rspa.royalsocietypublishing.org

Research

Cite this article: Higueras M, Puig P, Ainsbury EA, Rothkamm K. 2015 A new inverse regression model applied to radiation biodosimetry. *Proc. R. Soc. A* **471**: 20140588. http://dx.doi.org/10.1098/rspa.2014.0588

Received: 1 August 2014
Accepted: 4 December 2014

Subject Areas:
statistics, radiation biophysics

Keywords:
Bayesian calibration, biological dosimetry, radiotherapy, calibrative density, compound Poisson distribution, Hermite distribution

Author for correspondence:
Manuel Higueras
e-mail: manuel.higueras-hernaez@phe.gov.uk

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspa.2014.0588 or via http://rspa.royalsocietypublishing.org.

THE ROYAL SOCIETY PUBLISHING

A new inverse regression model applied to radiation biodosimetry

Manuel Higueras[1,2], Pedro Puig[2], Elizabeth A. Ainsbury[1] and Kai Rothkamm[1]

[1]Centre for Radiation, Chemical and Environmental Hazards, Public Health England, Chilton, Oxfordshire OX11 0RQ, UK
[2]Departament de Matemàtiques, Universitat Autònoma de Barcelona, Bellaterra, Barcelona 08193, Spain

Biological dosimetry based on chromosome aberration scoring in peripheral blood lymphocytes enables timely assessment of the ionizing radiation dose absorbed by an individual. Here, new Bayesian-type count data inverse regression methods are introduced for situations where responses are Poisson or two-parameter compound Poisson distributed. Our Poisson models are calculated in a closed form, by means of Hermite and negative binomial (NB) distributions. For compound Poisson responses, complete and simplified models are provided. The simplified models are also expressible in a closed form and involve the use of compound Hermite and compound NB distributions. Three examples of applications are given that demonstrate the usefulness of these methodologies in cytogenetic radiation biodosimetry and in radiotherapy. We provide R and SAS codes which reproduce these examples.

1. Introduction

In spite of strict safety measures and regulations, radiation accidents or unplanned exposures occur, for instance in radiology services and radiotherapy departments at hospitals, or using radiography cameras in industry. There have also been some major radiation/nuclear accidents, such as Chernobyl or Fukushima,

that have affected many people [1]. In the event of a radiation accident, biological dosimetry is essential for the timely determination of the radiation dose to which an individual has been exposed. On the other hand, radiotherapy is commonly used to treat cancerous tumours, and it is important to know the total absorbed blood dose to prevent possible complications or side effects. Biological dosimetry relies on quantifying the amount of damage induced by radiation at a cellular level, for instance by counting dicentrics or micronuclei. These aberrations appear because when cells are exposed to radiation, breaks are induced in the chromosomal DNA and the broken fragments may rejoin incorrectly. Therefore, the frequency of chromosome aberrations increases with the amount of radiation and is a reliable and very well-established biological indicator of radiation absorbed dose. Such information supports the clinical management of a patient, enables rapid triage in the case of a large-scale radiation incident and reassures the 'worried well' that they have not received a severe radiation dose. At high acute whole body doses above 2 Gy, haematopoietic failure (or myelodysplasia) is the primary threat associated with acute radiation syndrome which can be supported by early treatment with cytokines or, at very high doses, bone marrow transplants [2]. To estimate the dose absorbed by an individual, dose–effect calibration curves are required which are produced by irradiating peripheral blood lymphocytes to a range of doses. The protocol and methodology for such calibration experiments is described in a recent manual of the International Atomic Energy Agency [3].

The usual approach for constructing the calibration curve is to irradiate n blood samples from various healthy donor with several doses x_i, $i = 1, \ldots, n$. Then, for each irradiated sample, n_i cells are examined and the numbers of observed chromosomal aberrations y_{ij}, $j = 1, \ldots, n_i$ is recorded. For the dicentric assay, it is usually assumed that the counts y_{ij} follow a Poisson distribution [4] or a compound Poisson distribution [5] whose mean is a function of x_i and a set of parameters β, i.e. $\mathrm{E}(y_{ij}) = \mathrm{f}(x_i, \beta)$. From the point of view of IAEA [3], β are the calibration coefficients and $\mathrm{f}(x_i, \beta)$ is the mean of aberrations per cell (called yield or frequency of aberrations per cell, in the cytogenetics field). The parameters of this regression model are usually estimated by maximum likelihood [6], and the MLE and its estimated variance–covariance matrix are calculated and recorded. Therefore, in the case of an irradiated patient, a blood sample is taken and m lymphocytes are scored obtaining the counts $\tilde{y}_1, \ldots, \tilde{y}_m$. The classical approach to estimate the absorbed dose x and its confidence limits is to use the inverse regression method of Merkle [7], also described as a standard procedure in [3]. An improved classical inverse regression method applied to Electron Paramagnetic Resonance dosimetry is found in [8].

Bayesian approaches allow simple incorporation of prior information concerning the circumstances of the exposure. Groer & Pereira [9] were the first to investigate the use of Bayesian models in chromosome dosimetry, for neutron exposure, and since then several researchers have used Bayesian methods in radiation biodosimetry. For instance, Di Giorgio & Zaretzky [10] used a Bayesian approach to present the uncertainty on a biological dose estimate for a radiation overexposed patient in Latin America: a Poisson model with a Jeffrey's prior was used and it was further demonstrated that the Bayesian approach allows presentation of probabilities for dose ranges, which leads to a much more intuitive interpretation of the biological dosimetry results. A review of these methods can be found in [11]. There is also one recent program, CytoBayesJ [12], which provides some basic software tools for Bayesian analysis of cytogenetic radiation dosimetry data.

In this paper, we present a new Bayesian-type method to use cytogenetic data to estimate the dose to which a patient has been exposed. This method uses dose–effect calibration curves estimated by the classical (frequentist) approach suggested in the IAEA manual. Therefore, our new method has the advantage that allows reanalysis of many of the published examples of radiation exposures that were studied using the classical methods. In addition, the method is in fact a general inverse regression model for count responses that could also be applied in contexts other than radiation biodosimetry.

For the three routines implemented in the R statistical software (v. 3.1.1) run in the examples (§§3a,b and 4a) see the electronic supplementary material. A SAS (v. 9.3) routine for model (a) (see table 3) in §3a is also provided. A new R package called 'radir', which implements the Poisson response models presented here, is available under request to the corresponding author.

2. A Bayesian-type inverse regression model

The Poisson distribution is usually used to describe the distribution of dicentric chromosomes per cell when the patient has been irradiated with small doses and with a low linear energy transfer (low-LET radiation). However, after exposure to high-LET, acute radiation, the distribution of dicentrics per cell often presents overdispersion and therefore compound Poisson distributions are preferred. The commonly compound Poisson distributions in biodosimetry are the Neyman A (NA) [13], the negative binomial (NB) [14] and recently the family of rth-order univariate Hermite distributions [15]. These compound Poisson distributions, also known as stopped-Poisson distributions, can be justified by a simple physical model of chromosomal aberration formation: the particles traverse the cell nucleus following a Poisson process and, for each particle, there is a probability (the generalizing distribution) to produce k aberrations. Then the number of aberrations follows a compound Poisson distribution. In other words, a random variable Y follows a compound probability distribution if it can be represented by

$$Y = \sum_{i=1}^{N} \xi_i, \tag{2.1}$$

where N is a count data random variable and $\xi_1, \xi_2, \ldots$ are independent, identically distributed random variables that are also independent of N. In the case where N is Poisson, Y is said to follow a compound Poisson distribution. The distribution of ξ_i is called the generalizing distribution. In particular when the distribution of ξ_i is Poisson, the distribution of Y is an NA, when ξ_i follows a logarithmic distribution, Y is NB distributed, and when ξ_i is distributed as a binomial with a number of trials equal to 2, then Y follows a Hermite distribution [16]. This can be expressed according to the Gurland's notation [16,17] as $N \bigvee \xi$. In particular, parametrizing with respect to the population mean μ and dispersion index δ (the ratio of the variance to the mean σ^2/μ) we have the symbolic representation,

— $\mathrm{NA}(\mu, \delta) \sim \mathrm{Pois}(\mu/(\delta - 1)) \bigvee \mathrm{Pois}(\delta - 1)$
— $\mathrm{NB}(\mu, \delta) \sim \mathrm{Pois}(\mu \log(\delta)/(\delta - 1)) \bigvee \log((\delta - 1)/\delta)$
— $\mathrm{Herm}(\mu, \delta) \sim \mathrm{Pois}(\mu/2(\delta - 1)) \bigvee \mathrm{Bin}(2, \delta - 1)$.

Properties, formulae and algorithms to calculate the probabilities of these distributions can be found in [16]. In brief, they are partially closed under addition [18], the maximum-likelihood estimator of the population mean is the sample mean and they are also members of the discrete exponential dispersion family of distributions. These properties are shared with other distributions potentially useful in biodosimetry, such as Polya Aeppli or Poisson-inverse Gaussian. See [18] for more properties and characterizations of these distributions. In particular, given a random variable Y (with mean μ and dispersion index δ) belonging to one of these models, the sum of n independent copies of Y also belongs to the same model having the same dispersion index and a mean equal to $n\mu$. Moreover, if δ is known, the sum of the observations is a sufficient statistic for μ, containing all the information of the model. This is an important property that will be used in §4.

Let $D = \{(x_i, y_{ij})\}$, $i = 1, \ldots, n$, $j = 1, \ldots, n_i$ be a calibration dataset where each y_{ij} represents a count data observation which will be assumed to follow a Poisson distribution or a two-parameter compound Poisson distribution. Here x_i are the values of the independent variable, dose in the case of cytogenetic radiation biodosimetry. The number of different exposed doses is n and n_i is the sample, the number of blood cells for the ith dose. For all the models, we define the regression

function $\mathrm{E}(y_{ij}) = \mathrm{f}(x_i, \beta)$, $\beta \in \mathbb{R}^p$. Moreover, for compound Poisson modelling, we assume that the dispersion index is a constant (δ). In practice, this assumption could be verified by plotting the empirical values of the dispersion index ($s^2_{y_i}/\bar{y}_{i\cdot}$) against the x_i. However, we could assume another relationship between the independent variable and the dispersion index. Therefore, from now, we will consider the dispersion coefficient δ not to depend on x_i, and then the domain of the parameters is $\Theta = \{\beta, \delta\}$. Note that for the Poisson model $\delta = 1$ and the domain of the parameters is just $\Theta = \{\beta\}$.

Let $\mathrm{p}(y_{ij} = k) = \mathrm{p}(k|\mu, \delta)$ be the probability mass function of the model, parametrized in terms of its population mean and dispersion index. It is clear that $\mathrm{p}(y_{ij} = k) = \mathrm{p}(k|f(x_i, \beta), \delta) = \mathrm{p}(k|x_i, \Theta)$, and then the likelihood function of the calibration data D becomes

$$\mathrm{L}(D|\Theta) = \prod_{\substack{i=1,\ldots,n \\ j=1,\ldots,n_i}} \mathrm{p}(y_{ij}|x_i, \Theta). \tag{2.2}$$

According to the IAEA manual, the parameters are estimated by maximizing the likelihood function (2.2), obtaining $\hat{\Theta} = \{\hat{\beta}, \hat{\delta}\}$. It is well known that for large data samples, the distribution of $\Theta \in \mathbb{R}^{p+1}$ can be approximated by a multivariate Gaussian distribution $\mathrm{N}_{p+1}(\hat{\Theta}, \hat{\Sigma}_{\hat{\Theta}})$, where $\hat{\Sigma}_{\hat{\Theta}}$ is its estimated variance–covariance matrix, that is, the inverse of the estimated Fisher information matrix of the model. Note, however, that in the frequentist framework $\hat{\Theta} \sim \mathrm{N}_{p+1}(\Theta, \hat{\Sigma}_{\hat{\Theta}})$. It is important to remark that the laboratory providing the outputs of the calibration curve, that is $\hat{\Theta}$ and $\hat{\Sigma}_{\hat{\Theta}}$, could be different from the one analysing the patient sample; even though for a consistent assay, the calibration curve should be constructed with the data provided by the same laboratory that will analyse the patient data to guarantee that the scoring criteria applied for the construction of the curve are the same as those applied for patient analysis.

From here, the distribution of the expected count of dicentrics and dispersion index for a given dose of x, $(\mu, \delta)|x$ can be approximated by a bivariate normal distribution. This is a straightforward consequence of the multivariate delta method [19]

$$(\mu, \delta)|x \sim \mathrm{N}_2((\mathrm{f}(x, \hat{\beta}), \hat{\delta}), \nabla \cdot \hat{\Sigma}_{\hat{\Theta}} \cdot \nabla^t), \tag{2.3}$$

where ∇ denotes the derivative of $(\mathrm{f}(x, \beta), \delta)$ at $(\hat{\beta}, \hat{\delta})$, that is,

$$\nabla = \begin{pmatrix} \dfrac{\partial \mathrm{f}}{\partial \beta_0} & \cdots & \dfrac{\partial \mathrm{f}}{\partial \beta_p} & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}.$$

Following these arguments, note that for the Poisson model the distribution of $\mu|x$ is approximated by a univariate normal distribution with expectation $\mathrm{f}(x, \hat{\beta})$ and variance equal to $\mathrm{v}(x, \hat{\beta}) = \nabla \cdot \hat{\Sigma}_{\hat{\Theta}} \cdot \nabla^t$, where ∇ is now the gradient of $\mathrm{f}(x, \beta)$ at $\hat{\beta}$. The bivariate normal density in (2.3) will be denoted as $\phi(\mu, \delta|x)$ and $\phi(\mu|x)$ will be the normal univariate density used for the Poisson model. In some situations, the use of a bivariate or univariate normal could be incompatible with the fact that $\mu > 0$, and in general $\delta > 1$. Then, some approximations have to be carried restricting the parameters' domain. For the univariate normal distribution, one solution is to replace it by a gamma density with the same mean and variance. It is well known that a larger gamma distribution shape parameter (i.e. the ratio of the square of the mean to the variance) implies a better normal approximation. As we will see in the next sections, the normal approximation can be used in a wide range of situations, and it also will be compared with the gamma approximation. For our purposes $\mu|x$ will be called the *mean prior distribution*, because it will act as a prior for the inverse regression estimation problem.

Consider the test (patient) data $\tilde{y} = \{\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_m\}$, formed by m count data observations depending on an unknown regressor x that we aim to estimate. The likelihood function of the test data becomes

$$\mathrm{L}(\tilde{y}|\mu, \delta) = \prod_{i=1}^{m} \mathrm{p}(\tilde{y}_i|\mu, \delta). \tag{2.4}$$

Note that, because the knowledge of μ implies the knowledge of x, then we can write $L(\tilde{y}|\mu,\delta) = L(\tilde{y}|\mu,\delta,x)$. Therefore, an application of Bayes' theorem shows the expression of the posterior density of the parameters given the test data

$$f(\mu,\delta,x|\tilde{y}) = \frac{L(\tilde{y}|\mu,\delta)p(\mu,\delta,x)}{\int L(\tilde{y}|\mu,\delta)p(\mu,\delta,x)\,d\mu\,d\delta\,dx},$$

where $p(\mu,\delta,x)$ is the joint prior density of μ, δ and x. But, $p(\mu,\delta,x) = \phi(\mu,\delta|x)p(x)$, where $p(x)$ summarizes the prior information for x. This prior information can come from the characteristics of the radiation accident, such as the source and the duration of the exposure, etc.

Therefore, marginalizing over μ and δ we obtain the *calibrative density* of x, that it is the solution of the inverse regression problem

$$f(x|\tilde{y}) \propto p(x)\int L(\tilde{y}|\mu,\delta)\phi(\mu,\delta|x)\,d\mu\,d\delta. \tag{2.5}$$

As shown in §3, this calibrative density can be exactly calculated for the Poisson model, solving completely the problem of the absorbed dose estimation in the most frequent situation.

However, for the two-parameter compound Poisson models the integral in (2.5) does not have a closed form, thus some approximations are required such as numerical integration or simulation methods. For this reason, the model will be simplified in §4.

3. The Poisson model

When data are Poisson distributed, the likelihood function of the test data has the form

$$L(\tilde{y}|\mu) \propto \prod_{i=1}^{m} p(\tilde{y}_i|\mu) \propto e^{-m\mu}\mu^{\sum_{i=1}^{m}\tilde{y}_i}.$$

Because the sum of the observations is a sufficient statistic for the parameter of Poisson data, and the sum of independent Poisson random variables is also Poisson distributed, this likelihood function is equivalent to the probability function of one Poisson observation evaluated at s, that is,

$$L(\tilde{y}|\mu) \propto p(s|m\mu) \propto e^{-m\mu}(m\mu)^s,$$

where $s = \sum_{i=1}^{m}\tilde{y}_i$. Therefore, the calibrative density (2.5) remains

$$f(x|\tilde{y}) = p(x)q_s(x), \tag{3.1}$$

where

$$q_s(x) = \int_{-\infty}^{\infty} p(s|m\mu)\phi(\mu|x)\,d\mu. \tag{3.2}$$

Note that (3.2) represents the probability function of a mixed Poisson–normal distribution evaluated at s. Of course, strictly speaking, it is not possible to mix a Poisson with a normal distribution because the Poisson parameter always has to be positive. However, understanding this mixture as a purely formal operation, Kemp & Kemp [20] showed that this mixed Poisson distribution, provided the population mean of the mixing normal is greater than its variance, is just the Hermite distribution. Specifically, using Gurland's notation ([16,17]) we have the symbolic representation

$$\mathrm{Pois}(m\mu) \bigwedge_{\mu} \mathrm{N}(a,b^2) \sim \mathrm{Herm}\left(ma, 1+\frac{mb^2}{a}\right).$$

This notation means that the μ parameter in the Poisson distribution (left part) is normally distributed (right part). This representation is valid only for $a \geq mb^2$.

Consequently, (3.2) is the probability that a Hermite random variable takes a value equal to s. Specifically, it can be directly shown that the probability (3.2) can be obtained from the Hermite probability recursion described in [21]

$$(r+1)\mathrm{q}_{r+1}(x) = (m\mathrm{f}(x,\hat{\beta}) - m^2\mathrm{v}(x,\hat{\beta}))\mathrm{q}_r(x) + m^2\mathrm{v}(x,\hat{\beta})\mathrm{q}_{r-1}(x),$$

with $\mathrm{q}_0(x) = \exp(-m\mathrm{f}(x,\hat{\beta}) + m^2\mathrm{v}(x,\hat{\beta})/2)$ and defining $\mathrm{q}_{-1}(x) = 0$, provided that $\mathrm{f}(x,\hat{\beta}) - m\mathrm{v}(x,\hat{\beta}) \geq 0$. This last inequality is achieved for most of the studied examples, for the range of interest of the absorbed dose x. In a hypothetical situation where this inequality was not achieved, that is $\mathrm{f}(x,\hat{\beta}) - m\mathrm{v}(x,\hat{\beta}) < 0$, expression (3.2) mathematically does not make sense (the dispersion coefficient cannot be greater than 2) and it is therefore better to replace the mean prior normal density $\phi(\mu|x)$ by a gamma density $\Gamma(\mu|x)$ with the same mean $\mathrm{f}(x,\hat{\beta})$ and variance $\mathrm{v}(x,\hat{\beta})$. Then, expression (3.2) would remain

$$\mathrm{q}_s(x) = \int_0^\infty \mathrm{p}(s|m\mu)\Gamma(\mu|x)\,\mathrm{d}\mu. \tag{3.3}$$

Because mixing a Poisson with a gamma produces an NB distribution, it can be shown that $\mathrm{q}_s(x)$ in (3.3) is the probability that an NB random variable, with mean $m\mathrm{f}(x,\hat{\beta})$ and variance $m^2\mathrm{v}(x,\hat{\beta}) + m\mathrm{f}(x,\hat{\beta})$, takes a value equal to s.

The method presented here for the Poisson model, using the gamma distribution as a mean prior, is exactly the same as the full Bayesian method of Groer & Pereira [9] for the simple case where $\mathrm{f}(x,\beta) = \beta x$. However for other dose–response curves both methods differ. For this simple linear dose–response case, considering a uniform dose prior, direct calculations show that

$$\mathrm{f}(x|\tilde{y}) = \frac{m^{s+1}(\sum n_i x_i)^{\sum y_i}}{\mathcal{B}(s+1, \sum y_i - 1)} \frac{x^s}{(mx + \sum n_i x_i)^{s+\sum y_i}};$$

with mean, mode and variance of

$$\mathrm{M}_{|\tilde{y}} = \frac{s}{m}\frac{\sum n_i x_i}{\sum y_i},$$

$$\mathrm{E}_{|\tilde{y}} = \frac{\sum n_i x_i}{m}\frac{\mathcal{B}(s+2, \sum y_i - 2)}{\mathcal{B}(s+1, \sum y_i - 1)}$$

and

$$\mathrm{V}_{|\tilde{y}} = \frac{\mathrm{E}_{|\tilde{y}}}{m} \cdot \left[\left(\sum n_i x_i - 2\right)\mathcal{B}\left(s+2, \sum y_i - 2\right) + 2\mathcal{B}\left(s+3, \sum y_i - 3\right)\right];$$

according to notation in §2, where $\mathcal{B}(\cdot)$ denotes Euler's Beta function. The distribution function of this calibrative density can be expressed in terms of the hypergeometric function.

The following example illustrates how this methodology is applied to a real dataset.

(a) Example: Cobalt-60 gamma rays irradiation

Here we consider data from an inter-laboratory comparison for the semi-automated dicentric assay undertaken as part of the Multibiodose project (a large-scale European biodosimetry project) [22]. This dataset (table 1) is based on blood samples from eight healthy donors which were irradiated *in vitro* with cobalt-60 gamma rays at a high-dose rate of 0.27 Gy min^{-1} simulating acute whole body exposure. The data presented here were collated and analysed using the Metafer 4 automated analysis system (MetaSystems, Altlussheim, Germany) at a single participating laboratory, using the 'BfS' image analysis classifier (system settings—further information in Romm *et al.* [22]).

The u figures shown in table 1 are the values of the u-test statistic of Rao & Chakravarti [23], which is a normalized sample dispersion index

$$u = (d-1)\sqrt{\frac{n-1}{2(1-1/z)}},$$

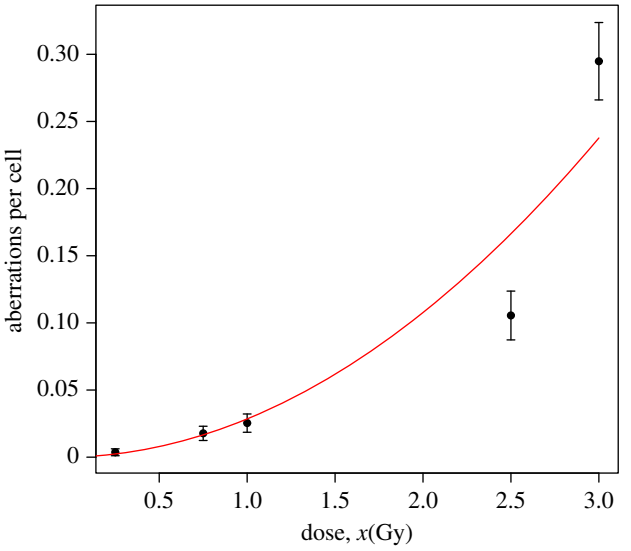

Figure 1. Observed means (dots), plus/minus twice their standard errors (error bars), and predicted means (solid line) of the number of dicentrics for Poisson fitting, based on the data in table 1, omitting the 1.5 Gy test data. (Online version in colour.)

Table 1. Frequency distributions of the number of dicentrics after exposure to six doses of gamma rays, and the sample means, dispersion coefficients and *u* values for each distribution. Test data in italics.

	no. dicentrics							
dose (Gy)	0	1	2	3	4	$\bar{y}$	*d*	*u*
0.25	2185	8				0.004	0.997	−0.113
0.75	2550	44	1			0.018	1.026	0.952
1.00	2231	54	2			0.025	1.044	1.503
1.50	*1712*	*96*	*3*			*0.056*	*1.003*	*0.092*
2.50	1196	123	7	1		0.105	1.038	0.985
3.00	1070	320	41	6	1	0.295	1.012	0.334

where $d = s_{\bar{y}}^2/\bar{y}$ is the sample dispersion coefficient, n the sample size (number of cells) and $z = n\bar{y}$ the total number of count events (number of dicentrics). When d is close to 1 then the data follow an equidispersed distribution. If the value of the u statistic is higher (lower) than (-)2, the distribution can be considered over- (under-) dispersed. The u-test is suggested by the IAEA [3] and in fact it is equivalent to the classical Fisher dispersion test. According to the u values shown in table 1, equidispersion of the calibration data can be assumed, thus justifying the use of a Poisson regression model.

The 1.5 Gy row was removed from the calibration dataset to be used as test data. This means that the true dose is known and it is possible to compare it with the resulting calibrative density. Following notation in §3, $s = 102$ and $m = 1811$, i.e. 102 scored dicentrics in 1811 blood cells.

In this example, for high-dose rate gamma-radiation exposure, an appropriate dose–response curve, i.e. the regression model, is a second degree polynomial without intercept [3], $f(x, \beta) = \beta_2 x^2 + \beta_1 x$ (figure 1). In biodosimetry, this is called the *linear-quadratic* dose–response curve. The intercept has been removed because we assume that for a dose $x = 0$ the expected number of dicentrics will be zero (for the 0 Gy sample there was only 1 dicentric in a total of 2592 blood cells). In general regression modelling, to analyse count data using a second degree polynomial

Table 2. BIC values using a second degree polynomial dose–response curve without constant term for the different models.

model	NB	Hermite	NA	Poisson
BIC	4088.834	4085.594	4085.524	4079.639

mean response is not common, and a log-link mean response is the usual approach. However, in biodosimetry, the linear-quadratic dose–response curve has a biophysical interpretation [3] and is one of the most frequently employed in practice. Some problems could occur maximizing the likelihood function because β_1 and β_2 have to be necessarily positive. To ensure this, it is sometimes necessary to use numerical algorithms allowing constrains in the parameter domain.

Table 2 shows the Bayesian information criterion (BIC) values for the four different response distributions treated in this work from the calibration data. These values support the use of the Poisson model. So for a Poisson response the maximum-likelihood parameter estimates and their estimated covariance matrix are the following:

— Fitted coefficients:

$$\hat{\beta}_1 = 3.126 \times 10^{-3} \quad \text{and} \quad \hat{\beta}_2 = 2.537 \times 10^{-2}.$$

— Estimated covariance matrix:

$$\hat{\Sigma}_{\hat{\beta}} = \begin{pmatrix} 7.205 & -3.438 \\ -3.438 & 2.718 \end{pmatrix} \times 10^{-6}.$$

As has been commented in §2, $\mu|x$ will follow a normal or a gamma distribution with mean $\mathrm{f}(x, \hat{\beta}) = \hat{\beta}_2 x^2 + \hat{\beta}_1 x$ and variance $\mathrm{v}(x, \hat{\beta}) = \nabla \cdot \hat{\Sigma}_{\hat{\beta}} \cdot \nabla^t$, where

$$\nabla = \left(\frac{\partial \mathrm{f}}{\partial \beta_1}, \frac{\partial \mathrm{f}}{\partial \beta_2} \right) = (x, x^2),$$

and therefore $\mathrm{v}(x, \hat{\beta}) = \hat{\Sigma}_{22} x^4 + 2\hat{\Sigma}_{21} x^3 + \hat{\Sigma}_{11} x^2$.

According to (3.2) and (3.3), for a normal or a gamma mean prior, the predictive posterior distribution $\mathrm{q}_{102}(x)$ represents the probability of a Hermite or NB random variable taking a value of 102 counts, both with same mean $45.939x^2 + 5.661x$ and variance $8.913x^4 - 22.553x^3 + 69.571x^2 + 5.661x$.

Despite the real dose being known, firstly, a non-informative prior dose distribution is chosen in order to not take advantage of this fact, so $\mathrm{p}(x) \propto 1$. Secondly, for our purposes of comparing results, we define an informative prior dose distribution assuming we do not know the real dose of the test data, but we observe a mean of 0.056 dicentrics per cell, then by comparison with table 1 it can reasonably be estimated that the dose is between 1 and 2.5 Gy. A simple informative prior could be a gamma whose mean is in the midpoint of this interval, i.e. 1.75, and whose standard deviation is in the halfway from the mean to cover this interval, i.e. 0.375. For a gamma distribution with this mean and standard deviation, the 95.67% of the values fall in the region of $1.75 \pm 2 \times 0.375$.

Figure 2 shows the plot of the three densities of the estimated dose for the data test. Note how these results incorporate the real dose (1.5 Gy) and show the similarities found using both mean priors. Note that the gamma mean prior is moderately more conservative.

To use the normal mean prior (3.2) for this calibration set, the following condition must be satisfied: $\mathrm{f}(x, \hat{\beta}) - m\mathrm{v}(x, \hat{\beta}) \geq 0$. It holds when $x \leq 3.337$ Gy, and this could also be used as prior information about the dose, that is, $\mathrm{p}(x) \sim \mathcal{U}(0, 3.337)$. For the range of the likely doses studied, the minimum value of the shape parameter of the mean prior gamma is 328.616, so the gamma or normal mean priors are practically indistinguishable.

The statistics of the three calibrative densities calculated in this example are shown in table 3.

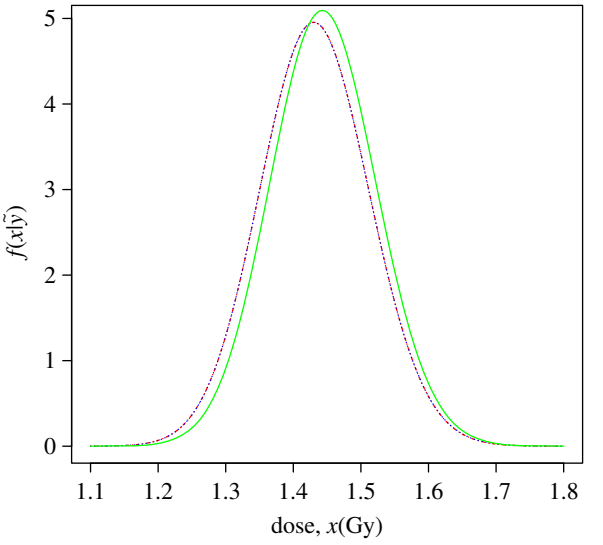

Figure 2. Calibrative densities of the 1.5 Gy test data calculated from a normal (blue/dotted line) and a gamma (red/dashed-dotted line) mean prior with non-informative prior dose distribution, and for a gamma mean prior with informative prior dose distribution (green/solid line). Red and blue curves are indistinguishable.

Table 3. Statistics summary of the calibrative densities for a normal (*a*) and a gamma (*b*) mean prior with non-informative prior dose distribution, and for a gamma mean prior with informative prior dose distribution (*c*).

model	mode	expected	s.d.	95% CI
(*a*)	1.430	1.432	0.081	(1.277, 1.594)
(*b*)	1.430	1.432	0.081	(1.277, 1.593)
(*c*)	1.443	1.445	0.078	(1.294, 1.602)

(b) Example: analysis of doses in thyroid cancer patients

This example illustrates how our methodology can be applied having only the fitted parameters of the dose–response curve, without knowing the calibration points. Serna *et al.* [24] studied chromosomal damage in lymphocytes of thyroid cancer patients after radioiodine treatment. The authors did a micronuclei assay in binucleated cells of blood samples from 25 patients 3 days after Iodine-131 (3.7 GBq) exposure.

The *in vitro* calibration curve was fitted by a linear-quadratic model with intercept, $\mathrm{f}(x, \beta) = G\beta_2 x^2 + \beta_1 x + \beta_0$ according to Poisson's law, and the estimate of β_0 was not taken into account, because the authors in [24] argued that the intercept could change for each patient. Constant G is the Lea–Catcheside generalized dose-protraction factor, which modifies the quadratic term according to the temporal pattern of exposure, being $G = 1$ for the *in vitro* assay. The authors calculated the following parameter estimates ($\hat{\beta}_i \pm \mathrm{SE}(\hat{\beta}_i)$)

$$\hat{\beta}_1 = (13.6 \pm 5.5) \times 10^{-3}, \quad \hat{\beta}_2 = (3.7 \pm 1.6) \times 10^{-2}, \quad \rho = -0.89,$$

where ρ is the correlation coefficient for $\hat{\beta}_1$ and $\hat{\beta}_2$. The patients were subjected to ablative radioiodine treatments for post-surgical thyroid remnants. Consequently, they had a prolonged exposure lasting several days and which means, the temporal pattern of exposure was different than that of the *in vitro* assay. Taking into account the exposure profile of the Iodine-131 treatment, the authors in [24] found the factor G to be close to 0.1.

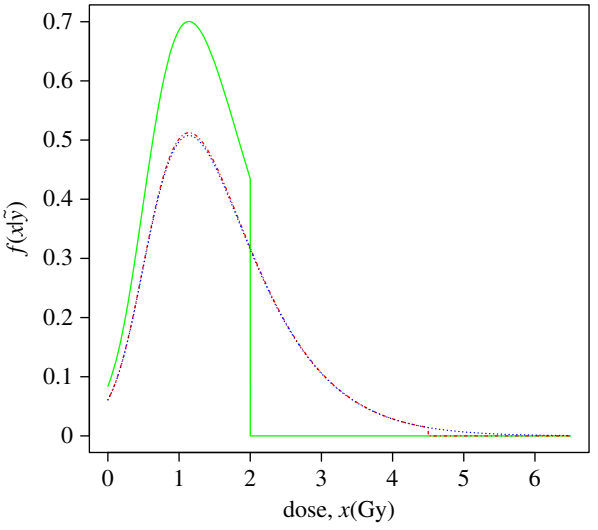

Figure 3. Calibrative densities of [24] Patient 1 test data calculated from a gamma mean prior density, with a $\mathcal{U}(0, 2)$ (green/solid line), a $\mathcal{U}(0, 4.5)$ (red/dashed-dotted line) prior dose distribution and a improper $\mathcal{U}(0, +\infty)$ (blue/dotted line) prior dose distribution.

Then β_0, the background for each patient, can be estimated counting the micronuclei of the patient from a blood sample taken before the treatment, information provided in [24]. This leads to the fitted regression model $\mathrm{f}(x, \hat{\beta}) = G\hat{\beta}_2 x^2 + \hat{\beta}_1 x + \hat{\beta}_0$ with a covariance matrix that incorporates the variance of $\hat{\beta}_0$ without correlation with $\hat{\beta}_1$ and $\hat{\beta}_2$.

To illustrate our techniques we are going to estimate the absorbed dose for Patient 1, but the same can be done for the others. Patient 1 presented 487 normal cells and 13 cells with just one micronucleus each. Before the treatment five micronuclei where found in 500 blood cells, thus $\hat{\beta}_0 = (10 \pm 4.450) \times 10^{-3}$. The u-statistic of the test data is -0.395, so this is compatible with the Poisson model.

Therefore, $\mu|x$ will be considered to follow a distribution with mean $\mathrm{f}(x, \hat{\beta}) = G\hat{\beta}_2 x^2 + \hat{\beta}_1 x + \hat{\beta}_0$ and variance $\mathrm{v}(x, \hat{\beta}) = \nabla \cdot \hat{\Sigma}_{\hat{\beta}} \cdot \nabla^t$, where

$$\nabla = \left(\frac{\partial \mathrm{f}}{\partial \beta_0}, \frac{\partial \mathrm{f}}{\partial \beta_1}, \frac{\partial \mathrm{f}}{\partial \beta_2} \right) = (1, x, Gx^2).$$

The condition $\mathrm{f}(x, \hat{\beta}) - m\mathrm{v}(x, \hat{\beta}) \geq 0$ is held when $x \leq 0.880\,\mathrm{Gy}$. This range of doses is very small for our purposes and consequently a gamma mean prior is preferred instead of a normal.

According to (3.3), for a gamma mean prior, the predictive posterior distribution $\mathrm{q}_{13}(x)$ represents the probability of an NB random variable taking a value of 13 counts, with mean $0.185x^2 + 6.8x + 5$ and variance $0.006x^4 - 0.399x^3 + 7.987x^2 + 6.8x + 9.95$.

Three calibrative densities have been calculated applying two different proper uniform prior dose distributions, both using information given in [24]. An administered radioiodine activity that produces a blood dose less than 2 Gy is considered safe, so we could take a uniform dose prior distribution from 0 to 2, assuming that doctors use prudent doses. On the other hand, the calibration curve was calculated up to a dose of 4.5 Gy, so another uniform dose prior distribution could be from 0 to 4.5. An improper uniform prior dose distribution from 0 to $+\infty$ is also applied.

Figure 3 shows the plot of the three densities of the estimated dose for the data test. Their statistics are indicated in table 4. These results agree with those displayed in [24], where the dose estimate for Patient 1 was 1.14 Gy.

Table 4. Statistics summary of the calibrative densities for two proper and one improper uniform dose priors.

prior dose distribution	mode	expected	s.d.	95% CI
$\mathcal{U}(0, 2)$	1.140	1.141	0.481	(0.203, 1.945)
$\mathcal{U}(0, 4.5)$	1.140	1.561	0.858	(0.203, 3.615)
$\mathcal{U}(0, +\infty)$	1.140	1.593	0.921	(0.253, 3.829)

4. The simplified compound Poisson calibration model

We now consider a dataset that follows a compound Poisson distribution. The likelihood function of the test data has been previously described in (2.4), and the calculation of the calibrative density (2.5) requires to use numerical integration or Monte Carlo methods. However, the model can be simplified by replacing δ in $\mathrm{L}(\tilde{y}|\mu, \delta)$ with the MLE $\hat{\delta}$ obtained from the calibration data. The performance of this simplification is analysed and compared in the example §3*a*. Then the likelihood function $\mathrm{L}(\tilde{y}|\mu, \hat{\delta})$, which we prefer to denote as $\mathrm{L}(\tilde{y}, \hat{\delta}|\mu)$, is equivalent to the probability function of the sum of the observations, that is the probability function of a compound Poisson observation,

$$\mathrm{L}(\tilde{y}, \hat{\delta}|\mu) \propto \mathrm{p}(s, \hat{\delta}|m\mu),$$

where $s = \sum_{i=1}^{m} \tilde{y}_i$. Then, the calibrative density is as described in (3.1) with

$$\mathrm{q}_s(x) = \int_{-\infty}^{\infty} \mathrm{p}(s, \hat{\delta}|m\mu)\phi(\mu|x)\,\mathrm{d}\mu \tag{4.1}$$

if the mean prior is a normal density, or

$$\mathrm{q}_s(x) = \int_{0}^{\infty} \mathrm{p}(s, \hat{\delta}|m\mu)\Gamma(\mu|x)\,\mathrm{d}\mu \tag{4.2}$$

when the mean prior is gamma distributed. Expressions (4.1) and (4.2) correspond to the probability function of mixed compound Poisson random variables, where the mixing density is respectively normal or gamma, evaluated at s. The operations of compounding and mixing are interchangeable for these models [16,17], e.g. mixing an NA with a normal result in the following:

$$\begin{aligned}
&\mathrm{NA}(m\mu, \hat{\delta}) \bigwedge_{\mu} \mathrm{N}(\mathrm{f}(x, \hat{\beta}), \mathrm{v}(x, \hat{\beta})) \\
&= \mathrm{Pois}\left(\frac{m\mu}{\hat{\delta} - 1}\right) \bigvee \mathrm{Pois}(\hat{\delta} - 1) \bigwedge_{\mu} \mathrm{N}(\mathrm{f}(x, \hat{\beta}), \mathrm{v}(x, \hat{\beta})) \\
&= \mathrm{Pois}\left(\frac{m\mu}{\hat{\delta} - 1}\right) \bigwedge_{\mu} \mathrm{N}(\mathrm{f}(x, \hat{\beta}), \mathrm{v}(x, \hat{\beta})) \bigvee \mathrm{Pois}(\hat{\delta} - 1) \\
&= \mathrm{Herm}\left(\frac{m\mathrm{f}(x, \hat{\beta})}{\hat{\delta} - 1}, 1 + \frac{m\mathrm{v}(x, \hat{\beta})}{(\hat{\delta} - 1)\mathrm{f}(x, \hat{\beta})}\right) \bigvee \mathrm{Pois}(\hat{\delta} - 1).
\end{aligned} \tag{4.3}$$

This is providing that (4.1) and (4.2) are, respectively, the probability functions of compound Hermite and compound NB random variables. Therefore, according to the different choices of

the compound Poisson distribution we obtain the following compound distributions for $q_s(x)$:

$$\left.\begin{aligned} &\text{NA}: \quad \mathcal{F}\left(\frac{m\mathrm{f}(x,\hat{\beta})}{\hat{\delta}-1}, 1+\frac{m\mathrm{v}(x,\hat{\beta})}{(\hat{\delta}-1)\mathrm{f}(x,\hat{\beta})}\right) \bigvee \mathrm{Pois}(\hat{\delta}-1), \\ &\text{NB}: \quad \mathcal{F}\left(\frac{m\mathrm{f}(x,\hat{\beta})\log(\hat{\delta})}{\hat{\delta}-1}, 1+\frac{m\mathrm{v}(x,\hat{\beta})\log(\hat{\delta})}{(\hat{\delta}-1)\mathrm{f}(x,\hat{\beta})}\right) \bigvee \log\left(\frac{\hat{\delta}-1}{\hat{\delta}}\right) \\ \text{and} \quad &\text{Hermite}: \quad \mathcal{F}\left(\frac{m\mathrm{f}(x,\hat{\beta})}{2(\hat{\delta}-1)}, 1+\frac{m\mathrm{v}(x,\hat{\beta})}{2(\hat{\delta}-1)\mathrm{f}(x,\hat{\beta})}\right) \bigvee \mathrm{Bin}(2,\hat{\delta}-1). \end{aligned}\right\} \tag{4.4}$$

Here $\mathcal{F}(\mu_{\mathcal{F}},\delta_{\mathcal{F}})$ indicates a Hermite or an NB distribution, according to (4.1) or (4.2), parametrized by its population mean and dispersion index. When $\mathcal{F}$ is the Hermite distribution, these representations make sense only when $\mathrm{f}(x,\hat{\beta})(\hat{\delta}-1) \geq m\mathrm{v}(x,\hat{\beta})$ for the NA, $\mathrm{f}(x,\hat{\beta})(\hat{\delta}-1) \geq m\mathrm{v}(x,\hat{\beta})\log(\hat{\delta})$ for the NB and $2\mathrm{f}(x,\hat{\beta})(\hat{\delta}-1) \geq m\mathrm{v}(x,\hat{\beta})$ for the Hermite.

Compound NB distributions have been studied and applied in several publications. Properties, characterizations and references can be found in [16]. Compound Hermite distributions are less common, so far there is one recent publication [25] that studies the continuous compound Hermite gamma distribution.

When $\mathcal{F}(\mu_{\mathcal{F}},\delta_{\mathcal{F}})$ is NB, the probabilities of the associated compound distributions can be calculated using the Panjer recursion formula [26]. This formula is based on the fact that the probabilities $p_n = \mathrm{P}(X=n)$ of a random variable X distributed as a $\mathrm{NB}(\mu_{\mathcal{F}},\delta_{\mathcal{F}})$ satisfy a first-order recurrence relation $p_n = p_{n-1}(a+b/n)$, where $a=(\delta_{\mathcal{F}}-1)/\delta_{\mathcal{F}}$ and $b=(\mu_{\mathcal{F}}-\delta_{\mathcal{F}}+1)/\delta_{\mathcal{F}}$. Then, if the probabilities of the generalizing distribution are denoted as f_k, the probabilities q_i of the corresponding NB compound distribution satisfy the recursion [26]

$$q_0 = \frac{p_0}{(1-f_0 a)^{1+b/a}} \quad \text{and} \quad q_i = \sum_{j=1}^{i}\left(a+\frac{bj}{i}\right) f_j q_{i-j}, \quad i \geq 1. \tag{4.5}$$

Expression (4.5) can be efficiently used to calculate (4.2). The values of a and b will be taken according to the chosen distribution of the observations, using the corresponding expression of $\mu_{\mathcal{F}}$ and $\delta_{\mathcal{F}}$ of the NB ($\mathcal{F}$) indicated in (4.4). In the next section we will give an example of application.

When $\mathcal{F}$ is Hermite, the probabilities of a Hermite compound distribution cannot be calculated using the Panjer recursion formula because the probabilities of the Hermite do not follow a linear recursion. To calculate the probabilities in this case we state and prove (in appendix A) the following proposition:

Proposition 4.1. *Let q_n, $n=0,1,2\ldots$ be the probabilities of a compound Hermite distribution of the form $\mathrm{Herm}(\mu_h,\delta_h)\bigvee\mathcal{P}$, where $\mathcal{P}$ is a count distribution with probabilities f_k, $k=0,1,2\ldots$. We define $r_j=\sum_{i=0}^{j} f_i f_{j-i}$, $j=0,1,2\ldots$, then*

$$q_n = \frac{\mu_h}{n}\sum_{i=0}^{n-1}(n-i)q_i\left\{(2-\delta_h)f_{n-i} + \frac{(\delta_h-1)}{2}r_{n-i}\right\}, \tag{4.6}$$

and $q_0 = \exp(\mu_h((2-\delta_h)(f_0-1)+(\delta_h-1)(f_0^2-1)/2))$.

It is important to remark that, to calculate $q_s(x)$ in (4.1) and (4.2), a computationally intensive direct numerical integration can be done instead to use the Panjer recursion or proposition 4.1. To this end, it would be enough to obtain numerically the probabilities which are available for a more wide range of models than those studied in this paper.

The use of (4.6) will be illustrated with a real data analysis in the next section.

Table 5. Frequency distributions of the number of micronuclei after exposure to 11 doses of gamma rays, and the sample means, dispersion coefficients and *u* values for each distribution. Test data in italics.

	no. micronuclei										
dose (Gy)	0	1	2	3	4	5	6	7	$\bar{y}$	*d*	*u*
0.00	4887	106	5	2					0.024	1.156	7.839
0.10	*4773*	*206*	*19*	*2*					*0.050*	*1.150*	*7.526*
0.25	4261	324	41	12	2				0.090	1.306	15.306
0.50	4536	364	76	17	7				0.119	1.449	22.484
0.75	4383	512	85	18	2				0.149	1.257	12.876
1.00	4225	636	115	19	5				0.189	1.240	12.009
1.50	4018	805	139	26	9	1	2		0.243	1.270	13.495
2.00	3499	1194	238	45	13	10	1		0.383	1.209	10.471
2.50	3171	1313	393	94	24	3	2		0.501	1.201	10.077
3.00	2582	1575	598	190	44	9	2	6	0.722	1.206	10.307
4.00	1974	1674	869	342	102	26	13	2	1.013	1.172	8.628

(a) Example: high linear energy transfer exposure

Puig & Valero [18] studied the fitting of an experiment of 11 samples of peripheral blood exposed to different doses of γ-rays (table 5), where the dose rate was 0.93 cGy min^{-1}. For each sample, approximately 5000 binucleated cells were inspected, and the numbers of micronuclei were counted.

The *u* values shown in table 5 confirm the overdispersion, thus Poisson regression is not adequate.

Similar to the example analysed in §3*a* the 0.1 Gy data will be removed to be used as test data. This distribution has a total of 250 micronuclei in a total of 5000 cells so $s = 250$ and $m = 5000$.

The appropriate dose–response curve, i.e. the regression model, is again a linear-quadratic model with intercept, $f(x, \beta) = \beta_2 x^2 + \beta_1 x + \beta_0$ (figure 4). Table 6 shows the BIC values for the four different models studied in this work. Note how these values support the use of the NB model.

Using the NB model, the maximum-likelihood estimation provides the following results:

— Fitted coefficients:

$$\hat{\beta}_0 = 3.639 \times 10^{-2}, \quad \hat{\beta}_1 = 1.156 \times 10^{-1}, \quad \hat{\beta}_2 = 3.241 \times 10^{-2}, \quad \hat{\delta} = 1.231.$$

— Estimated covariance matrix:

$$\hat{\Sigma}_{\hat{\theta}} = \begin{pmatrix} 73.749 & -115.908 & 29.210 & 13.976 \\ -115.908 & 373.338 & -110.398 & 36.919 \\ 29.210 & -110.398 & 38.102 & -3.625 \\ 13.976 & 36.919 & -3.625 & 1133.825 \end{pmatrix} \times 10^{-7}.$$

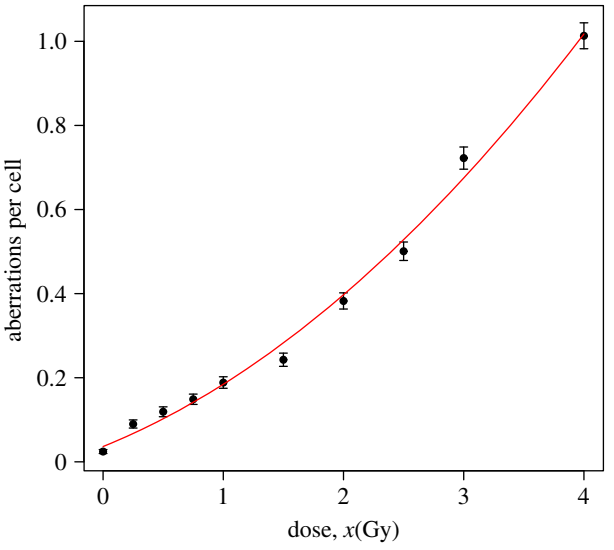

Figure 4. Observed means (dots), plus/minus twice their standard errors (error bars), and predicted means (solid line) of the number of micronuclei for NB fitting, based on the data in table 5, omitting the 0.1 Gy test data. (Online version in colour.)

Table 6. BIC values using a second degree polynomial dose–response curve for the different models.

model	Poisson	Hermite	NA	NB
BIC	67360.01	66537.46	66467.85	66437.93

Then, the prior densities are:

— *Complete Model*: According to (2.3), $(\mu, \delta)|x$ follows a bivariate normal distribution with mean $(\hat{\beta}_2 x^2 + \hat{\beta}_1 x + \hat{\beta}_0, \hat{\delta})$ and variance–covariance $\nabla \cdot \hat{\Sigma}_{\hat{\Theta}} \cdot \nabla^t$, where

$$\nabla = \begin{pmatrix} \frac{\partial f}{\partial \beta_0} & \frac{\partial f}{\partial \beta_1} & \frac{\partial f}{\partial \beta_2} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & x & x^2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

so the variance-covariance is

$$\begin{pmatrix} \hat{\Sigma}_{33}x^4 + 2\hat{\Sigma}_{32}x^3 + 2\hat{\Sigma}_{31}x^2 + \hat{\Sigma}_{22}x^2 + 2\hat{\Sigma}_{21}x + \hat{\Sigma}_{11} & \hat{\Sigma}_{43}x^2 + \hat{\Sigma}_{42}x + \hat{\Sigma}_{41} \\ \hat{\Sigma}_{43}x^2 + \hat{\Sigma}_{42}x + \hat{\Sigma}_{41} & \hat{\Sigma}_{44} \end{pmatrix}.$$

For this example, the calibrative density (2.5) is calculated via numerical integration in order to be compared with those calculated using the simplified models.

— *Simplified Models*: According to the arguments given in §4, $\mu|x$ follows a gamma or a normal distribution with mean $f(x, \hat{\beta}) = \hat{\beta}_2 x^2 + \hat{\beta}_1 x + \hat{\beta}_0$ and variance $v(x, \hat{\beta}) = \nabla \cdot \hat{\Sigma}_\beta \cdot \nabla^t$, where

$$\nabla = \left(\frac{\partial f}{\partial \beta_0}, \frac{\partial f}{\partial \beta_1}, \frac{\partial f}{\partial \beta_2} \right) = (1, x, x^2),$$

so the variance is $\hat{\Sigma}_{33}x^4 + 2\hat{\Sigma}_{32}x^3 + 2\hat{\Sigma}_{31}x^2 + \hat{\Sigma}_{22}x^2 + 2\hat{\Sigma}_{21}x + \hat{\Sigma}_{11}$. According to (4.4), for a normal or a gamma mean prior, the predictive posterior distribution $q_{250}(x)$ represents respectively the probability of a compound Hermite- or compound NB-Logarithmic random variable taking a value of 250 counts, both with same $f(x, \hat{\beta}) = 0.032x^2 + 0.116x + 0.036$, $v(x, \hat{\beta}) = 3.81 \times 10^{-6}x^4 + 1.525 \times 10^{-5}x^3 + 5.842 \times 10^{-6}x^2 - 2.318 \times 10^{-5}x + 7.375 \times 10^{-6}$ and $\hat{\delta} = 1.231$.

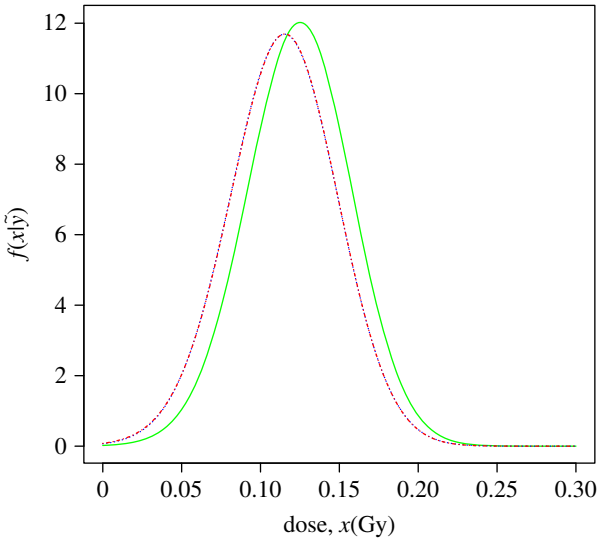

Figure 5. Calibrative densities of the 0.1 Gy test data using the complete model (2.5) (green/solid line), and the simplified ones with a normal (blue/dotted line) and a gamma (red/dashed-dotted line) mean prior density; all with a uniform prior dose distribution. Blue and red curves are indistinguishable.

Table 7. Statistics summary of the calibrative densities for the complete model, and the simplified models using a gamma and a normal mean prior with a uniform prior dose distribution.

model	complete	S. Norm. p.	S. Gam. p.
mode	0.125	0.115	0.115
expected	0.124	0.114	0.114
s.d.	0.033	0.034	0.034
95% CILB	0.059	0.047	0.047
95% CIUB	0.190	0.182	0.181

To use the normal mean prior (4.1) in this calibration set for NB responses, there is a condition to be satisfied: $\mathrm{f}(x,\hat{\beta})(\hat{\delta}-1) - m\mathrm{v}(x,\hat{\beta})\log(\hat{\delta}) \geq 0$. It is satisfied when $x \leq 4.294$ Gy. In this example, this is not a problem and it could be used as prior information about the dose, that is $\mathrm{p}(x) \sim \mathcal{U}(0, 4.294)$. For the range of the likely doses studied, the minimum value of the shape parameter of the mean prior gamma is 179.605, and consequently both gamma and normal mean priors are almost indistinguishable (red and blue curves in figure 5).

Figure 5 shows the plot of the three densities (one from the complete model and two from the simplified ones) of the estimated dose for the data test. Note that both calibrative densities from the simplified models are practically the same. The statistics of these densities are shown in table 7. These results incorporate the real dose (0.1 Gy) and also show their similarities, chiefly between the simplified models.

5. Conclusion

In this paper, we have presented several Bayesian-type methods for count data inverse regression, showing its application in the field of cytogenetic dosimetry. First, in §2 we defined our methodology for inverse regression, where responses are either Poisson or two-parameter compound Poisson. We have assumed that the dispersion index is constant along the different

doses. This methodology leads to a bivariate normal prior density when the responses follow a two-parameter compound Poisson distribution, and an univariate normal or gamma mean prior density when the responses follow a Poisson distribution. To use our methodology, only the estimates of the parameters and covariance matrix of the dose–response curve are required. This information is available from the standard frequentist analysis suggested by the IAEA manual, with many examples published by other researchers or laboratories. Therefore, our method is not a full Bayesian approach because the dose–response curve is estimated using frequentist analysis. MCMC methods could be used if the models were more complex or the prior densities more complicated. They might also be used for model averaging, since one might aim to avoid choosing one of the presented four models, preferring to use a weighted amalgam of them.

The Poisson model is developed in §3, leading to a closed form of the calibrative density. Two examples of dose estimation based on the dicentric assay are reported.

In §4, we treated two-parameter compound Poisson models, simplifying them to get the calibrative densities into a closed form. For this purpose, we have presented a method which involves calculating the probabilities of compound NB distributions, using Panjer's recursion [26], and compound Hermite distributions, using a recursion relation described in proposition 4.1. Another example of dose estimation is shown, based on data obtained with the micronucleus assay. We have assumed a constant dispersion coefficient, but our methods could be also extended to dose-dependent dispersion models of the form $\delta_{ij} = g(x_i, \gamma)$, $\gamma \in \mathbb{R}^q$.

The illustrative examples show applications using the most frequent calibrative curves, that are second-order polynomials (the linear-quadratic model). However, other response functions can be directly analysed using the same methodology. It should be noted that the approaches presented here may also prove useful in areas other than biological dosimetry.

Disclaimer. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

Funding statement. This work was funded by the National Institute for Health Research. The work carried out at UAB was funded by the grant MTM2012-31118 and by the grant UNAB10-4E-378 co-funded by ERDF 'A way of making Europe'.

Appendix A. Proof of proposition 4.1

First of all, let us recall some topics related to the probability-generating function (pgf). Given a count random variable X, its pgf $\Phi_X(s)$ is defined as

$$\Phi_X(s) = \sum_{k=0}^{\infty} p_k s^k,$$

where the coefficients of this power series are the probabilities $p_k = \mathrm{P}(X = k)$ and consequently the derivatives at $s = 0$ divided by $k!$ provide the probability mass function of X. The pgf of a compound probability distribution described in (2.1) is

$$\Phi_X(s) = \Phi_N(\Phi_\xi(s)), \tag{A 1}$$

where $\Phi_N(s)$ is the pgf of N and $\Phi_\xi(s)$ is the common pgf of the ξ_i [16].

One property of pgf's is that the sum of independent random variables is a random variable whose pgf is the product of the pgf's of the summed variables; e.g. given X and Y independent random variables with pgf's $\Phi_X(s)$ and $\Phi_Y(s)$ respectively, the pgf of $X + Y$ results

$$\Phi_{X+Y}(s) = \Phi_X(s)\Phi_Y(s). \tag{A 2}$$

According to Kemp & Kemp [20] the pgf of a random variable X Hermite distributed with mean μ_h and dispersion coefficient δ_h is

$$\Phi_X(s) = \mathrm{e}^{\mu_h\{(2-\delta_h)(s-1)+(\delta_h-1)(s^2-1)/2\}}, \tag{A 3}$$

therefore, according to (A 1), the pgf of a Herm$(\mu_h, \delta_h) \bigvee \mathcal{P}$ distribution, being $\psi(s)$ the pgf of $\mathcal{P}$, is

$$\phi(s) = \mathrm{e}^{\mu_h\{(2-\delta_h)(\psi(s)-1)+(\delta_h-1)(\psi^2(s)-1)/2\}}, \tag{A 4}$$

thus the probability in 0 is

$$q_0 = \phi(0) = \mathrm{e}^{\mu_h\{(2-\delta_h)(f_0-1)+(\delta_h-1)(f_0^2-1)/2\}}.$$

Note that $\psi^2(s)$ is the pgf of a sum of two independent identically distributed random variables having both a pgf equal to ψ, so

$$\varphi(s) = \psi^2(s) = \sum_{n=0}^{\infty} r_n s^n, \quad r_n = \sum_{i=0}^{n} f_i f_{n-i}.$$

The derivative of ϕ is

$$\phi'(s) = \left[\mu_h \left\{(2-\delta_h)(\psi'(s)-1) + \frac{(\delta_h-1)(\varphi'(s)-1)}{2}\right\}\right]\phi(s),$$

therefore,

$$\begin{aligned}\sum_{n=1}^{\infty} n q_n s^{n-1} &= \mu_h \left\{(2-\delta_h)\sum_{n=1}^{\infty} n f_n s^{n-1} + \frac{(\delta_h-1)}{2}\sum_{n=1}^{\infty} n r_n s^{n-1}\right\}\sum_{n=0}^{\infty} q_n s^n \\ &= \mu_h \sum_{n=1}^{\infty} n\left\{(2-\delta_h) f_n + \frac{(\delta_h-1)}{2} r_n\right\} s^{n-1} \sum_{n=0}^{\infty} q_n s^n,\end{aligned}$$

matching the coefficients with same degree in s in both sides leads to

$$q_n = \frac{\mu_h}{n}\sum_{i=0}^{n-1}(n-i) q_i \left\{(2-\delta_h) f_{n-i} + \frac{(\delta_h-1)}{2} r_{n-i}\right\}, \quad n \geq 1,$$

and this finishes the proof.

References

1. Suto Y, Hirai M, Akiyama M, Kobashi G, Itokawa M, Akashi M, Sugiura N. 2013 Biodosimetry of restoration workers for The Tokyo Electric Power Company (TEPCO) fukushima daiichi nuclear power station accident. *Health Phys.* **105**, 366–373. (doi:10.1097/HP.0b013e3182995e42)
2. DiCarlo AL, Maher C, Hick JL, Hanfling D, Dainiak N, Chao N, Bader JL, Coleman CN, Weinstock DM. 2011 Radiation injury after a nuclear detonation: medical consequences and the need for scarce resources allocation. *Disaster Med. Public Health Prep.* **5**, S32–S44. (doi:10.1001/dmp.2011.17)
3. IAEA. 2011 *Cytogenetic dosimetry: applications in preparedness for and response to radiation emergencies*. International Atomic Energy Agency: Vienna. See http://www-pub.iaea.org/MTCD/publications/PDF/EPR-Biodosimetry%202011_web.pdf.
4. Edwards AA, Lloyd DC, Purrott RJ. 1979 Radiation induced chromosome aberrations and the Poisson distribution. *Radiat. Environ. Biophys.* **16**, 89–100. (doi:10.1007/BF01323216)
5. Nelson SJ. 1984 A stochastic model of the effects of ionizing radiation on mammalian cells *in vitro*. *Bull. Math. Biol.* **46**, 423–446. (doi:10.1007/BF02462017)
6. Frome EL, DuFrain RJ. 1986 Maximum likelihood estimation for cytogenetic dose–response curves. *Biometrics* **42**, 73–84. (doi:10.2172/5652085)
7. Merkle W. 1983 Statistical methods in regression and calibration analysis of chromosome aberration data. *Radiat. Environ. Biophys.* **21**, 217–233. (doi:10.1007/BF01323412)
8. Demidenko E, Williams BB, Flood AB, Swartz HM. 2012 Standard error of inverse prediction for dose–response relationship: approximate and exact statistical inference. *Stat. Med.* **32**, 2048–2061. (doi:10.1002/sim.5668)
9. Groer PG, Pereira CABD. 1987 Calibration of a radiation detector: chromosome dosimetry for neutrons. In *Probability and Bayesian statistics* (ed. R Viertl), pp. 225–232. New York, NY: Plenum Publishing Corporation.

10. Di Giorgio M, Zaretzky A. 2011 Biological dosimetry—a Bayesian approach for presenting uncertainty on biological dose estimates. *Annals of 'II Encuentro de Docentes e Investigadores de Estadística en Psicología'*. University of Buenos Aires. See http://www.iaea.org/inis/collection/NCLCollectionStore/_Public/44/096/44096783.pdf.
11. Ainsbury EA, Vinnikov VA, Puig P, Higueras M, Maznyk NA, Lloyd DC, Rothkamm K. In press. Review of Bayesian statistical analysis methods for cytogenetic radiation biodosimetry, with a practical example. *Radiat. Prot. Dosim.* (doi:10.1093/rpd/nct301)
12. Ainsbury EA, Vinnikov VA, Puig P, Maznyk NA, Rothkamm K, Lloyd DC. 2013 CytoBayesJ: software tools for bayesian analysis of cytogenetic radiation dosimetry data. *Mutat. Res./Genet. Toxicol. Environ. Mutagen.* **756**, 184–191. (doi:10.1016/j.mrgentox.2013.06.005)
13. Virsik RP, Harder D. 1981 Statistical interpretation of overdispersed distribution of radiation-induced dicentric chromosome aberrations at high LET. *Radiat. Res.* **85**, 13–23. (doi:10.2307/3575434)
14. Brame RS, Groer PG. 2002 Bayesian analysis of overdispersed chromosome aberration data with the negative binomial model. *Radiat. Prot. Dosim.* **102**, 115–119. (doi:10.1093/oxfordjournals.rpd.a006079)
15. Puig P, Barquinero JF. 2011 An application of compound Poisson modelling to biological dosimetry. *Proc. R. Soc. A* **467**, 897–910. (doi:10.1098/rspa.2010.0384)
16. Johnson NL, Kemp AW, Kotz S. 2005 *Univariate discrete distributions*, 3rd edn. NJ: John Wiley & Sons.
17. Gurland J. 1957 Some interrelations among compound and generalized distributions. *Biometrika* **44**, 265–268. (doi:10.2307/2333264)
18. Puig P, Valero J. 2006 Count data distributions: some characterizations with applications. *J. Am. Stat. Assoc.* **101**, 332–340. (doi:10.1198/016214505000000718)
19. Serfling RJ. 1980 Transformations of given statistics. In *Approximation theorems of mathematical statistics*, pp. 122–124. New York, NY: John Wiley & Sons. (doi:10.1002/9780470316481)
20. Kemp AW, Kemp CD. 1966 An alternative derivation of the hermite distribution. *Biometrika* **53**, 627–628. (doi:10.1093/biomet/53.3-4.627)
21. Kemp CD, Kemp AW. 1965 Some properties of the 'Hermite' distribution. *Biometrika* **52**, 381–394. (doi:10.1093/biomet/52.3-4.381)
22. Romm H *et al.* 2013 Automatic scoring of dicentric chromosomes as a tool in large scale radiation accidents. *Mutat. Res. Toxicol. Environ. Mutagen.* **756**, 174–183. (doi:10.1016/j.mrgentox.2013.05.013)
23. Rao CR, Chakravarti IM. 1956 Some small sample tests of significance for a Poisson distribution. *Biometrics* **12**, 264–282. (doi:10.2307/3001466)
24. Serna A, Alcaraz M, Navarro JL, Acevedo C, Vicente V, Canteras M. 2008 Biological dosimetry and Bayesian analysis of chromosomal damage in thyroid cancer patients. *Radiat. Prot. Dosim.* **129**, 372–380. (doi:10.1093/rpd/ncm444)
25. Hürlimann W. 2013 A characterization of the compound multiparameter hermite gamma distribution via Gauss's principle. *Sci. World J.* **2013**, 468418. (doi:10.1155/2013/468418)
26. Panjer HH. 1981 Recursive evaluation of a family of compound distributions. *ASTIN Bull.* **12**, 22–26.