Quasi-Poisson regression models for radiation dose estimation from biomarkers

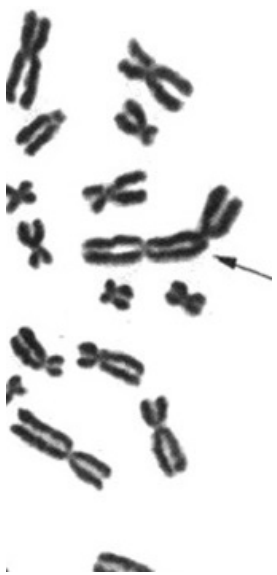
Jochen Einbeck

joint work with Rachel Sales, Liz Ainsbury, Stephen Barnard, Manuel Higueras

December 15, 2018

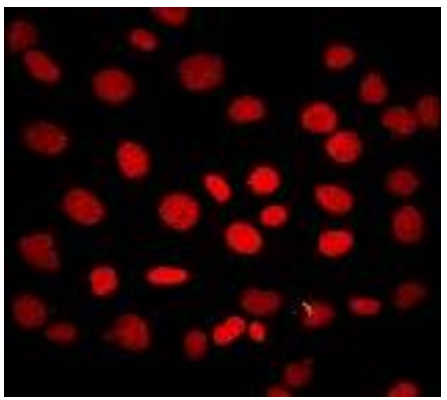
Radiation biodosimetry

- Radiation incident leading to (potentially) exposed individuals
- Contracted radiation dose can be estimated retrospectively by exploiting the radiation-induced change in certain biomarkers
- 'Gold–standard': Dicentric chromosomes (resulting from unsuccessful DNA–damage response)
- However: time– and work–intensive and expensive methodology



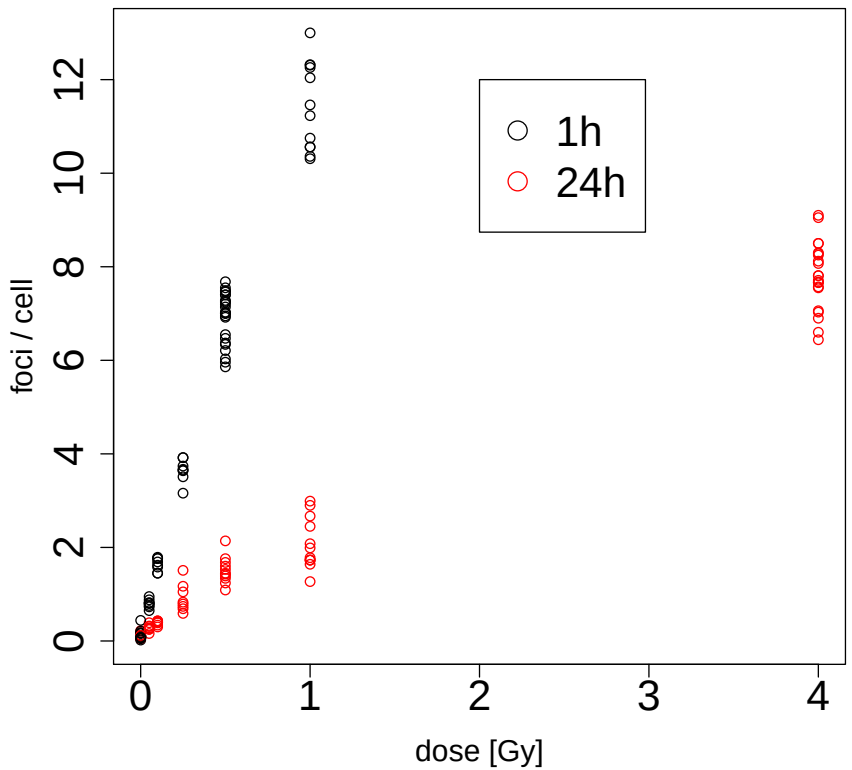
γ-H2AX as radiation biomarker

- Biomarkers based on proteins have recently emerged as a quicker and cheaper alternative
- The H2AX–histone responds to radiation–induced double strand breaks with phosphorylation, in this state then referred to as γ–H2AX
- γ–H2AX foci can be counted manually (immunofluorescence microscopy) or automated (flow cytometers)



H2AX calibration data from PHE

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

Calibration curve estimation

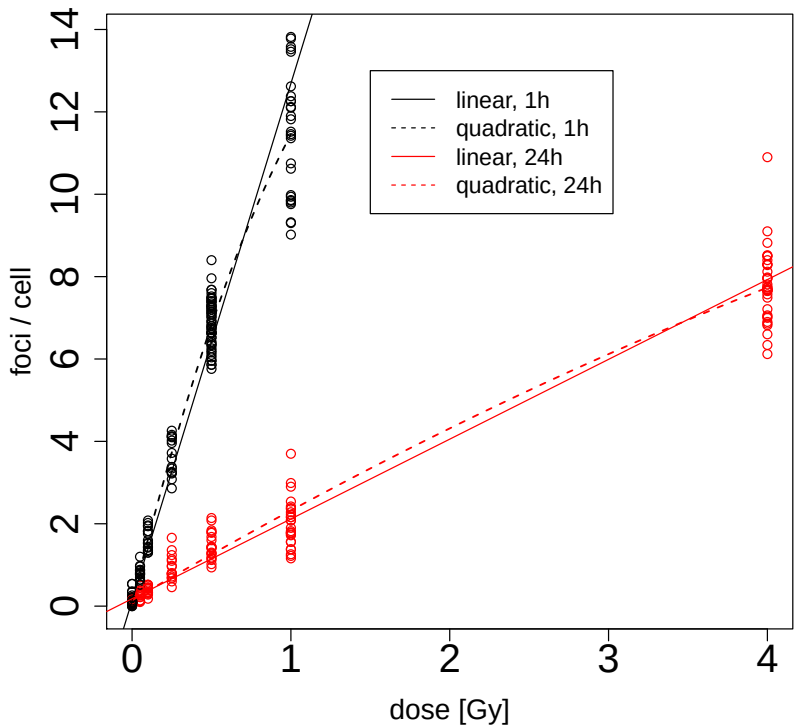
- For fixed time after exposure, calibration data (x_i, y_{ij}), $j = 1, \ldots, n_i$, with y_{ij} "yield of the j-th sample (of n cells) for dose x_i".

- Count data (Poisson) regression with 'identity link'

$$\mu_i \equiv E(y_{ij}|x_i) = A + Bx_i$$

(Quadratic models also considered, but discarded)

- Huge overdispersion ($\hat{\phi} \approx 60$!)



Quasi–Poisson regression

- Overdispersed Poisson model for $Y_{ij} = ny_{ij}$,

$$E(Y_{ij}|x_i) = nA + B(nx_i); \qquad \text{Var}(Y_{ij}|x_i) = \phi E(Y_{ij}|x_i)$$

Score equations for this model (with $\mu_i = A + Bx_i$),

$$\frac{1}{\phi}\sum_{i=1}^{d}\sum_{j=1}^{n_i}\begin{pmatrix} 1 \\ x_i \end{pmatrix}(Y_{ij} - n\mu_i)/\mu_i = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

... so the estimates of A and B do not depend on ϕ!

- However, standard errors do depend on ϕ, namely

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

Estimating dispersion

- 'Quick-and-dirty', from estimated Poisson model

$$\hat{\phi} = \frac{\text{Deviance}}{N-2}$$

- We use

$$\hat{\phi} = \frac{1}{N-2}\sum_{i=1}^{d}\sum_{j=1}^{n_i}\frac{(Y_{ij}-n\hat{\mu}_i)^2}{n\hat{\mu}_i} = \frac{n}{N-2}\sum_{i=1}^{d}\sum_{j=1}^{n_i}\frac{(y_{ij}-\hat{\mu}_i)^2}{\hat{\mu}_i}$$

  where $\hat{\mu}_i = \hat{A} + \hat{B}x_i$ and $N = \sum_{i=1}^{d} n_i$.

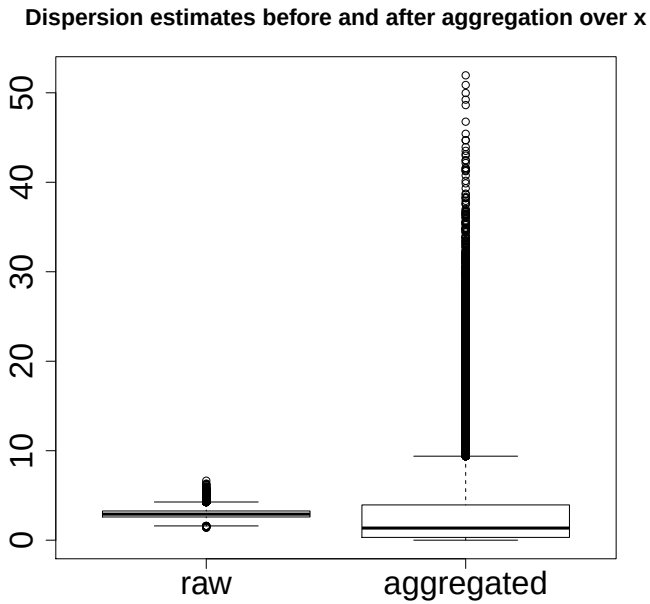
- Each additional covariate will increase the value '2' in the denominator by 1.

Estimating dispersion: Warning!

- Dispersion estimates can become poor if data are aggregated.

- Simulation example:
  10000 data sets of size 100 from
  $y_i \sim \text{NB1}(\mu_i, 2)$, with
  $\mu_i = 2 + 5x_i$ and
  $x_i \sim \text{Bin}(2, 0.5)$.

- That is, $\phi = 3$ known.



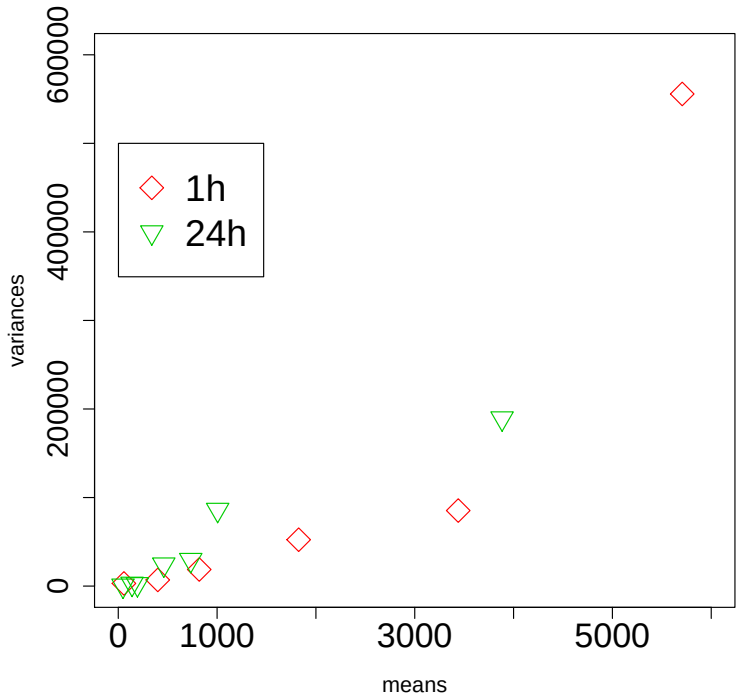
- Even though the average of estimates $\hat{\phi}$ are almost identical (2.96 and 2.98, resp.), *each individual dispersion estimate from the aggregated data is close to useless.*

Dispersion from calibration data

- Our calibration data are 'half–aggregated' (over cells; not over dose).

- Estimated dispersions:

$\hat{\phi}$	1h	24h
'quick-and-dirty'	57.91	54.26
GLM–based	59.56	57.55

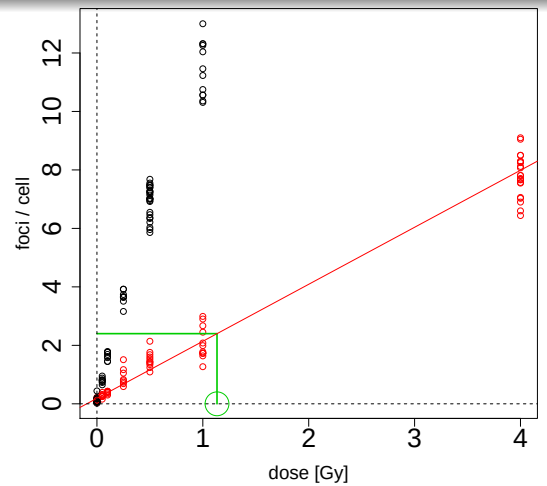


- $\hat{\phi} \approx 60$ appears a reasonable assumption.

Dose estimation

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

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



- UQ via delta–method:

$$\begin{aligned} SE^2(x_*) &= \left(\frac{\partial x_*}{\partial \hat{A}}\right)^2 SE^2(\hat{A}) + \left(\frac{\partial x_*}{\partial \hat{B}}\right)^2 SE^2(\hat{B}) + \left(\frac{\partial x_*}{\partial y_*}\right)^2 SE^2(y_*) \\ &= \frac{1}{\hat{B}^2} SE^2(\hat{A}) + \frac{(y_* - \hat{A})^2}{\hat{B}^4} SE^2(\hat{B}) + \frac{1}{\hat{B}^2} \frac{\hat{\phi} y_*}{n_*} \end{aligned}$$

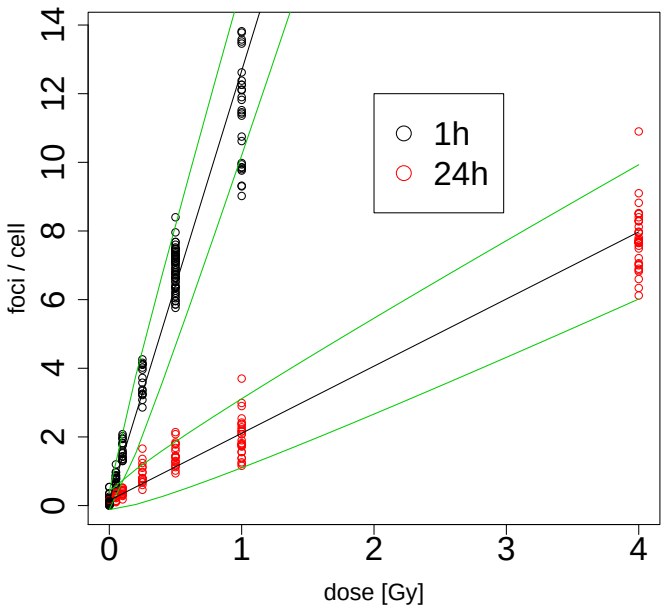
- This accounts for intra– and inter–individual variation, but still requires calibration curve to be ‘correct’

Curve validation

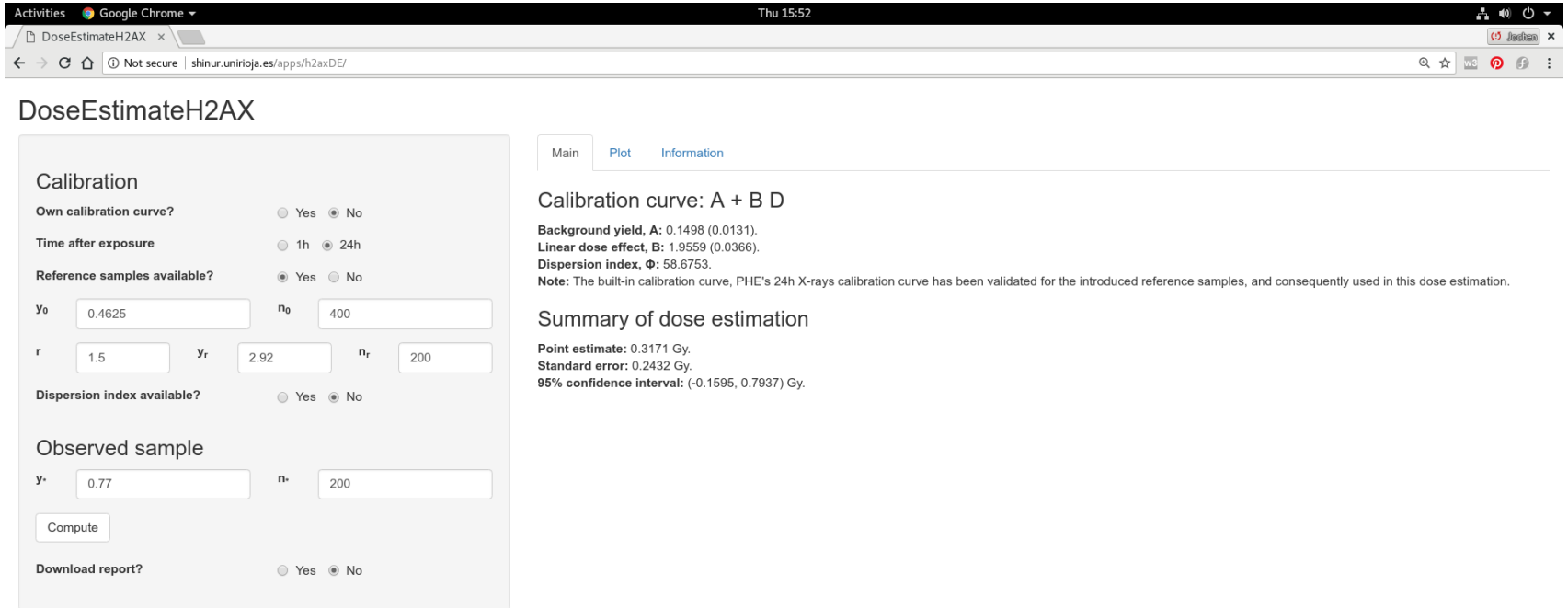
Complication: The calibration curve may vary with laboratory, scorer, equipment etc. Hence, a given calibration curve needs to be validated before use.

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

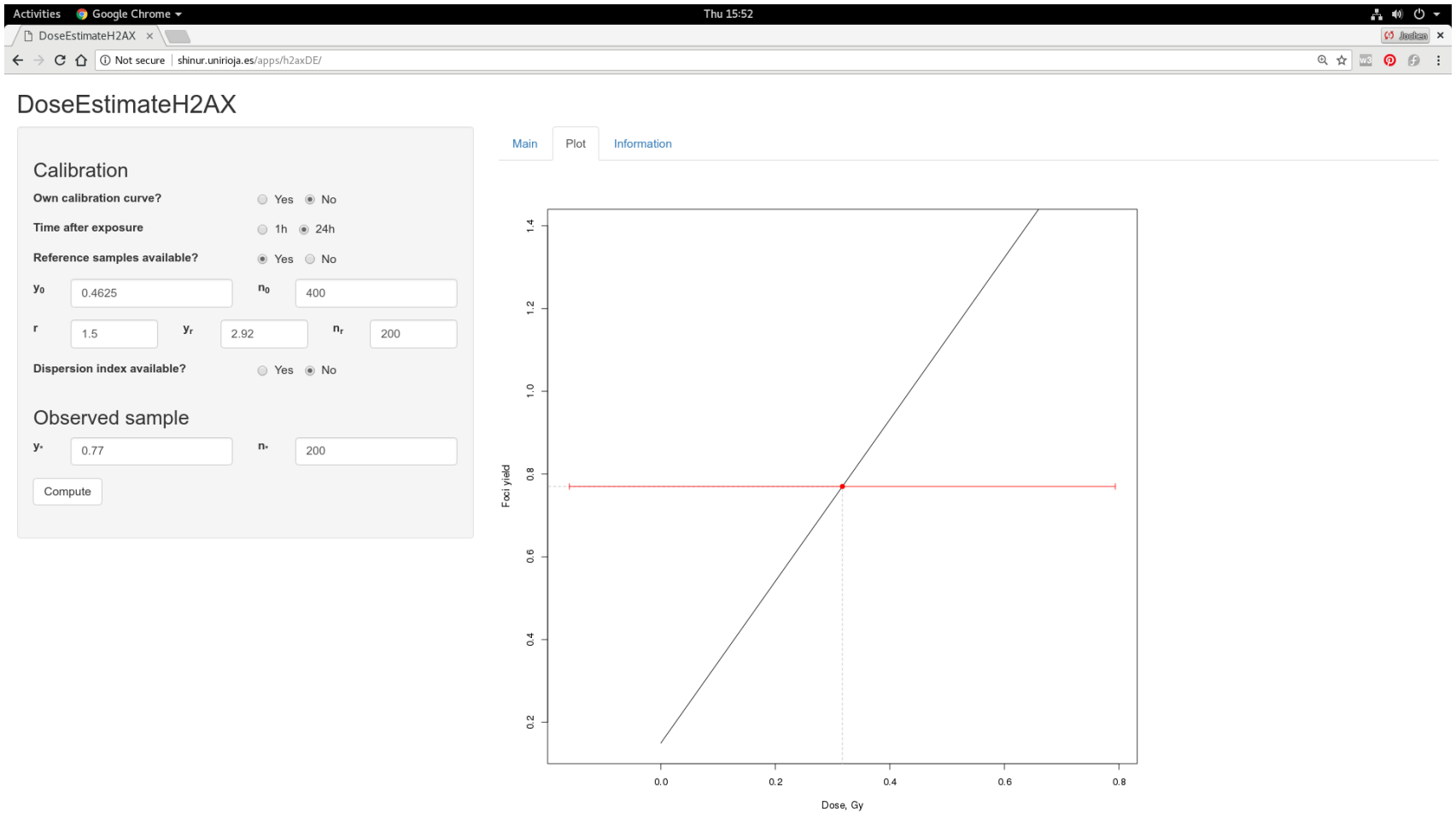
- If inside, validated
- If outside, a new calibration curve can be computed from the reference samples which still allows dose estimation, albeit at a higher variance



Web applet


Activities
Google Chrome
Thu 15:52
DoseEstimateH2AX
Jochen
Not secure
shinur.unirioja.es/apps/h2axDE/
DoseEstimateH2AX
Calibration
Own calibration curve?
Yes
No
Time after exposure
1h
24h
Reference samples available?
Yes
No
y0
0.4625
n0
400
r
1.5
yr
2.92
nr
200
Dispersion index available?
Yes
No
Observed sample
y*
0.77
n*
200
Compute
Download report?
Yes
No
Main
Plot
Information
Calibration curve: A + B D
Background yield, A: 0.1498 (0.0131).
Linear dose effect, B: 1.9559 (0.0366).
Dispersion index, Φ: 58.6753.
Note: The built-in calibration curve, PHE's 24h 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.

Web applet


Activities
Google Chrome
Thu 15:52
DoseEstimateH2AX
Not secure
shinur.unirioja.es/apps/h2axDE/
DoseEstimateH2AX
Calibration
Own calibration curve?
Yes
No
Time after exposure
1h
24h
Reference samples available?
Yes
No
y0
0.4625
n0
400
r
1.5
yr
2.92
nr
200
Dispersion index available?
Yes
No
Observed sample
y.
0.77
n.
200
Compute
Main
Plot
Information
Foci yield
Dose, Gy
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0.0
0.2
0.4
0.6
0.8

Summary

We believe to have solved the questions regarding

- ... the incorporation of overdispersion
- ... the validation of the calibration curve
- ... the quantification of uncertainty in this process

Open questions are

- Is the dispersion a 'universal property of foci counts', or a feature of the scoring mechanism?
- Partial body exposure cannot be easily identified since overdispersion will be present either way

Use for other (count data) radiation biomarkers

- Micronuclei: Overdispersion usually present but small (≈ 1.5), risk of incorrect estimation of ϕ possibly larger than benefits. If full count data distributions available then NB model preferable.
- Dicentric chromosomes
  - under partial body exposure: Overdispersion is due to zero–inflation; so direct use of ZIP models preferable.
  - under densely ionizing radiation: Overdispersion present but situation similar to micronuclei.
- Other protein-based biomarkers, such 53PB1
  - here probably useful, but yet to be tried...

Future directions

- Combinations of Biomarkers
  Idea: Use quick and cheap (but potentially high variance) biomarkers such as H2AX for the triage step, and a more precise biomarker (such as the dicentric assay) as a second step, depending on the first outcome. Bayesian approach favorable here.

- Why always estimate 'dose'?
  It seems to be an irrevocable standard to always estimate *dose*. Why not triage directly based on the H2AX count? This removes need for inverse regression, and reduces uncertainties. Could be dealt with easily through ordinal logistic regression.

References

Ainsbury EA et al (2017). Uncertainty of fast biological radiation dose assessment for emergency response scenarios. *International Journal of Radiation Biology* **93**, 127–135.

Einbeck J et al (2017). On the Use of Random Effect Models for Radiation Biodosimetry. In: Extended Abstracts Fall 2015. Ainsbury EA et al, *Research Perspectives CRM Barcelona* **7**, 89–94, Springer.

Einbeck J et al (2018). A statistical framework for radiation dose estimation from the γ–H2AX assay. PLoS ONE **13**(11):e0207464.

Oliveira M et al (2016). Zero–inflated regression models for radiation–induced chromosome aberration data: A comparative study. *Biometrical Journal* **58**, 259-79.