Bayesian solutions to biodosimetry count data problems, and software solutions

M Higueras, P Puig, EA Ainsbury, D Moriña, VA Vinnikov and K Rothkamm
The Bayesian approach

A Bayesian approach is highly applicable to ionising radiation dosimetry data. It allows the cytogenetic experts to consider prior knowledge surrounding an overexposure scenario.

This approach implies an accurate measure of the uncertainty of dose estimates.

The calibrative density is the solution of the Bayesian inverse regression problem,

$$P(x|y) \propto P(x) \int L(y|\Theta)P(\Theta|x)d\Theta.$$
Groer & Pereira (1987)

Poisson responses and just linear dose-response (without intercept).

\( Y_i \sim \text{Pois}(\alpha d_i), \ \alpha \sim \text{Gamma}(a, b) \implies \alpha|Y \sim \text{Gamma}(a + S, b + N), \)

\( S \) is the total number of aberrations and

\[
N = \sum n_i d_i.
\]

The absorbed dose

\( D \sim \text{Gamma}(A, B), \)

and the calibrative dose density results

\[
p(D|Y) \propto \frac{D^{A+s-1}}{e^{BD}(nD + a + N)^{S+s+b}}.
\]
Brame & Groer (2002)

NB responses with just linear dose-response (without intercept)

\[ Y_i \sim \text{NB}(\alpha d_i, \Psi), \quad \Psi \sim \text{Gamma}, \quad P(\alpha, \Psi) = P(\alpha)P(\Psi) \]

and the priors for \( \alpha \) are Uniform or Normal.

The NB model is compared to the Poisson one using the Bayes Factor

\[
BF = \frac{\int L_{\text{NB}}(Y|\alpha, \Psi)P(\alpha)P(\Psi)d\alpha d\Psi}{\int L_{\text{Pois}}(Y|\alpha)P(\alpha)d\alpha}.
\]

The calibrative dose density results

\[
P(D|Y) \propto P(D) \int \frac{\Gamma(\Psi^{-1} + s)}{\Gamma(s + 1)\Gamma(\Psi^{-1})} \left( \frac{\Psi n\alpha D}{1 + \Psi n\alpha D} \right)^s \left( \frac{1}{1 + \Psi n\alpha D} \right)^{1/\Psi} P(\alpha, \Psi|Y)d\alpha d\Psi.
\]

The integrals in this methodology are done using numerical integration.
Other Bayesian works in biodosimetry

- **Madruga et al. 1994, 1996**: log-normal model.

- **Kottas et al. 2002**: non-parametric model.

- **Serna et al. 2008**: Jeffrey’s prior for analyzing the background distribution.
Whole body exposure

Poisson models

The posterior of the population mean is approximated to a normal, by asymptotic normality of the posterior distribution for large samples and the delta-method, i.e.:

\[ \mu | D \sim N \left( f(D, \hat{\beta}), \nabla \cdot \hat{\Sigma} \cdot \nabla^T \right) . \]

The calibrative density results

\[ P(D | Y) \propto P(D) P(X_D = s), \]

where \( X_D \) is Hermite distributed. If \( \mu | D \) is approximated by a gamma, \( X_D \) is NB distributed.
Example: Romm et al. (2013)

- Blood samples from 8 healthy donors were irradiated in vitro with $^{60}$Co gamma-rays at a high-dose rate simulating acute whole body exposure.
- Dicentrics assay: Poisson responses and quadratic dose-response (without intercept).
- The 1.5 Gy sample is removed from the calibration dataset to be used as test data.
Example: Romm et al. (2013)

<table>
<thead>
<tr>
<th>Model</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode</td>
<td>1.430</td>
<td>1.430</td>
<td>1.443</td>
</tr>
<tr>
<td>Expected</td>
<td>1.432</td>
<td>1.432</td>
<td>1.445</td>
</tr>
<tr>
<td>SD</td>
<td>0.081</td>
<td>0.081</td>
<td>0.078</td>
</tr>
<tr>
<td>95% CI LB</td>
<td>1.277</td>
<td>1.277</td>
<td>1.294</td>
</tr>
<tr>
<td>95% CI UB</td>
<td>1.594</td>
<td>1.593</td>
<td>1.602</td>
</tr>
</tbody>
</table>

(a): normal mean prior, \( U(0, \infty) \) dose prior.

(b): gamma mean prior, \( U(0, \infty) \) dose prior.

(c): gamma mean prior, \( \text{Ga}(21.8, 12.4) \) dose prior.
Compound Poisson models

The joint posterior of the population mean and the dispersion index is defined:

$$(\mu, \delta) | D \sim N_2 \left( (f(D, \hat{\beta}), \hat{\delta}), \nabla \cdot \hat{\Sigma} \cdot \nabla^T \right).$$

The calibrative density can be defined directly and calculated by numerical integration (not computationally intensive, always bivariate: $D$ and $\delta$).

Fixing $\delta$ by its MLE, the model is reduced and the mean prior is applied like in the Poisson models. The resulting calibrative density is in terms of Compound-Hermite (-NB) mass function.
Example: Puig & Valero (2006)

- 11 samples of peripheral blood exposed to different doses of γ-rays (0.93 cGy/min \(^{-1}\) dose rate). For each sample, approximately 5000 binucleated cells were inspected.
- MN assay: NB responses and quadratic dose-response.
- The 0.1 Gy sample is used as test data.

<table>
<thead>
<tr>
<th>dose (Gy)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>(\bar{y})</th>
<th>(d)</th>
<th>(u)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>4887</td>
<td>106</td>
<td>5</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.024</td>
<td>1.556</td>
<td>7.839</td>
</tr>
<tr>
<td>0.10</td>
<td>4773</td>
<td>206</td>
<td>19</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.050</td>
<td>1.150</td>
<td>7.526</td>
</tr>
<tr>
<td>0.25</td>
<td>4261</td>
<td>324</td>
<td>41</td>
<td>12</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>0.090</td>
<td>1.306</td>
<td>15.306</td>
</tr>
<tr>
<td>0.50</td>
<td>4536</td>
<td>364</td>
<td>76</td>
<td>17</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td>0.119</td>
<td>1.449</td>
<td>22.484</td>
</tr>
<tr>
<td>0.75</td>
<td>4383</td>
<td>512</td>
<td>85</td>
<td>18</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>0.149</td>
<td>1.257</td>
<td>12.876</td>
</tr>
<tr>
<td>1.00</td>
<td>4225</td>
<td>636</td>
<td>115</td>
<td>19</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>0.189</td>
<td>1.240</td>
<td>12.009</td>
</tr>
<tr>
<td>1.50</td>
<td>4018</td>
<td>805</td>
<td>139</td>
<td>26</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td></td>
<td>0.243</td>
<td>1.270</td>
<td>13.495</td>
</tr>
<tr>
<td>2.00</td>
<td>3499</td>
<td>1194</td>
<td>238</td>
<td>45</td>
<td>13</td>
<td>10</td>
<td>1</td>
<td></td>
<td>0.383</td>
<td>1.209</td>
<td>10.471</td>
</tr>
<tr>
<td>2.50</td>
<td>3171</td>
<td>1313</td>
<td>393</td>
<td>94</td>
<td>24</td>
<td>3</td>
<td>2</td>
<td></td>
<td>0.501</td>
<td>1.201</td>
<td>10.077</td>
</tr>
<tr>
<td>3.00</td>
<td>2582</td>
<td>1575</td>
<td>598</td>
<td>190</td>
<td>44</td>
<td>9</td>
<td>2</td>
<td>6</td>
<td>0.722</td>
<td>1.206</td>
<td>10.307</td>
</tr>
<tr>
<td>4.00</td>
<td>1974</td>
<td>1674</td>
<td>869</td>
<td>342</td>
<td>102</td>
<td>26</td>
<td>13</td>
<td>2</td>
<td>1.013</td>
<td>1.172</td>
<td>8.628</td>
</tr>
</tbody>
</table>

Bayesian solutions to biodosimetry count data problems, and software solutions
Example: Puig & Valero (2006)

\[ \mathbf{D} \sim \mathcal{U}(0, \infty) \]

- **(a):** simplified with normal mean prior.
- **(b):** simplified gamma mean prior.
- **(c):** complete.

<table>
<thead>
<tr>
<th>Model</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode</td>
<td>0.115</td>
<td>0.115</td>
<td>0.125</td>
</tr>
<tr>
<td>Expected</td>
<td>0.114</td>
<td>0.114</td>
<td>0.124</td>
</tr>
<tr>
<td>SD</td>
<td>0.034</td>
<td>0.034</td>
<td>0.033</td>
</tr>
<tr>
<td>95% CI LB</td>
<td>0.047</td>
<td>0.047</td>
<td>0.059</td>
</tr>
<tr>
<td>95% CI UB</td>
<td>0.182</td>
<td>0.181</td>
<td>0.190</td>
</tr>
</tbody>
</table>
Conclusions of Higueras et al. (2015a)

- New whole body cytogenetic dose estimation Bayesian-like models for Poisson and two parameter compound Poisson responses are presented.
- The approach is valid for any given dose-response function one time differentiable in the parameter set domain.
- To use this methodology, only the estimates of the parameters and covariance matrix of the dose-response curve are required.

https://cran.r-project.org/web/packages/radir/
Bayesian solutions to biodosimetry count data problems, and software solutions
Video tutorial

http://polimedia.uab.cat/#v_592

Bayesian solutions to biodosimetry count data problems, and software solutions
Bayesian solutions to biodosimetry count data problems, and software solutions
Example: Pujol et al. (2014)

The mean of dicentrics per cell as a function of the absorbed dose is

\[
f(x; \beta) = \beta_0 e^{-\beta_1} e^{-\beta_2 x} \left( 1 + \frac{\beta_3 x (2 \beta_0 e^{-\beta_1} e^{-\beta_2 x} + 1)}{1 + \beta_3 x (\beta_0^2 (e^{-\beta_1} e^{-\beta_2 x})^2 + \beta_0 e^{-\beta_1} e^{-\beta_2 x})} \right).
\]

We assume Poisson responses. The 17 Gy test sample consists in 914 dicentrics scored in 100 cells. Assuming the gamma mean prior and a $U(0, \infty)$ prior dose.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Dose (Gy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode</td>
<td>16.143</td>
</tr>
<tr>
<td>Expected</td>
<td>16.814</td>
</tr>
<tr>
<td>SD</td>
<td>1.778</td>
</tr>
<tr>
<td>95% CI LB</td>
<td>14.148</td>
</tr>
<tr>
<td>95% CI UB</td>
<td>19.953</td>
</tr>
</tbody>
</table>

**Probability Density**

Dose, $D$, Gy
Partial body exposure

Bayes factor

The Bayesian alternative for PBI:

- Decision: does the sample come from a partial body exposure? In contrast to the frequentist $u$-test, the Bayarri et al. 2008 Bayes factor is proposed (ZIP vs. Poisson):

$$BF = \frac{n_0!}{(n + 1)!} \sum_{j=0}^{n_0} \frac{(n - j)!}{(n_0 - j)!} \left(1 - \frac{j}{n}\right)^{-\left(s + \frac{1}{2}\right)},$$

where $n$, $n_0$ and $s$ are respectively the sample size and frequency of zeros, and the sum of the total number of chromosomal aberrations.

- Once the ZIP assumption is supported, a new Bayesian model is proposed for the dose and fraction of the body irradiated estimation.
The model

The frequency of aberrations per cell is

\[ Z \sim ZIP \left( \mu, \frac{1 - F}{Fe^{-D/d_0} - F + 1} \right), \]

where analogously to Higuera et al. (2015a)

\[ \mu | D \sim \text{Gamma} \left( \frac{f(D, \hat{\beta})^2}{v(D, \hat{\beta})}, \frac{f(D, \hat{\beta})}{v(D, \hat{\beta})} \right) \]

an application of Bayes’ theorem shows the expression of the likelihood of \( D, F \) and \( d_0 \) for the given test data

\[ L(y|D, F, d_0) \propto (Fe^{-D/d_0} - F + 1)^{-n} \sum_{j=1}^{n_0} \binom{n_0}{j} \frac{F^{n-j}(1-F)^j}{(n-j)^s} P(X_j = s|D), \]

where \( X_j \) is a random variable negative binomial distribution with mean and variance depending on \( j \) and \( D \).
The model

Considering $D$, $F$ and $d_0$ as independent random variables, and

$$D \sim \text{Gamma} \left( \frac{\hat{D}^2}{\sigma_D^2}, \frac{\hat{D}}{\sigma_D^2} \right); \quad F \sim \mathcal{U}(0, 1); \quad d_0 \sim \mathcal{U}(2.7, 3.5);$$

the joint posterior density,

$$
P(D, F, d_0|y) = \frac{L(y|D, F, d_0)P(D, F, d_0)}{\int L(y|D, F, d_0)P(D, F, d_0)\,dDdFd\!d_0},$$

has a non-tractable form. The acceptance-rejection sampling is used to simulate the posterior distribution.
Example

In a recent experiment to simulate PBI, unirradiated and irradiated blood at each dose was mixed. For instance, the exposed blood fraction comprised 10% for 2 Gy, whose distribution of dicentrics plus centric rings is

<table>
<thead>
<tr>
<th>#Dic+CR</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>#cells</td>
<td>1043</td>
<td>16</td>
<td>3</td>
</tr>
</tbody>
</table>

The Bayes factor value for this sample gives ‘strong’ evidence in support to the ZIP assumption, because $6 < 2 \log BF = 9.41 < 10$.

The calibration dose response data is taken from Vinnikov et al. 2013.
Marginal posteriors
Joint posterior and statistic summary

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Dose (Gy)</th>
<th>Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode</td>
<td>1.39</td>
<td>0.11</td>
</tr>
<tr>
<td>Expected</td>
<td>1.50</td>
<td>0.13</td>
</tr>
<tr>
<td>SD</td>
<td>0.50</td>
<td>0.05</td>
</tr>
<tr>
<td>95% CI LB</td>
<td>0.66</td>
<td>0.07</td>
</tr>
<tr>
<td>95% CI UB</td>
<td>2.58</td>
<td>0.25</td>
</tr>
</tbody>
</table>
Conclusions

- Novel solutions for statistical analysis of cytogenetic biological dosimetry data have been developed.
- New Bayesian models have been created and applied in practical cytogenetic dose estimation.
- Some of these models have been implemented in the R statistical software for biodosimetry laboratory researchers.
- These new solutions lead to more accurate quantification of statistical uncertainty associated with cytogenetic dose estimates.


28 Bayesian solutions to biodosimetry count data problems, and software solutions
Acknowledgements
...and all the pieces matter.

Thanks for your attention!