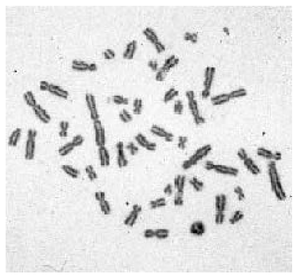
Zero–inflated models for radiation–induced chromosome aberration data

María Oliveira[1], Liz Ainsbury[2], Jochen Einbeck[1], Kai Rothkamm[2], Pere Puig[3], Manuel Higueras[2,3]

[1] *Department of Mathematical Sciences, Durham University, Durham, UK.*
[2] *Public Health England Centre for Radiation, Chemical and Environmental Hazards (PHE CRCE), Chilton, Didcot, UK*
[3] *Departament de Matematiques, Universitat Autonoma de Barcelona, Bellaterra (Barcelona), Spain.*

Introduction

The formation of chromosome aberrations is a well–established biological marker of exposure to ionizing irradiation. Accordingly, radiation–induced aberrations are studied extensively to provide data for biological dosimetry, where the measurement of chromosome aberration (dicentrics, micronuclei, centric rings, etc) frequencies in human lymphocytes is used for assessing absorbed doses of ionizing radiation to individuals.

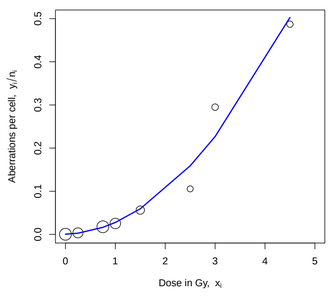


A lymphocyte chromosome spread with dicentric chromosomes (IAEA, 2011).

The Standard Procedure
The usual approach for constructing the calibration curve is to irradiate some blood samples from a healthy donor with several doses. Then, for each irradiated sample, the cells are examined and the numbers of observed chromosomal aberrations are recorded.

	y_{ij}							
x_i	0	1	2	3	4	5	n_i	y_i
0.00	2591	1	0	0	0	0	2592	1
0.25	2185	8	0	0	0	0	2193	8
0.75	2550	44	1	0	0	0	2595	46
1.00	2231	54	2	0	0	0	2287	58
1.50	1712	96	3	0	0	0	1811	102
2.50	1196	123	7	1	0	0	1327	140
3.00	1070	320	41	6	1	0	1438	424
4.50	895	360	110	25	5	1	1396	680

Frequency of dicentrics after exposure to Co–60 gamma rays.



Dose–response curve fitted according to the Poisson model with identity link.

- It is assumed that $y_i \sim \text{Poisson}(n_i\mu_i)$ where

$$\mu_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 \qquad (1)$$

  where n_i is the no. assessed cells, y_i is the no. chromosome aberrations, $x_i$is prefixed doses.
- The linear–quadratic relationship is well established.
- "Usually" not log–linear model. The identity link is the accepted standard in biodosimetry (IAEA, 2011).

Drawbacks: The assumption of equidispersion implicit in the Poisson model is often violated due to unobserved heterogeneity in the cell population, which will render the variance of observed aberration counts larger than their mean, and/or the frequency of zero counts greater than expected for the Poisson distribution.

Goal: Study the performance of zero–inflated models for modelling such data by comparing their behavior with other models already used in biodosimetry. The models are applied to several datasets with different radiation exposure patterns.

Zero–inflated regression models applied to biodosimetry

When overdispersion is attributed to the large number of zeros with respect to the Poisson model, a ZIP model may provide a good fit. Let Y_i, $i = 1, \ldots, n$ be the response variable. A ZIP regression model is defined as

$$\text{P}(Y_i = y_i) = \begin{cases} p_i + (1-p_i)\exp(-\lambda_i), & y_i = 0, \\ (1-p_i)\exp(-\lambda_i)\lambda_i^{y_i}/y_i!, & y_i > 0, \end{cases}$$

where $0 \leq p_i \leq 1$ and $\lambda_i > 0$. For the ZIP, $\text{E}(Y_i) = (1-p_i)\lambda_i = \mu_i$ and $\text{Var}(Y_i) = \mu_i(1+p_i\lambda_i)$.

When the overdispersion is both due to the heterogeneity of data and the excess of zeros, the ZINB regression model often is more appropriate than the ZIP. For the ZINB regression model, the probability mass function of the response variable Y_i $(i = 1, \ldots, n)$ is given by

$$\text{P}(Y_i = y_i) = \begin{cases} p_i + (1-p_i)(1+\alpha\lambda_i^c)^{-\lambda_i^{1-c}/\alpha}, & y_i = 0, \\ (1-p_i)\frac{\Gamma(y_i+\lambda_i^{1-c}/\alpha)}{y_i!\Gamma(\lambda_i^{1-c}/\alpha)}(1+\alpha\lambda_i^c)^{-\lambda_i^{1-c}/\alpha}(1+\lambda_i^{-c}/\alpha)^{-y_i}, & y_i > 0, \end{cases}$$

where $\alpha > 0$ is an overdispersion parameter and the index $c \in \{0, 1\}$ identifies the form of the underlying negative binomial distribution (ZINB1 and ZINB2, respectively). For the ZINB, $\text{E}(Y_i) = (1-p_i)\lambda_i = \mu_i$ and $\text{Var}(Y_i) = \mu_i(1+p_i\lambda_i+\alpha\lambda_i^c)$.

Both the mean λ_i of the underlying Poisson distribution and the extra–zeros proportion p_i parameters can depend on vectors of covariates, i.e.,

$$g(\lambda_i) = \boldsymbol{x_i}^t\boldsymbol{\beta} \text{ and } h(p_i) = \boldsymbol{z_i}^t\boldsymbol{\gamma},$$

where $\boldsymbol{x_i}$ and $\boldsymbol{z_i}$ are vectors of covariates, t denotes the transpose vector operator, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the corresponding vectors of regression coefficients and g and h are link functions.

Application to biological dosimetry

- Which mean should be modelled, the mean of the zero-inflated distribution, μ_i, or the mean of the underlying Poisson or negative binomial distribution, λ_i, which are relate via $\lambda_i = \mu_i(1-p_i)$?

  For compliance with formulation (1) and with practice in this particular field, it is adequate to model the mean of the corresponding zero–inflated distribution, μ_i, via the linear predictor in (1).
- How to model p_i?
  The mixture parameter p_i will be modelled as usual with logistic regression, where three different scenarios will be investigated:

  $(a)\ \text{logit}(p_i) = \gamma_0; \quad (b)\ \text{logit}(p_i) = \gamma_1 x_i; \quad (c)\ \text{logit}(p_i) = \gamma_0 + \gamma_1 x_i; \quad i = 1, \ldots, d.$

Comparative study

Zero–inflated models to describe the number of chromosome aberrations in biological dosimetry are compared with the Poisson, negative binomial, Neyman type A, Hermite, Pólya–Aeppli and Poisson–inverse Gaussian models. The models have been applied to four real datasets obtained under different scenarios:

(A) Whole body exposure – sparsely ionising radiation:
Frequency of dicentrics after exposure to Co–60 gamma rays (Romm *et al.* 2013).

(B) Whole body exposure – densely ionising radiation:
Number of dicentrics after exposure of blood samples to 1480 MeV oxygen ions (Di Giorgio *et al.* 2004).

(C) Partial body exposure – sparsely ionising radiation:
Frequencies of dicentrics + centric rings aberrations after partial body (50%) exposure with 200 kV X-rays (Heimers, 2006).

(D) Partial body exposure – densely ionising radiation:
Frequencies of dicentrics + centric rings aberrations after partial body (50%) exposure with 2.1 MeV neutrons (Heimers, 2006).

		(A)			(B)*			(C)			(D)*		
Models	k	loglik	AIC	BIC	loglik	AIC	BIC	loglik	AIC	BIC	loglik	AIC	BIC
Poisson	3	-3748.6	7503.2	7526.1	-2855.9	5717.7	5738.7	-3526.9	7059.8	7079.7	-2302.1	4608.2	4621.5
NB1	4	-3742.8	7493.7	7524.3	-2800.3	5608.6	5636.6	-3011.9	6031.7	6058.2	-2148.7	4303.3	4323.3
NB2	4	-3739.2	7486.5	7517.1	-2807.5	5623.0	5651.0	-2939.5	5887.0	5913.5	-2151.6	4309.3	4329.3
Neyman A	4	-3743.0	7493.9	7524.5	-2799.7	5607.5	5635.5	-3021.1	6050.1	6076.7	-2147.0	4299.9	4319.9
Hermite (r=2)	4	-3743.1	7494.2	7524.8	-2802.2	5612.3	5640.3	-3122.5	6253.0	6279.6	-2164.8	4335.6	4355.6
Polya-Aeppli	4	-3742.9	7493.8	7524.4	-2799.8	5607.6	5635.6	-3007.0	6022.0	6048.6	-2146.7	4299.3	4319.3
PIG	4	-3742.8	7493.5	7524.1	-2801.9	5611.8	5639.8	-3035.9	6079.8	6106.3	-2155.0	4316.1	4336.1
ZIP (a)	4	-3739.8	7487.6	7518.2	-2814.5	5637.1	5665.1	-2852.6	5713.3	5739.8	-2155.2	4316.3	4336.3
ZIP (b)	4	-3741.3	7490.5	7521.2	-2805.4	5618.7	5646.7	-2844.9	5697.8	5724.3	-2172.8	4351.6	4371.6
ZIP (c)	5	-3739.2	7488.4	7526.7	-2800.6	5611.2	5646.2	-2842.4	5694.8	5727.9	-2147.0	4302.1	4328.7
ZINB1 (a)	5	-3739.7	7489.4	7527.7	-2797.4	5604.8	5639.9	-2852.7	5715.3	5748.5	-2143.5	4294.9	4321.6
ZINB1 (b)	5	-3741.3	7492.5	7530.8	-2797.3	5604.6	5639.6	-2844.2	5698.4	5731.6	-2143.6	4295.2	4321.9
ZINB1 (c)	6	-3742.8	7497.7	7543.6	-2797.3	5606.7	5648.7	-2842.4	5696.8	5736.6	-2143.4	4296.8	4330.1
ZINB2 (a)	5	-3739.1	7488.3	7526.6	-2807.5	5624.9	5660.0	-2852.6	5715.3	5748.4	-2150.7	4309.4	4336.0
ZINB2 (b)	5	-3739.1	7488.2	7526.5	-2800.1	5610.1	5645.2	-2844.9	5699.8	5732.9	-2147.3	4302.6	4329.3
ZINB2 (c)	6	-3738.2	7488.3	7534.3	-2798.6	5609.2	5651.2	-2842.4	5696.8	5736.6	-2144.9	4299.9	4333.2
Hermite (r=3)	5	-3742.5	7495.0	7533.3	-2799.7	5609.5	5644.5	-3204.5	6419.0	6452.1	-2146.5	4301.1	4327.8
Hermite (r=4)	6	-3743.1	7498.1	7544.1	-2799.7	5611.4	5653.4	-3048.9	6109.8	6149.6	-2145.4	4300.7	4334.1

Results of fitting various models to datasets (A), (B), (C) and (D). The winning model in each column is underlined.
k is the sum of regression and model parameters.
* The number of parameters for datasets (B) and (D) is $(k-1)$ since fits where obtained by using a linear predictor without quadratic term.

Conclusions and further results

- Zero–inflated models behave well in several scenarios, especially for partial body exposure. Therefore, although for whole–body exposure and sparsely ionising radiation it is usually assumed that data follow a Poisson model, data under this scenario may depart from the Poisson model due to other circumstances (e.g., the scoring procedure).
- Results by modelling the mean yield of aberrations through a log–link were also obtained and they were similar to these ones obtained by using the identity link. The use of the log–link for the mean enabled the use of the score test for model testing which showed the adequacy of zero–inflated models.

References

1. *Cytogenetic Dosimetry: Applications in Preparedness for a Response to Radiation Emergencies*. International Atomic Energy Agency. Vienna.

2. Heimers *et al.* (2006). Chromosome aberration analysis and the influence of mitotic delay after simulated partial–body exposure with high doses of sparsely and densely ionising radiation. *Radiation and Environmental Biophysics*, **45**, 45–54.

3. Di Giorgio *et al.* (2004). Chromosome aberrations induced in human lymphocytes by heavy charged particles in track segment mode. *Radiation protection dosimetry*, **108**, 47–53.

4. Romm, H. et al. (2013). Automatic scoring of dicentric chromosomes as a tool in large scale radiation accidents. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, **756**, 174–183.

Acknowledgements

This poster presents independent research supported by the National Institute for Health Research, Research Methods Opportunity Funding Scheme entitled "Random effects modelling for radiation biodosimetry" (NIHR-RMOFS-2013-03-4). 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.