Statistical modelling in biological dosimetry

Jochen Einbeck

with thanks to E. Ainsbury, S. Barnard, D. Endesfelder, M. Higueras, and Y. Cai

Limerick, 8 November 2024









Radiation is energy in the form of waves or particles that travels through space or some material (includes heat, radio waves, light,...)



When we talk about radiation, we often mean ionizing radiation (α and β particles, γ -rays, X-rays, neutrons...), which carries enough energy to ionize atoms or molecules. Ionizing radiation can cause serious damage to cells, tissues, and DNA.

Dosimetry is the measurement of the absorbed dose delivered by ionizing radiation. The absorbed dose is measured in Gy (Joules per kg).

This talk will focus on biological dosimetry, short biodosimetry.

Biological dosimetry is based on biomarkers that measure either the damage caused by ionizing radiation (such as counts of aberrant chromosomes per cell) or the (damage-repair) response by the cell to the radiation exposure (through proteins). In either case, this (typically) produces count data.

Kayle B

Dicentric chromosomes

Frequencies of dicentric chromosomes in n = 4400 lymphocyte cells after *in vitro* 'whole body' exposure with 200 kV X-rays.

	y_{ij}								
x_i	0	1	2	3	4	5	6	7	n_i
1	1715	268	15	2	0	0	0	0	2000
2	638	298	56	8	0	0	0	0	1000
3	247	225	85	37	6	0	0	0	600
4	99	129	92	52	21	5	2	0	400
5	48	88	97	99	36	25	5	2	400

- x_i : dose (in Gy) used to irradiate blood sample $i, i = 1, \dots 5$.
- y_{ij} : counts of dicentric aberrations in *j*-th cell of blood sample *i*, $j = 1, \ldots n_i$.

These are count data, so a natural choice for the response distribution is Poisson, that is

$$f(y_{ij}|x_i) = e^{-\lambda_i} \frac{\lambda_i^{y_{ij}}}{y_{ij}!}$$

where

$$\lambda_i = E(y_{ij}|x_i)$$

is some parametric model for the dose-response relationship. For parameter estimation, firstly set up the likelihood function:

$$L = \prod_{i,j} f(y_{ij}|x_i) = \prod_{i,j} e^{-\lambda_i} \frac{\lambda_i^{y_{ij}}}{y_{ij}!} \propto \prod_i e^{-n_i \lambda_i} \lambda_i^{\sum_j y_{ij}}$$

Hence, one can conveniently work at the aggregated data level, with data $(x_i, Y_i) = (x_i, \sum_{j=1}^{n_i} y_{ij}).$

Aggregated data

Let $Y_i = \sum_j y_{ij}$. Then the aggregated data are:

Graphically, with circle size $\propto n_i$:



'Empirical dose-response curve'; this may be linear or quadratic, but in this context never exponential...

Model in terms of aggregated data: $Y_i \sim \mathsf{Pois}(n_i \lambda_i)$ with

$$\lambda_i \equiv E(Y_i|x_i)/n_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2.$$

Fitted dose-response curve from Poisson regression,

$$\hat{\lambda}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 x_i^2.$$

This curve (estimated from laboratory data) serves as calibration curve for the dose-estimation problem.



Dose estimation is an inverse regression problem:

- We have a model for the dicentric count, Y_i , given dose x_i .
- In practice, we want to estimate x_i given Y_i .

For instance, assume a patient has been admitted to hospital due to potential radiation exposure. A sample of $n_* = 200$ lymphocytes was analyzed, yielding $Y_* = 150$ dicentrics.



Mathematically, this is not a big problem. Assume the observed ratio of dicentrics ('yield') is $y_* = Y_*/n_*$. Then we have

$$y_* = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2$$

which can be solved wrt x as

$$\hat{x}_* = \frac{-\hat{\beta}_1 + \sqrt{\hat{\beta}_1^2 - 4\hat{\beta}_2(\hat{\beta}_0 - y_*)}}{2\hat{\beta}_2}$$

With $y_* = 150/200 = 0.75$, this gives

$$\hat{x}_* = 2.745.$$

Uncertainty

Of course, this estimation is not exact. There is uncertainty...

- ...due to the estimation of the calibration curve;
- ...due to random variation of Y_{*}, given x_{*}.



Combine the two sources of uncertainty ('Merkle's method', 1983):



Here, a 95% confidence interval for the 'true' dose, x_* , is given as [2.04, 3.33].

Official uncertainty assessment routine suggested by the International Atomic Energy Agency (IAEA, 2011).

Estimate *in vitro* dose-response curve $\hat{\lambda}(x)$ as before.

For the (potentially) exposed patient, count dicentrics, Y_* , in a sample of n_* cells, yielding 'test data likelihood'

$$L(Y_*|\lambda, n_*) \propto e^{-n_*\lambda} \lambda^{Y_*}$$

where $\lambda = \beta_0 + \beta_1 x + \beta_2 x^2$, and x representing dose. Assume a prior density $p(x,\lambda) = \phi(\lambda|x)p(x)$, where $\phi(\lambda|x) \sim N(\hat{\lambda}(x), \operatorname{Var}(\hat{\lambda}(x)))$. Use Bayes's theorem to obtain posterior density for (λ, x) :

$$p(\lambda, x|Y_*) \propto L(Y_*|\lambda, n_*)p(x, \lambda)$$

Integration over λ gives calibrative density of x:

$$p(x|Y_*) \propto p(x) \int L(Y_*|\lambda, n_*) \phi(\lambda|x) \, d\lambda$$

Integral has explicit solution via Hermite distribution.

Consider again the example before: Patient sample with $n_* = 200$, $Y_* = 150$.

Use the same estimated doseresponse curve, $\hat{\lambda}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 x_i^2$, as before:



A (semi-)Bayesian approach to UQ (Higueras et al, 2015)

Calibrative density for 'true' dose x, using R package radir:



A (semi-)Bayesian approach to UQ (Higueras et al, 2015)

Dose estimate: Mode of calibrative density:



A (semi-)Bayesian approach to UQ (Higueras et al, 2015)

Uncertainty assessment: 95% Credible intervals



CI = [2.48, 3.01].

Compare Merkle: [2.04, 3.33].

Generally considered as "gold-standard" for biological dosimetry.

- Little inter-individual or inter-lab variation;
- Little overdispersion (hence Poisson model is appropriate);
- can be adapted to deal with partial body exposures;
- well supported by software.

However, some disadvantages:

- Need to wait 2 to 3 days until metaphase in mitosis;
- Need experienced cytogeneticists for the 'scoring' of aberrations;
- Potential biases (cell death, repair).

Alternative biomarker: The γ -H2AX foci assay

- Histones are proteins which help to package the DNA double helix.
- Following radiation-induced double strand breaks, the H2AX histone phosphorylates, in this state referred to as γ-H2AX.
- The resulting foci can be counted manually or in a semi-automated way, using immunofluorescence microscopy.
- Typically, one examines a sample of 500-2000 (blood) cells on a given 'slide' and then records the number of foci per cell.



Foci yield (that is foci/cell, out of 500 sample cells) versus design dose:



- Strong (linear?) dose-response relationship; strong decay from 1h to 24h after exposure
- H2AX-based dose estimation has to happen within 24 hours of exposure!
- Considerable variation, so Uncertainty Quantification crucial

Denote again Y_i the total foci count from n_i cells for slide i, i = 1, ..., k.

As before, one can easily fit linear [or quadratic] curves

$$\lambda_i \equiv E(Y_i|x_i)/n_i$$

= $A + Bx_i [+Cx_i^2].$

Linear calibration curves appear sufficient for this assay.



However, variance-to-mean ratio (dispersion) now >> 1!

In fact, linear-model based dispersion estimate

$$\hat{\phi} = {\sf Deviance}/(N-2),$$

with $N = \sum_{i=1}^{k} n_i$, gives: $\frac{\text{time } 1h \quad 24h}{\hat{\phi} \quad 57.91 \quad 54.26}$



Quasi-Poisson regression

Overdispersed linear Poisson model

$$E(Y_i|x_i) = An_i + B(n_i x_i); \qquad \qquad \mathsf{Var}(Y_i|x_i) = \phi E(Y_i|x_i)$$

Score equations for this model

$$\frac{1}{\phi} \sum_{i=1}^{k} \begin{pmatrix} 1 \\ x_i \end{pmatrix} (Y_i - n_i \lambda_i) / \lambda_i = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

 \ldots so the estimates of A and B do not depend on $\phi!$

- However, standard errors do depend on ϕ , namely

$$SE(\hat{A}) = \sqrt{\hat{\phi}} \operatorname{SE}_P(\hat{A});$$
 $\operatorname{SE}(\hat{B}) = \sqrt{\hat{\phi}} \operatorname{SE}_P(\hat{B}).$

Dose estimation

Inverse regression: For a new yield y_* , one has

$$\hat{x}_* = \frac{y_* - \hat{A}}{\hat{B}}.$$

UQ via delta-method:



5

foci / cell 4 6 8

dose [Gv]

This accounts for intra- and inter-individual variation, but still requires calibration curve to be 'correct'

Complication: The calibration curve may vary with laboratory, scorer, equipment etc. Hence, a given calibration curve needs to be validated before use.

Before examining a patient sample, lab should irradiate two reference samples at 0Gy and 1.5Gy and compare yields with prediction interval:

- If inside, validated
- If outside, a new calibration curve can be computed from the reference samples which still allows dose estimation, albeit at a higher variance (Einbeck et al, 2018)



Web applet

Activities 🍵 Google Chrome 🕶	Thu 15:52	7 # Q *
DoseEstimateH2AX ×		(1) Jackan ×
← → C ☆ ③ Not secure shinur.unitigia.es/taps/h2axDE/		R 🕁 🖾 🗿 🗗 🛛

DoseEstimateH2AX



yo 0.4625

Main Plot Information

Calibration curve: A + B D

Background yield, Ar. 0.1489 (0.011); Liberar dose effect, 19659 (0.0158); Dispersion index, & 158.753. More: The builth an cateful on une, PFFS 240 X-rays calibration curve has been validated for the introduced reference samples, and consequently used in this dose estimation.

Summary of dose estimation

Point estimate: 0.3171 Gy. Standard error: 0.2432 Gy. 95% confidence interval: (-0.1595, 0.7937) Gy.

Dispersion index available? Observed sample

Reference samples available?

у.	0.77	n,	200
Cor	npute		
Down	load report?	③ Yes	* No

¥r 2.92

Yes
No

○ 1h ※ 24h

⊛ Yes ⊙ No

n₀ 400

Yes @ No

n_r 200

Web applet

Activities 🛛 🏮 Google Chrome 🕶 Thu 15:52 A # 0 -DoseEstimateH2AX × (1) Jackson ← → C ☆ ③ Not secure | shinur.unitioja.es/apps/h2acDE/ DoseEstimateH2AX Main Plot Information Calibration Own calibration curve? Yes ® No Time after exposure ○ 1h ※ 24h Reference samples available? ⊛ Yes ⊙ No ne 400 yo 0.4625 ¥r 2.92 n_r 200 f 1.5 Dispersion index available? Yes @ No 2 Observed sample y- 0.77 n- 200 ool yield 5 Compute 8 * 8 0.2 0.4 0.6 0.8 Dose Gy

Dosimetry with the γ -H2AX assay



RESEARCH ARTICLE

A statistical framework for radiation dose estimation with uncertainty quantification from the γ-H2AX assay

Jochen Einbeck¹, Elizabeth A. Ainsbury², Rachel Sales¹, Stephen Barnard², Felix Kaestle³, Manuel Higueras^{4,5}

1 Department of Mathematical Sciences, Dunham University, Dunham, United Kingdom, 2 Public Her, England, Chemical and Environmental Hazards, Chilton, Didoct, United Kingdom, 3 Bundesamt für Strahlenschrutz, Fachberrech Strahlenschutz und Gesundheit, Oberschleissheim, Gemany, 4 Depa de Matemäticas y Comuzacion, Universidad de La Rioja, Logrofio, La Rioja, Spain, 5 Basque Cente Applied Mathematics, Bibao, Basque Country, Spain

• jochen.einbeck@durham.ac.uk

Abstract

OPEN ACCESS

Pilation Einhack I Ainchus EA Calas D Damard

Over the last decade, the $\gamma\!-\!H2AX$ focus assay, which exploits the phosphorylation c H2AX histone following DNA double–strand–breaks, has made considerable progre



Why is the dispersion value (≈ 60) such high?

Calibration data from dosimetry units usually come in one of two forms:

Raw data (columns with all counts for a given slide)



Aggregated data (one row for slide, with averaged foci count)

Donor	Cond	n	Average	dose
	L B	50	7.52	0.5
	4 C	10	8.4	4
1	5 B	50	7.28	0.5
	5 C	40	10.9	4
	7 C	10	6.9	4
	9 D	50	0.14	0
1	θE	50	7.1	0.5
1	9 F	46	8.5	4
10	D	26	0.538462	0
10	DD	50	7.6	0.5
10	D F	30	7.66	4
1	L D	50	0.14	0

Accordingly one would use raw or aggregated data for the model fitting including the dispersion estimation.

Of course, if one has the raw data available, then we can also aggregate them, and fit both (raw and aggregated data models) to compare.

	Ra	aw	Aggregated		
	PHE-Foci1	BfS-Foci	PHE-Foci1	BfS-Foci	
$(\hat{eta}_0,\hat{eta}_1)$	(0.766, 1.700)	(2.011, 5.746)	(0.766, 1.700)	(2.011, 5.746)	
$(SE(\hat{\beta}_0), SE(\hat{\beta}_1))$	(0.042, 0.058)	(0.009, 0.023)	(0.213, 0.298)	(0.102, 0.248)	
$\hat{\phi}$	1.444	1.223	37.70	147.99	
$SE[\hat{\phi}]$	0.049	0.004	37.70	0.16	
ν	1198	233218	2	114	
$\chi^{2}_{ u,0.95}/ u$	1.068	1.005	2.996	1.227	

Table 6.1 Parameter estimates along with their associated standard errors and dispersion estimates obtained from each model. The last row gives the critical value that $\hat{\phi}$ would be compared with in a Poisson goodness-of-fit test at the 5% level of significance.

While these high dispersion estimates are clearly biased, they do correct a problem which sits elsewhere, namely unaccounted correlations in the raw data structure. Fitting raw data models naively will lead to incorrect results:



We suggest to either use the aggregated model for calibration curve fitting, or a raw data model with random effects for slides.

Data aggregation and dispersion

DE GRUYTER

Int. J. Biostat. 2022; 18(1): 183-202

ิล

Adam Errington*, Jochen Einbeck, Jonathan Cumming, Ute Rössler and David Endesfelder

The effect of data aggregation on dispersion estimates in count data models

https://doi.org/10.1515/ijb-2020-0079 Received May 29, 2020: accepted April 21, 2021: published online May 7, 2021

Abstract: For the modelling of count data, aggregation of the raw data over certain subgroups or predictor configurations is common practice. This is, for instance, the case for count data biomarkers of radiation exposure. Under the Poisson law, count data can be aggregated without loss of information on the Poisson parameter, which remains true if the Poisson assumption is relaxed towards quasi-Poisson. However, in biodosimetry in particular, but also beyond, the question of how the dispersion estimates for quasi-Poisson models behave under data aggregation have received little attention. Indeed, for real data sets featuring unexplained heterogeneities, dispersion estimates can increase strongly after aggregation, an effect which we will demonstrate and quantify explicitly for some scenarios. The increase in dispersion estimates implies an inflation of the parameter standard errors, which, however, by comparison with random effect models, can be shown to serve a corrective purpose. The phenomena are illustrated by *r*-H2AX foci data as used for instance in radiation biodosimetry for the calibration of dose-response curves.

- The γ-H2AX assay has a strong time-dependency. Ideally, we could estimate simultaneously exposure time and dose.
- Basic idea: If one had TWO measurements of foci yield at unknown points but with known time distance, one has two pieces of information to work out dose and exposure time from the relevant equations.
- Our work takes the premise that, for a given lab, two lab-specific calibration curves at *any* time points are available, and that the physical decay mechanism between measurement points is universal (not lab-specific).

We know from the literature that γ -H2AX foci decay follows a double-exponential law.

Hence a suitable model is

$$y = a + bt + d(Ae^{ut} + Be^{vt}) \tag{1}$$

where a, b, A, B, u, v are constants.

Assume further that we have two patient samples available (considered as realizations of model (1)), at time points t_0 and $t_0 + \Delta$, with known Δ .

The values t_0 are d are to be inferred.

The parameters u and v are determined by the physical decay mechanism and hence we fix them at u = -0.35 and v = -0.018 (Horn et al, 2011).

The parameters a, b, A, B are related to the absolute magnitude of foci observed, and hence are lab-specific. We can obtain them by equating the calibration curves $y_t = a_t + d \times b_t$, for $t = t_1, t_2$ to the model equation:

For the 'background radiation'

$$a_t = a + bt$$

For the dose-dependent part

$$b_t = Ae^{-0.35t} + Be^{-0.018t}$$

Note these are two equations each, at $t = t_1, t_2$.

For the actual dose and time estimation,

- set up a system of two equations of type (1) with our two yield measurements at times t_0 and $t_0 + \Delta$,
- solve this system for d and t₀, for which we use the Mathematica solver "NSolve".

The calibration curves of the previously shown PHE data

 $y_1 = 0.13 + 12.56d$

$$y_{24} = 0.18 + 1.94d$$

lead to the system of equations

$$y_t = 0.13 + 0.0021t + d(13.62e^{-0.35t} + 3.01e^{-0.018t})$$

$$y_{t+\Delta} = 0.13 + 0.0021(t+\Delta) + d(13.62e^{-0.35(t+\Delta)} + 3.01e^{-0.018(t+\Delta)})$$

which can be solved for any two yields y_t , $y_{t+\Delta}$ at unknown times t, $t+\Delta$ with known Δ .

Simulation (from MSc dissertation Y. Cai)

100 NB count data sets simulated from y_1, y_{24} ($\phi = 50, n = 500$)



Impact of number of cells on dose estimates



Impact of number of cells on time estimates



Estimating dose and time

Home > Developments in Statistical Modelling > Conference paper

Estimating Dose and Time of Exposure from a Protein-Based Radiation Biomarker



radiation, one needs to have an estimation of the dose of radiation received by the individual. In the context of a protein-based biomarker for radiation exposure, we present here a new method that, unlike the approaches that produce the estimation with data collected at a predetermined time after exposure, allows us to estimate the dose at any time within a reasonable time interval after exposure, as well as determine the time of exposure if needed. Namely, we take existing calibration curves and generalize them using the decay mechanism of y-H2AX foci to build a model that describes the functional relationship between the count of γ -H2AX foci in exposed blood cells and the time and dose of exposure. This model is illustrated using both real and simulated data.



Estimating dose and time

Home > Developments in Statistical Modelling > Conference paper **Estimating Dose and Time of Exposure** from a Protein-Based Radiation Biomarker Conference paper | First Online: 12 July 2024 pp 239-245 | Cite this conference paper (IWSM 2024) Yilun Cai 🖂, Jochen Einbeck, Stephen Barnard & Elizabeth Ainsbury Acce Part of the book series: Contributions to Statistics ((CONTRIB.STAT.)) Login Included in the following conference series: International Workshop on Statistical Modelling Chapter Available as PI a 309 Accesses Read on any de Instant downli Abstract Own it forever In order to analyze the potential damage to the human body caused by exposure to ionizing radiation, one needs to have an estimation of the dose of radiation received by the ✓ eBook

individual. In the context of a protein-based biomarker for radiation exposure, we present here a new method that, unlike the approaches that produce the estimation with data collected at a predetermined time after exposure, allows us to estimate the dose at any time within a reasonable time interval after exposure, as well as determine the time of exposure if needed. Namely, we take existing calibration curves and generalize them using the decay mechanism of y-H2AX foci to build a model that describes the functional relationship between the count of γ -H2AX foci in exposed blood cells and the time and dose of exposure. This model is illustrated using both real and simulated data

Developments i

✓ Hardcover Be

Institutional subs

Tax calculation

Purchases

(Experimental UQ in progress... but encounters some difficulties as numerical

solvers are involved...).

Biodose Tools

biodosetools 3.6.1.9000

Overview ₂₀

Reference Articles - Changelog

Biodose Tools



Links View on CRAN Browse source code Report a bug Read documentation

Search for

License Full license

Citation Citing biodosetools

Developers

Author, maintainer 💿

Biodose Tools (<u>Hernández et al. 2023</u>) is an open source project that aims to be a tool to perform all different tests and calculations needed by biological dosimetry laboratories. The app is developed using the <u>R</u> programming language and <u>Shiny</u> as a framework to offer an online, easyto-use solution. Although the intention is to provide the application as a website, all R routines are available as an R package, which can be downloaded for improvement or personal use.

We also aim to clarify and explain the tests used and to propose those considered most appropriate. Each laboratory in its routine work should choose the most suitable method, but the project aims to reach a consensus that will help us in case of mutual assistance or intercomparisons.

The project is initially developed by <u>RENEB</u> association, but contributions are always welcome.

Original Articles

Biodose Tools: an R shiny application for biological dosimetry



Abstract

Introduction

In the event of a radiological accident or incident, the aim of biological dosimetry is to convert the yield of a specific biomarker of exposure to ionizing radiation into an absorbed dose. Since the 1980s, various tools have been used to deal with the statistical procedures needed for biological dosimetry, and in general those who made several

Related Re



RENEB/EURAL robust dose es conditions bas

- available as R Shiny app, can be downloaded from CRAN or Github
- hosted and managed by the BfS, Munich
- currenty supports
 - Dicentrics
 - Translocations
 - Micronuclei
- Durham-based RA (Y. Zhang) currently working on $\gamma\text{-H2AX}$ extension

Software for biodosimetric analysis:

- **BioDoseTools** (Frequentist) dose estimation for dicentric chromosomes, micronuclei, and translocations. Available as R Shiny App: https://aldomann.shinyapps.io/biodosetools-v3/
- radir (Bayesian) dose estimation, mainly for dicentric chromosomes. R package Available on CRAN: https://cran.r-project.org/package=radir
- **DoseEstimateH2AX** (Frequentist) dose estimation for the γ -H2AX assay. Available as R Shiny App: https://shinur.unirioja.es/apps/h2axDE/

- IAEA (2011) International Atomic Energy Agency, Cytogenetic Dosimetry: Applications in Preparedness for and Response to Radiation Emergencies, IAEA, Vienna.
- **Horn S et al** (2011). Gamma-H2AX-based dose estimation for whole and partial body radiation exposure. *PloS ONE*, **6**(9): e25113.
- Higueras M et al. (2015). A new inverse regression model applied to radiation biodosimetry. *Proc's of the Royal Society* **471**.