Supplementary materials 1. Statistical methods

1.1. *Poisson distribution*

In a wide range of situations, the Poisson distribution is used to model count data. Rutherford and Geiger, for example, utilised it in their famous experiment in 1910, in which they counted the number of α-particles emitted from a polonium source over a period of time. In Biodosimetry, the Poisson distribution is commonly employed.

If $Y \sim \text{Pois}(\lambda)$, *i.e.* a discrete random variable, Y, follows a Poisson distribution with rate $\lambda > 0$, then its probability mass function is:

$$\Pr(Y = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k \in \{0, 1, 2, \dots\} \tag{S1.1}$$

Both the expected value and the variance of Y are equal to the rate λ, *i.e.*:

$$\text{E}(Y) = \lambda, \quad \sigma^2 = \text{Var}(Y) = \lambda. \tag{S1.2}$$

For real count data, the usual estimators of the rate parameter and the variance are,

$$\hat{\lambda} = \bar{y} = \frac{X}{N}, \tag{S1.3}$$

$$\hat{\sigma}^2 = \frac{1}{N-1}\left[\sum_{k=1}^{M} k^2 C_k - N\bar{y}^2\right] = \frac{1}{N-1}\left[\sum_{k=1}^{M} k^2 C_k - \frac{1}{N}\left(\sum_{k=1}^{M} k C_k\right)^2\right], \tag{S1.4}$$

where C_k is the count of cells where k chromosomal aberrations were detected, M is the maximum count realisation, $N = \sum_{k=0}^{M} C_k$ is the total number of cells analysed, and $X = \sum_{k=1}^{M} kC_k$ is the total number of chromosomal aberrations.

1.2. *Goodness of fit*

The goodness of fit of the fitted curve and significance of estimated coefficients should then be tested, for instance using an appropriate form of the F-test, z-test or t-test. **biodosetools** implements the t-test.

Let $\hat{\theta}$ be an estimator of the parameter $\theta \in \{\alpha, \beta, C\}$ in the fit model. Then the t-statistic for this parameter is defined as

$$t_{\hat{\theta}} = \frac{\hat{\theta}}{\widehat{se}(\hat{\theta})}, \tag{S1.5}$$

where $\widehat{se}(\hat{\theta})$ is the standard error of $\hat{\theta}$.

1.3. *Uncertainty on dose estimation*

1.3.1. Whole-body assessment: Merkle's method

The simplest solution for whole-body assessment was proposed by Merkle (1983), it allows both the Poisson error on the yield and the errors on the calibration curve to be taken into account.

Merkle's approach, illustrated in Figure S1.1, involves the following steps:

1. Assuming the Poisson (or quasi-Poisson) distribution, calculate the yields corresponding to lower and upper confidence limits on the observed yield λ (λ_L and λ_U).
2. Calculate the confidence limits of the dose-effect calibration curve according to:

$$\lambda = C + \alpha D + \beta G(x) D^2 \pm R\sqrt{\sigma_C^2 + \sigma_\alpha^2 D^2 + \sigma_\beta^2 G(x) D^4 + 2\sigma_{C,\alpha} D + 2\sigma_{C,\beta} G(x) D^2 + 2\sigma_{\alpha,\beta} G(x) D^3}, \tag{S1.6}$$

   where R^2 is the confidence factor, defined as an upper-confidence limit of a chi-square distribution, $\chi^2(\nu)$, with 2 or 3 degrees of freedom (ν).
3. Calculate the dose at which λ crosses the dose-effect calibration curve. This is the estimated dose (D). For this we can simply take the inverse of the LQ dose-effect calibration curve, as follows:

$$D = \frac{-\alpha + \sqrt{\alpha^2 + 4\beta G(x)(\lambda - C)}}{2\beta G(x)}. \tag{S1.7}$$

4. Calculate the dose at which λ_L crosses the upper curve. This is the lower confidence limit of the dose (D_L).
5. Calculate the dose at which λ_U crosses the lower curve. This is the upper confidence limit of the dose (D_U).

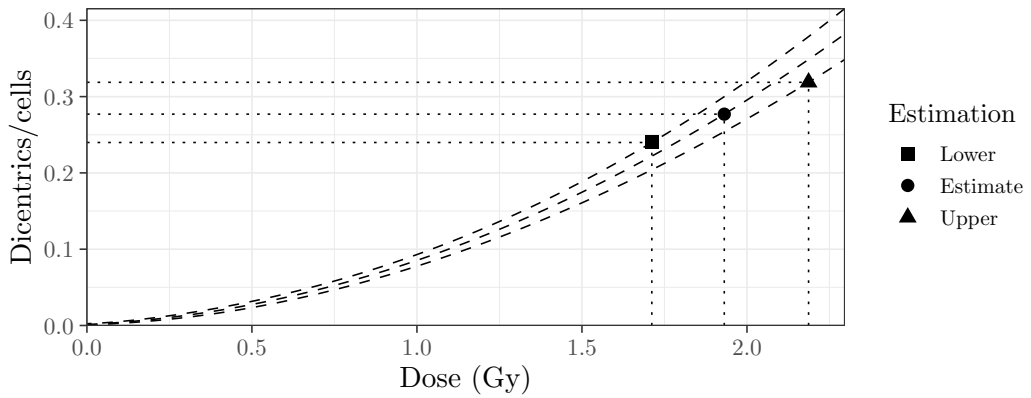

Figure S1.1. A dose-effect calibration curve with its 83% confidence limits, used to estimate dose uncertainties using Merkle's method.

As suggested by some authors, in order to reduce a possible overestimation of the uncertainty (Schenker and Gentleman 2001; Austin and Hux 2002), it is recommended to use an 83% confidence limit of the regression curve when overlapped with the Poisson

83% confidence limits, as this tends to yield a global 95% confidence interval for the dose estimation.

1.3.2. Whole-body assessment: delta method

Another approach for whole-body assessment is using the delta method (**(alias?)**). It also allows both the Poisson error on the yield and the errors on the calibration curve to be taken into account.

The delta method expands a function of a set of random variables $f(X_1, X_2, \ldots, X_n)$ about its mean, usually with a first-order Taylor expansion, and then takes the variance. Using the delta method, the variance σ_f^2 of $f(X_1, X_2, \ldots, X_n)$ can be expressed as follows (Klein 1953):

$$\sigma_f^2 = \sum_i^n \left(\frac{\partial f}{\partial X_i}\right)^2 \sigma_{X_i}^2 + \sum_i^n \sum_{j \neq i}^n \frac{\partial f}{\partial X_i}\frac{\partial f}{\partial X_j}\sigma_{X_i,X_j}. \tag{S1.8}$$

The approach using the delta method, illustrated in Figure S1.2, involves the following steps:

1. Calculate the dose at which λ crosses the dose-effect calibration curve. This is the estimated dose (D). For this we can simply take the inverse of the LQ dose-effect calibration curve, as shown in (S1.7).
2. By differentiation of the above equation, express the variance on the estimated dose (σ_D^2) in terms of the variances and co-variances of C, α, β, and λ. The formal expression is as follows:

$$\begin{aligned}\sigma_D^2 = &\left(\frac{\partial D}{\partial C}\right)^2 \sigma_C^2 + \left(\frac{\partial D}{\partial \alpha}\right)^2 \sigma_\alpha^2 + \left(\frac{\partial D}{\partial \beta}\right)^2 \sigma_\beta^2 + \left(\frac{\partial D}{\partial \lambda}\right)^2 \sigma_\lambda^2 \\ &+ 2\frac{\partial D}{\partial C}\frac{\partial D}{\partial \alpha}\sigma_{C,\alpha} + 2\frac{\partial D}{\partial C}\frac{\partial D}{\partial \beta}\sigma_{C,\beta} + 2\frac{\partial D}{\partial \alpha}\frac{\partial D}{\partial \beta}\sigma_{\alpha,\beta}.\end{aligned} \tag{S1.9}$$

   Note that this derivation assumes that the covariances $\sigma_{\lambda,C}$, $\sigma_{\lambda,\alpha}$, and $\sigma_{\lambda,\beta}$ are all zero. This is appropriate because, in general, the measurements used to calculate the calibration curve and those used to determine the patient's yield are independent. The variance and co-variances on C, α, and β are derived from the fitted calibration curve, whereas the variance on λ is derived according to (S1.4).
3. The lower and upper 95% confidence limits of the dose estimation (D_L and D_U) and the yield (λ_L and λ_U) are then calculated as follows:

$$D_{L,U} = D \pm 1.96\sigma_D, \tag{S1.10}$$

$$\lambda_{L,U} = \lambda \pm 1.96\sigma_\lambda. \tag{S1.11}$$

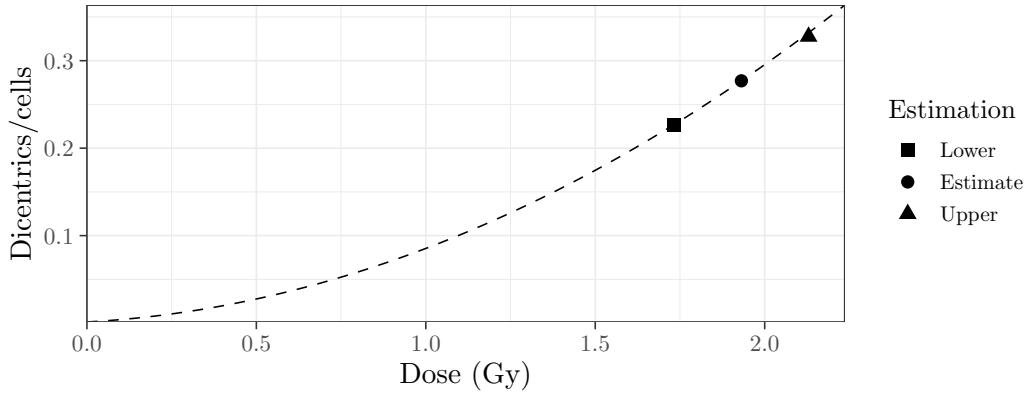

Figure S1.2. A dose-effect calibration curve used to estimate dose uncertainties using delta method.

1.3.3. Partial-body assessment: Dolphin's method

This method was first proposed by Dolphin (1969) and is based on a contaminated Poisson method. The chromosomal aberrations distribution of a partial-body irradiation is expected to be overdispersed ($u > 1.96$). The observed distribution is considered to be the mixture of (a) a Poisson distribution which represents the irradiated fraction of the body and (b) the remaining unexposed fraction. This method assumes that the background level in the unirradiated part is zero. Undamaged cells will comprise two subpopulations: those from the unexposed fraction and irradiated cells which received no damage.

The distribution of the damage in the irradiated cells can be described by a zero-truncated Poisson distribution with probability function,

$$\Pr(Y = k \mid Y > 0) = \frac{P(Y = k)}{1 - P(Y = 0)} = \frac{\lambda^k}{(e^\lambda - 1)k!}, \quad k \in \{1, 2, \dots\} \tag{S1.12}$$

where $Y \sim \text{Pois}(\lambda)$. The distribution of the total number of dicentrics in all the cells can be understood as a zero-inflated Poisson distribution $Z \sim \text{ZIP}(\lambda, \pi)$ with probability function,

$$P(Z = k) = \begin{cases} \pi + (1 - \pi)e^{-\lambda}, & k = 0 \\ (1 - \pi)\dfrac{\lambda^k e^{-\lambda}}{k!}, & k \in \{1, 2, \dots\} \end{cases} \tag{S1.13}$$

where π is the proportion of extra zeros.

The approach using Dolphin's method, illustrated in Figure S1.3, involves the following steps:

1. Calculate the yield of dicentrics (λ) in the irradiated fraction by solving the equation,

$$\frac{X}{N - C_0} = \frac{\lambda}{1 - e^{-\lambda}}, \tag{S1.14}$$

where X is the total number of dicentrics observed, N is the total number of cells, and C_0 is the number of cells free of dicentrics. The solution λ of this equation is

the maximum likelihood estimator (MLE) of the mean yield of dicentrics λ of the zero-truncated Poisson distribution (S1.12).

2. Calculate the fraction of cells scored which were irradiated ($f = 1 - \pi$) as follows:

$$f = \frac{X}{N\lambda}. \tag{S1.15}$$

3. Calculate the variance on the yield (σ_λ^2), the fraction of cells scored which were irradiated (σ_f^2), and their covariance ($\sigma_{f,\lambda}$) by inverting the observed Fisher information matrix (I) of the zero-inflated model:

$$\begin{pmatrix} \sigma_\lambda^2 & \sigma_{f,\lambda} \\ \sigma_{f,\lambda} & \sigma_f^2 \end{pmatrix} = I^{-1}(\lambda, \pi = 1 - f), \tag{S1.16}$$

where

$$I(\lambda, \pi = 1 - f) = \begin{pmatrix} Nf\left(\dfrac{f-1}{f+(1-f)e^\lambda} + \dfrac{1}{\lambda}\right) & \dfrac{N}{f+(1-f)e^\lambda} \\ \dfrac{N}{f+(1-f)e^\lambda} & N\dfrac{e^\lambda - 1}{f(f+(1-f)e^\lambda)} \end{pmatrix}. \tag{S1.17}$$

4. Calculate the dose at which λ crosses the dose-effect calibration curve. This is the estimated dose (D). For this we can simply take the inverse of the LQ dose-effect calibration curve, as shown in (S1.7).
5. Calculate the variance on the estimated dose (σ_D^2) using the delta method, as shown in (S1.9). The variance and co-variances on C, α, and β are derived from the fitted calibration curve, whereas the variance on λ is derived from the inverse of the observed Fisher information matrix (S1.16).
6. Calculate the initial fraction of irradiated cells (F):

$$F = \frac{f}{f + (1-f)e^{-D/D_0}}, \tag{S1.18}$$

where f is the fraction of cells scored which were irradiated, D is the expression of the estimated dose (S1.7), and D_0 is the survival coefficient with experimental values between 2.7 and 3.5 (Lloyd, Purrott, and Dolphin 1973; Barquinero et al. 1997).

7. Calculate the variance on the initial fraction of irradiated cells (σ_F^2) using the delta method:

$$\begin{aligned} \sigma_F^2 = &\left(\frac{\partial F}{\partial f}\right)^2 \sigma_f^2 + \left(\frac{\partial F}{\partial C}\right)^2 \sigma_C^2 + \left(\frac{\partial F}{\partial \alpha}\right)^2 \sigma_\alpha^2 + \left(\frac{\partial F}{\partial \beta}\right)^2 \sigma_\beta^2 + \left(\frac{\partial F}{\partial \lambda}\right)^2 \sigma_\lambda^2 \\ &+ 2\frac{\partial F}{\partial f}\frac{\partial F}{\partial \lambda}\sigma_{f,\lambda} + 2\frac{\partial F}{\partial C}\frac{\partial F}{\partial \alpha}\sigma_{C,\alpha} + 2\frac{\partial F}{\partial C}\frac{\partial F}{\partial \beta}\sigma_{C,\beta} + 2\frac{\partial F}{\partial \alpha}\frac{\partial F}{\partial \beta}\sigma_{\alpha,\beta}. \end{aligned} \tag{S1.19}$$

Note that this derivation assumes that the covariances $\sigma_{\lambda,C}$, $\sigma_{\lambda,\alpha}$, $\sigma_{\lambda,\beta}$, $\sigma_{f,C}$, $\sigma_{f,\alpha}$, and $\sigma_{f,\beta}$ are all zero. The variance and co-variances on C, α, and β are derived from the fitted calibration curve, whereas the variances on λ and f and covariance $\sigma_{f,\lambda}$ are derived from the inverse of the observed Fisher information matrix (S1.16).

8. The lower and upper 95% confidence limits of the dose estimation (D_L and D_U), the yield (λ_L and λ_U) are calculated following (S1.10) and (S1.11), respectively, and the initial fraction of irradiated cells (F_L and F_U) is then calculated as follows:

$$F_{L,U} = F \pm 1.96\sigma_F. \tag{S1.20}$$

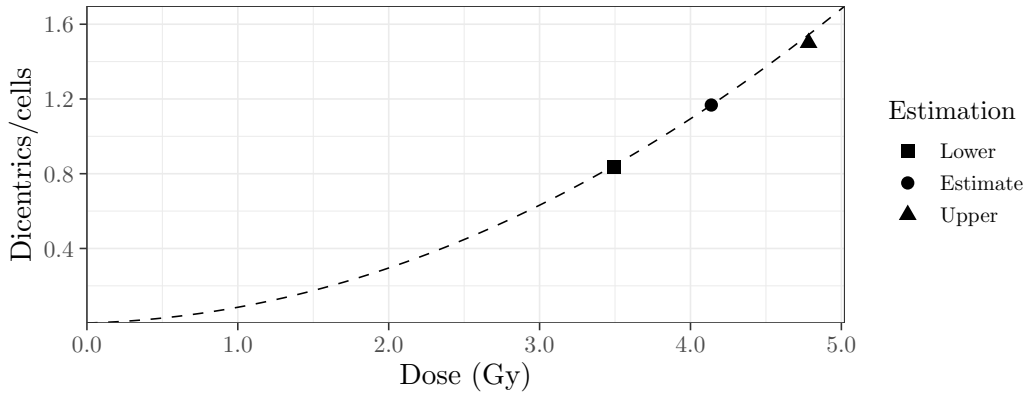

Figure S1.3. A dose-effect calibration curve used to estimate dose uncertainties using Dolphin's method.

1.3.4. Heterogeneous assessment: mixed Poisson model

So far, all non-homogeneous exposures have been handled as partial-body exposures with one fraction of the body uniformly irradiated at a certain dose, while the rest of the body is not exposed and hence without chromosome aberrations. This, however, represents a rather idealised situation, since the majority of accidents involve non-uniform exposures, where mixing of almost homogeneously irradiated and nonirradiated blood is extremely unlikely.

To remedy this, a mathematical approach based on a mixed Poisson model (S1.21) that can be used in cases of suspected non-homogeneous exposures with two different doses was proposed by Pujol et al. (2016). This model allows to infer two different distributions from an observed dicentric cell distribution.

For a heterogeneous exposure with two radiation doses x_1 and x_2, the distribution outcome of dicentrics is a mixture of two Poisson distributions (S1.1). A random variable Y distributed as a mixture of two independent Poisson distributions with rates λ_1 and λ_2 has the following probability mass function:

$$\Pr(Y = k) = \omega \frac{\lambda_1^k e^{-\lambda_1}}{k!} + (1 - \omega) \frac{\lambda_2^k e^{-\lambda_2}}{k!}, \tag{S1.21}$$

where λ_1 represents the yield of dicentrics for the dose x_1, λ_2 represents the yield for the dose x_2 and ω, a parameter between 0 and 1, represents the population proportion of scored cells that have received a dose x_1. Similarly, $1 - \omega$ can be understood as the population proportion of scored cells that have received a dose x_2.

The approach using the mixed Poisson model, illustrated in Figure S1.4, involves the following steps:

1. Calculate the maximum likelihood estimates for the yields (λ_1 and λ_2) and the fraction of scored cells that have received a dose D_1 ($f_1 = \omega$) using an optimisation method, such as Limited-memory BFGS-B (Byrd et al. 1995).

2. Calculate the variances on the yields and the fraction of scored cells that have received a dose D_1 ($\sigma^2_{\lambda_1}$, $\sigma^2_{\lambda_2}$, and $\sigma^2_{f_1}$) by inverting the observed Fisher information matrix (I) resulting from the aforementioned optimisation method:

$$\begin{pmatrix} \sigma^2_{f_1} & \sigma_{f_1,\lambda_1} & \sigma_{f_1,\lambda_2} \\ \sigma_{f_1,\lambda_1} & \sigma^2_{\lambda_1} & \sigma_{\lambda_1,\lambda_2} \\ \sigma_{f_1,\lambda_2} & \sigma_{\lambda_1,\lambda_2} & \sigma^2_{\lambda_2} \end{pmatrix} = I^{-1}. \tag{S1.22}$$

3. Calculate doses at which λ_1 and λ_2 cross the dose-effect calibration curve. These are the estimated doses (D_1 and D_2). For this we can simply take the inverse of the LQ dose-effect calibration curve, as shown in (S1.7).
4. Calculate the variance on the estimated doses ($\sigma^2_{D_i}$) using the delta method, as shown in (S1.9). The variances and co-variances on C, α, and β are derived from the fitted calibration curve, whereas the variances on λ_i are derived from the inverse of the observed Fisher information matrix (S1.22).
5. Calculate the initial fraction of irradiated cells (F_1):

$$F_1 = \frac{f_1}{f_1 + (1 - f_1)e^{-\gamma(D_1 - D_2)}}, \quad F_2 = 1 - F_1, \tag{S1.23}$$

where f_1 is the fraction of cells scored which were irradiated at the highest dose, D_1 and D_2 are the estimated doses, $\gamma = 1/D_0$ is the survival coefficient which is a constant value calculated experimentally from each culture treatment.
6. Calculate the variance on the initial fraction of irradiated cells ($\sigma^2_{F_1} = \sigma^2_{F_2}$) using the delta method:

$$\sigma^2_{F_1} = \left(\frac{\partial F_1}{\partial \gamma}\right)^2 \sigma^2_\gamma + \left(\frac{\partial F_1}{\partial f_1}\right)^2 \sigma^2_{f_1} + \left(\frac{\partial F_1}{\partial D_1}\right)^2 \sigma^2_{D_1} + \left(\frac{\partial F_1}{\partial D_2}\right)^2 \sigma^2_{D_2}, \tag{S1.24}$$

where the variance on γ is obtained experimentally, the variance on f_1 is derived from the optimisation method, and the variances on D_i are obtained using the delta method.
7. The lower and upper 95% confidence limits of the dose estimations ($D_{i,L}$ and $D_{i,U}$), the yields ($\lambda_{i,L}$ and $\lambda_{i,U}$), and the initial fractions of irradiated cells ($F_{i,L}$ and $F_{i,U}$) are calculated following (S1.10), (S1.11), and (S1.20), respectively.

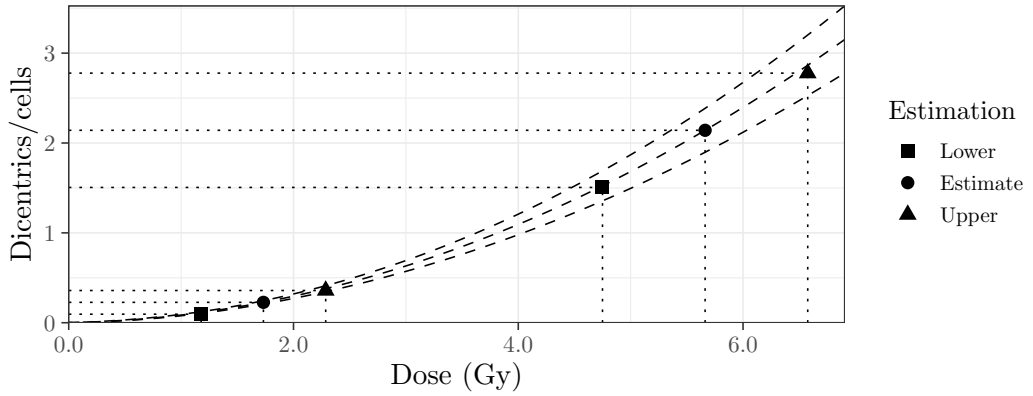

Figure S1.4. A dose-effect calibration curve used to estimate dose uncertainties using heterogeneous method.

References

Austin, Peter C., and Janet E. Hux. 2002. "A brief note on overlapping confidence intervals." *Journal of Vascular Surgery* 36 (1): 194–195. https://linkinghub.elsevier.com/retrieve/pii/S0741521402000307.

Barquinero, J. F., L. Barrios, M. R. Caballin, R. Miro, M. Ribas, and J. Egozcue. 1997. "Biological dosimetry in simulated in vitro partial irradiations." *International Journal of Radiation Biology* 71 (4): 435–440. https://doi.org/10.1080/095530097144058.

Byrd, Richard H., Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. 1995. "A Limited Memory Algorithm for Bound Constrained Optimization." *SIAM Journal on Scientific Computing* 16 (5): 1190–1208. https://doi.org/10.1137/0916069.

Dolphin, G.W. 1969. "Biological Dosimetry with Particular Reference to Chromosome Aberration Analysis: A Review of Methods." In *Handling of Radiation Accidents*, edited by International Atomic Energy Agency, 215–224. Vienna. https://inis.iaea.org/search/search.aspx?orig{_}q=RN:45029080.

International Atomic Energy Agency. 2001. *Cytogenetic Analysis for Radiation Dose Assessment.* Technical Reports Series 405. Vienna: International Atomic Energy Agency. https://www.iaea.org/publications/6303.

Klein, L.R. 1953. *A Textbook of Econometrics.* Row, Peterson. https://books.google.co.uk/books?id=uzwiAAAAMAAJ.

Lloyd, D C, R J Purrott, and G W Dolphin. 1973. "Chromosome aberration dosimetry using human lymphocytes in simulated partial body irradiation." *Physics in Medicine and Biology* 18 (3): 421–431. https://doi.org/10.1088/0031-9155/18/3/007.

Merkle, W. 1983. "Statistical methods in regression and calibration analysis of chromosome aberration data." *Radiation and Environmental Biophysics* 21 (3): 217–233. http://link.springer.com/10.1007/BF01323412.

Pujol, Mònica, Leonardo Barrios, Pedro Puig, María Rosa Caballín, and Joan-Francesc Barquinero. 2016. "A New Model for Biological Dose Assessment in Cases of Heterogeneous Exposures to Ionizing Radiation." *Radiation Research* 185 (2): 151–162. http://www.bioone.org/doi/10.1667/RR14145.1.

Schenker, Nathaniel, and Jane F. Gentleman. 2001. "On Judging the Significance of Differences by Examining the Overlap Between Confidence Intervals." *The American Statistician* 55 (3): 182–186. http://www.tandfonline.com/doi/abs/10.1198/000313001317097960.