Estimating exposure fraction from radiation biomarkers: a comparison of frequentist and Bayesian approaches

Adam Errington, Jochen Einbeck, Jonathan Cumming

Department of Mathematical Sciences, Durham University, UK

Abstract. If individuals are exposed to ionising radiation, due to some radiation accident, for medical reasons, or during spaceflight, there is often a need to estimate the contracted radiation dose. The field of biodosimetry is concerned with estimating the dose retrospectively, based on certain biomarkers, which are typically based on counts of some cytogenetic or biomolecular features of the cell arising after radiation-induced double-strand-breaks. Such techniques face particular challenges when the exposure is only partial rather than whole-body, which, when unaccounted for, may lead to grossly inaccurate dose estimates. For biomarkers which are overdispersed, there are currently no procedures available for the detection of partial-body exposures. We consider the question of estimating the exposure fraction as well as quantifying its uncertainty, using Bayesian and frequentist methods, by means of simulation scenarios which are motivated by foci count data as arising for the $\gamma-$H2AX protein biomarker.

Keywords: Dispersion · Zero-Inflation · Uncertainty Quantification.

1 Introduction

The major goal of space missions is to allow human exploration without exceeding a certain risk level from exposure to space radiation. Clearly, the understanding of human exposure to this ionising radiation in the aircraft environment is of great importance in the field of aerospace. Since human response to ionising radiation is both individual and variable, one approach is to explore radiation effects on a cellular level. Individual radiation sensitivity can provide the basis for personalised countermeasures against key environmental factors in long-term missions. Radiation biomarkers, which try to quantify the radiation dose through the damage that has been caused on a cellular level, are necessary in order to determine the radiation sensitivity in a blood sample.

Most radiation biomarkers come in the form of count data, including cytogenetic biomarkers (dicentric chromosomes, micronuclei [2]), or protein-based biomarkers such as the γ-H2AX assay [1]. The latter biomarker considers counts of foci which appear after phosphorylation of the H2AX histone following double-strand breaks. While this biomarker motivates our work, the principles are applicable to other biomarkers and also beyond the field of biodosimetry. The Poisson

model is a natural choice for the analysis of count data, and has been successfully applied for the dicentric assay under full-body exposure. Laboratory data (with known doses) are used to fit a linear or quadratic model to the measured yields (counts per cell), resulting in a 'calibration curve'. Following exposure of an individual, the observed yield of a blood sample is equated to the calibration curve, and dose estimated via inverse regression [2].

However, the Poisson assumption of equidispersion (variance = mean) may be violated. Firstly, unobserved heterogeneity in the cell population or aspects of the scoring procedure may cause deviation from this property. While early evidence suggests that the distribution of γ-H2AX foci among the scored blood cells adheres well to the Poisson assumption and hence can be analysed by employing methods used for the dicentric assay [1], practical γ-H2AX data sets almost always exhibit overdispersion. Secondly, individuals are often only partially exposed, in which case their blood will contain a mixture of cells showing no radiation impact at all (*structural zeros*), and cells featuring a distribution of counts according to dose of exposure. It is important to detect partial-body exposure, and to quantify the fraction of exposure, as otherwise the resulting dose estimates will be incorrect, with potentially severe consequences.

For the dicentric assay, where whole-body counts are usually equidispersed, any significant overdispersion serves as evidence for the presence of partial exposure. It is well established how to adjust for partial-body exposure for this biomarker through the 'contaminated Poisson method' [3, 4]. However, for biomarkers such as micronuclei or the γ-H2AX, which become overdispersed even in the case of whole body exposure, no such mechanism is yet known. In this paper we focus on the problem of estimating the exposure fraction and quantifying its uncertainty, for such scenarios. Bayesian and frequentist techniques will be employed and compared.

2 Methodology

To represent partial-body exposure in the case of an overdispersed distribution of foci counts, we require models which can handle both overdispersion and excess zero counts in data. To account for the extra zero counts, zero-inflated models describe the data as a combination of two distributions: a distribution which takes a single value at zero and a count distribution such as the Poisson or NB1. For a sample consisting of n cells, we define Y_j to be the response variable representing the observed number of foci for cell j ($j = 1, .., n$). The probability mass function for the zero-inflated Poisson (ZIP) model is:

$$P(Y_j = y_j|\lambda, p) = \begin{cases} p + (1-p)e^{-\lambda}, & \text{for } y_j = 0 \\ (1-p)\frac{e^{-\lambda}\lambda^{y_j}}{y_j!}, & \text{for } y_j > 0 \end{cases} \quad (1)$$

where $0 \leq p \leq 1$ and $\lambda > 0$, possibly depending on covariates such as dose. Here, λ refers to the mean of the underlying Poisson distribution and p is the zero-inflation parameter. The ZIP model has the properties: $\mathrm{E}(Y_j) = (1-p)\lambda = \mu$

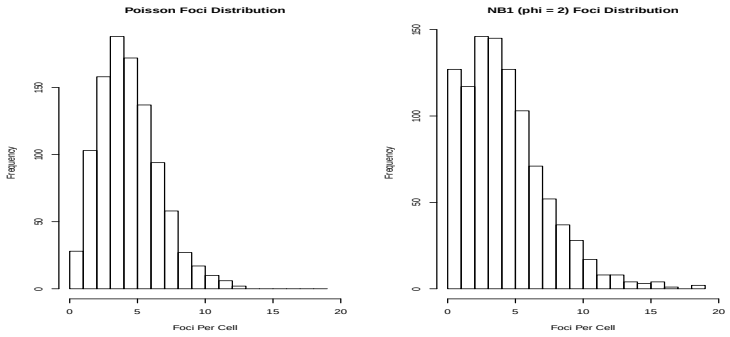

Fig. 1. A comparison of the individual number of foci per cell produced in an equidispersed and overdispersed whole-body sample.

and $\mathrm{Var}(Y_j) = (1-p)\lambda(1+p\lambda)$ and reduces to a Poisson when $p = 0$. Since $\mathrm{Var}(Y_j) \geq \mu$, zero-inflation can be seen as a special form of overdispersion.

In our context, for data which stem from full- or partial-body exposure, it is sensible to consider overdispersion and zero–inflation as two separately identifiable model properties. A suitable model for this purpose is the ZINB-1 model, which shares same the mean as the ZIP, but has variance $\mathrm{Var}(Y_j)=(1-p)\lambda(1+\alpha+p\lambda)$, where α is a dispersion parameter. This variance suggests that the ZINB-1 exhibits overdispersion when $\alpha > 0$ and $p > 0$. For $\alpha = 0$, the ZINB-1 reduces to the ZIP. We estimate the model parameters α and p in two ways:

- *Maximum Likelihood estimation.* This will implicitly produce standard errors of $\hat{\alpha}$ and $\hat{p}$ through the Fisher information matrix.
- *Bayesian estimation.* Using priors $p \sim U[0,1]$, $\alpha \sim \Gamma(0.01, 0.01^{-1})$, and a prior for λ which is determined by linear transformation of a dose prior according to the known calibration curve, the posterior distribution is computed via a MCMC Gibbs sampling algorithm. Uncertainty Quantification is based on this posterior distribution.

[subsection] The exposure fraction, $F = 1-p$, and dispersion, $\phi = 1+\alpha$, can be estimated via $\hat{F} = 1-\hat{p}$ and $\hat{\phi} = 1+\hat{\alpha}$, respectively, where clearly $SE(\hat{F}) = SE(\hat{p})$, and $SE(\hat{\phi}) = SE(\hat{\alpha})$. We note that $F = 1-p$ is a simplifying assumption as it ignores certain effects (such as cell death) which prevent irradiated cells being observable at the time of scoring. However, this effect is considered minor for the γ-H2AX biomarker where cells are typically scored after a few hours, in contrast to the dicentric biomarker where at least 48 hours need to pass until mitosis [3].

Table 1. MLE and posterior mean estimates for the zero-inflation and dispersion parameter based on an average of the 100 Poisson simulation runs. For reference, the posterior median and mode are given in brackets (median, mode).

	Frequentist		Bayesian	
	ZIP	ZINB-1	ZIP	ZINB-1
$\hat{p}$	0.4998 ± 0.0113	0.4997 ± 0.0113	0.4998 ± 0.0113	0.4993 ± 0.0113
			(0.4998, 0.4999)	(0.4993, 0.4992)
$\hat{\alpha}$		0.0148 ± 0.3266		0.0549 ± 0.0223
				(0.0486, 0.0334)

3 Simulation

To generate H2AX-type foci count samples, we make use of a whole-body calibration curve reported previously in the literature [5]:

$$\mu = 0.35 + 1.48D. \tag{2}$$

Assuming a fixed and known dose of $D \equiv 3$Gy for this simulation, $n = 1000$ observations were taken separately from two scenarios:

A. Poi($\lambda = \mu = 4.79$)
B. NB1($\lambda = \mu = 4.79$; $\phi = 2$)

(with 'base' dispersion $\phi = 1+\alpha$), providing an equidispersed (A) and an overdispersed (B) whole-body sample (see Figure 1). In order to mimic a 50% partial exposure scenario, 1000 zeros were manually added to the above samples. The whole process was repeated 100 times. Hereafter, information regarding dose level and fraction used to generate this data is assumed to be unknown.

Scenario A. One finds from Table 1 that, for the ZIP model, the Bayesian and frequentist estimates (and their standard errors) of p are identical. It follows that an estimate for the exposed fraction, F, is found through $\hat{F} = 1 - \hat{p} = 0.5002 \pm 0.0113$. Estimates of exposure fraction under ZINB-1 are very similar to those under ZIP. While the frequentist ZINB-1 estimate obtained for α suggests equidispersion, the Bayesian confidence interval for α does not cover the true value $\alpha = 0$. From Figure 2, we see that the Bayesian versions tend to skew the estimates away from the true values, which is related to the choice of priors.

Scenario B. From fitting ZIP and ZINB-1 models, it is clear from the MLE estimates presented in Table 2 (ZIP: $\hat{F} = 0.4851 \pm 0.0113$, ZINB-1: $\hat{F} = 0.4995 \pm 0.0118$) and the boxplots in Figure 3 that the ZINB-1 was able to account for overdispersion due to zero-inflation and sampling and was therefore the preferred model in estimating the exposed fraction. The corresponding fraction estimates from the Bayesian methods appear to show that the ZIP deviates the most from the true value (ZIP: $\hat{F} = 0.4852 \pm 0.0112$, ZINB-1: $\hat{F} = 0.4996 \pm 0.0118$). The true value of α was found to be within 1 standard error in both the frequentist and Bayesian ZINB-1, with the latter producing a slightly closer estimate. It

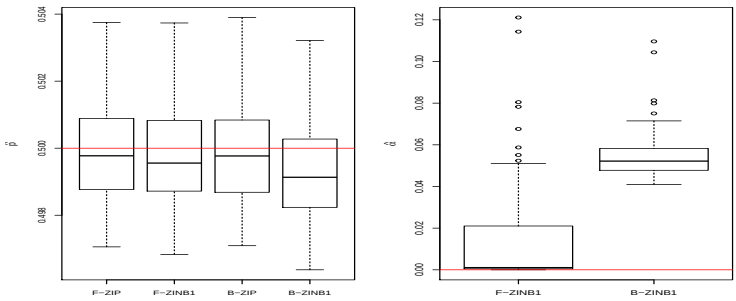

Fig. 2. A comparison of the frequentist (F-ZIP/F-ZINB1) and Bayesian (B-ZIP/B-ZINB1) distributions of the model parameters resulting from 100 Poisson simulations. The red lines at $p = 0.5$ and $\alpha = 0$ indicate the true parameter values.

appears that there is no strong preference for estimating the exposed fraction and its uncertainty utilising a Bayesian approach over the standard maximum likelihood method.

Table 2. MLE and posterior mean estimates for the zero-inflation and dispersion parameter based on an average of the 100 NB1 simulation runs.

	Frequentist		Bayesian	
	ZIP	ZINB-1	ZIP	ZINB-1
$\hat{p}$	0.5149 ± 0.0113	0.5005 ± 0.0118	0.5148 ± 0.0112	0.5004 ± 0.0118
			(0.5148, 0.5148)	(0.5004, 0.5007)
$\hat{\alpha}$		0.9849 ± 0.1096		0.9898 ± 0.1096
				(0.9853, 0.9781)

Acknowledgement. We thank Dr David Endesfelder, BfS, Neuherberg, Germany, for providing some `rjags` code for zero-inflated models.

References

1. Horn S, Barnard S, and Rothkamm K: Gamma-H2AX-based dose estimation for whole and partial body radiation exposure. *PLoS One*, 6:e25113 (2011)
2. IAEA. Cytogenetic dosimetry: applications in preparedness for and response to radiation emergencies. International Atomic Energy Agency: Vienna (2011)
3. Hilali A et al.: An Appraisal of the Value of the Contaminated Poisson Method to Estimate the Dose Inhomogeneity in Simulated Partial-Body Exposure. *Radiat. Res.* 128, 108–111 (1991)
4. Szluinska M, Edwards AA, and Lloyd DC.: Statistical Methods for Biological Dosimetry. Health Protection Agency, Centre for Radiation, Chemical and Environmental Hazards, Radiation Protection Division (2005)

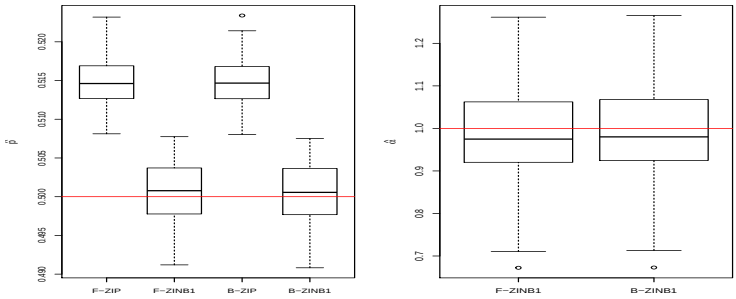

Fig. 3. A comparison of the frequentist (F-ZIP/F-ZINB1) and Bayesian (B-ZIP/B-ZINB1) distributions of the model parameters resulting from 100 NB1 simulations.The red lines at $p = 0.5$ and $\alpha = 1$ indicate the true parameter values.

5. Rothkamm K et al.: Manual versus automated gamma–H2AX foci analysis across five European laboratories: can this assay be used for rapid biodosimetry in a large scale radiation accident? *Mutat. Res.* 756, 170–173 (2013)