

## Improving the accuracy of dose estimates from automatically scored dicentric chromosomes by accounting for chromosome number

David Endesfelder , Ulrike Kulka , Jochen Einbeck & Ursula Oestreicher

To cite this article: David Endesfelder , Ulrike Kulka , Jochen Einbeck & Ursula Oestreicher (2020) Improving the accuracy of dose estimates from automatically scored dicentric chromosomes by accounting for chromosome number, International Journal of Radiation Biology, 96:12, 1571-1584, DOI: [10.1080/09553002.2020.1829152](https://doi.org/10.1080/09553002.2020.1829152)

To link to this article: <https://doi.org/10.1080/09553002.2020.1829152>



Copyright © 2020 The Author(s). Published with license by Taylor & Francis Group, LLC.



[View supplementary material](#)



Published online: 16 Oct 2020.



[Submit your article to this journal](#)



Article views: 337



[View related articles](#)



[View Crossmark data](#)

## Improving the accuracy of dose estimates from automatically scored dicentric chromosomes by accounting for chromosome number

David Endesfelder<sup>a</sup>, Ulrike Kulka<sup>a,b</sup> , Jochen Einbeck<sup>c</sup>, and Ursula Oestreicher<sup>a</sup>

<sup>a</sup>Department of Effects and Risks of Ionising and Non-Ionising Radiation, Federal Office for Radiation Protection, Neuherberg, Germany;

<sup>b</sup>Department of National and International Cooperation, Federal Office for Radiation Protection, Neuherberg, Germany; <sup>c</sup>Department of Mathematical Sciences, Durham University, Durham, UK

### ABSTRACT

**Purpose:** The traditional workflow for biological dosimetry based on manual scoring of dicentric chromosomes is very time consuming. Especially for large-scale scenarios or for low-dose exposures, high cell numbers have to be analyzed, requiring alternative scoring strategies. Semi-automatic scoring of dicentric chromosomes provides an opportunity to speed up the standard workflow of biological dosimetry. Due to automatic metaphase and chromosome detection, the number of counted chromosomes per metaphase is variable. This can potentially introduce overdispersion and statistical methods for conventional, manual scoring might not be applicable to data obtained by automatic scoring of dicentric chromosomes, potentially resulting in biased dose estimates and underestimated uncertainties. The identification of sources for overdispersion enables the development of methods appropriately accounting for increased dispersion levels.

**Materials and methods:** Calibration curves based on *in vitro* irradiated (137-Cs; 0.44 Gy/min) blood from three healthy donors were analyzed for systematic overdispersion, especially at higher doses (>2 Gy) of low LET radiation. For each donor, 12 doses in the range of 0–6 Gy were scored semi-automatically. The effect of chromosome number as a potential cause for the observed overdispersion was assessed. Statistical methods based on interaction models accounting for the number of detected chromosomes were developed for the estimation of calibration curves, dose and corresponding uncertainties. The dose estimation was performed based on a Bayesian Markov-Chain-Monte-Carlo method, providing high flexibility regarding the implementation of priors, likelihood and the functional form of the association between predictors and dicentric counts. The proposed methods were validated by simulations based on cross-validation.

**Results:** Increasing dose dependent overdispersion was observed for all three donors as well as considerable differences in dicentric counts between donors. Variations in the number of detected chromosomes between metaphases were identified as a major source for the observed overdispersion and the differences between donors. Persisting overdispersion beyond the contribution of chromosome number was modeled by a Negative Binomial distribution. Results from cross-validation suggested that the proposed statistical methods for dose estimation reduced bias in dose estimates, variability between dose estimates and improved the coverage of the estimated confidence intervals. However, the 95% confidence intervals were still slightly too permissive, suggesting additional unknown sources of apparent overdispersion.

**Conclusions:** A major source for the observed overdispersion could be identified, and statistical methods accounting for overdispersion introduced by variations in the number of detected chromosomes were developed, enabling more robust dose estimation and quantification of uncertainties for semi-automatic counting of dicentric chromosomes.

### ARTICLE HISTORY

Received 15 July 2020

Revised 11 September 2020

Accepted 17 September 2020

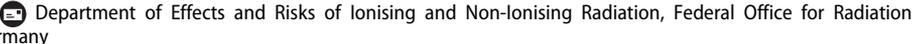
### KEYWORDS

Dicentric chromosome; automatic scoring; overdispersion; chromosome number; biological dosimetry

## Introduction

Radiation specific biomarkers provide an important tool to obtain a retrospective assessment of the blood dose in cases of unclear overexposures to ionizing radiation. Due to its low background level, the high specificity to ionizing radiation, the good reproducibility, the comparability of *in vivo*

and *in vitro* results (Romm et al. 2009), relatively good stability over time and its ability to detect partial body irradiations the dicentric assay is still considered as the ‘gold standard’ in biological dosimetry. Nevertheless, due to the huge time and effort required for the conventional, manual scoring of dicentric chromosomes by skilled human scorers

**CONTACT** David Endesfelder  [dendesfelder@bfs.de](mailto:dendesfelder@bfs.de) 

 Supplemental data for this article can be accessed [here](#).

Copyright © 2020 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

(Oestreicher et al. 2018), it has only limited suitability for large scale radiation accidents with high numbers of potentially exposed individuals. Various approaches such as networking among international laboratories (Kulka et al. 2018), scoring in triage mode (Di Giorgio et al. 2011; Wilkins et al. 2011; Oestreicher et al. 2017) and scoring in a less restrictive manner (Flegal et al. 2010; Flegal et al. 2012) have been successfully developed in recent years to overcome these limitations. In addition, technical developments in automatic image analysis (Schunck et al. 2004) enable accelerated counting of dicentric chromosomes by semi-automated scoring strategies (Vaurijoux et al. 2009; Romm et al. 2013; Oestreicher et al. 2018). Due to the shorter human intervention, automatic counting of dicentric chromosomes provides a valuable tool not only in large scale emergency scenarios but also for low dose research, where very high cell numbers are required to decrease the detection limit (Oestreicher et al. 2018). However, there are considerable differences between the conventional and the semi-automatic scoring approach (Romm et al. 2013) and statistical assumptions usually used for dose estimation and quantification of uncertainties might not generally be applicable for semi-automatic scoring.

The conventional scoring procedure has been successfully harmonized in recent years (IAEA 2011; ISO 2014) and it is generally accepted that dicentric chromosome counts in complete cells with 46 centromeres follow a Poisson distribution for low-LET acute whole body exposures. One of the central properties of Poisson distributed random variables is equidispersion, i.e. the variance equals the mean. For manually counted dicentric chromosomes and low-LET irradiation, overdispersion, i.e. variance > mean indicates partial body irradiation (IAEA 2011) and requires special models for dose estimation (Sasaki and Miyata 1968; Dolphin 1969; Higuera et al. 2016). For semi-automatically counted dicentric chromosomes contradictory results have been published. While some authors do not detect overdispersion for acute whole-body exposures (Vaurijoux et al. 2012; Gruel et al. 2013), others suggest that overdispersion can be observed at higher doses (Vaurijoux et al. 2009; Romm et al. 2013) or that models accounting for overdispersion provide better fits than Poisson models (Oliveira et al. 2016). Data from semi-automatic counting for low-LET radiation doses >2 Gy are sparsely available and overdispersion at higher doses has therefore rarely been analyzed. Nevertheless, the observations of increased overdispersion at higher doses suggest that technical reasons introduce overdispersion. In the context of Poisson regression models, overdispersion can be observed due to 1. omitted predictor variables, 2. incorrectly specified functional forms, 3. random variation in conditional expectations and 4. dependence between events constituting a count (Berk and MacDonald 2008). Regarding semi-automatic counting of dicentric chromosomes, a possible explanation for the observed overdispersion might be, that the number of chromosomes detected by the image analysis software can vary considerably between cells (Romm et al. 2013), e.g. due to differences in the quality of slide preparations (background, cell density or chromosome

spreading). In conventional scoring mode only cells with 46 centromeres are used for dicentric counting (Romm et al. 2013). Therefore, all cells have the same probability for the detection of dicentric chromosomes. In contrast, for semi-automatic counting, cells with low numbers of detected chromosomes have a lower probability for the detection of dicentric chromosomes. Thus, the number of chromosomes influences the expected value of the counts from dicentric chromosomes and acts like an omitted predictor. If dicentric counts from low-LET acute whole-body exposure scenarios are systematically overdispersed and overdispersion is not appropriately accounted for in the statistical models, the number of falsely detected partial body irradiations will be inflated, potentially causing wrong conclusions regarding further treatment of individuals. Furthermore, the estimated confidence intervals might be too permissive, and in case of omitted predictor variables, dose estimates might be biased.

The quantification of uncertainties is a central part of dose reconstruction based on biological markers (IAEA 2011; Ainsbury et al. 2018) and requires an appropriate experimental design, reasonable models, knowledge about the sources contributing to the uncertainty budget as well as assumptions on the underlying statistical distributions. Typically, for conventional manual scoring of dicentric chromosomes, the uncertainties resulting from the Poisson model for the calibration data are combined with uncertainties from the Poisson distributed test data (Merkle 1983; Savage and Papworth 2000; IAEA 2011; Higuera et al. 2015). For overdispersed data, relaxation of the Poisson assumption by Quasi-Poisson models (IAEA 2011; Einbeck et al. 2018) was suggested as well as the use of Negative Binomial, Neyman A or Hermite distributions (Puig and Barquinero 2011; Ainsbury et al. 2013). However, although the latter models improve the estimation of confidence intervals, potentially biased dose estimates resulting from omitted explanatory variables cannot appropriately be accounted for by these models. In such cases, it is therefore crucial to identify missing predictors and adjust the statistical models accordingly (Berk and MacDonald 2008; Hilbe 2011). For instance, if the mean number of detected chromosomes per cell for the test data is different from the calibration data, the resulting dose estimates will be biased, if chromosome number is not accounted for in the statistical models. Current models for semi-automatically counted chromosomes do not account for such situations and it is therefore necessary to analyze the impact of the number of detected chromosomes on the resulting numbers of dicentric chromosomes.

Here, semi-automatically counted calibration curves based on Cs-137 in vitro irradiated blood samples are presented for three healthy individuals, enabling the analysis of inter-individual variation for automatically counted dicentric chromosomes. Inter-individual variation of dicentric counts between adults is relatively low (Pajic et al. 2015) and is commonly assumed to be almost negligible for conventional manual counting. For semi-automatic scoring the situation is less clear, as differences in quality of slide preparation

of blood samples might have a considerable impact on the output produced by the algorithm. To determine whether inter-individual variation and/or overdispersion occurs for semi-automatically counted dicentric chromosomes, 12 doses ranging from 0 to 6 Gy were evaluated. Several doses >2 Gy have been chosen to determine the association between overdispersion and dose. The influence of the automatically detected number of chromosomes per cell on dicentric counts, dispersion levels and the resulting dose and uncertainty estimates was analyzed in detail. A new statistical model for calibration curve estimation including interaction effects between dose and chromosome number is introduced and validated. Furthermore, a Bayesian approach to solve the inverse regression problem for dose estimation is proposed, accounting for the influence of chromosome number. The methods were validated based on simulations for all three individuals separately and for the pooled dataset. Until now, chromosome number has not been considered in statistical methods for biological dosimetry based on semi-automatically counted dicentric chromosomes. The results suggest that the quantification of uncertainties can be improved, biased dose estimates partly corrected and individual variability reduced.

## Material and methods

### Blood donors and irradiations

Blood samples (10 ml heparinized tubes) from 3 healthy donors, 2 females (44 and 52 years) and one male person (19 years) were irradiated with  $^{137}\text{Cs}$  gamma rays (dose rate 0.44 Gy/min at 37°C in a HWM D 2000 unit; Wälischmiller Engineering GmbH, Markdorf, Germany). Peripheral blood samples were obtained, with informed consent, from healthy adult donors, in accordance with the local ethics commission of the Bayerischen Landesärztekammer approved procedure. For the establishment of the dose response curves whole blood samples were irradiated with low-LET doses of 0.1, 0.25, 0.5, 0.75, 1, 1.5, 2.0, 3.0, 4.0, 5.0 and 6.0 Gy. One unexposed sample served as control. For one dose response curve, the dose points 0.1 and 6.0 Gy were supplemented to an already existing curve at a later date using blood from the same donor. Cell culturing and preparation were performed according to well established procedures provided by the IAEA (2011) recommendations and ISO (2014). After irradiation and a repair time of 2 h at 37°C, 0.5 ml whole blood was transferred to culture tubes containing RPMI-1640 culture medium (Biochrom, Berlin) at a blood/media ratio of 0.5 ml/5 ml supplemented with 10% FCS (Biochrom, Berlin), 2% PHA (Biochrom, Berlin) and antibiotics (Penicillin–Streptomycin, 10,000 U/ml Penicillin, 10 mg/ml Streptomycin, Biochrom, Berlin). For cell cycle controlled scoring long term Colcemid treatment (Roche, Mannheim) with a final concentration in culture of 0.08 µg/ml was added 24 h after culture set up. Blood samples were cultured in total for 48 h. The hypotonic treatment of cells was carried out with 75 mM KCl. Cells were then fixed in methanol:acetic acid (3:1) three times and the suspension was stored in the freezer (−18°C) until further use. For slide

preparation the cell solution was concentrated according to the cell yield. The quality and quantity of the metaphases were checked under the microscope prior to Giemsa staining. The slides were covered by cover slips and fixed with Eukitt.

### Dicentric analysis

First, all slides were analyzed using the automatic scoring system Metafer 4 by MetaSystems (Altlußheim, Germany) including the software modules for metaphase finding (MSearch) to detect the metaphase spreads. In the second step, additional software tools were applied for auto-capturing of high resolution images at 63× magnification with oil (AutoCapt) and automatic detection of dicentric candidates (DCScore). In comparison to a fully automatic scoring strategy, the applied semi-automatic scoring of dicentric chromosomes includes the third step, where a human scorer evaluates the automatically detected dicentric candidates on the screen of a PC (Romm et al. 2013). In comparison to the conventional manual scoring, the scoring procedure for dicentric chromosomes in the automatic mode is restricted to evaluate dicentric candidates detected by the software. Thus, the human intervention of an experienced scorer is very short and involves only the decision if the detected dicentric candidate (which is marked by the software with a red frame and thus easy to recognize), should be confirmed as a dicentric chromosome or rejected as a false positive. Further information such as undetected dicentrics (false negatives), numbers of acentric fragments, or completeness of cells is not recorded. However, the number of detected chromosomes and objects per cell is recorded by the software and can be utilized for downstream analysis. In summary, the experiments and the scoring of dicentric chromosomes were performed as described in (Oestreicher et al. 2018).

### Statistical methods

Quasi-Poisson generalized linear regression models (glm) were estimated based on dicentric counts from doses  $D_i$  to obtain linear-quadratic calibration curves from in vitro data without accounting for chromosome number ( $N_C$ )

$$E[Y_i] = \lambda_i = C + aD_i + bD_i^2 \quad (1)$$

where  $Y_i = \frac{R_i}{N_i}$ . The variable  $R_i$  represents the dicentric counts for  $N_i$  cells at dose  $D_i$  and the index  $i$  indicates the  $i$ -th design dose. The random variable  $R_i$  was described by a Quasi-Poisson model with expected value  $E[R_i] = \lambda_i N_i$ , variance  $\sigma^2(R_i) = \theta \lambda_i N_i$  and dispersion parameter  $\theta$ . Coefficients of regression curves were compared using two sample  $Z$ -tests. To test the effect of donor, for each single dose  $D \geq 1$  Gy Quasi-Poisson glms were separately estimated with donor as the only predictor. The null hypothesis that there is no reduction in deviance compared to the intercept only (null) model, i.e. adding donor does not improve the model fits, was tested by  $F$  tests. Additionally,

to test the null hypothesis of equal means, all combinations of pairwise comparisons of dicentric counts between donors were performed using Wald test. Generally, for calibration curve estimation based on glms, an identity link function was used as commonly suggested for biological dosimetry (IAEA 2011). For glm based comparisons at single doses, logarithmic link functions were used, providing the natural link function for Poisson regression models.

The dispersion index  $\delta_i$  was defined as  $\delta_i = \frac{s_i^2}{Y_i}$ , where  $s_i^2 = \frac{1}{N_i-1} \sum_{k=1}^{N_i} (Y_{i,k} - \bar{Y}_i)^2$  is the sample variance for a set of  $k = 1, \dots, N_i$  cells at a given dose  $D_j$ . For each dose, the null hypothesis that the data is equidispersed was tested using Papworth's U test (Papworth 1983) and was rejected for  $|U| > 1.96$ .

To test whether overdispersion occurs in published datasets, as well, linear-quadratic glms based on a Negative Binomial Type 2 (NB2) or a Poisson distribution were compared via Likelihood-Ratio tests (LRTs). To account for the full distribution across cells, each cell was included as one single data point for the estimation of regression models. A modified version of R code from (Oliveira et al. 2016) was used to perform constrained Maximum-Likelihood (ML) estimation based on the R package *maxLik* (Henningsen and Toomet 2011). To test the null hypothesis that there is no difference in  $N_C$  between the three donors, a Kruskal-Wallis test (Kruskal and Wallis 1952) was applied.

To account for the influence of chromosome number on dicentric counts,  $N_C$  was divided into four chromosome classes

$$T = \{N_C | T_1 : 25 < N_C \leq 30, T_2 : 30 < N_C \leq 35, T_3 : 35 < N_C \leq 40, T_4 : 40 < N_C \leq 50\}.$$

NB2 regression models were estimated for each chromosome class separately to show the relationship of dicentric counts and  $N_C$ . To determine whether the binning of data in chromosome classes has an influence on the dispersion index,  $\delta$  was calculated for each chromosome class separately and averaged. The effect of different slides on  $\delta$  was analyzed for each dose by comparing the following fixed or mixed effects regression models for all slides containing  $>100$  metaphases:

1. Intercept only (no effects of  $N_C$  or slide)
2.  $N_C$  as fixed effect (effect of  $N_C$ , no slide effect)
3.  $N_C$  as fixed, slide as random effect on intercept (effect of  $N_C$  and slide)

Negative Binomial Type 1 (NB1) models were used for the latter comparisons, assuming that the dispersion within each dose and chromosome class is relatively constant. Estimates were obtained using the R package *glmmTMB* (Brooks et al. 2017).

To account for the remaining overdispersion and the effect of  $N_C$ , the following NB2 glms with interaction terms between chromosome class and dose were applied:

$$E[Y_j] = \lambda_j = \mathbf{X}_j^T \boldsymbol{\beta} \quad (2)$$

where

$$\mathbf{X}_j = \begin{pmatrix} 1 \\ D_j \\ D_j^2 \\ I_{T_{2,j}} D_j \\ I_{T_{2,j}} D_j^2 \\ I_{T_{3,j}} D_j \\ I_{T_{3,j}} D_j^2 \\ I_{T_{4,j}} D_j \\ I_{T_{4,j}} D_j^2 \end{pmatrix} \text{ and } \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \\ \beta_8 \\ \beta_9 \end{pmatrix} \quad (3)$$

where  $I_{T_{m,j}}$  represents the indicator function ( $I_{T_{m,j}}=1$ , if the corresponding metaphase belongs to chromosome class  $T_m$ , otherwise  $I_{T_{m,j}}=0$ ). The dicentric counts  $Y_j$  from metaphase  $j$ , dose  $D_j$  and chromosome class  $T_{k,j}$  follow a NB2 distribution with mean  $\lambda_j$ , dispersion parameter  $\alpha$  and dispersion index  $\delta = 1 + \alpha\lambda$ . As  $N_C$  had relatively little impact on dicentric counts for low doses, it was assumed that the chromosome classes have a common intercept, resulting in more robust estimates at lower doses. Again, a modified version of R code from (Oliveira et al. 2016) was used to obtain calibration curve coefficients based on constrained ML optimization.

For a given, estimated calibration curve with an estimated coefficient vector  $\hat{\boldsymbol{\beta}}$ , dispersion parameter  $\hat{\alpha}$  and corresponding standard errors  $\sigma_{\hat{\boldsymbol{\beta}}}$  and  $\sigma_{\hat{\alpha}}$  the joint posterior density for the parameters given the data can be obtained by

$$p(D, \boldsymbol{\beta}, \alpha | y_j) \propto p(y_j | D, \boldsymbol{\beta}, \theta) p(D, \boldsymbol{\beta}, \theta) \quad (4)$$

with likelihood (Hilbe 2011)

$$p(y_j | D, \boldsymbol{\beta}, \alpha) = \frac{\Gamma(y_j + \alpha^{-1})}{y_j! \Gamma(\alpha^{-1})} \cdot \frac{\alpha^{y_j} \lambda_j^{y_j}}{(1 + \alpha \lambda_j)^{y_j + \alpha^{-1}}} \quad (5)$$

To estimate the Bayesian posterior density based on Markov-Chain-Monte-Carlo (MCMC) sampling, the following model was implemented in JAGS (Plummer 2003):

$$\beta_i \sim N(\hat{\beta}_i, \sigma_{\hat{\beta}_i}) \quad (\text{priors for } \beta_i)$$

$$\alpha \sim \text{Gamma}\left(\frac{\hat{\alpha}}{\sigma_{\hat{\alpha}}^2}, \frac{\hat{\alpha}}{\sigma_{\hat{\alpha}}}\right) \quad (\text{priors for } \alpha)$$

$$D \sim \text{Unif}(0, 10) \quad (\text{prior for dose in Gy})$$

$$y_j \sim \text{NB2}(\lambda_j, \alpha) \quad (\text{likelihood}),$$

assuming independence of the coefficients  $\beta_i$ .

To test the effect of including  $N_C$  into the models for calibration curve and dose estimation, simulations based on cross-validation were performed by using each of the slides with a cell number  $>100$  and dose  $D \geq 1$  Gy in turn as test data for dose estimation. In each simulation run, the remaining slides were used as training data for the

estimation of the calibration curve. Three different models were compared:

- $M_1$ : Quasi-Poisson models without accounting for  $N_C$  on the raw data (standard approach used in biological dosimetry)
- $M_2$ : Quasi-Poisson models after exclusion of metaphases with  $N_C \leq 25$
- $M_3$ : NB2 interaction models accounting for  $N_C$ .

For models  $M_1$  and  $M_2$ , Quasi-Poisson models based on Equation (1) were used for calibration curve estimation and error propagation based on the Delta method was applied for dose estimation (Savage and Papworth 2000). The latter approach is similar to the implementation in the currently most commonly used software for biological dosimetry (Ainsbury and Lloyd 2010). The calibration curves for  $M_3$  were obtained based on Equations (2) and (3). Dose estimates for  $M_3$  were obtained by the proposed Bayesian MCMC method described above. The latter accounts for the interaction of chromosome number and dose as well as for the remaining overdispersion. The convergence of the chains for the simulation runs was assessed via automatic convergence diagnostics implemented in the R package *runjags* (Denwood 2016). For the pooled dataset, the effective sample size for the posterior MCMC samples of the dose ranged from  $n_{\text{eff}}=7513$  to  $n_{\text{eff}}=12,722$  with a median of  $n_{\text{eff}}=10,839$ , where the total number of MCMC samples was 20,000 for all simulation runs. The mean autocorrelation was 0.29 for a lag of one and 0.005 for a lag of five MCMC samples. On the whole, these results suggested good convergence of the MCMC chains. The deviation of the estimated dose  $\hat{D}$  from the true dose  $D^*$  was defined as  $\Delta = \hat{D} - D^*$  and the percentage deviation from true dose as  $\tilde{\Delta} = 100 \cdot \frac{|\hat{D} - D^*|}{D^*}$ . All calculations were performed in R version 3.6.1 and all R code can be found in the [Supplementary documents](#). The data used for the statistical analysis will be provided by the authors upon request.

## Results

### Calibration curves from semi-automatic scoring

Calibration curves from semi-automatic scoring of dicentric chromosomes were estimated based on in vitro irradiated blood samples from each of the three donors  $C_1$ ,  $C_2$  and  $C_3$  (Figure 1(A)). In total, 12 dose points in the range from 0 to 6 Gy were evaluated for each individual. The number of analyzed metaphases ranged between 363 and 4743 for  $C_1$ , 1261 and 8774 for  $C_2$  and 1254 and 5648 for  $C_3$  (Table 1). The linear coefficients  $a_0$  ranged between 0.0114 and 0.0217 and the quadratic coefficients  $b_0$  ranged between 0.0121 and 0.0267 (Table 2). The index '0' indicates that the calibration curves were estimated based on the raw data. While none of the coefficients was significantly different between  $C_1$  and  $C_2$ , a highly significant difference was observed when the quadratic coefficient ( $b_0$ ) from the raw data calibration curve of  $C_3$  was compared to the corresponding coefficients of  $C_1$

and  $C_2$  ( $p < .001$ ). Adding donor as a predictor to Quasi-Poisson glms estimated for each dose  $D \geq 1$  Gy separately improved the models significantly ( $p < .001$  for all  $D \geq 2$  Gy;  $p < .05$  for all  $D \geq 1$  Gy), indicating considerable differences between the dicentric counts of the individuals at higher doses (Figure 1(A)). Especially, dicentric counts of  $C_3$  showed highly significant differences compared to  $C_1$  and  $C_2$  for all high doses ( $D \geq 2$  Gy;  $p < .001$  for 9/10 comparisons). In contrast, dicentric counts of  $C_1$  and  $C_2$  were relatively similar and only one dose with a weakly significant difference was observed ( $D = 5$  Gy,  $p = .01$ ).

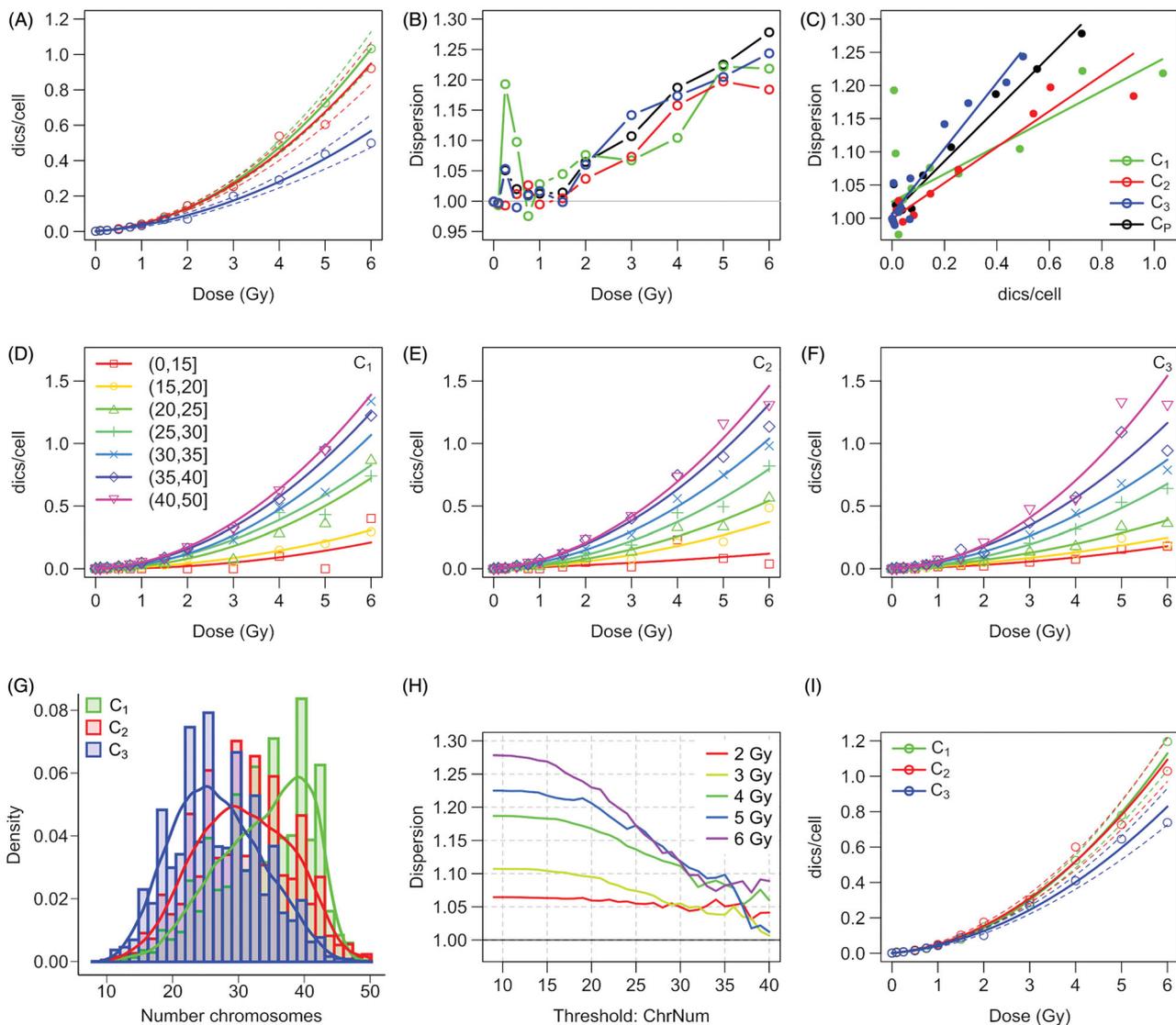
### Overdispersion for high doses

For all three donors the Dispersion Index ( $\delta$ ) increased with increasing dose (Figure 1(B)) and significant overdispersion ( $U$ -test,  $p < .05$ ) was observed for all  $D \geq 2$  Gy for  $C_3$  and for 80% of the dose points  $D \geq 2$  Gy for each of  $C_1$  and  $C_2$  (Table 1). Thus, for 13/15 dose points with  $D \geq 2$  Gy the data wrongly suggested partial body exposure although the experiment simulated a whole body exposure. For the pooled dataset  $C_p$ , the dispersion index  $\delta(C_p)$  was very similar to  $\delta(C_3)$  and for most  $D \geq 2$  Gy slightly higher than  $\delta(C_1)$  and  $\delta(C_2)$ . For all individuals  $\delta$  increased approximately linear with the mean number of dicentric chromosomes (Figure 1(C)). The linear increase of  $\delta$  with the expected value is a typical feature of the NB2 distribution, where  $\delta(Y) = \frac{\text{Var}(Y)}{E(Y)} = 1 + \alpha\lambda$  depends on a dispersion parameter  $\alpha$  and increases linearly with the expected value  $\lambda$ . Therefore, it was assumed that the NB2 distribution can be used to model persisting Poisson overdispersion after possible apparent overdispersion (Hilbe 2011) caused by variations in chromosome number has been dealt with.

To test whether overdispersion for semi-automatic counting is an intrinsic property of this dataset, NB2 based regression models were compared with Poisson based models on published data from Romm et al. (2013) via LRTs. The latter tests the null hypothesis that the dispersion parameter  $\alpha = 0$ , indicating equidispersion. If we reject the null hypothesis it can be assumed that the NB2 model is superior to the Poisson model, indicating a significant degree of overdispersion. The null hypothesis could be rejected at a significance level  $\alpha = 0.05$  for all three individuals  $C_1$ ,  $C_2$  and  $C_3$  and for 7/8 datasets from Romm et al. (2013) (Table 3), suggesting that overdispersion is a common property of semi-automatic scoring across various datasets.

### Overdispersion and chromosome number

One possible reason for the observed overdispersion is that dicentric counts  $Y|D$  at each dose point are not independent, e.g. due to one or more explanatory variables that are not considered in standard models used in biological dosimetry. Overdispersion caused by omitted explanatory variables is a typical case of apparent overdispersion and can be eradicated from the model by including the missing predictor variables (Hilbe 2011). It can be expected that the probability for the detection of dicentric chromosomes will



**Figure 1.** Individual calibration curves, chromosome number and the association to dispersion levels. (A) Semi-automatically scored calibration curves from the raw data of three individuals, including all metaphases identified by the software, estimated based on linear-quadratic Quasi-Poisson regression models. (B) Dose (x-axis) vs dispersion index (y-axis) for three individuals and the pooled dataset. (C) Mean number of dicentrics (x-axis) vs dispersion index (y-axis). Regression lines were fitted via ordinary least-squares linear regression models. (D-F) Linear-quadratic Negative Binomial regression curves for donors  $C_1$ ,  $C_2$ , and  $C_3$  for each chromosome class separately. (G) Histograms of the number of detected chromosomes (x-axis) for each individual. (H) Threshold for excluding metaphases based on the number of detected chromosomes (x-axis) vs dispersion index (y-axis) for all doses  $D \geq 2$  Gy. (I) Semi-automatically scored calibration curves after excluding all metaphases with chromosome number  $\leq 25$  of three individuals estimated based on linear-quadratic Quasi-Poisson regression models.

decrease with decreasing number of chromosomes ( $N_C$ ) detected by the automatic scoring system. The latter can be a potential cause for the violation of the independence assumption for  $Y|D$  and for the observed overdispersion and was therefore investigated in detail. Typical reasons for low chromosome numbers detected by the software are highly condensed chromosomes (Figure 2(A)), overlapping chromosomes and cytoplasmic background due to low quality spreads (Figure 2(B)) or overlapping non-stimulated lymphocytes (Figure 2(C)). Examples for high quality metaphases with relatively high numbers of detected chromosomes are shown in Figures 2(D-F).

To evaluate the influence of  $N_C$  on dicentric counts, calibration curves were estimated for different classes of  $N_C$  for each of the individuals (Figures 1(D-F)). As expected,

dicentric counts increased rapidly with increasing  $N_C$ , especially for  $D \geq 1$  Gy (Supplementary Figure 1). Particularly, the quadratic calibration curve coefficient showed a considerable increase from the lowest ( $N_C < 15$ ) to the highest chromosome class ( $40 < N_C \leq 50$ ) for  $C_1$  (0.006 vs 0.036),  $C_2$  (0.001 vs 0.035) and  $C_3$  (0.004 vs 0.040). The distribution of  $N_C$  revealed substantial differences between the three donors ( $p < .0001$ , Figure 1(G)). Here,  $C_3$  had the lowest median chromosome number ( $N_C = 26$ ), followed by  $C_2$  ( $N_C = 31$ ) and  $C_1$  ( $N_C = 35$ ).

To test whether variations in  $N_C$  introduce overdispersion, different thresholds for the exclusion of chromosomes were applied for each dose  $D \geq 2$  Gy. Exclusion of metaphases with low chromosome number substantially decreased the degree of overdispersion for all of the high

**Table 1.** Data for calibration curves of individuals C<sub>1</sub>, C<sub>2</sub> and C<sub>3</sub>.

Donor	Dose (Gy)	N <sub>C</sub>	Cells	DC	Dic/cell	SE	Dicentric distribution							δ	U	
							0	1	2	3	4	5	6			7
C <sub>1</sub>	0	34	2498	4	0.0016	0.0008	2494	4							1	-0.05
	0.1	29	1679	12	0.0071	0.0021	1667	12							0.99	-0.2
	0.25	35	2596	20	0.0077	0.0019	2578	16	2						1.19	<b>7.12</b>
	0.5	34	2569	36	0.014	0.0024	2535	32	2						1.1	<b>3.54</b>
	0.75	34	3135	78	0.0249	0.0028	3057	78							0.98	-0.98
	1	35	4741	200	0.0422	0.003	4548	186	7						1.03	1.37
	1.5	34	3870	287	0.0742	0.0045	3599	256	14	1					1.04	<b>1.96</b>
	2	35	2570	372	0.1447	0.0078	2234	304	29	2	1				1.08	<b>2.73</b>
	3	33	962	244	0.2536	0.0168	752	180	27	2	1				1.07	1.47
	4	33	917	447	0.4875	0.0242	579	248	74	14	1	1			1.1	<b>2.24</b>
5	35	478	347	0.7259	0.0431	246	154	51	19	6	2			1.22	<b>3.43</b>	
6	30	363	375	1.0331	0.0589	141	123	66	18	9	6			1.22	<b>2.94</b>	
C <sub>2</sub>	0	31	4541	5	0.0011	0.0005	4536	5							1	-0.05
	0.1	32	7172	31	0.0043	0.0008	7141	31							1	-0.25
	0.25	32	6869	50	0.0073	0.001	6819	50							0.99	-0.42
	0.5	31	8528	142	0.0167	0.0014	8388	138	2						1.01	0.76
	0.75	30	7641	194	0.0254	0.0018	7452	184	5						1.03	1.63
	1	30	5468	225	0.0411	0.0027	5247	217	4						0.99	-0.28
	1.5	29	4966	412	0.083	0.0041	4572	376	18						1	0.23
	2	30	4102	601	0.1465	0.0061	3553	500	46	3					1.04	1.67
	3	31	3883	982	0.2529	0.0084	3049	697	127	9	1				1.07	<b>3.23</b>
	4	32	1721	928	0.5392	0.019	1042	487	147	37	5	2	1		1.16	<b>4.62</b>
5	30	1261	762	0.6043	0.024	740	333	144	37	5	2			1.2	<b>4.95</b>	
6	32	1418	1306	0.921	0.0277	615	468	208	94	26	6	1		1.18	<b>4.9</b>	
C <sub>3</sub>	0	28	2305	2	0.0009	0.0006	2303	2							1	-0.02
	0.1	29	4464	14	0.0031	0.0008	4450	14							1	-0.14
	0.25	27	5648	34	0.006	0.0011	5615	32	1						1.05	2.86
	0.5	28	5421	57	0.0105	0.0014	5364	57							0.99	-0.54
	0.75	29	2259	58	0.0257	0.0034	2202	56	1						1.01	0.31
	1	25	4889	162	0.0331	0.0026	4731	154	4						1.02	0.82
	1.5	26	4939	334	0.0676	0.0037	4616	312	11						1	-0.08
	2	24	3741	262	0.07	0.0045	3492	239	8	1	1				1.06	<b>2.6</b>
	3	27	5238	1048	0.2001	0.0066	4345	760	113	18	2				1.14	<b>7.26</b>
	4	27	3142	914	0.2909	0.0104	2408	581	130	20	2	1			1.17	<b>6.87</b>
5	25	1254	547	0.4362	0.0205	848	293	90	19	3	1			1.2	<b>5.13</b>	
6	25	1757	878	0.4997	0.0188	1113	475	122	35	8	3	0	1	1.24	<b>7.22</b>	

The Dispersion Index  $\delta$  and the Papworth U statistic are shown in the last two columns. Significant overdispersion is indicated by bold entries in column U.

doses (>2 Gy), particularly for  $N_C > 35$  ( $\delta \approx 1.05$ , Figure 1(H)). However, for such a stringent  $N_C$  threshold only 47% (C<sub>1</sub>), 28% (C<sub>2</sub>) and 12% (C<sub>3</sub>) of the initially accepted metaphases would remain for further analysis. To avoid extensive loss of data, a less stringent threshold was applied to exclude only metaphases with a very low probability of the detection of dicentric chromosomes ( $N_C \leq 25$ ). Using this less stringent approach 86% (C<sub>1</sub>), 74% (C<sub>2</sub>) and 54% (C<sub>3</sub>) of the initially accepted metaphases could be used for further analysis and calibration curves of the individuals became already much more similar (Figure 1(I)), e.g. the difference between the expected number of dicentrics per cell between C<sub>1</sub> and C<sub>3</sub> at dose 6 Gy decreased from 0.47 before to 0.30 after exclusion of metaphases with  $N_C \leq 25$ . The linear coefficients ( $a_{25}$ ) ranged between 0.0112 and 0.0249 and the quadratic coefficients ( $b_{25}$ ) ranged between 0.0189 and 0.0294 (Table 2). However, there was still a significant difference between  $b_{25}(C_3)$  versus  $b_{25}(C_1)$  ( $p < .0001$ ) and  $b_{25}(C_2)$  ( $p = .0018$ ). Here, the index '25' indicates that all metaphases with  $N_C \leq 25$  were excluded prior to the analysis. In general, the calibration curve coefficients indicated an increased number of detected dicentrics after exclusion of metaphases with low  $N_C$ .

For each of the three individuals and for all three individuals combined,  $\delta$  already decreased by excluding cells with  $N_C \leq 25$  and further decreased by taking the average

**Table 2.** Calibration curve coefficients and standard errors for individuals C<sub>1</sub>, C<sub>2</sub>, C<sub>3</sub>.

	C <sub>1</sub>		C <sub>2</sub>		C <sub>3</sub>	
	Est	SE	Est	SE	Est	SE
C <sub>0</sub>	0.0028	0.0010	0.0013	0.0011	0.0006	0.0012
a <sub>0</sub>	0.0114	0.0032	0.0190	0.0046	0.0217	0.0049
b <sub>0</sub>	0.0267	0.0013	0.0232	0.0017	0.0121	0.0015
C <sub>25</sub>	0.0027	0.0009	0.0012	0.0011	0.0005	0.0010
a <sub>25</sub>	0.0112	0.0030	0.0243	0.0049	0.0249	0.0050
b <sub>25</sub>	0.0294	0.0013	0.0263	0.0018	0.0189	0.0016

The subscripts '0' indicate coefficients based on the raw data and subscripts '25' indicate coefficients after the exclusion of metaphases with  $\leq 25$  detected chromosomes.

dispersion  $\bar{\delta}$  calculated within each chromosome class (Figure 3(A-D)). The difference between the dispersion levels of the pooled dataset to each of the three individual datasets was relatively low, indicating that donor specific effects besides variations in  $N_C$  had little influence on  $\delta$ . However, although overdispersion was reduced by accounting for  $N_C$ , the data remained slightly overdispersed ( $\approx 1.1$  at D = 6 Gy).

To test whether the persisting overdispersion was due to variations of the mean number of dicentric chromosomes between slides, regression models including chromosome class as fixed effect and slide as a random effect were estimated for the pooled dataset after exclusion of metaphases

with low chromosome numbers for each dose  $D \geq 1$  Gy and compared to the same model without accounting for slide effect. As already discussed above, including chromosome class reduced the dispersion considerably. Including slide into the regression model had almost no effect on the estimated dispersion indices (Supplementary Figure 2). Hence, it is very likely that the remaining overdispersion is not caused by slide effects.

**Table 3.** Comparison of Negative Binomial and Poisson regression models by LRTs.

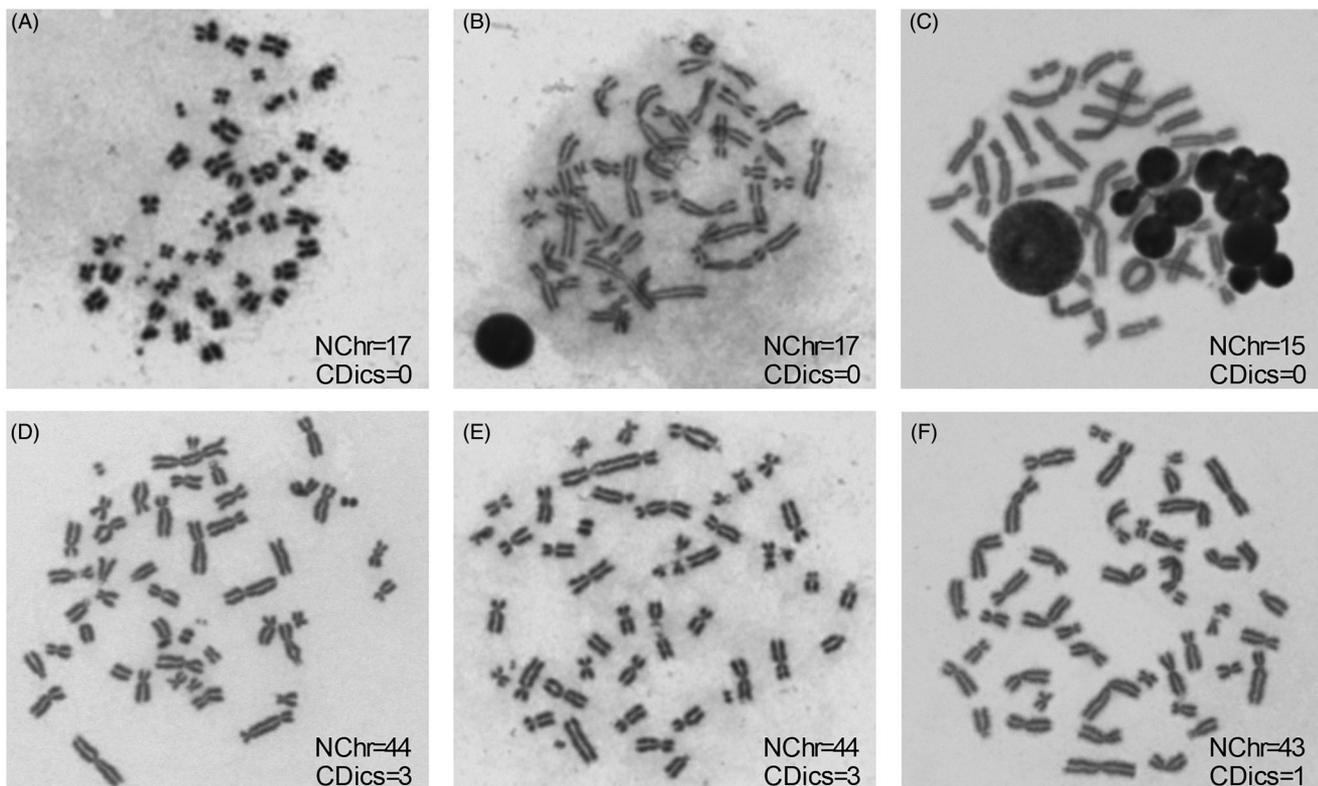
Curve	Classifier	$-2 \times \log(\text{LR})$	p Value	$\alpha$ (NB2)
C <sub>1</sub>	BFS	31	$1.2 \times 10^{-8}$	0.28
C <sub>2</sub>	BFS	77	$1.1 \times 10^{-18}$	0.28
C <sub>3</sub>	BFS	141	$7.8 \times 10^{-33}$	0.56
lab1*	IRSN	0.48	.24	0.06
lab2*	IRSN	20	$3.5 \times 10^{-6}$	0.29
lab3*	IRSN	13	<b>.00015</b>	0.24
lab4*	IRSN	11	<b>.00037</b>	0.17
lab1*	BFS	3.9	<b>.024</b>	0.14
lab3*	BFS	19	$7.6 \times 10^{-6}$	0.31
lab5*	BFS	3.8	<b>.026</b>	0.14
lab6*	BFS	21	$1.9 \times 10^{-6}$	0.21

Calibration curves have been estimated for individuals C<sub>1</sub>, C<sub>2</sub> and C<sub>3</sub> and for data from Romm et al. (2013) based on Poisson and NB2 regression models. Small *p* values indicate that the NB2 model outperformed the Poisson model, indicating a certain degree of overdispersion. Significant *p* values are shown in bold. The parameter  $\alpha$  indicates the degree of overdispersion.

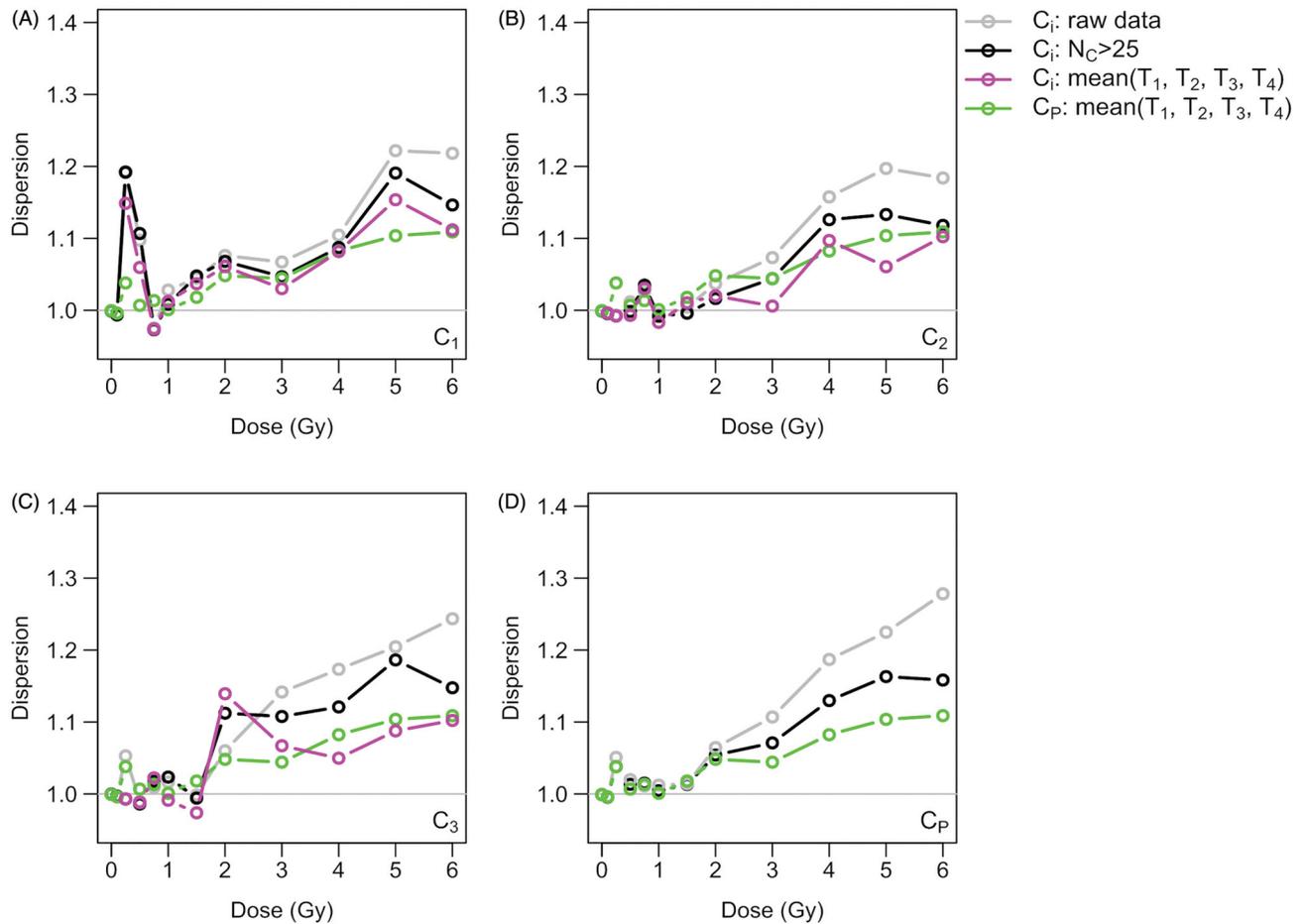
\*Data from Romm et al. (2013).

### Interaction models for calibration curve estimation

To account for the influence of  $N_C$  without losing too many data points, a single calibration model from the pooled dataset of all donors was estimated based on a multiplicative regression model with interaction between dose and  $N_C$ . To model the overdispersion persisting after accounting for chromosome number, NB2 based regression models were used. The resulting coefficients of the fit for the pooled dataset showed that mainly the quadratic interaction coefficients ( $\beta_5, \beta_7, \beta_9$ ) from Equation (3) increased with increasing chromosome number, i.e.  $0 < \beta_5 < \beta_7 < \beta_9$  (Table 4). The resulting curves for each of the four chromosome classes fitted very well to the data points of all three individuals (Figure 4(A-D)). Compared to a linear-quadratic NB2 regression model without accounting for chromosome number, the interaction model accounting for chromosome number (Equations (2) and (3)) showed lower AIC (Akaike's information criterion) and BIC (Bayesian information criterion) values and a LRT suggested that accounting for chromosome number by the suggested interaction model significantly ( $p < .0001$ ) improved the fit (Table 5). The semi-automatically scored data for high chromosome numbers resembled the calibration curve scored in full manual mode closest (Supplementary Figure 3). However, for all chromosome classes, the number of dicentric per cell was much higher for manual compared to semi-automatic scoring. In summary, these results suggested that accounting for  $N_C$  substantially improved the fit for the pooled dataset.



**Figure 2.** Examples for automatically identified metaphases with variable numbers of detected chromosomes. Low number of detected chromosomes due to small, condensed chromosomes (A), low metaphase spread, overlapping chromosomes and high background intensity (B), overlapping non-stimulated lymphocytes (C). Good examples for automatically identified metaphases with reasonable chromosome numbers (D-F).



**Figure 3.** Dose (x-axis) vs dispersion index (y-axis). Dispersion index for the raw data, including all metaphases identified by the software, only metaphases with chromosome number >25, average over dispersion indices for each chromosome class for donors  $C_1$ ,  $C_2$  and  $C_3$  (A–C) and average over dispersion indices for each chromosome class for the pooled dataset  $C_P$  (A–D).

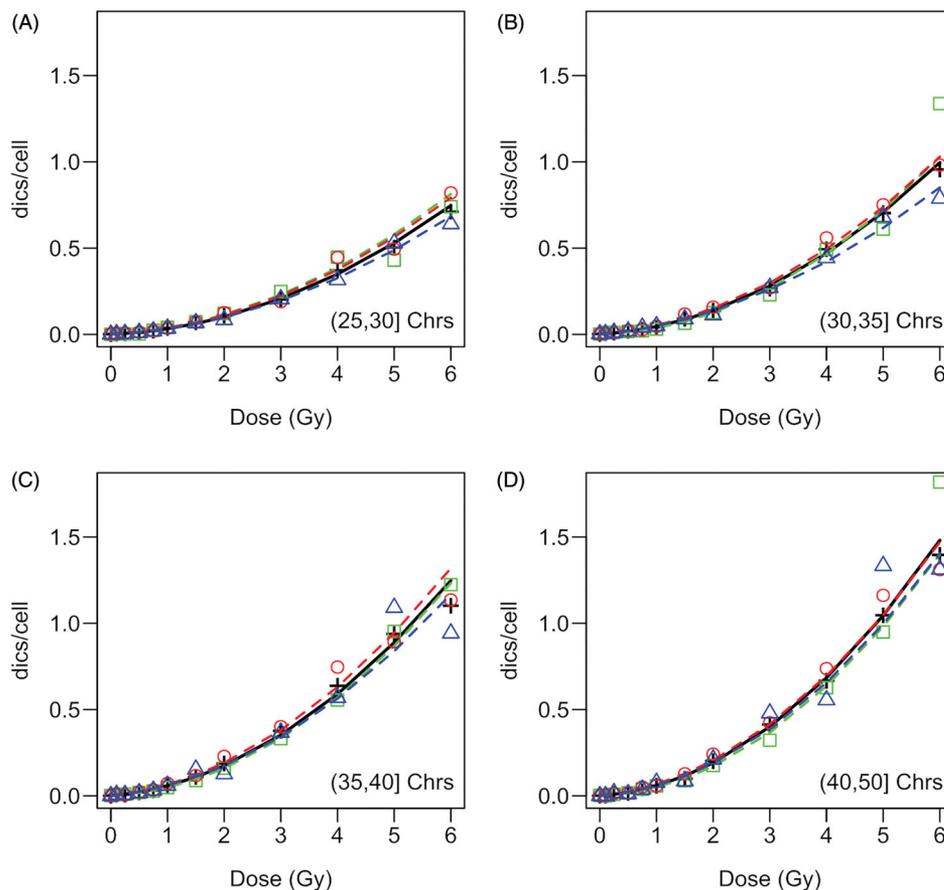
**Table 4.** Calibration curve coefficients for the pooled dataset  $C_P$  based on NB2 regression models with interaction between dose and chromosome number.

	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$	$\beta_9$
X	1	D	D <sup>2</sup>	$I_2D$	$I_2D^2$	$I_3D$	$I_3D^2$	$I_4D$	$I_4D^2$
Estimate	0.0017	0.0126	0.0186	0.0071	0.0057	0.0132	0.0117	0.0064	0.0193
SE	0.0004	0.0021	0.0008	0.0030	0.0012	0.0036	0.0014	0.0044	0.0021

### Improved dose estimation and quantification of uncertainties

Ignoring chromosome number can potentially cause biased point estimates of the dose, increased variance between dose estimates and underestimation of uncertainties. To determine the influence of  $N_C$  on dose estimates, models  $M_1$ ,  $M_2$  and  $M_3$  were compared using simulations based on the available data from the three donors, where  $M_1$  represents the standard model applied in biological dosimetry,  $M_2$  the standard model after exclusion of metaphases with low chromosome numbers and  $M_3$  the interaction model accounting for chromosome number. A cross-validation strategy was applied where one slide was used as test data for dose estimation and the remaining slides were used as training data for calibration curve estimation (see Materials and Methods for details).

Compared to  $M_1$ , the point estimates of the resulting dose estimates showed a lower percentage deviation from the true dose  $D^*$  for  $M_2$  and decreased further for  $M_3$  (Figure 5(A)). For  $C_1$  a non-significant decreasing trend was observed, for  $C_2$  ( $p < .05$ ,  $M_3$  vs  $M_1$  and  $M_2$ ),  $C_3$  ( $p < .05$ ,  $M_3$  vs  $M_1$  and  $M_2$ ) and the pooled dataset ( $p < .001$ ,  $M_3$  vs  $M_1$  and  $M_2$ ) the decrease was tested significant by paired Wilcoxon tests. The latter results indicated that accounting for  $N_C$  corrects biased point estimates of the dose to some degree. Furthermore, the variance ( $\sigma_\Delta^2$ ) of the deviation from the true dose decreased markedly from  $M_1$  to  $M_3$  (Figure 5(B)) for  $C_1$  ( $\sigma^2(M_1) = 0.09$ ;  $\sigma^2(M_3) = 0.06$ ),  $C_2$  ( $\sigma^2(M_1) = 0.24$ ;  $\sigma^2(M_3) = 0.15$ ) and especially for  $C_3$  ( $\sigma^2(M_1) = 0.35$ ;  $\sigma^2(M_3) = 0.17$ ) and  $C_P$  ( $\sigma^2(M_1) = 0.42$ ;  $\sigma^2(M_3) = 0.14$ ). In addition, the percentage of slides with extreme estimates ( $\hat{\Delta} > 25\%$  or  $|\Delta| > 1$  Gy) decreased by including  $N_C$  for  $C_2$ ,  $C_3$  and the pooled dataset (Figure 5(C)). Coverage ( $\kappa$ ) was defined as the percentage of estimated 95% confidence intervals including the true dose  $D^*$ . Coverage and, hence, uncertainty estimation improved for  $C_1$  ( $\kappa(M_1) = 86\%$ ;  $\kappa(M_3) = 98\%$ ),  $C_2$  ( $\kappa(M_1) = 70\%$ ;  $\kappa(M_3) = 87\%$ ),  $C_3$  ( $\kappa(M_1) = 65\%$ ;  $\kappa(M_3) = 90\%$ ) and  $C_P$  ( $\kappa(M_1) = 56\%$ ;  $\kappa(M_3) = 84\%$ ). Nevertheless, although the underestimation of uncertainties was substantially improved, in some cases  $\kappa$  was still slightly lower than the nominal level of 95%



**Figure 4.** Calibration model accounting for chromosome number. The calibration model for the pooled dataset ( $C_p$ , solid line) was estimated based on a multiplicative regression model with interaction between dose and chromosome class. The lines were fitted based on a generalized linear regression model using a Negative Binomial distribution of type 2 with identity link. The predicted regression lines for the pooled dataset and the mean dicentric counts for individuals  $C_1$  (squares),  $C_2$  (circles) and  $C_3$  (triangles) and the pooled dataset (crosses) are shown for each chromosome class separately (A–D).

**Table 5.** Comparison of linear-quadratic NB2 regression models with and without accounting for the number of detected chromosomes.

Model	AIC	BIC	Resid. df	2xlog-lik	df	LR statistic	$p$ Value (Chi)
Linear-quadratic	48090	48128	90,049	−48,082			
Linear-quadratic accounted for chromosome number	47591	47685	90,043	−47579	6	502.5	<.0001

$p$  values were assessed based on a Chi-squared test statistic, small  $p$  values indicate that the model accounting for chromosome number outperformed the model without accounting for chromosome number.

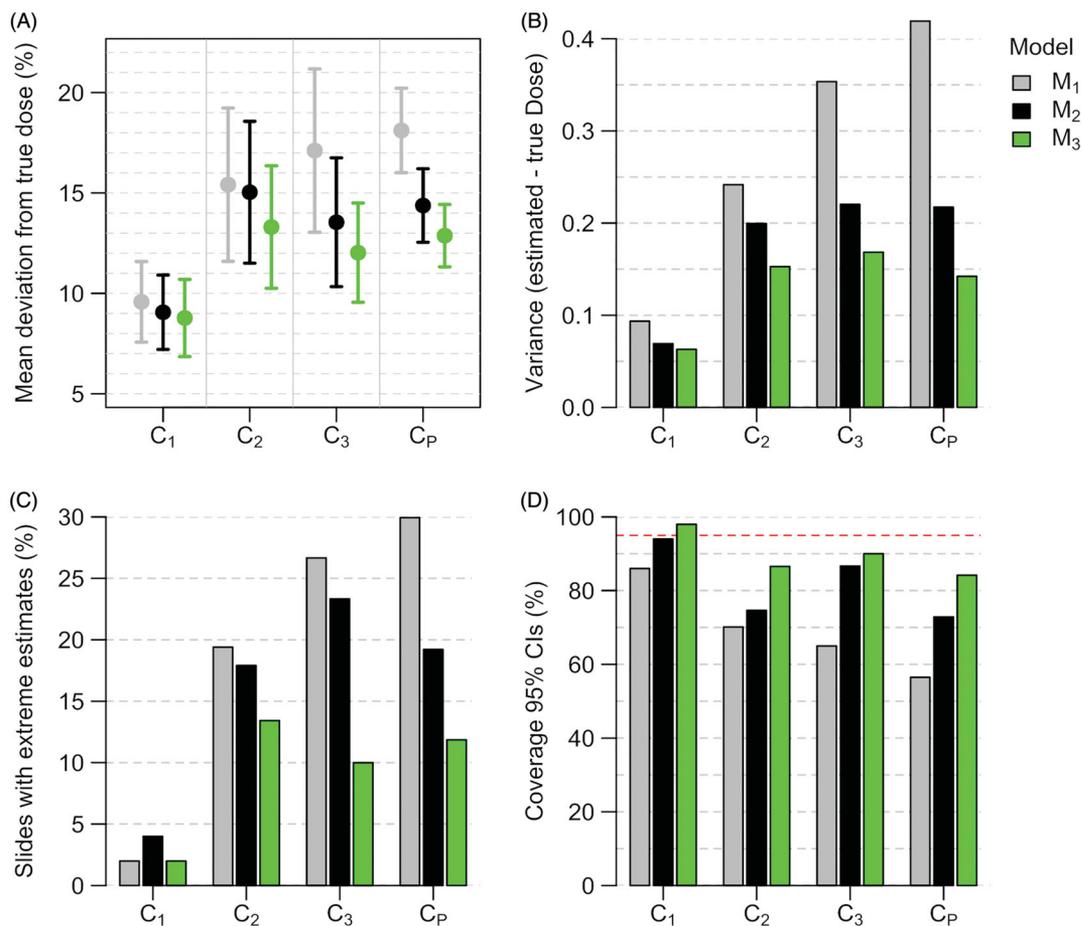
(Figure 5(D)), pointing to slightly underestimated uncertainties, in particular for  $C_p$ .

## Discussion

The scoring procedure for semi-automatically counted dicentric chromosomes can introduce systematic overdispersion and standard statistical methods for dose estimation applied for manual scoring might therefore not generally be applicable. To obtain more reliable estimates of the dose and the corresponding confidence intervals, it is essential to understand the mechanistic sources introducing overdispersion and to account for overdispersion by adjusting the models appropriately. Depending on the source of overdispersion it might not be sufficient to adjust the assumptions about the distribution of the dicentric counts. In particular, this is the case if apparent overdispersion (Hilbe 2011) occurs due to additional explanatory variables that were not

considered in the statistical model. In such cases, the models have to be adjusted accordingly and missing parameters have to be identified and included into the models. A possible mechanism introducing overdispersion is the variability in chromosome numbers between metaphases. The latter violates the assumption that all metaphases at a given dose have the same dicentric rate, and variability in the rate will lead to overdispersion. Especially for samples with low quality, the number of detected chromosomes can be significantly lower than for samples with better quality, because of difficulties in the identification of the individual chromosomes by the software tool. Determining the impact of differences in chromosome number on dispersion levels enables the development of new statistical models for improved dose and uncertainty estimation.

In total, 129,005 metaphases from three different donors were scored semi-automatically at 12 dose points ranging from 0 to 6 Gy of low LET radiation. The data was analyzed



**Figure 5.** Validation of interaction model for dose and uncertainty estimation based on cross-validation. Mean deviation from true dose (in %) (A), variance of the difference estimated vs true dose (in Gy) (B), percentage slides with extreme dose estimates (C) and coverage of 95% confidence intervals (D) for the Quasi-Poisson model applied on the raw data (M<sub>1</sub>), after exclusion of metaphases with chromosome number  $\leq 25$  (M<sub>2</sub>) and the interaction model accounting for chromosome number where the posterior distribution of the dose was estimated based on the Bayesian MCMC method (M<sub>3</sub>). The nominal coverage level in (D) is 95%.

for systematic overdispersion and for the impact of the variability of the number of detected chromosomes on dispersion levels, uncertainty and dose estimates. For all three donors overdispersion was observed for higher doses and dispersion levels increased approximately linearly with the mean number of dicentrics. The detection of the systematic overdispersion at higher doses was possible as several doses  $>2$  Gy were analyzed. For most doses  $>2$  Gy the data wrongly suggested a partial body instead of a whole body irradiation scenario. Moreover, the data suggested that variations in chromosome numbers are the major source for the observed overdispersion. Based on these findings, statistical models for improved estimation of calibration curves and for the inverse regression problem of dose estimation were developed to avoid misclassification of blood samples in the frame of biological dosimetry. For the estimation of the calibration curve, a constrained Maximum-Likelihood approach was used to estimate coefficients and the corresponding standard errors. This approach was chosen to resemble the current approach for calibration curve estimation (IAEA 2011) as close as possible. For dose estimation, a Bayesian model based on MCMC sampling was developed. For cross-validation, convergence was assessed based on automatic

criteria (Denwood 2016). However, for dose estimations of real cases convergence and autocorrelation of the MCMC chains should be verified manually for each case. The Bayes model was implemented in JAGS (Plummer 2003) as it provides very high flexibility regarding the implementation of the functional relationship of dicentric counts and other parameters as well as the likelihood and the priors. Recently, Bayes methods have been developed and successfully applied in different scenarios of dose reconstructions in biological dosimetry and provide a versatile tool to quantify uncertainties for inverse regression problems (Higuera et al. 2015; Higuera et al. 2016; Słonecka et al. 2018, 2019). To test whether the proposed model improves dose estimation, the dataset was divided into several sets of training and validation data by a cross-validation strategy. Compared to models currently applied in biological dosimetry, the proposed model showed less biased dose estimates, less variability between simulation runs and significantly improved estimation of uncertainties.

Currently, there are relatively few studies reporting dispersion levels from exposures  $>2$  Gy for automatically counted dicentric chromosomes and published results are contradictory. The automatically scored calibration curve

shown in Vaurijoux et al. (2009) revealed significant (U test,  $p < .05$ ) overdispersion at 3 Gy and the dispersion level of  $\delta = 1.1$  was comparable to results from the three donors of the current study. However, as no other dose points  $>2$  Gy were reported it was not possible to analyze the relationship between dose and dispersion levels at high doses. The authors from Romm et al. (2013) provided eight semi-automatically counted calibration curves from six laboratories based on two different classifiers. In total, 24 dose points  $>2$  Gy were reported and 50% of these dose points showed significant overdispersion (U test,  $p < .05$ ). Reanalysis of the data from Romm et al. (2013) based on LRTs suggested that the NB2 model accounting for overdispersion performed significantly ( $p < .05$ ) better than the conventional Poisson model for 7/8 curves (Table 3). In line with this result, Oliveira et al. (2016) suggested that overdispersed models outperformed the Poisson models for one of the curves from Romm et al. (2013) based on AIC (Akaike 1974), BIC (Schwarz 1978) and a score test comparing a Poisson to a Negative Binomial model by modeling the mean via a log-link function (Dean and Lawless 1989). In contrast, Gruel et al. (2013) reported 14 samples from three whole body doses  $>2$  Gy and found only one sample showing significant overdispersion (U test,  $p < .05$ ). Possible explanations might be differences in protocols for sample preparation, criteria for metaphase selection or classifiers for metaphase, chromosome and dicentric chromosome detection.

Although chromosome number seems to be the major cause for the observed overdispersion, the data remained overdispersed after chromosome number was accounted for. As accounting for slide effects did not further reduce the amount of the remaining overdispersion significantly, it is not very likely that dependencies within slides are the cause for the remaining overdispersion. Furthermore, the remaining overdispersion had approximately similar values for all of the three probands and for the pooled dataset, suggesting that this effect is independent of the proband. Overdispersion within bins of chromosome classes could also be due to different chromosome numbers within one chromosome class. However, if bins of width one were selected, i.e. each chromosome number was considered separately, the remaining overdispersion remained at approximately similar levels (data not shown due to low sample size within each chromosome number) as for the wider chromosome classes used for the models described above. It is therefore likely, that other sources lead to random variability in the dicentric rates which introduce additional overdispersion. Here, the remaining overdispersion was modeled by the Negative Binomial distribution of Type 2 as the relationship between the mean number of dicentric chromosomes and the dispersion index was approximately linear. However, although the proposed model improved the results of dose estimation considerably, simulations by cross-validation suggested that the uncertainty is still slightly underestimated for two of the probands and especially for the pooled dataset. The latter indicates that apparent overdispersion from unknown sources rather than 'real Poisson overdispersion' (Hilbe 2011) might be the source for the remaining

overdispersion. Thus, further research might be required to investigate additional sources introducing uncertainties for semi-automatic counting of dicentric chromosomes.

Ideally, only metaphases where the software detected  $>40$  chromosomes would be used for the scoring of dicentric chromosomes. However, if too stringent thresholds were applied, very few metaphases would remain for scoring. The automatic method has the main advantage that very high numbers of metaphases can be scored in a short amount of time. This advantage would be partly lost by too stringent thresholds. Here, a mixed strategy was suggested by firstly excluding metaphases with very low chromosome numbers ( $N_C \leq 25$ ) and, secondly, by accounting for chromosome number in the statistical models. While metaphases with a very low probability for the detection of dicentric chromosomes are not considered, the number of analyzed metaphases still remains reasonably high and the bias introduced due to chromosome number can be corrected by the model. These steps can be performed after the usual steps applied for the semi-automatic detection of dicentric chromosomes based on the output data of the software tool.

The comparison of calibration curves from three different probands enabled the determination of individual effects. Due to lower quality of the metaphase spreads proband C<sub>3</sub> had comparatively low chromosome numbers and the observed dicentric rates were therefore much lower than for probands C<sub>1</sub> and C<sub>2</sub>. If data from C<sub>3</sub> had been used for dose estimation based on calibration curves from C<sub>1</sub> and C<sub>2</sub>, doses would have been massively underestimated, leading to wrong conclusions. The proposed method reduced sample bias in dose estimates introduced by differences in chromosome numbers considerably. Variations in chromosome numbers were also relatively high within a single proband and the proposed method also improved dose estimation within a single proband.

Here, it was shown that the proposed models improve dose estimation and quantification of uncertainties considerably for the given experimental setup. However, different experimental setups from other laboratories, as e.g. different SOPs, classifiers, exclusion criteria for metaphases and scoring strategies might lead to variable patterns of the observed dispersion levels. As discussed above, the reasons why some studies do not detect overdispersion at higher doses remain unknown and laboratories performing automatic scoring should therefore thoroughly check whether overdispersion occurs at higher doses using their own SOPs and classifier settings. Several doses  $>2$  Gy should be considered to enable the detection of overdispersion at higher doses and the analysis of the functional relationship between dose and dispersion levels. Furthermore, to obtain calibration curves that are representative for the whole range of chromosome numbers, cell numbers must be reasonably high for each of the chromosome classes. To reduce individual bias, e.g. due to sample quality, samples from several probands should be included for the estimation of semi-automatic calibration curves. The proposed model was tested based on a cross-validation strategy based on the data from the three probands

included in this study. Although this strategy suggested that the models are relatively robust against sample specific bias, it might theoretically be possible that other variables can introduce further bias for samples from other probands. Therefore, the proposed models should be tested based on independent data in future, e.g. during inter-laboratory comparisons. Generally, automatic counting of dicentric chromosomes is more susceptible to low quality of metaphase spreads than the conventional manual scoring strategy as the scorer can correct bias introduced due to quality issues to some degree because of better visual possibilities to identify individual chromosomes. Problems can potentially be caused by low quality metaphase spreads, overlapping and twisted chromosomes, highly condensed chromosomes, high background levels and overlap with non-stimulated lymphocytes. The suggested approach will not be able to correct bias introduced by very low quality of slide preparations with a very low number of metaphases and/or very bad quality of the metaphase spreads. Those slides should always be excluded rigorously in advance.

Overdispersion in the context of low-LET acute whole-body exposures is usually a strong indication for partial body or heterogeneous exposures. However, overdispersion of the whole body data will lead to inflated false positive calls of partial body exposures in the frame of biological dosimetry applications and therefore wrong conclusions regarding the treatment of exposed individuals. Improved methods for the correct detection of partial body exposures will have to be developed in future to overcome the intrinsic overdispersion caused by variable chromosome numbers for semi-automatically scored dicentric data. For these methods, overdispersion introduced by the number of detected chromosomes will have to be considered as well as the persisting overdispersion from unknown sources. Furthermore, the results presented here will have to be validated by other laboratories and in inter-laboratory comparisons and should be included in software tools for biological dose reconstruction in future. Validation experiments might also shed light on the sources for the remaining overdispersion and for the reasons of the slight underestimation of uncertainties that was still present after accounting for chromosome numbers. Performing experiments with variations in the experimental setup, classifier settings or scoring strategies might also enable the detection of the sources for the observed overdispersion. Low-quality samples can strongly bias the output of the algorithms. To minimize the risk of biased dose estimates, objective criteria should be defined to exclude low quality samples prior to the automatic counting of dicentric chromosomes.

## Conclusions

The presented results demonstrate that systematic overdispersion occurs at high doses after low LET radiation exposure for semi-automatic scoring of dicentric chromosomes. It could be shown that variable numbers of detected chromosomes by the software tool are a major source for the observed overdispersion. Statistical methods were introduced

that account for the overdispersion introduced by variable chromosome numbers per cell as well as for persisting overdispersion. Simulations suggested that the proposed models significantly improve the quantification of uncertainties and reduce bias in dose estimates for data from single probands and for the pooled dataset. Compared to conventional manual counting, automatic scoring of dicentric chromosomes enables much faster scoring of samples for biological dosimetry and the proposed approach provides a new opportunity to obtain more robust dose estimates. However, the uncertainty was still slightly underestimated and the data remained slightly overdispersed after accounting for chromosome number. The reasons will have to be investigated in future studies.

## Acknowledgments

The authors thank Martina Denk and Claudia Kerscher, Biological Dosimetry Section, Federal Office for Radiation Protection, Neuherberg, Germany, for excellent technical assistance.

## Author contributions

UK and UO designed and performed the biological experiments. DE developed the mathematical models and wrote the manuscript. JE provided important advice for the development of the mathematical models. All authors reviewed and revised the manuscript.

## Disclosure statement

The authors declare that they have no conflict of interest.

## Notes on contributors

*David Endesfelder*, PhD, Biomathematician, Federal Office for Radiation Protection (BfS), Unit: Biological Dosimetry, Oberschleissheim, Germany.

*Ursula Oestreicher*, Dipl. Biol., Biologist and head of the Unit: Biological Dosimetry, Federal Office for Radiation Protection (BfS), Unit: Biological Dosimetry, Oberschleissheim, Germany.

*Jochen Einbeck*, PhD, Associate Professor for Statistics in the Department of Mathematical Sciences, Durham University and Co-Director (Health Data Science) in the Durham research Methods Center

*Ulrike Kulka*, PhD, Biologist and formerly head of the Unit: Biological Dosimetry, Federal Office for Radiation Protection (BfS), Oberschleissheim, Germany.

## ORCID

Ulrike Kulka  <http://orcid.org/0000-0002-7734-3162>

## References

- Ainsbury EA, Lloyd DC. 2010. Dose estimation software for radiation biodosimetry. *Health Phys.* 98(2):290–295.
- Ainsbury EA, Samaga D, Della Monaca S, Marrale M, Bassinet C, Burbidge CI, Correcher V, Discher M, Eakins J, Fattibene P, et al. 2018. Uncertainty on radiation doses estimated by biological and

- retrospective physical methods. *Radiat Prot Dosimetry*. 178(4): 382–404.
- Ainsbury EA, Vinnikov VA, Maznyk NA, Lloyd DC, Rothkamm K. 2013. A comparison of six statistical distributions for analysis of chromosome aberration data for radiation biodosimetry. *Radiat Prot Dosimetry*. 155(3):253–267.
- Akaike H. 1974. A new look at the statistical model identification. *IEEE Trans Automat Contr*. 19(6):716–723.
- Berk R, MacDonald JM. 2008. Overdispersion and Poisson Regression. *J Quant Criminol*. 24(3):269–284.
- Brooks ME, Kristensen K, Benthem KJv, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Mächler M, Bolker BM. 2017. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R J*. 9(2):378.
- Dean C, Lawless JF. 1989. Tests for detecting overdispersion in Poisson regression models. *J Am Stat Assoc*. 84(406):467–472.
- Denwood MJ. 2016. runjags: an R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. *J Stat Soft*. 71(9):1–25.
- Di Giorgio M, Barquinero JF, Vallerga MB, Radl A, Taja MR, Seoane A, De Luca J, Oliveira MS, Valdivia P, Lima OG, et al. 2011. Biological dosimetry intercomparison exercise: an evaluation of triage and routine mode results by robust methods. *Radiat Res*. 175(5): 638–649.
- Dolphin G. 1969. Biological dosimetry with particular reference to chromosome aberration analysis: a review of methods. *Proceedings of the Symposium on the Handling of Radiation Accidents*; May 19–23; Vienna, (Austria). Vienna: IAEA; p. 215–224.
- Einbeck J, Ainsbury EA, Sales R, Barnard S, Kaestle F, Higuera M. 2018. A statistical framework for radiation dose estimation with uncertainty quantification from the  $\gamma$ -H2AX assay. *PLoS One*. 13(11):e0207464.
- Flegal FN, Devantier Y, Marro L, Wilkins RC. 2012. Validation of QuickScan dicentric chromosome analysis for high throughput radiation biological dosimetry. *Health Phys*. 102(2):143–153.
- Flegal FN, Devantier Y, McNamee JP, Wilkins RC. 2010. Quickscan dicentric chromosome analysis for radiation biodosimetry. *Health Phys*. 98(2):276–281.
- Gruel G, Gregoire E, Lecas S, Martin C, Roch-Lefevre S, Vaurijoux A, Voisin P, Voisin P, Barquinero JF. 2013. Biological dosimetry by automated dicentric scoring in a simulated emergency. *Radiat Res*. 179(5):557–569.
- Henningsen A, Toomet O. 2011. maxLik: a package for maximum likelihood estimation in R. *Comput Stat*. 26(3):443–458.
- Higuera M, Puig P, Ainsbury EA, Rothkamm K. 2015. A new inverse regression model applied to radiation biodosimetry. *Proc Math Phys Eng Sci*. 471(2174):20140588.
- Higuera M, Puig P, Ainsbury EA, Vinnikov VA, Rothkamm K. 2016. A new Bayesian model applied to cytogenetic partial body irradiation estimation. *Radiat Prot Dosimetry*. 168(3):330–336.
- Hilbe JM. 2011. *Negative binomial regression*. 2nd ed. Cambridge: Cambridge University Press.
- IAEA. 2011. *Cytogenetic dosimetry: applications in preparedness for and response to radiation emergencies*. Vienna (Austria): EPR-Biodosimetry, International Atomic Energy Agency.
- ISO. 2014. *International Organization for Standardization (ISO), ISO19238:2014. Radiation protection performance criteria for service laboratories performing biological dosimetry by cytogenetics*. Geneva: ISO.
- Kruskal WH, Wallis WA. 1952. Use of ranks in one-criterion variance analysis. *J Am Stat Assoc*. 47(260):583–621.
- Kulka U, Wojcik A, Di Giorgio M, Wilkins R, Suto Y, Jang S, Quingjie L, Jiayang L, Ainsbury E, Woda C, et al. 2018. Biodosimetry and biodosimetry networks for managing radiation emergency. *Radiat Prot Dosimetry*. 182(1):128–138.
- Merkle W. 1983. Statistical methods in regression and calibration analysis of chromosome aberration data. *Radiat Environ Biophys*. 21(3): 217–233.
- Oestreicher U, Endesfelder D, Gomolka M, Kesminiene A, Lang P, Lindholm C, Rößler U, Samaga D, Kulka U. 2018. Automated scoring of dicentric chromosomes differentiates increased radiation sensitivity of young children after low dose CT exposure in vitro. *Int J Radiat Biol*. 94(11):1017–1026.
- Oestreicher U, Samaga D, Ainsbury E, Antunes AC, Baeyens A, Barrios L, Beinke C, Beukes P, Blakely WF, Cucu A, et al. 2017. RENEB intercomparisons applying the conventional Dicentric Chromosome Assay (DCA). *Int J Radiat Biol*. 93(1):20–29.
- Oliveira M, Einbeck J, Higuera M, Ainsbury E, Puig P, Rothkamm K. 2016. Zero-inflated regression models for radiation-induced chromosome aberration data: A comparative study. *Biom J*. 58(2):259–279.
- Pajic J, Rakic B, Rovcanin B, Jovicic D, Novakovic I, Milovanovic A, Pajic V. 2015. Inter-individual variability in the response of human peripheral blood lymphocytes to ionizing radiation: comparison of the dicentric and micronucleus assays. *Radiat Environ Biophys*. 54(3):317–325.
- Papworth DG. 1983. Exact tests of fit for a poisson distribution. *Computing*. 31(1):33–45.
- Plummer M. 2003. JAGS: a program for analysis of bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*; Mar 20–22; Vienna (Austria).
- Puig P, Barquinero JF. 2011. An application of compound Poisson modelling to biological dosimetry. *Proc R Soc A*. 467(2127): 897–910.
- Romm H, Ainsbury E, Barnard S, Barrios L, Barquinero JF, Beinke C, Deperas M, Gregoire E, Koivistoinen A, Lindholm C, et al. 2013. Automatic scoring of dicentric chromosomes as a tool in large scale radiation accidents. *Mutat Res*. 756(1-2):174–183.
- Romm H, Oestreicher U, Kulka U. 2009. Cytogenetic damage analysed by the dicentric assay. *Ann Ist Super Sanita*. 45(3):251–259.
- Sasaki MS, Miyata H. 1968. Biological dosimetry in atomic bomb survivors. *Nature*. 220(5173):1189–1193.
- Savage JRK, Papworth DG. 2000. Constructing a 2B calibration curve for retrospective dose reconstruction. *Radiat Prot Dosim*. 88(1): 69–76.
- Schunck C, Johannes T, Varga D, Lorch T, Plesch A. 2004. New developments in automated cytogenetic imaging: unattended scoring of dicentric chromosomes, micronuclei, single cell gel electrophoresis, and fluorescence signals. *Cytogenet Genome Res*. 104(1-4):383–389.
- Schwarz G. 1978. Estimating the dimension of a model. *Ann Statist*. 6(2):461–464.
- Słonecka I, Łukasik K, Fornalski KW. 2018. Analytical and quasi-Bayesian methods as development of the iterative approach for mixed radiation biodosimetry. *Radiat Environ Biophys*. 57(3): 195–203.
- Słonecka I, Łukasik K, Fornalski KW. 2019. Simplified Bayesian method: application in cytogenetic biological dosimetry of mixed  $n + \gamma$  radiation fields. *Radiat Environ Biophys*. 58(1):49–57.
- Vaurijoux A, Gregoire E, Roch-Lefevre S, Voisin P, Martin C, Voisin P, Roy L, Gruel G. 2012. Detection of partial-body exposure to ionizing radiation by the automatic detection of dicentrics. *Radiat Res*. 178(4):357–364.
- Vaurijoux A, Gruel G, Pouzoulet F, Gregoire E, Martin C, Roch-Lefevre S, Voisin P, Voisin P, Roy L. 2009. Strategy for population triage based on dicentric analysis. *Radiat Res*. 171(5):541–548.
- Wilkins RC, Romm H, Oestreicher U, Marro L, Yoshida MA, Suto Y, Prasanna PG. 2011. Biological dosimetry by the triage dicentric chromosome assay – further validation of international networking. *Radiat Meas*. 46(9):923–928.