. Some results in that direction are presented in Branco [4] and Kottas, Branco and Gelfand [19]. The later one presents a Bayesian nonparametric approach for the multinomial calibration problem. However, in both cases easiness of computational implementation using BUGS is lost and more elaborated programs are required.

References

1. Albert, J.H. and Chib, S.. Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association 1993*, **88**: 669-679.

2. Bedrick, E.J., Cristensen, R. and Johnson, W.. A new perspective on priors for generalized linear models. *Journal of the American Statistical Association 1996*, **91**: 1450-1460.

3. Bedrick, E.J., Cristensen, R. and Johnson, W.. Bayesian binomial regression. *The American Statistician 1997*, **51**(3): 211- 218.

4. Branco, M.D.. Calibração: uma abordagem bayesiana. *Phd Thesis 1997* IME - USP (in portuguese).

5. Branco, M.D., Bolfarine, H. and Iglesias, P.. Bayesian calibration under a Student-t model. *Computational Statistics 1998*,**13**: 319-338.

6. Branco, M.D., Bolfarine, H., Iglesias, P. and Arellano-Valle, R.B.. Bayesian analysis of the calibration problem under elliptical distributions. *Journal of Statistical Planning and Inference 2000*,**90**: 69-85.

7. Brown, PJ. *Measurement, Regression and Calibration*. Oxford University Press, 1993.

8. Carlin, BP and Chib, S. Bayesia model choice via Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society 1995*, **B 57**: 473-484.

9. Chen, CF. On asymptotic normality of limiting density functions with bayesian implications. *Journal of the Royal Statistical Society 1985*, **B 97**: 540-546.

10. Chen, MH, Shao, A and Ibrahim, J . *Monte Carlo Methos in Bayesian Computation*. Springer-Verlag, N.Y., 2000.

11. Finney, D . *Probit Analysis*. Cambridge University Press, 1971

12. Geisser, S and Eddy, WF. A predictive approach to model selection. *Journal of the American Statistical Association 1979*,**74**: 153-160.

13. Gelfand, A, Dey, DK and Chang, H . Model determinating using predictive distributions with implementation via sampling-based methods (with discussion in *Bayesian Satistics 1992* 4, Eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith). Oxford: Oxford Univerty Press, pp. 147-167.

14. Gelfand, A and Dey, D . Bayesian model choice: asymptotic and exact calculations. *Journal of the Royal Statistical Society 1994*,**56 B**: 501-514.

15. Gelfand, A and Sahu, S . Gibbs sampling identifiability and improper priors in generalized linear mixed models. *Journal of the American Statistical Association 1999*, **94**: 247-253.

16. Geman, S and Geman, D . Stochastic relaxation Gibbs distributions, and the bayesian

restoration of imagens. *IEEE 1984* **PAMI-6**: 721-741.

17. Geweke, J. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4, 1992* Eds.JM Bernardo, JO Berger, AP Dawid and AFM Smith. Clarendon Press, Oxford, UK.

18. Gilks,WR and Wild, P . Adaptative rejection sampling for Gibbs sampling. *Applied Satistics 1992*, **41**: 337-348.

19. Kottas, A, Branco, MD and Gelfand, AE . A nonparametric Bayesian modelling approaches for cytogenetic dosimetry. *Technical Report 2001.* RT-MAE 2001-01

20. Lehmann, EL . *Elements of Large-Sample Theory.* Springer, N.Y., 1999.

21. Lloyd, CJ . *Statistical Analysis of Categorical Data.* Wiley, N.Y., 1999

22. Madruga, MR, Ochi-Lohmann, TH, Okazaki, K, Pereira, CA de B and Rabello-Gay, MN . Bayesian Dosimetry II: Credibility Intervals for Radiation Dose. *Environmetrics 1996*, 7, 325-331.

23. Mudholkar, GS and George, EO . A remark on the shape of the logistic distribution. *Biometrika 1978*, 667-668.

24. O'Hagan, A . Fractional Bayes factors for model comparison (with discussion). *Journal of the Royal Statistical Society 1995*, **B 57**: 99-138.

25. Osborne, C. Statistical calibration: a review. *Internat.Statist.Rev. 1991*, **59**: 309-336.

26. Robert, CP and Casella, G .*Monte Carlo Statistical Methods.* Springer Texts in Statistics, 1999.

27. De Santis, F and Spezzaferri, F . Alternative Bayes factor for model selection. *Canadian Journal of Statistics 1997*, **25(4)**: 503-515.

28. Spiegelhater, DJ, Thomas, A, Best, N and Gilks, WR . *BUGS. Bayesian inference using gibbs sampling. Version 0.5.* Medical research Council, Biostatistic Unit, Cambridge, U.K., 1995.

29. Wedderburn, RWM . On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika 1976*, **63**: 27-32.

30. Zellner, A and Rossi, P . Bayesian analysis of dichotomous quantal response models. Journal of Econometrics 1984, 365- 393.

Table 3.1: Frequency of MN for binucleated cells from healthy older subjects

Doses	20	50	100	200	300	400	500
y_i	49	70	146	243	268	363	470
k_i	1038	1003	1085	1037	951	1105	1241

Figure 3.1: Graphics (p_i versus d_i)

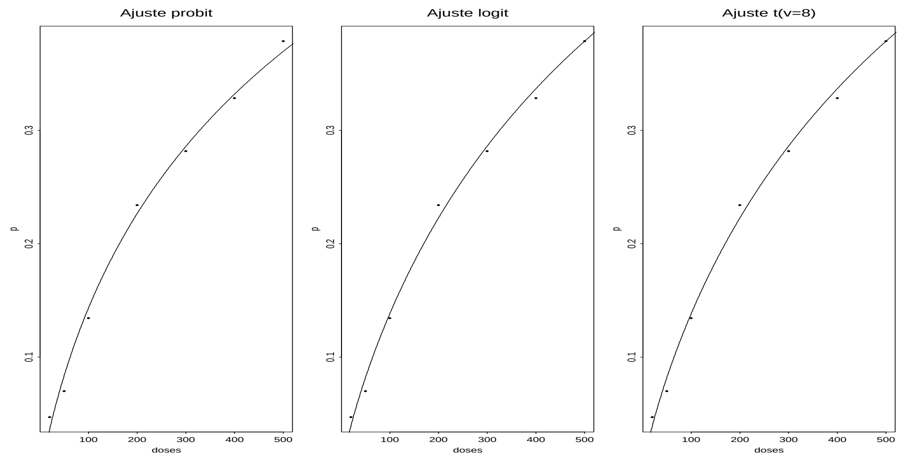

Table 3.2: Adjusted values for logit, probit and Student-t models

Dose d_i	probit $\hat{p}$ - $\tilde{p}$	logit $\hat{p}$ - $\tilde{p}$	Student-t $\hat{p}$ - $\tilde{p}$
20	0.011114	0.006919	0.007790
50	-0.013870	-0.012522	-0.012308
100	-0.008794	-0.002845	-0.003908
200	0.007646	0.0013799	0.011432
300	-0.004149	-0.001813	-0.004420
400	-0.003539	-0.005934	-0.008308
500	0.009051	0.001959	0.000013
mse	0.0000804	0.0000637	0.0000638

Table 3.3: MLE and 95% asymptotic C.I. for d_0 (Student-t link)

Dose	MLE	asymptotic C.I.
20	27.8563	(18.6603,41.5841)
50	36.7177	(25.4559,52.9618)
100	95.2111	(74.8464,121.1167)
200	220.3152	(182.4256,206.0745)
300	291.5146	(241.3704,352.0761)
400	377.3032	(315.6008,451.0689)
500	504.1587	(417.1400,609.3301)

Table 3.4. MLE, posterior mean and 95% C.I. for d_0, when $y_0 = 1117$

Link	MLE	Bayesian
Logit	757.21	761.4
	(649.7,882.5)	(654.7,891.9)
Probit	834.2	840.0
	(702.1,991.2)	(708.6,1001.0)
Student-t	754.6	—
	(646.6, 880.7)	—

Figure 3.2: Conditional predictive densities for the logit (dot) and the probit (triagle) links.