RENEB Inter-Laboratory Comparison 2021: The Cytokinesis-Block Micronucleus Assay

Vral A[1*], Endesfelder D[2*], Balázs K[3], Beinke C[4], Cuceu Petrenci C[5], Finot F[5], Garty G[6], Hadjiiska L[7], Hristova R[7], Ivanova I[7], Lee Y[8], Lumniczky K[3], Milanova M[9], Monteiro Gil O[10], Oestreicher U[2], Pajic J[11], Patrono C[12], Pham ND[13], Perletti G[1], Seong KM[8], Sommer S[14], Szatmári T[3], Testa A[12], Tichy A[9], Tran TM[13], Wilkins R[15], Port M[4], Abend M[4], Baeyens A[1]

[1]Faculty of Medicine and Health Sciences, Radiobiology Research Unit, Universiteit Gent, Gent, Belgium

[2]Bundesamt für Strahlenschutz, BfS, Oberschleissheim, Germany

[3]National Public Health Center, Department of Radiobiology and Radiohygiene, Unit of Radiation Medicine, Budapest, Hungary

[4]Bundeswehr Institute of Radiobiology, Munich, Germany

[5]Genevolution, Porcheville, France

[6]Center for Radiological Research, Columbia University, New York, US

[7]National Centre of Radiobiology and Radiation Protection, Sofia, Bulgaria

[8]Laboratory of Biological Dosimetry, National Radiation Emergency Medical Center, Korea Institute of Radiological & Medical Sciences, Seoul, Republic of Korea

[9]Department of Radiobiology, Faculty of Military Health Sciences, University of Defence, Hradec Králové, Czech Republic

[10]Centro de Ciências e Tecnologias Nucleares, Instituto Superior Técnico (IST), Universidade de Lisboa, Lisbon, Portugal

[11]Serbian Institute of Occupational Health, Belgrade, Serbia

[12]Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile, Rome, Italy

[13]Center Radiation technlogy & Biotechnology; Dalat Nuclear Research Institute; Dalat City, Vietnam

[14]Institute of Nuclear Chemistry and Technology, Warsaw, Poland

[15]Health Canada, Radiation Protection Building, Ottawa, Canada

*shared first author

Corresponding author:

Prof. Dr. Anne Vral
Radiobiology Research Unit
Dept. Human Structure and Repair
Faculty of Medicine and Health Sciences
Ghent University
Corneel Heymanslaan 10
Ghent, Belgium
email: anne.vral@ugent.be
phone number: +32-9-3325129

Running title: RENEB inter-laboratory comparison 2021 – micronucleus assay

Key words: micronucleus, biodosimetry, radiation accident, triage scoring, RENEB

Vral A, Endesfelder D, Balázs K, Beinke C, Cuceu Petrenci C, Finot F, Garty G, Hadjiiska L, Hristova R, Ivanova I, Lee Y, Lumniczky K, Milanova M, Monteiro Gil O, Oestreicher U, Pajic J, Patrono C,Pham ND, Perletti G, Seong KM, Sommer S, Szatmári T, Testa A, Tichy A, Tran TM, Wilkins R, Port M, Abend M, Baeyens A. RENEB Inter-Laboratory Comparison 2021: The Cytokinesis-Block Micronucleus Assay. *Radiat. Res.*

Abstract

Purpose

The goal of the RENEB inter-laboratory comparison 2021 exercise was to simulate a large-scale radiation accident involving a network of biodosimetry labs. Labs were required to perform their analyses using different biodosimetric assays in triage mode scoring and to rapidly report estimated radiation doses to the organizing institution. This article reports the results obtained with the cytokinesis-block micronucleus assay.

Methods

Three test samples were exposed to blinded doses of 0, 1.2 and 3.5 Gy X-ray doses (240 kVp, 13 mA, ~75 keV, 1 Gy/min). These doses belong to 3 triage categories of clinical relevance: a *low dose* category, for no exposure or exposures inferior to 1 Gy, requiring no direct treatment of subjects; a *medium dose* category, with doses ranging from 1 to 2 Gy, and a *high dose* category, following exposure to doses higher than 2 Gy, with the two latter requiring increasing medical attention. After irradiation the test samples (#1, #2 and #3) were sent by the organizing laboratory to 14 centers participating in the micronucleus assay exercise. Laboratories were asked to setup micronucleus cultures and to perform the micronucleus assay in triage mode, scoring 500 binucleated cells manually, or 1000 binucleated cells in automated/semi-automated mode. One laboratory received no blood samples, but scored pictures from another lab. Based on their calibration curves, laboratories had to provide estimates of the administered doses. The accuracy of the reported dose estimates was further

analyzed by the micronucleus assay lead.

Results

The micronucleus assay allowed classification of samples in the corresponding *clinical triage* categories (*low*, *medium*, *high dose* category) in 88% of cases (manual scoring, 88%; semi-automated scoring, 100%; automated scoring, 73%). Agreement between scoring laboratories, assessed by calculating the Fleiss' kappa, was excellent (100%) for semi-automated scoring, good (83%) for manual scoring and poor (53%) for fully automated scoring.

Correct classification into *triage scoring dose intervals* (reference dose $\pm$0.5 Gy for doses $\leq$ 2.5 Gy, or reference dose $\pm$1 Gy for doses >2.5 Gy), recommended for triage biodosimetry, was obtained in 79% of cases (manual scoring, 73%; semi-automated scoring, 100%; automated scoring, 67%).

The percentage of dose estimates whose 95% confidence intervals included the reference dose was 58% (manual scoring, 48%; semi-automated scoring, 72%; automated scoring, 60%). For the irradiated samples #2 and #3, a systematic shift towards higher dose estimations was observed. This was also noticed with the other cytogenetic assays in this intercomparison exercise.

Accuracy of the rapid triage modality could be maintained when the number of manually scored cells was scaled down to 200 binucleated cells.

Conclusions

In conclusion, the micronucleus assay, preferably performed in a semi-automated or manual scoring mode, is a reliable technique to perform rapid biodosimetry analysis in large-scale radiation emergencies.

Introduction

When large-scale radiological emergencies occur, initial triage using biodosimetric assays will support the classification of victims according to their degree of exposure. Such classification into categories is essential to guide appropriate clinical responses by distinguishing '*worried-well*' individuals, not requiring particular attention, from those who actually need immediate medical care (development of life-threatening acute health effects) or surveillance because of increased risk of later occurring health effects (cancer and non-cancer diseases). This screening is also necessary to prevent the overcrowding of healthcare facilities *(1 ,2)*. In a previous project *(3)* focusing on the development of multi-disciplinary biodosimetric tools to manage large-scale radiological casualties, 3 triage categories of clinical relevance were established: (i) a *low dose* category, for exposures inferior to 1 Gy (green code), requiring no direct treatment of subjects who were not irradiated or minimally irradiated; (ii) a *medium dose* category, with doses ranging from 1 to 2 Gy (yellow code), and (iii) a *high dose* category, following exposure to doses higher than 2 Gy (red code), with the two latter requiring increasing medical attention *(1 ,4)*.

Given that a single biodosimetry laboratory can analyze only few dozens of victims, laboratory networks must be established to manage large-scale radiation accidents. In 2012, the RENEB project (*Realizing* the European Network of Biodosimetry), funded by the European Commission (EURATOM, FP 7, GA 295513). was initiated to ensure the implementation and harmonization of different assays in a large number of European laboratories, equipped (expertise, infrastructure, etc.) for emergency response strategies *(5)*.

Since 2017, RENEB (*Running* the European Network of biological and physical retrospective dosimetry, https://www.reneb.net/) became a legal association *(6)* after the end of the RENEB project.

Within the RENEB network, regular inter-laboratory comparison exercises have been organized for harmonisation and training as well as to maintain the preparedness for large-scale emergency situations, using a broad variety of biodosimetric assays *(7-19)*.

Among these assays, the well-established cytokinesis-block micronucleus (CBMN) test *(20)* is widely used for biological dosimetry. Compared with the gold-standard, the dicentric chromosome assay, the scoring of micronuclei (MN) is easier and faster to perform, albeit it requires a slightly longer time in culture. The development of automated MN scoring methodologies, based on high-speed microscopy image analysis *(21-24)* or on flow cytometric technologies *(25)*, have made the CBMN assay suitable and attractive for the emergency triage of large-scale radiation accidents. In addition, the RABiT-II tool, a fully automated, miniaturized high-throughput version of the CBMN assay, allowing accelerated sample processing, has been described *(26)*. At present, several inter-laboratory comparison studies have confirmed the reliability of the automated CBMN assay for high throughput population triage *(9 ,27 ,28)*.

The goal of the RENEB inter-laboratory comparison (ILC) 2021 exercise was to simulate a large-scale radiation accident involving a network of biodosimetry labs. The labs were asked to process blood samples according to their standard procedure, to perform their analyses in a triage mode and to report the results - in the form of estimated radiation doses - as quickly as possible. Besides the assay-specific results which are described in separate papers, the ILC exercise allowed to compare the results obtained with the different assays, including the CBMN assay, on the basis of

reporting time, dose estimates, and identification of triage categories based on clinical relevance. These results are described in the Interassay paper of this exercise *(29)* .

Materials and methods

Setup of the ILC exercise: participating laboratories, irradiation and shipment of the blood samples

Blood samples from a single male healthy donor (age: 32 years), were taken after obtaining ethical approval and informed consent to be used for all biological assays. The blood samples were *in vitro* irradiated at the facility of the Bundeswehr Institute of Radiobiology (BIR), with 240 kVp X-rays (13 mA, ~75 keV) at a dose rate of 1 Gy/min. Irradiation was performed at room temperature and samples were kept at room temperature until shipment to all participating labs.

Three blood samples were included in the RENEB ILC 2021 and the corresponding doses were blinded and coded and the samples are referred to as test sample #1 (0 Gy), #2 (1.2 Gy) and #3 (3.5 Gy). Samples were irradiated with doses belonging, according to previous consensus, to the 3 main clinical triage categories described in the introduction section: *low*, *medium* and *high* dose (for details see *(29)*).

Samples were shipped to all partner labs for processing, blind scoring and triage dose assessment. For the CBMN assay and other cytogenetic assays, no calibration samples were sent on beforehand to the participants. Only participants who had previously established their own CBMN calibration curve could take part in the exercise. Details of the calibration curves and conditions are shown in Tables 1 and 2.

For the CBMN assay, coded blood samples were shipped to 14 laboratories (RENEB members and non-member institutions) in 13 different countries. Eleven labs were

from Europe, 1 from Asia, 1 from Canada and 1 from the US. An additional partner lab in Asia received no blood samples, but scored pictures from another lab. One lab could not establish cell cultures, due to technical problems. Taken together, a total of 14 laboratories provided micronuclei scoring data and radiation dose estimates.

Cytokinesis-block micronucleus assay

The 14 laboratories participating in the CBMN assay ILC (L1-L14) were requested to culture the 3 test samples (#1, #2, #3) immediately on the day of arrival according to the standard procedures which were also used to setup their calibration curves. The labs also received a scoring sheet, in which they were asked to provide information on the arrival time of the blood sample, on culture and preparation details, on the scoring mode utilized (manual, semi-automated or automated), as well as details on the lab-specific calibration curve. Labs were asked to perform MN analyses and dose estimations only in triage mode, since the aim of this exercise was to simulate a large-scale emergency situation. Labs were asked to score 1000 binucleated (BN) cells for automated/semi-automated scoring and 500 BN cells for manual scoring. Scorings should be ideally performed on a single slide to keep the time for analysis as short as possible. Labs performing manual scoring were also asked to provide data after scoring 200 BN cells, to compare the accuracy of dose estimation in both conditions (200 vs. 500 BN cells). For the automated or semi-automated scoring methods, cells were stained with DAPI and a Metafer platform (Metasystems, Altlussheim, Germany) was used. Semi-automated scoring was based on the visual rating of the sole MN-positive BN cells that were sorted automatically. In total, 11 labs performed manual MN scoring (i.e., detection of BN cells and MN counts performed at the microscope by a dedicated scorer), 5 labs performed combined automated and semi-automated

scoring and 1 lab performed only semi-automated scoring of cells stained with a centromere/telomere FISH protocol. Protocol details for the automated/semi-automated and manual CBMN assays are given in Table 3.

Once dose estimations were calculated, the labs were asked to immediately send their scoring sheets with the results to the coordinator of the ILC exercise, who registered the reporting time for each laboratory.

The dose estimates received from participating labs were centrally evaluated by ILC work package leaders using three research questions, addressing the accuracy of the dose estimation, going from low granularity to high granularity: (i) did the dose estimates allow correct allocation of samples to the triage groups based on the specific clinical relevance (green, yellow, red code)?, (ii) do the dose estimates fall within the ±0.5 Gy interval of the reference dose (for doses ≤2.5 Gy), or within the ±1.0 Gy interval of the reference dose (for doses >2.5 Gy), as indicated for triage scoring by Lloyd and coworkers *(30)*, (iii) do the 95% CIs of the estimated doses, calculated from the calibration curves of each lab, include the administered reference doses?

Data analysis

Fourteen labs performed a total of 22 scorings of MN for each of the three administered doses, using one or more scoring technique (manual, n=11; semi-automated; n=6, automated, n=5), summing up to total 66 assessments. To calculate each estimated dose and corresponding 95% confidence intervals (CIs), the Biodose Tools software *(31)* was used for 18 scorings, and other software packages ('Dose Estimate', *(32)*, n=3, CABAS, *(33)*, n=1) were used for 4 scorings in total.

The Fleiss' fixed-marginal multirater Kappa, and its 95% confidence interval, was calculated to assess the extent of agreement between the rating labs involved in the

ILC exercise *(34)*. The degree of agreement depended on the classification of an estimated dose into the corresponding triage groups according to clinical relevance (green, yellow or red codes).

Values of the Fleiss' kappa can range from -1.0 to 1.0, with -1.0 indicating total disagreement, 0.0 indicating agreement equal to chance, and 1.0 indicating total agreement. According to Fleiss, kappa values less than 0.4 indicate "poor agreement", values from 0.4 to 0.75 indicate "intermediate to good agreement" and values above 0.75 show "excellent agreement".

Median deviations of estimated doses from the administered reference doses were calculated and box-and-whisker plots were produced.

Differences between the absolute deviations of the estimated doses from the reference doses, obtained by scoring 200 or 500 binucleated cells, were analyzed using a Mann-Whitney-Wilcoxon test.

Results

Delivery and reporting times

Eight out of total 14 blood samples arrived at the partner labs within 24 hours, i.e., one day after sample irradiation. Two labs, located near the irradiation facility, received the samples directly, without shipment. Four labs (L3, L5, L8, L9) (2 in the EU and 2 outside the EU) received the samples 2 days after irradiation (range: 41-53h) and one lab (L11, a European country outside the EU) received the samples 69 hours after irradiation. Delivery and reporting times are listed in Table 4. Additional information about shipment details, reporting time and arisen problems is provided in the inter-assay, overview paper (*(29)*, submitted to this journal issue). The range between reporting times for the participating laboratories is rather broad, as it depends on the priority given to the CBMN assay by labs performing multiple assays. The earliest dose estimation reports for the CBMN assay were received by the organizer within 3 days, the latest was sent after 37 days, with a deadline of 42 days set by the organizing laboratory (Table 4).

MN calibration (dose-response) curves

Information about the characteristics of the MN calibration curves is shown in Tables 1 and 2. Figure 1 also shows the MN dose-response curves obtained by each lab, classified according to specific scoring methods (manual scoring, n=11; semi-automated scoring, n=6; automated scoring, n=5). Ten, 4 and 8 calibration curves were obtained by irradiating whole blood samples with Co-60 gamma rays, with Cs-137 gamma rays and with X-rays, respectively.

The procedures applied to set up MN cultures for the calibration curves were in general comparable (Table 3). Differences were especially observed with respect to the time

of cytochalasin B addition and the staining method. Laboratories performing both semi-automated and automated scoring added cytochalasin B 23-24 h after starting cell cultures. Among labs performing manual scoring, four added cytochalasin B after 23-24 h, and 7 after 44-45 h in culture. Three different staining methods were used for manual scoring (Giemsa, acridine orange and DAPI), while DAPI staining was used by all labs performing semi-automated and automated scoring. The degree of experience of the scorers varied between 2 years and over 30 years.

Heterogeneity of calibration curves was observed for all scoring methods but was more pronounced after manual scoring (Figure 1).

Dose estimations

Table 4 shows the dose estimates reported by each lab after analysis of the three test samples, grouped according to the three scoring methods.

As shown in the Methods section, three research questions addressed the accuracy of the dose estimation.

(i) do the dose estimates allow correct allocation of samples to the triage groups based on the specific clinical relevance (low [green code], medium [yellow code] or high dose [red code] category)?

The classification of test samples according to clinical relevance is shown in Table 4. The total number of samples assigned to the correct clinical category is also shown. The unirradiated sample (#1) was correctly classified in clinical category 1 (0-1 Gy, green code) by all participants who performed manual and semi-automated scoring, and 16/17 of the participants estimated a dose of exactly 0 Gy. Using the fully automated scoring procedure, 2 out of 5 labs misclassified the sample. Dose estimates of 6.9 Gy and 1.6 Gy were reported by L2 and L6.

Sample #2 (1.2 Gy) was classified in the correct clinical category (1-2 Gy, yellow code) by 7/11 participants using manual scoring, by 6/6 using semi-automated scoring and by 3/5 using automated scoring. Sample #3 (3.5 Gy) was classified in the category of high exposure (> 2 Gy, red code) by 22/22 of participants using any scoring method.

Table 5 summarizes the allocation to clinical relevance groups of test samples #1, #2 and #3, and shows the analysis of the agreement between scoring laboratories, performed by calculating the Fleiss' kappa. Results are presented for each scoring technique, or for merged techniques.

Analysis shows that the use of a semi-automated technique resulted in 100% inter-rater agreement, with a kappa value of 1 (95% CI, 0.79 to 1), which was rated as "excellent". The inter-rater agreement for manual MN scoring was high (83%) and the kappa value was rated as "good" (k=0.73; 95% CI, 0.58 to 0.88). When manual and semi-automated techniques were merged, 87% agreement was obtained, and the kappa was rated as "excellent" (k=0.81, 95%CI, 0.72 to 0.89). Automated scoring resulted in low inter-rater agreement (53%), with kappa equal to 0.27 (95% CI, -0.3 to 0.57), rated as "poor". Overall, the 80% agreement achieved by merging all available scoring techniques (n=66 assessments) resulted in a "good" kappa value equal to 0.69 (95% CI, 0.61-0.76).

(ii) do the dose estimates fall within the ± 0.5 Gy interval of the true doses for doses ≤ 2.5 Gy or within the ±1.0 Gy interval for doses >2.5 Gy?

Table 4 and Figure 2 provide information on the accuracy of classification of the test samples based on the error range accepted for triage scoring *(30)*.

For sample #1, only two estimates out of 22 exceeded the reference dose by more than 0.5 Gy, likely due to background noise.

Sample #2 was classified in the correct triage category (±0.5 Gy) by 7/11, 6/6 and 3/5 participants performing manual, semi-automated and automated scoring, respectively. The median deviation from the reference dose was 0.44 Gy for manual scoring, 0.33 Gy for semi-automated scoring and 0.42 Gy for automated scoring (plot is shown in Figure 2), in all cases inferior to 0.5 Gy.

Correct classification of sample #3 (within the ±1 Gy range) was achieved by 6/11, 6/6 and 4/5 participants performing manual, semi-automated and automated scoring, respectively. The median deviation from the true dose was 0.72 Gy for manual scoring, 0.27 Gy for semi-automated scoring and 0.23 Gy for automated scoring (Figure 2), in all cases inferior to 1.0 Gy. Results more adherent to the value of the reference dose were obtained by semi-automated scoring.

(iii) do the 95% CIs of the estimated doses include the real dose?

Twenty over 22 provided estimates included the 0 Gy dose in the 95% CI of the dose estimate of the non-irradiated sample (Table 4, Figure 3). Two laboratories (L2, L6), performing automated scoring, did not include the 0 Gy reference dose in the 95% CI of the estimated dose.

When scorings of MN yields performed with any technique were merged (22 scorings), 46% (10/22) and 36% (8/22) of the CIs of the estimated doses for samples #2 and #3, respectively, contained the value of the reference dose (Table 4, Figure 3). Ten over a total of 22 estimations (manual scoring, 6; semi-automated, 3; automated, 1) showed a 95% CI to the estimated dose whose lower limit was higher than the 1.2 Gy reference dose. A similar over-estimation (manual scoring, 8, semi-automated, 1, automated, 1) was observed for the 3.5 Gy reference dose.

Comparison of binucleated cell thresholds for triage scoring

The main aim of this ILC was to simulate a large radiation accident in which the labs are asked to perform their analysis in triage mode and to report the results as quickly as possible. This can be achieved by scoring 500 BN cells manually, or 1000 BN cells automatically/semi-automatically on a single slide. Notably, biodosimetry assessments following ISO standards are performed on two slides from duplicate cultures. In this ILC we investigated if a further reduction to 200 manually-scored BN cells, reducing the scoring time from approximately 30 to 15 minutes, would yield accurate dose estimates.

The results of this comparison are presented in Figure 4. In general, the 95% CIs are wider if only 200 BN cells are scored. However, the absolute deviations from the reference doses were not significantly different between scoring of 500 and 200 BN cells for any of the three test samples ($P>0.05$, Wilcoxon-Mann-Whitney test for all comparisons).

Discussion

The legal association RENEB, *Running the European NEtwork of Biological and retrospective physical dosimetry*, is a Network comprising 16 voting member organizations from 13 European countries. It is trained to act in a concerted way to supplement early triage of radiation victims in the case of large-scale radiological emergency. To ensure high and consistent quality of biodosimetry services, regular exercises are performed to train and maintain the capacities of the participating labs in providing accurate dose estimates. The aim of the 2021 ILC exercise was to simulate a large-scale radiation emergency in order to test the logistics of a large event and to compare the results obtained with various assays used in parallel. The endpoint

of this exercise was to assess dose estimates allowing the categorization of radiation victims into triage groups according to clinical relevance *(1 ,4)*.

For the cytokinesis-block micronucleus assay, 14 labs were involved, and 3 different MN scoring methods (manual/semi-automated/automated) were used.

Calibration (dose-response) curves

Previous ILCs have shown a large variation of MN calibration curves, which does not allow the use of a common curve *(9 ,27 ,35)*. Also in this ILC, differences between calibration curves were observed, with the highest heterogeneity shown by manual scoring (Figure 1). This may be partly due to the variety of cell culture conditions and staining methods adopted by labs performing manual MN scoring. In addition, when manual scoring is performed, the stringency in applying similar criteria for identifying binucleated cells and micronuclei is more prone to variability and is more subjective, compared to semi-automated or automated scoring. Moreover, in the latter two, most labs use the same protocol and MN classifiers.

Besides the differences discussed above, differences in radiation sources (viz., Co-60, Cs-137 gamma rays and X-rays) used to set up calibration curves can be partly responsible for the observed variability.

Variability was also observed with respect to how calibration curves were generated by each lab. According to the ISO standard for the CBMN assay (ISO-17099-2014 *(36)*), a calibration curve should preferentially be produced from at least 6 donors of varying age and gender, with comparable numbers of BN cells being scored from each donor. Moreover, at least 7 doses, including 0 Gy, should be administered. For this ILC, the number of donors used for constructing calibration curves varied among labs between 1 and 16 (Table 2). In addition, the number of BN cells scored per dose point

per donor varied markedly (between 200 and 10689 BN cells, Table 2). This exercise demonstrates that in various labs appropriate MN calibration curves, established according to ISO-17099-2014, are missing. To ensure high quality biodosimetry services in the future, the network should promote the setup of appropriate curves according to ISO standards. In order to harmonize the practice of participating labs, a quality manual based on ISO standards has been produced within the RENEB network by Voisin et al. *(37 ,38)*.

Reporting time of dose estimates

In this exercise, the dose estimation reporting time was recorded. For the CBMN assay, the first results were received by day 3 (76h:16min), which is the minimum reporting time technically possible, given that the standard culture time is 70-72 hours. Most labs had much longer reporting times. This is because all labs were also performing the dicentric assay and in some cases even additional assays. Thus, MN scoring and reporting of dose estimations have been performed with low priority in several cases. In the NATO biodosimetry study of Romm et al. *(27)* similar reporting times, ranging between 4 and 17 days, depending on the priority given to the CBMN assay, were obtained.

Allocating the correct clinical category and precision of dose estimates

When a large radiological incident occurs, the main task of a biodosimetry network is to perform a quick estimate of the dose, allowing the categorization of casualties in triage groups (green, yellow and red code) in order to inform medical decision makers and to reassure unexposed persons (the so-called ‘worried well’ population). Using

the CBMN assay, results were obtained by three scoring methods - manual, semi-automated and automated – and the suitability for triage dose estimation was evaluated (Tables 4 and 5).

Considering all dose estimates (n=66) obtained by using different scoring techniques, the agreement between labs in allocating the samples to the correct triage groups was good (80% achieved agreement).

When the different techniques were considered separately, the best results were obtained with the semi-automated scoring procedure (100% of samples correctly classified with 100% interscorer agreement), followed by the manual scoring procedure (88% of samples correctly classified with 83% agreement). Automated scoring performed less efficiently (73% of samples correctly classified, 53% interscorer agreement). The misclassifications which occurred using automated analysis of samples #1 (0 Gy) and #2 (1.2 Gy) were mainly due to the presence of falsely positive micronuclei, representing background noise. Such noise may result in the misclassification of non-exposed or minimally exposed individuals ('worried well' cases). With semi-automated scoring, where BN cells with MN are visually checked by the operator, all samples were correctly classified. This confirms what has been shown in previous exercises, namely that the semi-automated method is the most appropriate procedure *(11 ,27 ,28)*.

Semi-automated scoring also appears a preferable procedure, since it allowed dose estimations falling in all cases (100%) within the prescribed triage reference intervals (±0.5 Gy for doses <2.5 Gy, ±1.0 Gy for doses >2.5 Gy; Table 4, Figure 3).

The participating labs also provided data assessing the presence of reference doses in the 95% CIs of the estimated doses. This represents a stringent strategy to evaluate dose estimates, especially in a triage setting. Ninety-one percent of the confidence

intervals of the dose estimates correctly contained the 0 Gy dose. Thus, the measured micronucleus yield was not significantly different from the background yield present in the general population. The reference dose values of samples #2 (1.2 Gy) and #3 (3.5 Gy) were included in the 95% CIs of the estimated doses in 46% (sample #2) and 36% (sample #3) of cases. In 45% of cases, doses were overestimated, i.e. the reference dose was inferior to the lower limit of the 95% CI of the estimated dose. This was not only observed when the CBMN assay was used, but also when other cytogenetic methods (e.g., dicentric or translocation assays) were applied in this exercise.

The possible reasons for such a shift to higher dose estimations can be manifold: (i) although the test samples were irradiated with 240 kVp X-rays, many labs used calibration curves obtained by using gamma-rays. Since the latter are characterized by a lower relative biological effectiveness *(39)*, an overestimation of the real administered dose might have occurred; (ii) irradiation and further storage of the test samples at room temperature might have affected the efficiency of DNA damage repair, compared to irradiation and incubation performed at 37°C. This might have increased the MN yield, resulting in turn in an overestimation of the reference dose; (iii) the blood specimens which were blindly irradiated might have been sampled from a donor showing a mildly radiosensitive phenotype. The systematic shift to higher dose estimations is also discussed in-depth in the article by Endesfelder et al. *(40)*.

Whereas overestimation of the dose can be somewhat tolerated in a triage setting - as no subject will be dangerously assigned to a lower clinical relevance category - underestimation of the received dose may have an impact on the victim's prognosis and on the clinical attention given to the same person. Reassuringly, in our MN exercise, underestimation of the reference dose, based on 95% CIs, occurred in one case out of 22 (4%) relative to the 1.2 Gy dose and in 3 cases (13%) relative to the

3.5 Gy dose. More reassuringly, as shown above, in our exercise based on the CBMN assay, 58 estimated doses out of 66 (88%) were assigned to the correct clinical relevance category.

Comparison of binucleated cell thresholds

The ISO standard for the MN assay (ISO-17099-2014) *(36)* and the IAEA 2011 procedure manual *(41)* recommend the performance of *rapid* triage scoring in case of large radiation accidents. This can be implemented by scoring 200 BN cells manually, or by automated/semiautomated analysis of higher numbers (e.g., 1000) of BN cells. This procedure, however, needs further validation, as few studies have been published so far. Wilkins et al. *(35)* have shown that scorings performed on 200 BN cells result in good dose estimates. McNamee et al. *(25)* investigated counts performed on 50 to 500 BN cells, observing an increasing sensitivity of the CBMN assay with higher cell numbers. When 200 cells were scored, a dose of 1 Gy could be detected over background rates. However, no comparison was done with estimates obtained scoring higher numbers of BN cells. In the NATO study of Romm et al. *(27)*, one lab analyzed 200 and 2000 BN cells per dose point by manual scoring and found no improved accuracy with higher cell numbers. Depuydt et al. *(9)* compared the results obtained by manually scoring 500 vs. 1000 BN cells. No significant differences in estimated doses were observed. The data obtained in this exercise show that no significant differences in absolute deviations from the reference dose are observed when 200 or 500 BN cells are scored. If further validated, these results indicate that scoring 200 BN cells may be appropriate for rapid triage biodosimetry. Such procedure can reduce the scoring time from 30 min to approximately 15 min per slide, which is comparable with the time needed to score 1000 BN cells in semi-automated mode.

In conclusion, the cytokinesis-block micronucleus assay allowed classification of samples in the correct triage categories, based on clinical relevance, in 88% of cases. Agreement between scoring laboratories, assessed by calculating the Fleiss' kappa, was excellent for semi-automated scoring, good for manual scoring and poor for fully automated scoring. Accuracy of the rapid triage modality can be reasonably maintained when the number of manually scored cells is scaled down to 200 binucleated cells.

These data confirm that the CBMN assay, preferably performed in a semi-automated or manual scoring mode, is a reliable and suitable technique to achieve rapid biodosimetry in large-scale radiation emergencies.

Acknowledgment

We thank all the RENEB work-package leaders for their contribution to the discussion and organisation of the RENEB ILC 2021. We thank the blood donor for providing the blood samples. We thank all the technical staff from the participating labs that were involved in the processing and scoring of micronuclei.

This work was supported by RENEB and the German Ministry of Defense. The views expressed in this publication are those of the authors and not necessarily those of the funding bodies.

References

1. Jaworska A, Ainsbury EA, Fattibene P, Lindholm C, Oestreicher U, Rothkamm K, et al. Operational guidance for radiation emergency response organisations in Europe for using biodosimetric tools developed in EU MULTIBIODOSE project. Radiation Protection Dosimetry. 2014.164(1-2):165-69. doi: 10.1093/rpd/ncu294

2. Bushberg JT, Kroger LA, Hartman MB, Leidholdt EM, Miller KL, Derlet R, et al. Nuclear/radiological terrorism: Emergency department management of radiation casualties. The Journal of Emergency Medicine. 2007.32(1):71-85. doi: 10.1016/j.jemermed.2006.05.034

3. Jaworska A, Wojcik A, Ainsbury EA, Fattibene P, Lindholm C, Oestreicher U, et al. Guidance for using MULTIBIODOSE tools in Emergencies for adiation Emergency Response Organisations in Europe. RENEB: RENEB, 2017:52.

4. Ainsbury EA, Barnard S, Barrios L, Fattibene P, de Gelder V, Gregoire E, et al. Multibiodose Radiation Emergency Triage Categorization Software. Health Physics. 2014.107(1):83-89. doi: 10.1097/hp.0000000000000049

5. Kulka U, Ainsbury L, Atkinson M, Barquinero JF, Barrios L, Beinke C, et al. Realising the European Network of Biodosimetry (RENEB). Radiation Protection Dosimetry. 2012.151(4):621-25. doi: 10.1093/rpd/ncs157

6. Kulka U, Abend M, Ainsbury E, Badie C, Barquinero JF, Barrios L, et al. RENEB – Running the European Network of biological dosimetry and physical retrospective dosimetry. International Journal of Radiation Biology. 2016.93(1):2-14. doi: 10.1080/09553002.2016.1230239

7. Barnard S, Ainsbury EA, Al-hafidh J, Hadjidekova V, Hristova R, Lindholm C, et al. The first gamma-H2AX biodosimetry intercomparison exercise of the developing European

biodosimetry network RENEB. Radiation Protection Dosimetry. 2014.164(3):265-70. doi: 10.1093/rpd/ncu259

8. Oestreicher U, Samaga D, Ainsbury E, Antunes AC, Baeyens A, Barrios L, et al. RENEB intercomparisons applying the conventional Dicentric Chromosome Assay (DCA). International Journal of Radiation Biology. 2016.93(1):20-29. doi: 10.1080/09553002.2016.1233370

9. Depuydt J, Baeyens A, Barnard S, Beinke C, Benedek A, Beukes P, et al. RENEB intercomparison exercises analyzing micronuclei (Cytokinesis-block Micronucleus Assay). International Journal of Radiation Biology. 2016.93(1):36-47. doi: 10.1080/09553002.2016.1206231

10. Barquinero JF, Beinke C, Borràs M, Buraczewska I, Darroudi F, Gregoire E, et al. RENEB biodosimetry intercomparison analyzing translocations by FISH. International Journal of Radiation Biology. 2016.93(1):30-35. doi: 10.1080/09553002.2016.1222092

11. Terzoudi GI, Pantelias G, Darroudi F, Barszczewska K, Buraczewska I, Depuydt J, et al. Dose assessment intercomparisons within the RENEB network using G(0)-lymphocyte prematurely condensed chromosomes (PCC assay). International journal of radiation biology. 2017.93(1):48-57. doi: 10.1080/09553002.2016.1234725 [published Online First: 2016/11/04].

12. Moquet J, Barnard S, Staynova A, Lindholm C, Monteiro Gil O, Martins V, et al. The second gamma-H2AX assay inter-comparison exercise carried out in the framework of the European biodosimetry network (RENEB). International Journal of Radiation Biology. 2016.93(1):58-64. doi: 10.1080/09553002.2016.1207822

13. Manning G, Macaeva E, Majewski M, Kriehuber R, Brzóska K, Abend M, et al. Comparable dose estimates of blinded whole blood samples are obtained independently of culture conditions and analytical approaches. Second RENEB gene expression study.

International Journal of Radiation Biology. 2016.93(1):87-98. doi: 10.1080/09553002.2016.1227105

14. Trompier F, Baumann M, Barrios L, Gregoire E, Abend M, Ainsbury E, et al. Investigation of the influence of calibration practices on cytogenetic laboratory performance for dose estimation. International Journal of Radiation Biology. 2016.93(1):118-26. doi: 10.1080/09553002.2016.1213455

15. Wojcik A, Oestreicher U, Barrios L, Vral A, Terzoudi G, Ainsbury E, et al. The RENEB operational basis: complement of established biodosimetric assays. International Journal of Radiation Biology. 2016.93(1):15-19. doi: 10.1080/09553002.2016.1235296

16. Abend M, Amundson SA, Badie C, Brzoska K, Hargitai R, Kriehuber R, et al. Inter-laboratory comparison of gene expression biodosimetry for protracted radiation exposures as part of the RENEB and EURADOS WG10 2019 exercise. Scientific reports. 2021.11(1):9756-56. doi: 10.1038/s41598-021-88403-4

17. Endesfelder D, Oestreicher U, Kulka U, Ainsbury EA, Moquet J, Barnard S, et al. RENEB/EURADOS field exercise 2019: robust dose estimation under outdoor conditions based on the dicentric chromosome assay. International Journal of Radiation Biology. 2021.97(9):1181-98. doi: 10.1080/09553002.2021.1941380

18. Gregoire E, Barquinero JF, Gruel G, Benadjaoud M, Martinez JS, Beinke C, et al. RENEB Inter-Laboratory comparison 2017: limits and pitfalls of ILCs. International Journal of Radiation Biology. 2021.97(7):888-905. doi: 10.1080/09553002.2021.1928782

19. Waldner L, Bernhardsson C, Woda C, Trompier F, Van Hoey O, Kulka U, et al. The 2019–2020 EURADOS WG10 and RENEB Field Test of Retrospective Dosimetry Methods in a Small-Scale Incident Involving Ionizing Radiation. Radiation Research. 2020.195(3) doi: 10.1667/rade-20-00243.1

20. Fenech M, Morley AA. Measurement of micronuclei in lymphocytes. Mutation Research/Environmental Mutagenesis and Related Subjects. 1985.147(1-2):29-36. doi: 10.1016/0165-1161(85)90015-9

21. Varga D, Johannes T, Jainta S, Schuster S, Schwarz-Boeger U, Kiechle M, et al. An automated scoring procedure for the micronucleus test by image analysis. Mutagenesis. 2004.19(5):391-7. doi: 10.1093/mutage/geh047 [published Online First: 2004/09/25].

22. Decordier I, Papine A, Plas G, Roesems S, Vande Loock K, Moreno-Palomo J, et al. Automated image analysis of cytokinesis-blocked micronuclei: an adapted protocol and a validated scoring procedure for biomonitoring. Mutagenesis. 2009.24(1):85-93. doi: 10.1093/mutage/gen057 [published Online First: 2008/10/16].

23. Willems P, August L, Slabbert J, Romm H, Oestreicher U, Thierens H, et al. Automated micronucleus (MN) scoring for population triage in case of large scale radiation events. International Journal of Radiation Biology. 2010.86(1):2-11. doi: 10.3109/09553000903264481

24. Capaccio C, Perrier JR, Cunha L, Mahnke RC, Lörch T, Porter M, et al. CytoRADx: A High-Throughput, Standardized Biodosimetry Diagnostic System Based on the Cytokinesis-Block Micronucleus Assay. Radiation Research. 2021.196(5) doi: 10.1667/rade-20-00030.1

25. Rodrigues MA, Beaton-Green LA, Wilkins RC. Validation of the Cytokinesis-block Micronucleus Assay Using Imaging Flow Cytometry for High Throughput Radiation Biodosimetry. Health Physics. 2016.110(1):29-36. doi: 10.1097/hp.0000000000000371

26. Repin M, Pampou S, Garty G, Brenner DJ. RABiT-II: A Fully-Automated Micronucleus Assay System with Shortened Time to Result. Radiation research. 2019.191(3):232-36. doi: 10.1667/RR15215.1 [published Online First: 2019/01/18].

27. Romm H, Barnard S, Boulay-Greene H, De Amicis A, De Sanctis S, Franco M, et al. Laboratory Intercomparison of the Cytokinesis-Block Micronucleus Assay. Radiation Research. 2013.180(2):120-28. doi: 10.1667/rr3234.1

28. Thierens H, Vral A, Vandevoorde C, Vandersickel V, de Gelder V, Romm H, et al. Is a semi-automated approach indicated in the application of the automated micronucleus assay for triage purposes? Radiation Protection Dosimetry. 2014.159(1-4):87-94. doi: 10.1093/rpd/ncu130

29. Port M, Barquinero JF, Endesfelder D, Moquet J, Oestreicher U, Terzoudi G, et al. RENEB biological and physical dosimetry study. Laboratory intercomparison of eight dosimetry assays. Radiat Res. 2022.(submitted)

30. Lloyd DC, Edwards AA, Moquet JE, Guerrero-Carbajal YC. The role of cytogenetics in early triage of radiation casualties. Appl Radiat Isot. 2000.52(5):1107-12. doi: 10.1016/s0969-8043(00)00054-3

31. Hernández A, Barquinero JF, Endesfelder D, Puig P, Barrios L, Einbeck J, et al. Biodose Tools 2019 [Available from: https://aldomann.shinyapps.io/biodose-tools-beta/_w_7bf9e171/_w_b69c9547/.

32. Ainsbury EA, Lloyd DC. DOSE ESTIMATION SOFTWARE FOR RADIATION BIODOSIMETRY. Health Physics. 2010.98(2):290-95. doi: 10.1097/01.hp.0000346305.84577.b4

33. Deperas J, Szluinska M, Deperas-Kaminska M, Edwards A, Lloyd D, Lindholm C, et al. CABAS: a freely available PC program for fitting calibration curves in chromosome aberration dosimetry. Radiation Protection Dosimetry. 2007.124(2):115-23. doi: 10.1093/rpd/ncm137

34. Fleiss JL. Measuring nominal scale agreement among many raters. Psychological Bulletin. 1971.76(5):378-82. doi: 10.1037/h0031619

35. Wilkins RC, Beaton-Green LA, Lachapelle S, Kutzner BC, Ferrarotto C, Chauhan V, et al. Evaluation of the annual Canadian biodosimetry network intercomparisons. International Journal of Radiation Biology. 2015.91(5):443-51. doi: 10.3109/09553002.2015.1012305

36. BSI. ISO 17099:2014 Radiological protection – Performance criteria for laboratories using the cytokinesis block micronucleus (CBMN) assay in peripheral blood lymphocytes for biological dosimetry.: BSI:1--31.

37. Voisin P, Barquinero F, Blakely B, Lindholm C, Lloyd D, Luccioni C, et al. Towards a standardization of biological dosimetry by cytogenetics. Cell Mol Biol (Noisy-le-grand). 2002.48(5):501-4. [published Online First: 2002/07/31].

38. Voisin P. Standards in biological dosimetry: A requirement to perform an appropriate dose assessment. Mutation Research/Genetic Toxicology and Environmental Mutagenesis. 2015.793:115-22. doi: 10.1016/j.mrgentox.2015.06.012

39. ICRP. Relative Biological Effectiveness (RBE), Quality Factor (Q), and Radiation Weighting Factor (wR). ICRP Publication 92. 2003.Ann. ICRP 33 (4)

40. Endesfelder D, Oestreicher U, Bucher M, Beinke C, Siebenwirth C, Ainsbury E, et al. RENEB Inter-Laboratory Comparison 2021: The Dicentric Chromosome Assay. Radiation Research. 2022(submitted).

41. Cytogenetic Dosimetry: Applications in Preparedness for and Response to Radiation Emergencies. Vienna: INTERNATIONAL ATOMIC ENERGY AGENCY 2011.

Figure Legends

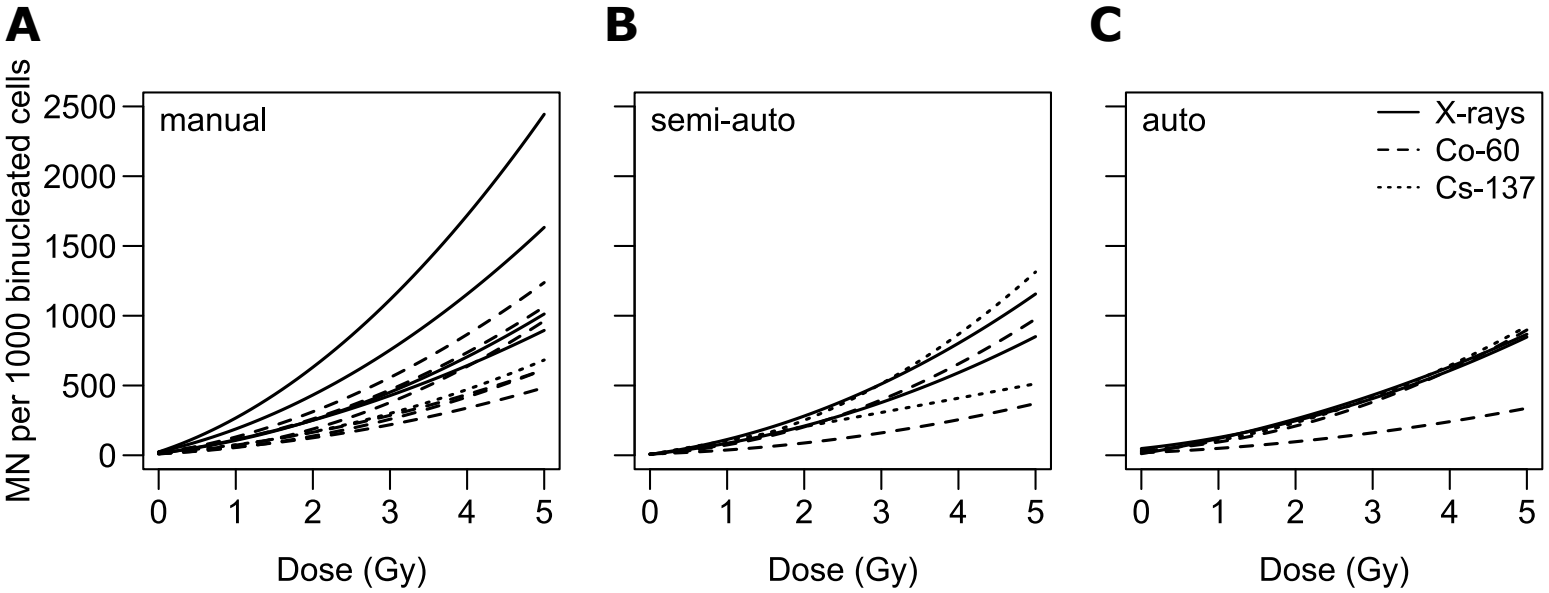
Figure 1: Calibration curves used by the 14 participating laboratories for manual (**A**), semi-automated (semi-auto) (**B**) and automated (auto) scoring (**C**) of micronuclei.

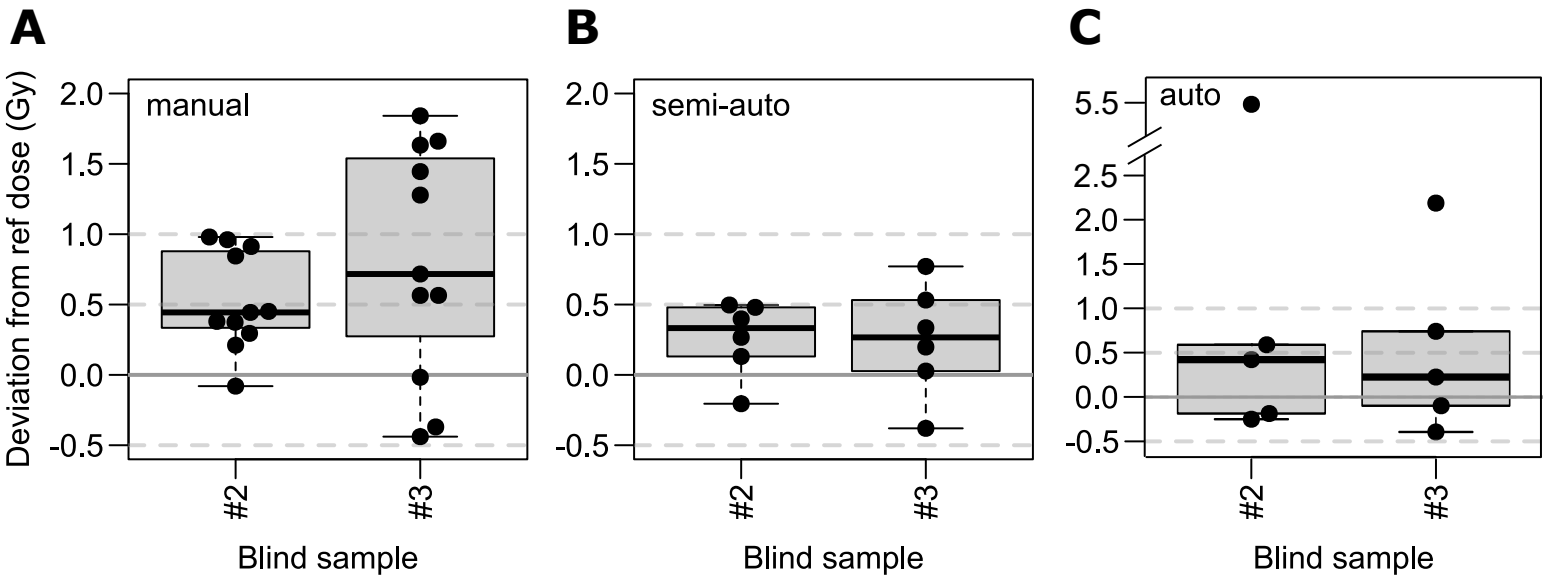
Figure 2: (**A-C**) Box and whisker plots showing the distributional characteristics of the median deviation (in Gy) of the estimated doses from the administered reference doses ('ref dose') for samples #2 (1.2 Gy) and #3 (3.5 Gy). Black dots show the deviation of the estimated doses from the reference doses reported by each lab. The solid black lines indicate median deviations. The solid gray line indicates a deviation of 0 Gy (no deviation), the gray dashed horizontal lines a deviation of ±0.5 Gy or ±1 Gy, respectively. All results shown in Figure 3 are based on the scoring of 500 (manual) or approximately 1000 (semi-automated & automated) binucleated cells.

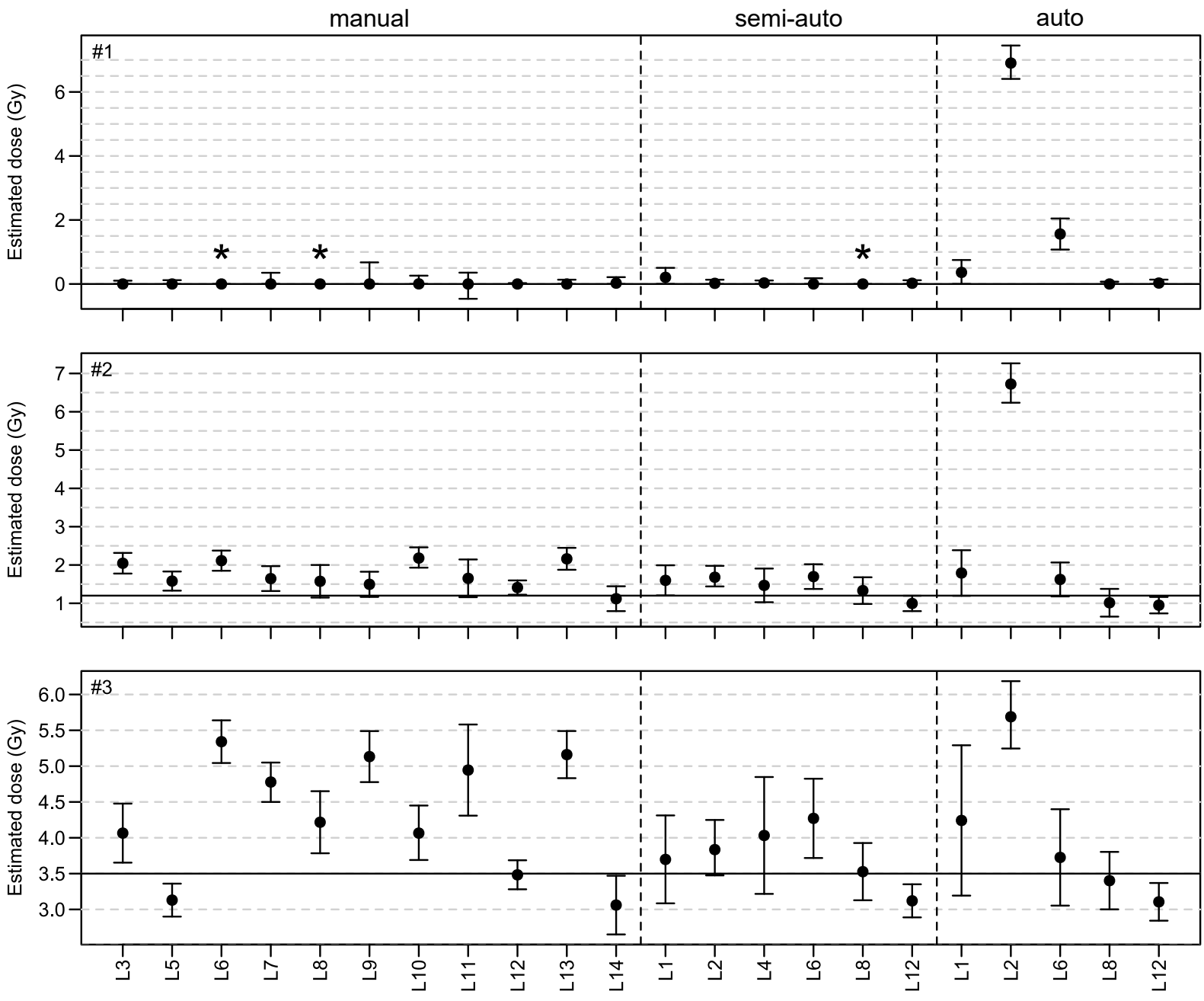
Figure 3: Dose estimates for the three test samples (#1: 0 Gy; #2: 1.2 Gy; #3: 3.5 Gy) based on manual scoring of 500 BN cells or on the semi-automated/automated scoring of ~1000 BN cells provided by each participating laboratory. The vertical bars show the 95% confidence intervals (CI) of the dose estimates provided by the participants. For some dose estimate values (indicated by an asterisk), accurate CIs could not be determined by the Biodose Tools software, due to the very low yield of MN scored. The horizontal black lines show the administered reference doses for each test sample. The results from manual, semi-automatic and automatic scoring are shown from left to right and separated by vertical dashed lines.

Figure 4: Comparison of dose estimates, and the corresponding 95% confidence

intervals (vertical bars), based on manual scoring of 200 (black) or 500 (gray) binucleated cells for each of the three test samples (#1: 0 Gy; #2: 1.2 Gy; #3: 3.5 Gy). One participant (L10) out of eleven did not provide results based on 200 binucleated cells. The horizontal black lines show the true administered doses for each of the test samples


A
B
C
MN per 1000 binucleated cells
2500
2000
1500
1000
500
0
manual
semi-auto
auto
X-rays
Co-60
Cs-137
0
1
2
3
4
5
Dose (Gy)


A
B
C
manual
semi-auto
auto
Deviation from ref dose (Gy)
2.0
1.5
1.0
0.5
0.0
-0.5
5.5
2.5
#2
#3
Blind sample


manual
semi-auto
auto
#1
#2
#3
Estimated dose (Gy)
L3
L5
L6
L7
L8
L9
L10
L11
L12
L13
L14
L1
L2
L4
L6
L8
L12
L1
L2
L6
L8
L12

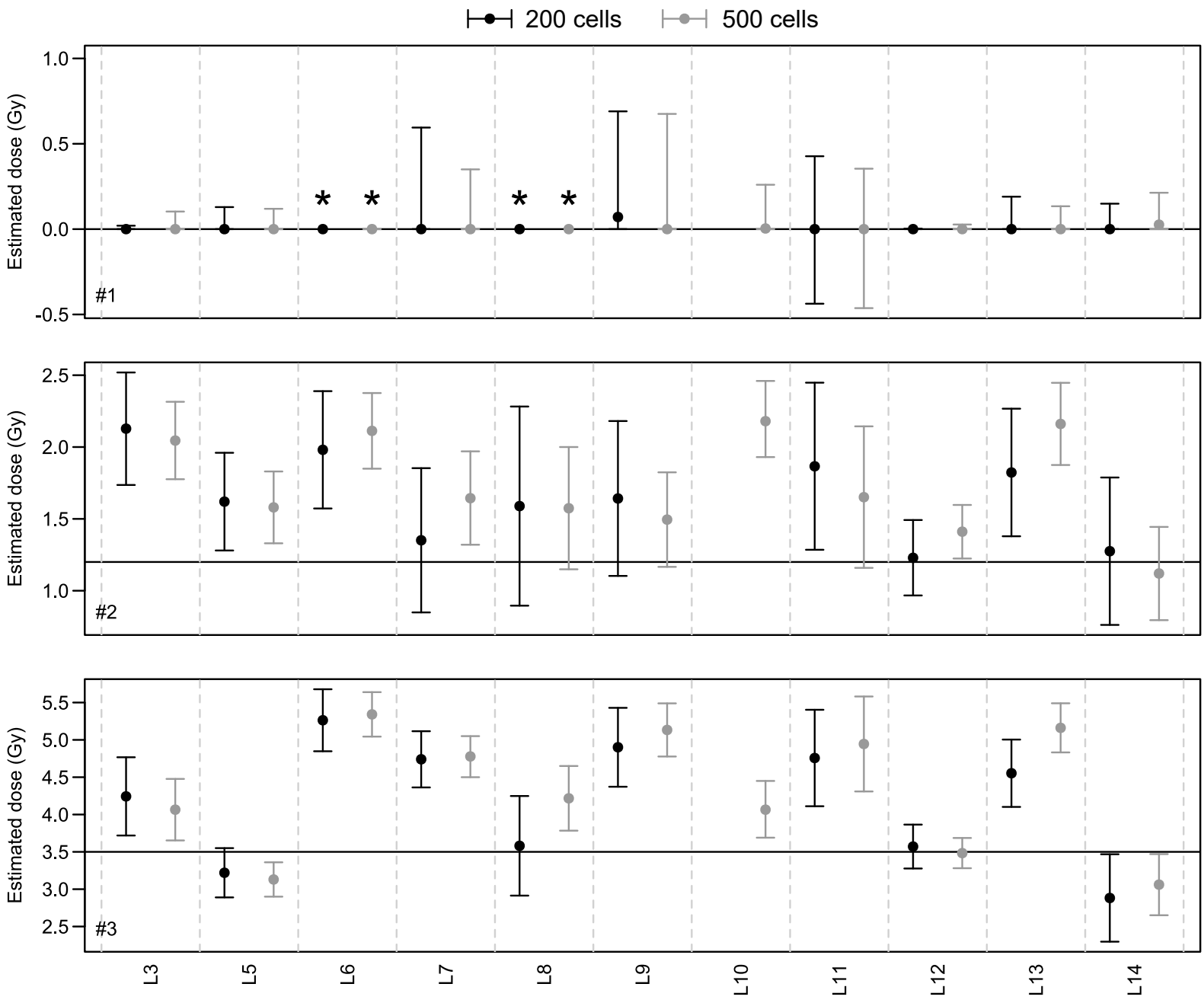

200 cells
500 cells
Estimated dose (Gy)
1.0
0.5
0.0
-0.5
#1
*
*
*
*
Estimated dose (Gy)
2.5
2.0
1.5
1.0
#2
Estimated dose (Gy)
5.5
5.0
4.5
4.0
3.5
3.0
2.5
#3
L3
L5
L6
L7
L8
L9
L10
L11
L12
L13
L14

Table 1. Information about the radiation qualities used and the coefficients of the MN calibration curves of the participating laboratories. Laboratories are ordered according to scoring method.

Lab Code	**Scoring**	**Source**	**C** x 10^{-2}	**α** x 10^{-2}	**β** x 10^{-2}	**SE(C)** x 10^{-2}	**SE(α)** x 10^{-2}	**SE(β)** x 10^{-2}
L3	manual	Co-60	1.29	8.61	3.18	0.48	1.77	0.67
L6	manual	Co-60	1.24	1.96	3.41	0.1	0.1	0.12
L7	manual	Co-60	1.22	2.41	1.93	0.1	0.23	0.07
L8	manual	Co-60	1.97	2.61	1.34	0.18	0.45	0.14
L13	manual	Co-60	2.41	5.85	2.99	0.53	1.14	0.29
L14	manual	Co-60	0.87	4.95	1.43	0.18	0.69	0.26
L9	manual	Cs-137	2.26	3.28	1.98	0.14	0.38	0.13
L11	manual	Orthovoltage	1.04	8.24	1.89	0.15	0.5	0.17
L10	manual	X-ray (200 kV)	1.55	6.44	2.7	0.25	0.99	0.37
L12	manual	X-ray (220 kV)	2.39	18.32	6.02	0.67	2.27	0.73
L5	manual	X-ray (250 kV)	2.32	12.6	3.92	0.75	2.15	0.54
L6	semi-auto	Co-60	0.92	3.3	3.21	0.28	0.88	0.38
L8	semi-auto	Co-60	0.96	1.71	1.1	0.12	0.32	0.1
L1	semi-auto	Cs-137	0.58	3	4.63	0.52	1.7	0.67
L4	semi-auto	Cs-137	0.61	9.85	0.06	0.24	1.77	0.41
L12	semi-auto	X-ray (220 kV)	0.64	7.57	3.08	0.25	1.21	0.41
L2	semi-auto	X-ray (240 kV)	0.79	5.58	2.25	0.46	1.52	0.17
L6	auto	Co-60	3.78	2.85	2.87	0.51	1.21	0.47
L8	auto	Co-60	1.79	2.35	0.81	0.16	0.39	0.12
L1	auto	Cs-137	2.91	5.13	2.56	0.85	1.93	0.63
L12	auto	X-ray (220 kV)	1.14	9.31	1.56	0.34	1.35	0.43
L2	auto	X-ray (240 kV)	4.81	5.9	2.02	1.2	1.4	0.16

Table 2. Characteristics of the calibration curves of the participating laboratories. Laboratories are ordered according to irradiation source.

Scoring method	Lab Code	Min (Gy)	Max (Gy)	Num-ber of doses	Source	Number of donors per dose point	Total number of BN cells scored per dose point	Irradiated in, Temperature (°C)	Dose rate (Gy/min)
manual	**L3**	0	4	9	Co-60	3	2009 - 5060	water, 37	0.26
manual auto/semi-auto	**L6**	0 0	4 3	19 9	Co-60 Co-60	7 1	1000-9784 1000	water, 37 water, 37	0.3 0.3
manual	**L7**	0	5	8	Co-60	16	12000 - 16000	air, 20	0.13-0.18
manual /auto / semi-auto	**L8**	0	4	9	Co-60	6	15863 - 18000	water, 37	0.5
manual	**L13**	0	5	7	Co-60	1	1000	air, 21	0.86
manual	**L14**	0	4	10	Co-60	3	1980-6106	air, 20	0.28
auto/semi-auto	**L1**	0	5	10	Cs-137	1	2408-12991	air, 37	0.49
semi-auto	**L4**	0	6	9	Cs-137	2	4555 - 21378	air, 20	0.6
manual	**L9**	0	4	11	Cs-137	7	12000-14000	air, 20	0.63
manual	**L11**	0	4	9	Orthovoltage	14	2800	air, 20	1
manual	**L10**	0	4	7	X-ray (200 kV)	1	546 - 2024	water, 37	0.49-0.99
manual auto/semi-auto	**L12**	0 0	4 4	6 8	X-ray (220 kV) X-ray (220 kV)	9 8	8578-9000 15385-16000	air, 20 air, 20	3 3
manual	**L5**	0	5	10	X-ray (250 kV)	2	1995 - 2034	air, 20	0.37
auto/semi-auto	**L2**	0	4	8	X-ray (240 kV)	10	29031- 73901	air, 20	1

Table 3. Differences in the standard protocols of the CBMN assay between the participating laboratories. Laboratories are ordered according to scoring method.

Lab Code	Scoring	Years of MN scoring experience*	Medium – FCS (%) - blood (%)	Culture time (h)	Time cyto B added (h)	KCl 0.075M + fixation methanol: acetic acid: ringer	Staining method	Notes
L3	manual	15, no	RPMI, 20%	72	24	KCl + 5:1	Giemsa	
L5	manual	10, yes	RPMI, 10%	72	44	KCl + 5:1	Acridine Orange	cells with > 3MN scored as 3 MN
L6	manual	>20, yes	RPMI, 20%	70	23	KCl + 4:1:5; 4:1	DAPI	
L7	manual	>30, yes	RPMI, 25%	72	44	*No KCl but distilled water:RPMI with 2%FCS (4:1)* + 3:1	Giemsa	cell smears were fixed
L8	manual	6, yes	RPMI, 20%	72	24	KCL + 10:1:11; 10:1	DAPI	
L9	manual	>10, yes	RPMI, 15%	72	44	KCl + 3:1	Giemsa	#2 scored by young scientist with 2 years experience
L10	manual	3, yes	RPMI, 10%	72	45	KCL + 3:1	Giemsa	
L11	manual	10-15, yes	PB-MAX karyotyping medium, no	72	44	KCL + 3:1 + 1% FA; 3:1	Giemsa	FA = formaldehyde
L12	manual	>20, no	RPMI, 10%	70	23	KCl + 4:1:5; 4:1	Acridine orange	
L13	manual	6, yes	RPMI, 20%	72	44	KCL + 3:1 + 1% FA; 3:1	Giemsa	
L14	manual	5, yes	RPMI, 15%	72	44	KCl + 3:1	Giemsa	manual scoring of pictures received from L8
L1	semi-auto	12, yes	RPMI, 10%	72	24	KCl + 4:1:5; 4:1	DAPI	
L2	semi-auto	>8, no	RPMI, 20%	70	23	KCl + 4:1:5	DAPI	
L6	semi-auto	>20, yes	RPMI, 20%	70	23	KCl + 4:1:5; 4:1	DAPI	
L8	semi-auto	6, yes	RPMI, 20%	72	24	KCL + 10:1:11; 10:1	DAPI	
L12	semi-auto	>30, no	RPMI, 10%	70	23	KCl + 4:1:5; 4:1	DAPI	
L4	semi-auto	>30, yes	RPMI, 20%	72	46	KCl + 3:1	DAPI + FISH (telomere-centromere)	fix: ethanol instead of methanol
L1	auto	-	RPMI, 10%	72	24	KCl + 4:1:5; 4:1	DAPI	no visual inspection
L2	auto	-	RPMI, 20%	70	23	KCl + 4:1:5	DAPI	no visual inspection
L6	auto	-	RPMI, 20%	70	23	KCl + 4:1:5; 4:1	DAPI	no visual inspection
L8	auto	-	RPMI, 20%	72	24	KCL + 10:1:11; 10:1	DAPI	no visual inspection
L12	auto	-	RPMI, 10%	70	23	KCl + 4:1:5; 4:1	DAPI	no visual inspection

*were the calibration curve samples and the exercise samples scored by the same researcher? Yes/no

Table 4. Dose estimate values reported by each laboratory for the 3 blind blood samples according to scoring method. Delivery times and reporting times are also presented.

Scoring method	**Lab Code**	**Delivery time (h)**	**Reporting time (h)**	**Estimated doses (Gy) with 95% CIs**		
Reference doses				**0 Gy**	**1.2 Gy**	**3.5 Gy**
Manual (n=11)	L3	45	174	0 (0-0.10)	2.05 (1.78-2.32)	4.07 (3.65-4.48)
	L5	53	143	0 (0-0.12)	1.58 (1.33-1.83)	3.13 (2.9-3.36)
	L6	21	238	0 (NA)*	2.11 (1.85-2.38)	5.34 (5.04-5.64)
	L7	24	191	0 (0-0.35)	1.64 (1.32-1.97)	4.78 (4.5-5.05)
	L8	41	289	0 (NA)*	1.57 (1.15-2.00)	4.22 (3.78-4.65)
	L9	45	147	0 (0-0.68)	1.50 (1.17-1.82)	5.13 (4.78-5.49)
	L10	21	356	0.00 (0-0.26)	2.18 (1.93-2.46)	4.07 (3.69-4.45)
	L11	69	178	0 (-0.46-0.35)	1.65 (1.16-2.14)	4.95 (4.31-5.58)
	L12	24	78	0 (0-0.27)	1.41 (1.22-1.60)	3.48 (3.28-3.69)
	L13	21	892	0 (0-013)	2.16 (1.88-2.45)	5.16 (4.83-5.49)
	L14		155**	0.03 (0-0.21)	1.12 (0.80-1.44)	3.06 (2.65-3.47)
No. (%) of estimates inside correct clinically relevant groups				*11 (100%)*	*7 (64%)*	*11 (100%)*
No. (%) of estimates inside ±0.5 Gy (#1,#2) or ±1 Gy (#3) intervals				*11 (100%)*	*7 (64%)*	*6 (55%)*
No. (%) of estimates for which the 95% CI includes the reference dose				*11 (100%)*	*4 (36%)*	*1 (9%)*
Semi-automated (n=6)	L1	3	76	0.21 (0-0.50)	1.60 (1.21-1.99)	3.70 (3.09-4.31)
	L2	3	93	0.02 (0-0.13)	1.68 (1.44-1.98)	3.84 (3.48-4.25)
	L4	22	314	0.03 (0-0.11)	1.47 (1.03-1.91)	4.03 (3.22-4.85)
	L6	21	238	0.00 (0.00-0.18)	1.70 (1.37-2.02)	4.27 (3.72-4.82)
	L8	41	289	0 (NA)*	1.33 (0.98-1.68)	3.53 (3.13-3.93)
	L12	24	78	0.02 (0-0.12)	1.00 (0.80-1.20)	3.12 (2.89-3.35)
No. (%) of estimates inside correct clinically relevant groups				*6 (100%)*	*6 (100%)*	*6 (100%)*
No. (%) of estimates inside ±0.5 Gy (#1,#2) or ±1 Gy (#3) intervals				*6 (100%)*	*6 (100%)*	*6 (100%)*
No. (%) of estimates for which the 95% CI includes the reference dose				*6 (100%)*	*3(50%)*	*4 (67%)*
Automated (n=5)	L1	3	76	0.36 (0-0.75)	1.79 (1.20-2.39)	4.24 (3.19-5.29)
	L2	3	93	6.90 (6.41-7.45)	6.72 (6.24-7.26)	5.69 (5.25-6.19)
	L6	21	238	1.56 (1.08-2.05)	1.62 (1.18-2.07)	3.73 (3.05-4.40)
	L8	41	289	0 (0-0.07)	1.01 (0.65-1.38)	3.40 (3.00-3.80)
	L12	24	78	0.03 (0-0.14)	0.95 (0.74-1.17)	3.11 (2.84-3.37)
No. (%) of estimates inside correct clinically relevant groups				*3 (60%)*	*3 (60%)*	*5 (100%)*
No. (%) of estimates inside ±0.5 Gy (#1 ,#2) or ±1 Gy (#3) intervals				*3 (60%)*	*3 (60%)*	*4 (80%)*
No. (%) of estimates for which the 95% CI includes the reference dose				*3 (60%)*	*3 (60%)*	*3 (60%)*

Grey shaded cells indicate dose estimates laying outside the clinically relevant group

* In case of very low MN yields, accurate CIs of the estimated dose could not be determined by the Biodose Tools software

** L14 received no blood samples but images from L8. Results were reported 155h after receiving the images

Table 5. Agreement between the radiation dose administered to blinded samples and the allocation to the three clinical relevance triage groups performed by each participating laboratory. Results achieved using single scoring techniques, or merged techniques, are shown.

Scoring method (labs involved)	Blind samples (dose administered)	Triage groups based on clinical relevance			Fleiss' fixed-marginal Kappa	95% confidence interval	Achieved agreement	Overall rating of achieved agreement
		Low dose exposure[a]	Medium dose exposure[b]	High dose exposure[c]				
Manual (n=11)	Sample 1 (0 Gy)	11 (100%)	0 (0%)	0 (0%)	0.73	0.58 to 0.88	83%	Good
	Sample 2 (1.2 Gy)	0 (0%)	7 (64%)	4 (36%)				
	Sample 3 (3.5 Gy)	0 (0%)	0 (0%)	11 (100%)				
Semi-automated (n=6)	Sample 1 (0 Gy)	(100%)	0 (0%)	0 (0%)	1.0	0.79 to 1.00	100%	Excellent
	Sample 2 (1.2 Gy)	0 (0%)	6 (100%)	0 (0%)				
	Sample 3 (3.5 Gy)	0 (0%)	0 (0%)	6 (100%)				
Automated (n=5)	Sample 1 (0 Gy)	3(60%)	1(20%)	1 (20%)	0.27	-0.03 to 0.57	53%	Poor
	Sample 2 (1.2 Gy)	1 (20%)	3 (60%)	1 (20%)				
	Sample 3 (3.5 Gy)	0 (0%)	0 (0%)	5 (100%)				
Manual and Semi-automated (n=14)	Sample 1 (0 Gy)	17 (100%)	0 (0%)	0 (0%)	0.81	0.72 to 0.89	87%	Excellent
	Sample 2 (1.2 Gy)	0 (0%)	13 (76%)	4 (24%)				
	Sample 3 (3.5 Gy)	0 (0%)	0 (0%)	17 (100%)				
All techniques (n=14)	Sample 1 (0 Gy)	20(91%)	1 (4,5%)	1 (4,5%)	0.69	0.61 to 0.76	80%	Good
	Sample 2 (1.2 Gy)	1 (4%)	16 (73%)	5 (23%)				
	Sample 3 (3.5 Gy)	0 (0%)	0 (0%)	22 (100%)				

(a), 0-1 Gy dose (green code)
(b), 1-2 Gy dose (yellow code)
(c), >2 Gy dose (red code)