

Variation in the measurement of DNA damage by comet assay measured by the ECVAG† inter-laboratory validation trial

Lykke Forchhammer, Clara Johansson¹, Steffen Loft, Lennart Möller¹, Roger W. L. Godschalk², Sabine A. S. Langie², George D. D. Jones³, Rachel W. L. Kwok³, Andrew R. Collins⁴, Amaya Azqueta⁴, David H. Phillips⁵, Osman Sozeri⁵, Maciej Stepiak⁶, Jadwiga Palus⁶, Ulla Vogel⁷, Håkan Wallin⁷, Michael N. Routledge⁸, Catherine Handforth⁸, Alessandra Allione⁹, Giuseppe Matullo⁹, João Paulo Teixeira¹⁰, Solange Costa¹⁰, Patrizia Riso¹¹, Marisa Porrini¹¹ and Peter Møller*

Department of Public Health, Section of Environmental Health, University of Copenhagen, Copenhagen, Denmark¹Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden,²Department of Health Risk Analysis and Toxicology, Maastricht University, Maastricht, The Netherlands,³Department of Cancer Studies and Molecular Medicine, University of Leicester, Leicester, UK,⁴Department of Nutrition, Faculty of Medicine, University of Oslo, Oslo, Norway,⁵Section of Molecular Carcinogenesis, Institute of Cancer Research, Sutton, UK,⁶Toxicology and Carcinogenesis Department, Nofer Institute of Occupational Medicine, Łódź, Poland,⁷National Research Centre for the Working Environment, Copenhagen, Denmark,⁸The Molecular Epidemiology Unit, Leeds Institute of Genetics, Health and Therapeutics, University of Leeds, Leeds, UK, ⁹Institute for Scientific Interchange Foundation, Turin, Italy, ¹⁰Environmental Health Department, National Institute of Health, Porto, Portugal and ¹¹Dipartimento di Scienze e Tecnologie Alimentari e Microbiologiche, Sezione Nutrizione Umana, Università degli Studi di Milano, Milano, Italy

The comet assay has become a popular method for the assessment of DNA damage in biomonitoring studies and genetic toxicology. However, few studies have addressed the issue of the noted inter-laboratory variability of DNA damage measured by the comet assay. In this study, 12 laboratories analysed the level of DNA damage in mono-cyte-derived THP-1 cells by either visual classification or computer-aided image analysis of pre-made slides, coded cryopreserved samples of cells and reference standard cells (calibration curve samples). The reference standard samples were irradiated with ionizing radiation (0–10 Gy) and used to construct a calibration curve to calculate the number of lesions per 10⁶ base pair. All laboratories detected dose–response relationships in the coded samples irradiated with ionizing radiation (1.5–7 Gy), but there were overt differences in the level of DNA damage reported by the different laboratories as evidenced by an inter-laboratory coefficient of variation (CV) of 47%. Adjustment of the primary comet assay end points by a calibration curve prepared in each laboratory reduced the CV to 28%, a statistically significant reduction ($P < 0.05$, Levene's test). A large fraction of the inter-laboratory variation originated from differences in image analysis, whereas the intra-laboratory variation was considerably smaller than the variation between laboratories. In summary, adjustment of primary comet assay results by

reference standards reduces inter-laboratory variation in the level of DNA damage measured by the alkaline version of the comet assay.

Introduction

The comet assay, also known as the single-cell gel electrophoresis assay, is a popular method in genetic toxicology and biomonitoring studies for assessing and measuring DNA damage at the level of individual cells. The damage detected is commonly referred to as strand breaks, although the alkaline version of the comet assay also detects alkali-labile sites and transient breaks created during DNA repair processes. It is probably the methodological simplicity of the comet assay, its low material requirement and the ability to focus on the level of damage in individual cells that are the main reasons for its widespread use.

Although the comet assay procedure is relatively simple, methodological differences exist and these have been discussed extensively elsewhere (1,2). The key methodological steps of the assay include (i) embedding single-cell suspensions in agarose, (ii) lysis that removes cellular and nuclear membranes and the majority of proteins, (iii) alkaline treatment, (iv) electrophoresis and (v) staining of the DNA comets formed. The extent of comet formation is visualized and assessed by fluorescence microscopy. Several international workgroups have made recommendations for the procedure (3–6). These will most likely develop into official guidelines issued, for instance, by the Organisation for Economic Co-operation and Development. However, laboratories will undoubtedly continue to use their own procedures and this will contribute to inter-laboratory variation in the measurement of DNA damage. Inter-laboratory differences can in principle be reduced by including reference standards in the analysis. However, these have not yet been developed for the comet assay, although many laboratories prepare their own reference standards as controls for intra-laboratory variation.

The expression of comet assay end points in different units has made it problematic even to begin to address the variation in DNA damage between laboratories. There are many ways of measuring the level of DNA damage with the comet assay, including continuous measurements (computerized scoring in different units such as percent DNA in the tail (%T), tail length and tail moment) and categorical measurements (visual scoring in arbitrary units), as well as various descriptions of the distribution of the images. DNA damage is often expressed as tail moment, which essentially is the product of the length and %T, although there are diverse formulae to calculate the

*To whom correspondence should be addressed. Øster Farimagsgade 5, Building 5, 2nd floor, PO Box 2099, DK-1014 Copenhagen K, Denmark. Tel: +45 3532 7654; Fax: +45 3525 7686; Email: p.moller@pubhealth.ku.dk

†ECVAG (the European Comet Assay Validation Group) was created within the Environmental Cancer Risk, Nutrition and Individual Susceptibility, a Network of Excellence, in order to validate the comet assay.

tail moment. At least in theory, it should be possible to make direct comparisons between laboratories using the same comet assay end point, but in practice this has been much more difficult than expected. It has been argued that the tail length is a poor end point because it is only linear with respect to dose within a narrow interval and the tail moment, although a valid measure, was judged to be problematical because it is difficult to visualize the appearance of the comets from tail moment values (7). Alternatively, one could argue that the most informative way of expressing comet assay results is as lesions per unaltered nucleotides or base pairs (bp). This can be done using a calibration curve determined by the dose–response relationship of ionizing radiation. Using this approach, it has been shown that intra-laboratory variation in DNA damage could be slightly reduced when different investigators from the same laboratory analysed the same samples (8,9). Previous inter-laboratory studies on the comet assay have documented a rather large variation (10–13). The European Standards Committee on Oxidative DNA Damage (ESCODD) attempted to decrease the variation using standardized methods and parallel analysis of reference samples, although that trial did not include laboratory-specific calibration curves. The aim of the present study was to investigate the inter-laboratory variation in DNA damage measured by the comet assay and to assess the possibility of reducing it by calibration with reference samples.

Materials and methods

Twelve laboratories participated in this inter-laboratory validation trial. All samples were prepared at the University of Copenhagen and distributed to the different laboratories as either pre-made slides or cryopreserved samples, the latter consisting of calibration curve or coded samples. Each laboratory also received a questionnaire to complete concerning its protocol for the comet assay (Table I).

Analysis of DNA damage in pre-made slides of gel-embedded cells

The purpose of this part of the study was to assess the variation in the image analysis of comet assay nucleoids. We used monocytic THP-1 cells from the American Type Culture Collection (Manassas, VA, USA). These cells were chosen because undamaged THP-1 cells have clear circular ‘comet’ images.

They were cultured in RPMI 1640 medium with 10% foetal bovine serum (Invitrogen A/S, Tåstrup, Denmark) and 1% penicillin–streptomycin solution (the stock solution from Invitrogen A/S contains 10 000 units/ml of penicillin G and 10 000 µg/ml of streptomycin in 0.85% saline). Cells were irradiated with 0, 2.5, 5 or 10 Gy of γ -rays from a Gamma Cell 2000 Cs¹³⁷ source (dose rate 3.77 Gy/min), suspended in phosphate-buffered saline (PBS). The cells were irradiated in ice-cold PBS to avoid generation of DNA damage during the subsequent processing of the cells.

The cells were processed by the alkaline comet assay in the same laboratory that cultured and irradiated the cells. Briefly, THP-1 cells were embedded in 0.75% low-melting-point agarose (Sigma-Aldrich A/S, Brøndby, Denmark) on GelBond® films (Lonza Copenhagen ApS, Vallenbæk Strand, Denmark) and lysed [1% Triton X-100, 2.5 M NaCl, 100 mM Na₂ ethylenediaminetetraacetic acid (EDTA), 10 mM Tris, pH 10] for a minimum of 1 h at 4°C. The GelBond films were then immersed in an alkaline solution (300 mM NaOH, 1 mM Na₂EDTA, pH 13) for 40 min and the duration of the subsequent electrophoresis was 20 min in the same solution at 1.30 v/cm (voltage across the platform) and 300 mA. After electrophoresis, the GelBond films were washed thrice for 5 min in Tris buffer (0.4 M Tris-HCl, pH 7.5), rinsed with milliQ® water, dried in 96% ethanol and distributed to the 12 participating laboratories by mail. The laboratories were instructed to stain the slides and analyse the gel-embedded comets. Each of the four coded slides consisted of nucleoids in two separate gels. We refer to these samples as pre-made slides in this article.

Unknown cryopreserved samples A–D

THP-1 cells suspended in ice-cold PBS were irradiated with 1.5, 2.5, 5 or 7 Gy of γ -rays from a Cs¹³⁷ source. The irradiated cells were then counted and aliquots frozen in a mixture containing 50% foetal bovine serum, 40% culture medium (RPMI 1640; Invitrogen) and 10% dimethylsulfoxide. Each participating laboratory received the cryopreserved coded samples to be analysed according to its own protocol.

Calibration curves

THP-1 cells suspended in ice-cold PBS were irradiated with 0, 2.5, 5 or 10 Gy of γ -rays from a Cs¹³⁷ source. The irradiated cells were then counted and aliquots frozen in a mixture containing 50% foetal bovine serum, 40% culture medium (RPMI 1640) and 10% dimethylsulfoxide. Each participating laboratory received the cryopreserved irradiated samples to be analysed according to the same protocol and in the same run as the coded samples. The laboratories obtained calibration curves by scoring slides of the 0-, 2.5-, 5- and 10-Gy irradiated THP-1 cells.

Reported end points

The laboratories reported the results in different primary comet assay end points, i.e. %T, tail moment, tail length and visual score. Two laboratories only reported data as %T because their software systems do not measure tail length

Table I. Assay conditions used by different laboratories^a

Laboratory	Agarose (%) ^b	Alkali (min)	Electrophoresis				Nuclei scored/gel (no. of gels/sample)	Stain ^d	Software
			Current (mA)	Voltage (V)	Voltage/cm (V/cm) ^c	Time (min)			
1	0.75	40	300	30	0.91	20	100 (4)	YOYO-1	Comet IV (Perceptive Instruments) ^e
2	0.75	40	300	24	1.53	30	50 (3)	EtBr	Komet 4.0 (Kinetic Imaging Ltd)
3	0.65	20	300	25	0.90	20	50 (2)	EtBr	Comet III (Perceptive Instruments) ^e
4	0.65	20	300	22	1.48	24	50 (1)	EtBr	Comet IV (Perceptive Instruments)
5	0.75	40	300	25	1.30	20	100 (2)	YOYO-1	Visual score
6	0.60	20	300	30	0.88	20	50 (4)	PI	Komet 5.5 (Kinetic Imaging Ltd)
7	1.00	40	260–300	25	1.60	30	30 (2)	DAPI ^f	Comet IV (Perceptive Instruments)
8	1.00	40	300	25	1.14	20	>50 (2)	EtBr	Cometa 1.5 (Immagini e Computer) ^e
9	1.00	20	300	25	0.76	30	50 (2)	DAPI	Lucia 4.61 (Preoptic C, Czech Republic)
10	0.80	40	300	23	1.24	20	50 (2)	EtBr	Komet 5.5 (Kinetic Imaging Ltd)
11	0.75	40	292–300	25	1.25	20	50 (2–3)	SYBR Gold	Komet 3.0 (Kinetic Imaging Ltd)
12	0.60	20	300	30	1.13	20	100 (1)	EtBr	Comet IV (Perceptive Instruments)

^aAll laboratories used the same solution in the alkaline unwinding and electrophoresis of the comet assay (0.3 M NaOH, 1 mM EDTA, pH > 13).

^bThe data represent the agarose density [% agarose in PBS or Tris base, acetic acid and EDTA (TAE)].

^cThe values represent the voltage across the platform of the electrophoresis tank.

^dThe dyes are 4',6-diamidino-2-phenylindole (DAPI), ethidium bromide (EtBr), propidium iodide (PI) and YOYO-1 iodide (YOYO-1).

^eThis laboratory also analysed the level of DNA damage in the pre-made slides by visual classification.

^fThis laboratory used DAPI for the cryopreserved samples. However, the staining generated a large background signal on the pre-made slides (GelBonds). This laboratory stained the pre-made slides with SYBR Gold.

or tail moment. The visual score was calculated by classification of images in five different categories (9). Usually, the visual score is reported in the range of 0–400 arbitrary units, but in this article, we convert the visual score to give a range of 0–100 arbitrary units so that it is comparable with %T. This makes it easier to compare the standard deviation (SD) between these end points. The primary comet assay end points were converted to lesions per 10^6 bp by means of the derived laboratory-specific calibration curves. We have used lesions per 10^6 bp as end point because the simple version of the comet assay, which is used in this study, detects a broad range of nucleotide lesions. Alternatively, the expression of the DNA damage as lesions per 10^9 dalton DNA is just as informative, although the nominator and denominator have different units. It is possible by calibration with ionizing radiation to transform primary comet assay end points to lesions per 10^9 bp because there is well-established relationship between the dose of ionizing radiation and yield of strand breaks in DNA. The yield of strand breaks per Gy has been found to be 0.27 and 0.31 breaks/ 10^9 dalton using alkaline sucrose sedimentation technique in two different studies, respectively (14,15); for our calibration, we have used the average of these values (0.29 lesions/ 10^9 dalton DNA). Furthermore, we assume that the average molecular weight of a DNA bp is 650 dalton. Based on dose equivalence in the alkaline sucrose sedimentation technique and comet assay, the slopes of the calibration curves (DNA damage/Gy) were used to calculate how many lesions per 10^6 bp a particular level of DNA damage corresponded to:

$$\frac{\text{DNA damage}_{\text{sample}}}{\alpha_{\text{calibration curve}}} \times \frac{0.29}{10^9} \times 650 \times 10^6 = \text{number of lesions}/10^6 \text{ bp}$$

The ‘DNA damage_{sample}’ is the primary comet assay end point, which can be the visual score, %T, tail length or tail moment; the ‘ $\alpha_{\text{calibration curve}}$ ’ is the specific calibration curve that corresponds to the primary comet assay end point.

The human genome contains 2.9×10^9 nucleotides (16), corresponding to approximately 6×10^9 bp (or approximately 4×10^{12} dalton DNA) per diploid cell in G₀ phase. The THP-1 cells in our calibration curve samples contained 6×10^9 dalton DNA (the DNA content was measured in THP-1 cells using an Eppendorf BioPhotometer). The DNA content of the THP-1 cells is somewhat higher than the content of human diploid cells because it is a tetraploid cell line (Japanese Collection of Research Bioresources; <http://cellbank.nibio.go.jp/celldata/jcrb0112.htm>, search August 19, 2009) and we used proliferating cell cultures.

Statistical analysis

All the statistical analyses were carried out by parametric tests using general linear model (GLM), repeated measures analysis of variance (ANOVA) and linear regression analysis (17). Differences in the distribution were assessed by Levene’s test. In all the tests, the level of significance was 5%. The analyses were carried out in Statistica version 5.5 for Windows (StatSoft, Inc., Tulsa, OK, USA). The inter-laboratory variation in DNA damage is mainly assessed as differences in SD of either the comet assay end points or the residuals (SD_{res}); this requires that the range in the data is the same. For this reason, we have calculated secondary comet assay end points using common calibration curves; these are generated as the mean of the laboratory-specific calibration curves. The data obtained from these calculations from common calibration curves are referred to as adjusted data.

Results

Variation in DNA damage measured by analysis of pre-made slides of gel-embedded comet nucleoids

The investigators in the different laboratories used different dyes to stain the slides (Table I). The dyes were propidium iodide, ethidium bromide, YOYO®-1 iodide, SYBR Gold and 4’,6-diamidino-2-phenylindole. In addition, the software systems differed and a few laboratories used visual scoring (Table I).

Figure 1 shows the results obtained from the analysis of the coded pre-made slides, which contained cells that had been exposed to 0, 2.5, 5 or 10 Gy (examples of images recorded from the staining of the slides in different laboratories are available as supplementary data in *Mutagenesis* online). From the results depicted in Figure 1, it can be seen that the tail length data are rather poor, there is a large degree of variation in the baseline levels of DNA damage (0 Gy) and the dose–response relationships are poor. The data expressed as tail

moment are much better than tail length in terms of low variation in the baseline level of DNA damage and dose–response relationships. In fact, linear dose–response relationships in terms of the tail moment ($R^2_{\text{mean}} = 0.94$, SD = 0.11) and %T ($R^2_{\text{mean}} = 0.96$, SD = 0.04) were not significantly different from each other ($P = 0.48$, paired *t* test of regression coefficients). The data expressed as visual score also increased dose dependently, although the slopes decrease at high doses as the assay reaches saturation.

Table II shows the mean level of DNA damage, SD and coefficient of variation (CV) calculated as SD/mean in the pre-made slides. Only the data expressed as arbitrary units (visual score) and %T are shown because it is possible to compare directly values of these end points. It can be seen that the CV decreases as the level of DNA damage increases when assessed as visual score. The same trend toward reduced CV is not as obvious in the slides scored by software systems (%T). For both the visual score and the %T data sets, the dose of ionizing radiation accounted for most of the overall variation (84.9% and 67.2%, respectively). The inter-laboratory variation was 17.0% for %T and 4.9% for visual scoring. However, it can also be seen in Table II that for the same level of DNA damage (i.e. 35.5 for visual scoring and 35.1 for %T), there were approximately the same SD values (12.0 and 12.6, respectively), indicating that the variation is the same for these end points at similar level of DNA damage. Overall, these data indicate that the variation in DNA damage measured by different laboratories is similar whether it is analysed as %T by computerized scoring or as arbitrary units by visual scoring. This means that neither of the end points appears to be superior, although they both appear to be better primary end points than the tail length.

Variation in DNA migration measured by analysis of cryopreserved samples

Figure 2 depicts the results from the calibration curve samples and the coded samples when the laboratories analysed cryopreserved samples by their own comet assay protocols. All 12 laboratories reported dose-dependent relationships between the dose of ionizing radiation and the level of DNA damage in both the cryopreserved calibration curve samples and the coded samples when measured as %T or visual score ($P < 0.05$, except in one laboratory where the linear relationship of calibration curve samples was of borderline statistical significance, $P = 0.054$, linear regression). As can be seen in Figure 2, there was a huge variation in the values of DNA damage obtained in different laboratories. This difference could be due to different assay procedures; however, even if this was the case, the relative level of DNA damage in the calibration curve and coded samples within each laboratory should be similar. The samples irradiated with 2.5 and 5 Gy were included as both calibration curve samples and coded samples. This means that differences between the samples irradiated with 2.5 Gy (and 5 Gy) can only be explained by variation between slides or bias in the analysis of uncoded samples. The latter issue is a controversial issue in comet assay analysis; it is hypothesized that knowledge about samples might cause investigators unintentionally to select or omit some images rather than to select them randomly. This bias is omitted by coding the samples prior to analysis. Figure 3A shows this relationship in the subset of the samples treated with 2.5 and 5 Gy. Samples irradiated with 5 Gy had—as expected—significantly higher levels of DNA damage than those irradiated with 2.5 Gy [$P < 0.05$, repeated measures

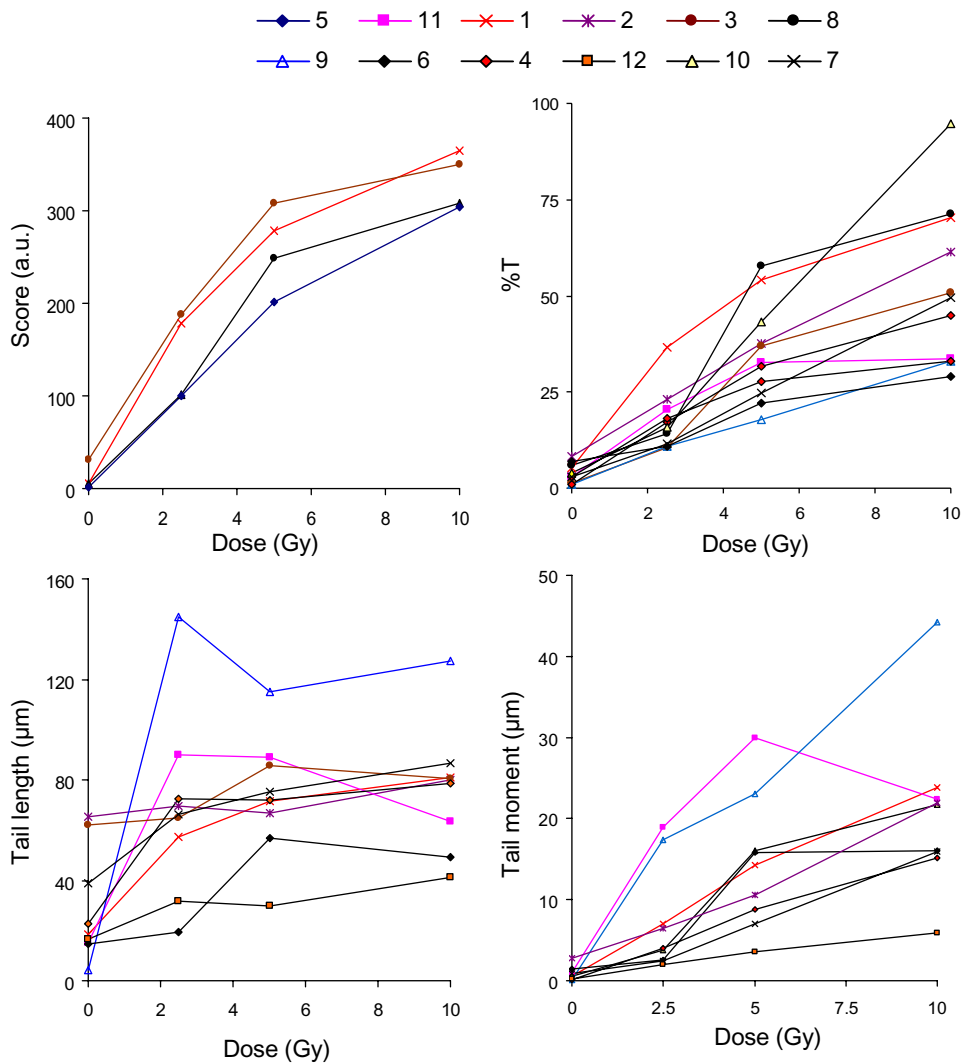


Fig. 1. Level of DNA damage in the coded pre-made slides expressed in the different primary end points: arbitrary units (a.u.; $n = 4$), %T ($n = 11$), tail length ($n = 9$) and tail moment ($n = 9$). Each line represents the result from one laboratory. The laboratories have the same number in Figures 1 and 2, and Tables I and II.

Table II. DNA damage measured in pre-made slides as visual score in arbitrary units (a.u.) or %T^a

Dose of ionizing radiation (Gy)	Visual score (a.u.), $n = 4$		%T, $n = 11$	
	Mean (SD)	CV (%)	Mean (SD)	CV (%)
0	2.7 (3.3)	136	3.9 (2.4)	61
2.5	35.5 (12.0)	34	17.2 (7.7)	45
5	64.8 (11.3)	17	35.1 (12.6)	36
10	82.9 (7.6)	9	51.9 (20.6)	40

^aThe contribution of the inter-laboratory variation and the dose of ionizing radiation to the overall variation was analysed by a GLM with the laboratory as random categorical variable and the dose of ionizing radiation as continuous variable. Statistical analysis of the visual score data set showed that the variation attributed to the dose of ionizing radiation and inter-laboratory variation explained 84.9% ($P < 0.001$) and 4.9% ($P = 0.21$), respectively, of the overall variation. The residual variation explained 10.2% of the overall variation ($SD_{res} = 10.4$). GLM analysis of the %T data set showed that the variation attributed to the dose of ionizing radiation and inter-laboratory variation explained 67.2% ($P < 0.001$) and 15.6% ($P < 0.01$) of the overall variation. The residual variation explained 17.0% of the overall variation ($SD_{res} = 9.1$).

ANOVA with the sample (2.5 or 5 Gy) as categorical variable]. There was no significant difference between calibration curve samples and the coded samples exposed to the same doses of radiation. This indicates that the investigators were not biased when scoring the calibration curve samples as compared with the coded samples. Figure 3B shows the relationship between the slopes obtained with the calibration curve samples and those obtained with the coded samples; the difference (calculated by subtracting the value of the slopes from each other) between the slopes of the calibration curve and coded samples from each laboratory did not depart significantly from zero [Δ value_{slope} = 0.04, 95% confidence interval (CI) -0.03 to 0.11]. This analysis indicates that participants did not treat the calibration curve and coded samples differently, although there clearly is variation between laboratories.

The total variation in the level of DNA damage in the cryopreserved samples can be explained by the following variables: the dose of ionizing radiation, different laboratory procedures, different software systems and unexplained variation (noise). This was assessed in a single-factor GLM analysis with the level of DNA damage in the coded samples as

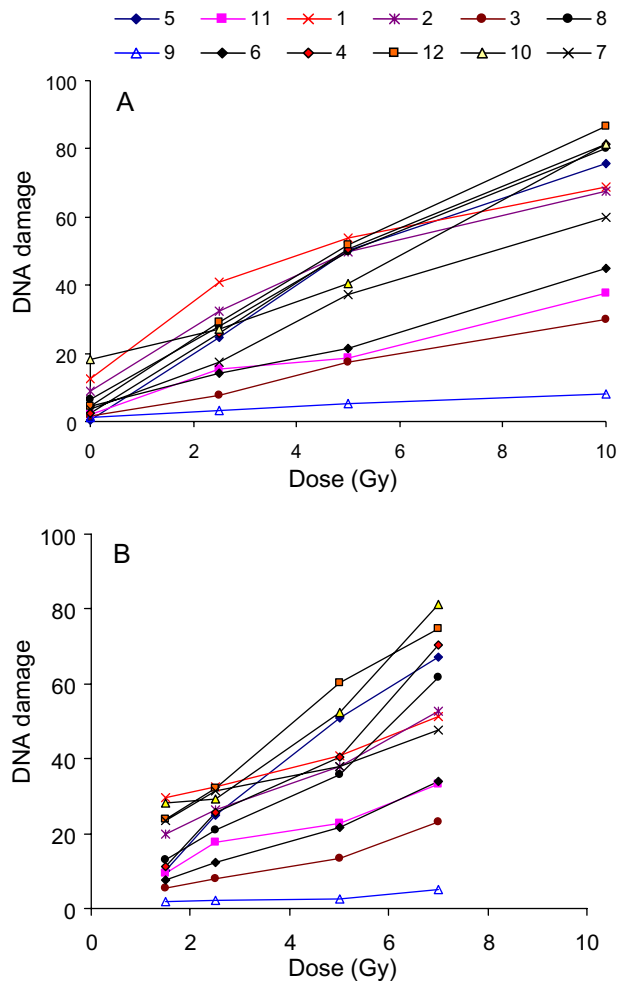


Fig. 2. Level of DNA damage in (A) calibration curve samples and (B) coded samples. The points and lines represent the data from different laboratories. The level of DNA damage is reported as visual score in arbitrary units ($n = 1$, range 0–100) and %T ($n = 11$). All laboratories reported dose-dependent relationships both for the calibration curve samples ($R^2_{\text{linear regression}}: 0.95\text{--}0.99$, $P < 0.05$; except data from one laboratory that was not statistically significant, $P = 0.054$) and for the coded samples ($R^2_{\text{linear regression}}: 0.96\text{--}0.99$, $P < 0.05$).

dependent variable and the overall laboratory procedures (categorical variable, $n = 12$), dose of radiation (continuous variable, $n = 4$) and level of DNA damage in the pre-made slides (continuous variable, $n = 4$) as independent variables. The contribution of the software systems to the overall variation in DNA damage reported by different laboratories can be estimated because the same samples of cells were used for the coded cryopreserved samples and the pre-made slides. This analysis showed that the laboratory procedures ($F = 5.2$, $P < 0.001$), dose of ionizing radiation ($F = 14.6$, $P < 0.001$) and the level of DNA damage in the pre-made slides ($F = 9.1$, $P < 0.01$) had statistically significant single-factor effects and these variables explained the majority of the variation ($R^2_{\text{model}} = 0.90$, GLM analysis). Not surprisingly, the dose of ionizing radiation was a strong predictor of the level of DNA damage in the coded samples. However, differences in both laboratory procedures and image analysis systems also contributed significantly to the variation in the coded cryopreserved samples. Figure 4 shows the relationship between the level of DNA damage in the pre-made slides

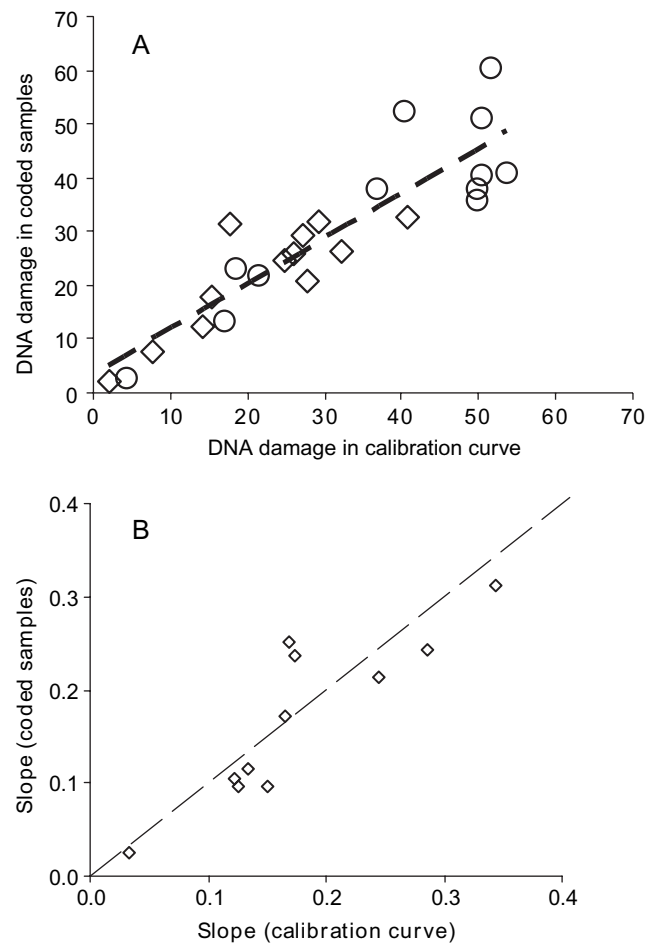


Fig. 3. Relationship between the levels of DNA damage in calibration curve samples and coded samples. (A) Levels of DNA damage in calibration curve samples versus coded samples treated with either 2.5 Gy (diamonds) or 5 Gy (circles) of ionizing radiation. Each diamond and circle represents the values reported from one laboratory. The level of DNA damage is reported as %T ($n = 11$) or as arbitrary units ($n = 1$, range 0–100). The correlation coefficient is $R^2 = 0.80$. The levels of DNA damage caused by the two treatments are significantly different ($P < 0.05$, repeated measures ANOVA with treatment as categorical variable). (B) Relationship between slopes of the calibration curve samples and coded samples. The slopes correspond to Gy equivalents originating from linear regression analyses between the level of DNA damage (dependent variable) and dose of ionizing radiation (independent variable). Each point represents the slope obtained by one laboratory ($n = 11$). One data point (0.91, 1.47) has been omitted from the graph because it was much higher than the other data points.

and the cryopreserved samples derived from the exact same exposure to ionizing radiation. The agarose density, duration of the alkaline treatment and electrophoresis, and strength of the electrophoretic field (V/cm) are considered to be the critical parameters that will affect the migration of DNA in the comet assay (outlined in Table 1). However, we did not find clear relationships between the agarose density, duration of the alkaline treatment and electrophoresis, and strength of the electrophoretic field (V/cm) in different laboratories and the level of DNA damage [linear regression analysis with the baseline (0 Gy) samples or the slope of the dose–response curve as dependent variable]. Also, different combinations of the four variables did not reveal any clear associations between the assay conditions and level of DNA damage. It should be emphasized that this analysis only implies that differences between assay conditions were not strong enough to explain

the differences of the level of DNA damage measured in different laboratories. In addition, the statistical power of this analysis was small considering that there were four variables (assay conditions) and only 12 laboratories.

The variation attributed to different laboratory procedures (inter-laboratory variation) and intra-laboratory variation can be assessed by an analysis of the calibration curve samples that were analysed in one to three separate experiments in the different laboratories. The period between the first and last experiment was about 6–10 months in each laboratory, indicating that it should be possible to detect if there was day-to-day variation. We did not attempt to standardize the assay conditions in the laboratories during the period of the

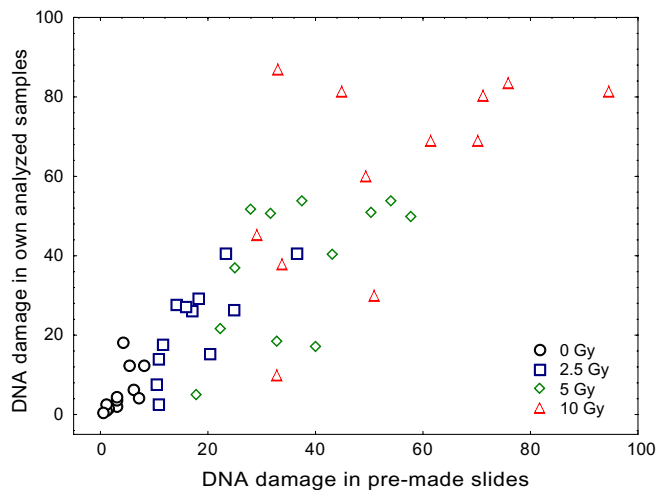


Fig. 4. Relationship between the level of DNA damage in coded cryopreserved samples (analysed in each laboratory by its own comet assay protocol) and pre-made slides (scored in each laboratory). The results represent laboratories that analysed the level of DNA damage as %T ($n = 11$) and arbitrary score ($n = 1$, range 0–100).

trial; this probably means that a number of assay conditions could have changed because of the new batches of chemicals and because of possible changes in the intensity of the fluorescent lamp in the microscopes. Ten laboratories analysed calibration curve samples on two or three different occasions, whereas two laboratories only analysed the calibration curve samples in one experiment. Table III outlines the statistical analysis of the data; in this analysis, the intra-laboratory variation is the same as the day-to-day variation and the residual variation is the unexplained variation. It is obvious that the dose of ionizing radiation explains most of the variation, whereas the magnitude of the residual variation is larger than the day-to-day variation.

Level of DNA damage in cryopreserved samples expressed as lesions per 10^6 bp

Figure 5 shows the number of lesions per 10^6 bp in the coded samples that represented cells that had been irradiated with 1.5, 2.5, 5 or 7 Gy of ionizing radiation. All laboratories reported the same ranking of the four coded samples (results not shown); this ranking was in accordance with the dose of ionizing radiation. The mean linear regression coefficient (R^2) for all laboratories was 0.97 (range 0.89–0.99). It can be seen in Figure 5 that the CV is reduced as the level of DNA damage increases.

The calibration curve generated by ESCODD has been used by some investigators to calculate the number of DNA breaks in samples (the actual ESCODD calibration curve was documented in the final ESCODD report to the European Union, but it was actually never reported in the published papers from ESCODD). In the present study, a ‘common calibration curve’ was obtained as the mean from all laboratory-specific calibration curves. Figure 6 shows a comparison between the ESCODD curve and the common calibration curve in this experiment (the mean level of DNA damage in the coded samples is also depicted in the figure, although they were not used in the common calibration curve).

Table III. Assessment of the contribution of intra-laboratory (day-to-day variation) variation to the overall variation in the 12 participating laboratories

Laboratory	Mean (SD) DNA damage in cryopreserved samples ^a				% of total variation ^b			SD _{res} ^c
	0 Gy	2.5 Gy	5 Gy	10 Gy	Dose of ionizing radiation	Intra-laboratory variation	Residual variation	
1	11.2 (1.8)	34.7 (8.7)	55.7 (2.8)	73.7 (6.8)	91.5***	0.1	8.4	7.4
2	9.7 (4.1)	30.5 (6.6)	51.0 (4.9)	72.4 (5.0)	93.7***	0.1	6.2	6.2
3	0.9 (0.7)	8.7 (1.2)	15.7 (5.1)	34.5 (10.9)	86.1***	5.7	8.2	4.0
4	0.5 (0.3)	26.3 (1.3)	51.0 (0.8)	83.3 (6.4)	97.7***	0.3	2.0	4.6
5	6.1 (4.0)	14.0 (4.1)	20.1 (1.5)	41.8 (5.3)	93.1***	0.5	6.4	3.6
6	2.0 (1.2)	23.8 (2.5)	44.4 (6.6)	80.4 (4.5)	98.3***	0.4	1.3	3.4
7	3.4 (0.6)	23.5 (5.9)	41.0 (4.3)	65.9 (12.0)	92.8***	3.2	4.0	4.9
8	6.5 (0.4)	27.5 (3.3)	50.3 (2.2)	81.7 (1.3)	98.8***	0.2	1.0	5.2
9	1.0 (0.1)	2.6 (0.8)	4.9 (0.6)	9.7 (2.1)	93.7***	0.1	6.2	0.9
10	16.2 (2.8)	30.7 (4.9)	43.8 (4.6)	73.0 (11.6)	94.9***	0.2	4.9	5.1
11	2.2	15.3	18.6	37.8	97.4***	NC	2.6	2.3
12	4.5	29.3	51.6	82.8	99.3***	NC	0.7	3.0

^aThe data are based on %T except for Laboratory 5, which measured DNA damage as visual score in arbitrary units (the data are the visual score in range of 0–100). The calculations are based on data from the calibration curve samples that had been analysed in one (Laboratories 11 and 12), two (Laboratories 1, 9 and 10) and three (Laboratories 2–8) separate experiments.

^bAnalysed by GLM with the experiment and dose of ionizing radiation as categorical (random factor) and continuous variables, respectively. The data are reported as the percentage of the total sum of squares. The statistical significance is as follows: *** $P < 0.001$. The GLM analysis was only used for data from laboratories that analysed the calibration curve samples in two or three independent experiments. Data from laboratories that only reported results from one experiment were analysed by linear regression and the percentage of variation corresponds to the regression coefficient (R^2) of the regression analysis. For these data, the contribution of the experiment was not calculated (NC). The term ‘intra-laboratory variation’ is the same as the day-to-day variation (variation from one experiment to another), whereas the residual variation is the unexplained fraction of the total variation in the study.

^cSD_{res} is the SD of the residuals.

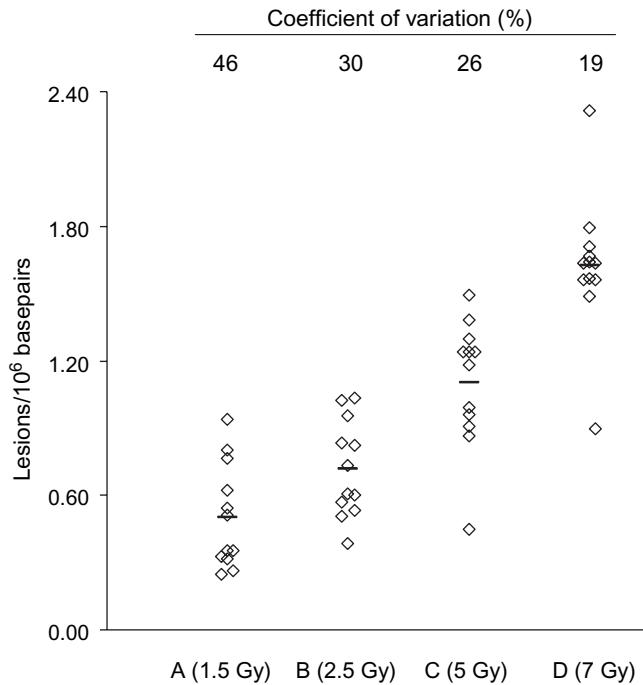


Fig. 5. Lesions per 10^6 bp in the coded THP-1 samples. The samples labelled A–D had been exposed to 1.5, 2.5, 5 and 7 Gy, respectively. The data are calculated from %T ($n = 11$) and visual score in arbitrary units ($n = 1$, range 0–100) by use of the laboratory-specific calibration curves. Each point represents the value from one laboratory. The CV represents the inter-laboratory variation in DNA damage reported by the different laboratories.

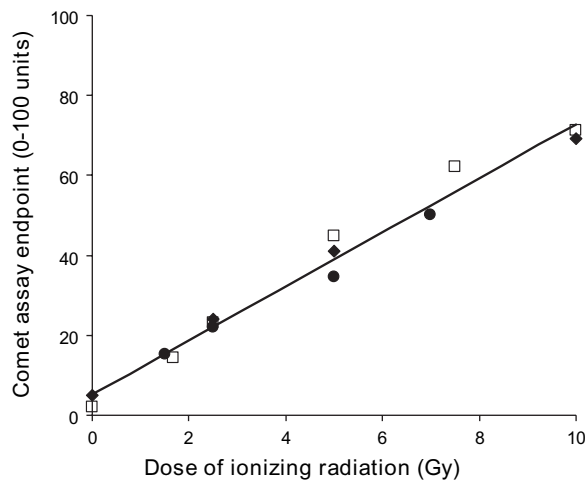


Fig. 6. Calibration curve outline. Based on data from the coded samples (solid circles), calibration curve samples (solid diamonds) and the ESCODD calibration curve (open squares). The line represents the regression line ($R^2 = 0.98$). Based on the linear regression analysis, 1 unit (%T or arbitrary score) corresponds to 0.1447 Gy equivalents. This corresponds to 0.042 lesions/ 10^9 dalton DNA or 0.0273 lesions/ 10^6 bp.

The ESCODD calibration curve and the common calibration curve are virtually the same.

Figure 7 depicts differences between results calculated by the common calibration curve and results from the laboratory-specific calibration curves (the data in figure 7 are the number of lesions per 10^6 bp per Gy and the symbols indicate the dose of ionizing radiation). It can be seen that there were larger differences in the estimated level of lesions per 10^6 bp per Gy

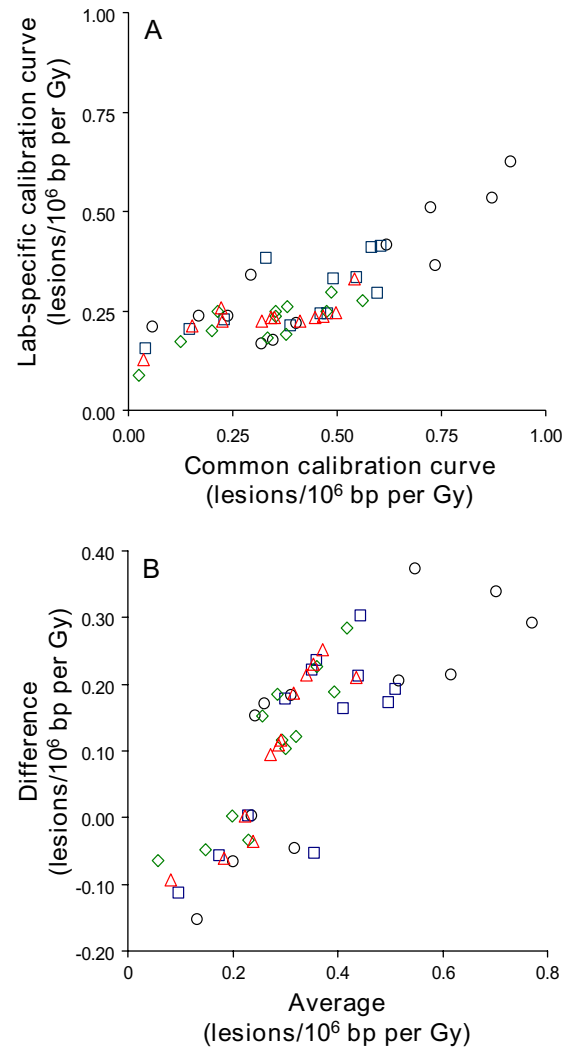


Fig. 7. (A) Comparison of DNA damage in terms of lesions per 10^6 bp per Gy calculated using a common calibration curve or laboratory-specific calibration curve. The common calibration curve was generated by first generating mean calibration curves from each laboratory (the laboratories had analysed the calibration samples on one to three different occasions) and then generating the common calibration curve as the mean of different laboratories. This means that the calibration curve from each laboratory contributed equally to the common calibration curve. The variation in DNA damage calculated by the common calibration curve is larger ($SD = 0.18$) than the variation obtained when using the laboratory-specific calibration curve ($SD = 0.08$, $P < 0.05$, Levene's test). (B) Bland–Altman plot of the difference in lesions per 10^6 bp per Gy calculated by the common calibration curve and laboratory-specific calibration curves. It can be seen that there is a systematic trend toward larger difference between the calibration curves as the level of DNA damage increases, indicating that the two ways of calibration generate different results that depend on the level of DNA damage in the sample. The symbols indicate samples irradiated with 1.5 Gy (circles), 2.5 Gy (squares), 5 Gy (diamonds) and 7 Gy (triangles).

when the data were calculated by the common calibration curve than by the laboratory-specific calibration curves (Figure 7A). Figure 7B outlines a Bland–Altman plot of the same data, which is commonly used to assess whether the results from different assays differ. The Bland–Altman plot (Figure 7B) shows a systematic trend in the data with the difference between the two ways of calibration becoming larger when the value (lesions per 10^6 bp per Gy) is large.

Figure 8 shows inter-laboratory variation in the data calculated as lesions per 10^6 bp per Gy. Most importantly, it

can be seen that using the laboratory-specific calibration curves reduced the variation between the laboratories (CV = 28%) as compared with the common calibration curve (CV = 47%, $P < 0.05$, Levene's test). Depicted in Figure 8 is an analysis where we have calculated the level of DNA damage based on tail moment as primary end point and laboratory-specific calibration curves on tail moment. This analysis provide the approximately same variation between laboratories (CV = 34%) as the data based on %T, which means that the variation between laboratories did not get better (or worse) using the tail moment as primary end point.

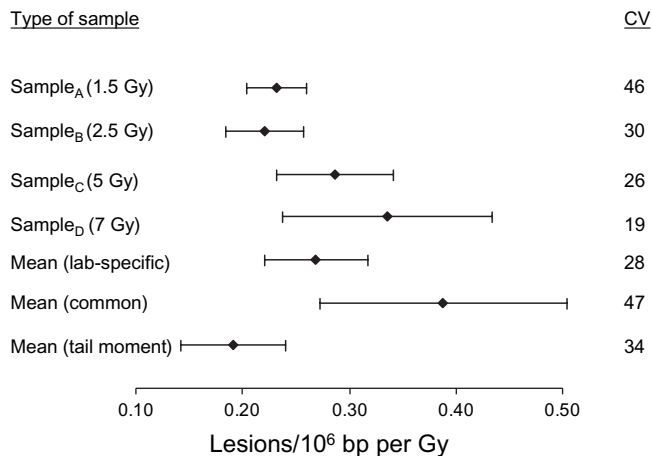


Fig. 8. Variation in the level of DNA damage per 10^6 bp per Gy in monocytic THP-1 cells. The data in samples A–D represent the level of lesions per 10^6 bp per Gy after adjustment of the primary comet assay end point (%T or visual score) with the laboratory-specific calibration curves. For each laboratory, we have calculated the mean level of lesions per 10^6 bp per Gy from the four samples A–D, using the laboratory-specific calibration curve: the ‘mean (lab-specific)’ row depicts this analysis ($n = 12$). The same adjustment and analysis was carried out with the common calibration curve that was mean of all the laboratory-specific calibration curves ($n = 12$). The ‘mean (tail moment)’ shows the results of adjustment of the primary comet assay end point in terms of the tail moment ($n = 9$). It can be seen that the inter-laboratory variation is lower when the data have been adjusted by the laboratory-specific calibration curve as compared with the common calibration curve. However, adjustment of primary comet assay end points in terms of either %T/visual score or tail moment with the corresponding calibration curves yield low inter-laboratory variation.

Assessment of components influencing the variation in DNA damage

Based on the data reported in previous tables and figures, it is possible to calculate the magnitude of variation between different experiments in each laboratory (intra-laboratory variation) and between laboratories (inter-laboratory variation) and the magnitude of residual variation. Table III outlines the calculation of the magnitude of the intra-laboratory variation, which is identical to the inter-experimental variation. These calculations were based on the data from the calibration curve samples that were analysed on 1–3 separate days in the laboratories. For each laboratory, we have calculated the contribution of the dose of ionizing radiation, intra-laboratory variation and residual variation (Table III). As can be seen, both the intra-laboratory (inter-experiment) variation and the residual variation are small compared with the variation related to the dose of ionizing radiation. The residual variation is, on average, higher than the intra-laboratory (inter-experiment) variation, indicating that there is a large part of the variation that cannot be explained by day-to-day variation.

Table IV outlines the results of statistical tests where the dose of ionizing radiation, intra-laboratory variation, inter-laboratory variation, primary comet assay end points (computerized image analysis versus visual score) and residual variation have been investigated. In all the tests, the dose of ionizing radiation was the variable that explained the largest part of the variation. The first row of Table IV shows the analysis of the cryopreserved calibration curve samples. It can be seen that intra-laboratory variation contributes very little to the overall variation, which is in accordance with the results obtained in Table III. In this analysis, 21.6% of the variation was explained by differences between laboratories. The second row of Table IV shows the results of the analysis of the pre-made slides. Here, it is possible to compare the variation of the laboratory procedures (staining and analysis) with the variation related to the use of different end points. As can be seen, the residual variation is almost the same in the first and second rows in Table IV (SD_{res} : 9.2 versus 11.1) and the contribution of the dose of ionizing radiation was also very similar. The sum of the inter-laboratory variation and residual variation in the cryopreserved calibration curve samples was 33.9%; this is only slightly higher than the same type of variation in the pre-made slides (27.9%). The variation in the pre-made slides can

Table IV. Assessment of variables contributing to the overall variation in DNA damage in monocytic THP-1 cells exposed to ionizing radiation

Sample	% of total variation ^a					SD_{res} ^b
	Dose of ionizing radiation	Intra-laboratory variation	Inter-laboratory variation	Primary comet assay end point	Residual variation	
Cryopreserved calibration curve samples ^c	66.0***	0.1	21.6***	NC	12.3	9.2
Pre-made slides	68.7***	NC	8.9	3.4**	19.0	11.1
Coded samples ^d	43.6***	NC	43.5***	NC	12.8	7.3
Coded samples (adjusted) ^e	77.1***	NC	16.3***	NC	6.6	3.8

^aThe percentage of total variation has been calculated from the sum of squares in GLM analysis with the level of DNA damage as dependent variable; the dose of ionizing radiation (continuous variable), laboratory (categorical random variable) and experiment (categorical random variable) were independent variables in the GLM analysis. The contribution of the intra-laboratory variation (day-to-day variation) was assessed by analysis of the calibration curve samples that were analysed on 1–3 different days of the analysis. Statistically significant effects are indicated as follows: ** $P < 0.01$; *** $P < 0.001$; NC, not calculated.

^bThe SD of the residuals (SD_{res}) did not deviate from the normal distribution in any of the statistical tests ($P > 0.05$, Shapiro–Wilks test).

^cThese are the calibration curve samples that have been analysed on one to three occasions by the 12 laboratories.

^dThe cryopreserved coded samples had received 1.5, 2.5, 5 or 7 Gy of ionizing radiation and have been analysed on one occasion in each laboratory.

^eThe values of DNA damage in the coded samples were adjusted according to the following equation: adjusted value = (value/slope_{laboratory-specific calibration curve}) × Slope_{mean of all calibration curves}.

only originate from differences in the staining of slides and analysis of the comet images. This implies that most of the inter-laboratory variation originates from different procedures in the staining and analysis of images. The contribution of the dose of ionizing radiation and laboratory procedures to the overall variation is shown in the third row of Table IV. These samples were coded, in contrast to the calibration curve samples (first row in Table IV), and there was a restricted dose range of ionizing radiation. The coding probably explains why the variation explained by the dose of ionizing radiation was lower in this analysis. However, it should also be emphasized that the residual variation was lower ($SD_{res} = 7.3$), which means that there was less random variation in the analysis of the coded samples compared with the uncoded cryopreserved calibration curve samples. The fourth row of Table IV shows the results of an analysis where the data analysed in the third row were adjusted with the laboratory-specific calibration curve; the residual variation (SD_{res}) was reduced and the variation explained by the dose of ionizing radiation increased. This means that some of the random variation, embedded as variation in the laboratory procedures in the analysis of the third row in Table IV can be eliminated by correcting for differences in laboratory procedures. Thus, a part of the random variation can be eliminated by adjusting data using the calibration curves (and thereby correcting for differences in laboratory procedures). This finding is in accordance with the assessment in Figure 8, which shows that the variation between laboratories was reduced using laboratory-specific calibration curves rather than a common calibration curve.

Discussion

This study shows that the comet assay is a reliable technique for detecting DNA damage in biological samples in a dose-dependent way. It is reassuring that all participating laboratories were able to detect dose-dependent relationships in coded samples. However, it is also clear that there is a large variation in the values reported by different laboratories. This variation arises from differences in comet assay protocols and systems of analysis.

The results clearly document that the tail length as a primary comet assay end point is rather poor because (i) there is a large variation in non-irradiated samples, (ii) the dose-response relationship saturates at low doses and (iii) the tail length does not even show a dose-dependent relationship with ionizing radiation. The differences in the tail length are probably the results of different calibration of the software systems and because some dyes may stain the comet tail better than others. However, tail moment, %T and visual score displayed better dose-response relationships with ionizing radiation, and calibration using tail moment yielded the same inter-laboratory variation as calibration using %T and visual score as primary comet assay end points. It is not surprising that the tail moment yields reliable results because several laboratories have previously reported linear relationships between ionizing radiation dose and comet tail moment (18–20). However, it should still be kept in mind that the main criticism of the tail moment as primary end point has been that it is difficult to compare the values between laboratories, rather than its reliability (7).

The values of DNA damage reported by different laboratories had a large variation. In the pre-made slides, there was a low inter-laboratory variation (determined as SDs) in the

unexposed samples but a large inter-laboratory variation with irradiated cells (Table II). This pattern is expected when analysis is by visual scoring because it is easier to discriminate circular (Class 0) and highly damaged (Class 4) images than those in the middle of the range. The same pattern was observed for the pre-made slides analysed by %T, suggesting that with image analysis, too, it is easier to measure nuclei without migration than images having migration. Most remarkable was the huge variation in %T in the pre-made slides. The SDs in the pre-made slides analysed as %T increased with increasing dose of ionizing radiation. The inter-laboratory variation in %T originates most likely from different settings of the image analysis systems or staining. However, it should also be highlighted that the analysis of the coded cryopreserved samples showed that the CV decreased with the frequency of lesions per 10^6 bp (Figure 5). This is most easily explained by the fact that the SDs remained relatively constant in the different samples of irradiated cells after calibration of the data, whereas the level of DNA damage increased with increasing dose of ionizing radiation.

From the assessment of the variation (Table IV) it can be inferred that the inter-laboratory variation is larger than the residual variation and that there is very little intra-laboratory variation. The study design used cannot discriminate between the intra-laboratory variation and day-to-day variation (this means that the variation attributed to the intra-laboratory variation also includes a contribution from the day-to-day variation and vice versa). The residual variation in normal studies originates from variation within the same day (e.g. variation in DNA damage between gels) and from day-to-day variation (e.g. variation from one batch of analysis to another). The analysis in Table IV indicates that day-to-day variation, rather than variation within the same day, is the most important contributor to the overall residual variation. This is in accordance with an earlier assessment where it was found that the assay variation (analysed in reference samples) was the same as the residual variation in a regression model of variables that contributed to the overall variation in DNA damage in mononuclear blood cells (21). Although the source of the residual variation is unknown, it can be speculated that a part of it originates from variation in DNA damage between gels. It has been argued that analysis of 25–50 images per gel is sufficient to obtain a reliable measurement of the variability in between gels, whereas scoring of 100 images from two to three gels is adequate to obtain an accurate estimate of the DNA damage in the sample (22).

Earlier studies on inter-investigator variation in visual score showed that the variation (expressed as CV) in formamido-pyrimidine DNA glycosylase (FPG)-sensitive sites could be reduced from 59% to 39% in mononuclear blood cells using investigator-specific calibration curves compared with a common calibration curve (8). Here, we show a similar reduction in the CV, which was reduced from 47% to 28% by this means (Figure 8). This adjustment of the data increased the percentage of the overall variation that could be attributed to the dose of ionizing radiation (Table IV). Overall, the data indicate that it is possible to reduce inter-laboratory variation by adjusting the primary comet assay data with the laboratory-specific calibration curves. This is supported by results from another study that showed that the CI was lower in mononuclear blood cells when using investigator-specific calibration curves than when using common calibration curves [the estimated geometric means were 0.43 FPG-sensitive

sites/10⁶ bp (95% CI 0.14–1.27) versus 0.31 FPG-sensitive sites/10⁶ bp (95% CI 0.11–0.86)] (9).

The results outlined in Table III provide the possibility of calculating the precision of DNA damage measurement by the comet assay. A collective analysis of biomonitoring studies indicated that the level of DNA damage in leukocytes from healthy humans is about 10 units (arbitrary units on a scale of 0–100) when using %T and visual score as primary comet assay end points (23). From the overall calibration curve in Figure 6, we can estimate the average level of DNA damage in human leukocytes to be 0.27 lesions/10⁶ bp. The variation in the measurement is outlined in Table III; as an example, we use the residual variation from all the laboratories to calculate the precision in the analysis ($SD_{res, mean} = 4.2$). This yields a 95% CI of 0.21–0.34 lesions/10⁶ bp for the mean level of 0.27 lesions/10⁶ bp. In other words, a single analysis of the level of DNA damage in a blood sample by the comet assay in a single experiment is associated with a rather large variation. This implies that to produce clinically meaningful results with the comet assay requires that samples are obtained on more than one occasion or that they are analysed on several occasions. However, based on the assumption that $\alpha = 5\%$ (i.e. probability of committing a type I error) and $\beta = 20\%$ (probability of committing a type II error), it would require group sizes of 13 or 46 subjects to detect a 50% or 25% difference in DNA damage between two groups, respectively. A similar power calculation, encompassing inter-individual variation, made more than a decade ago indicated that it would require 18–50 subjects to detect a 50% difference between two groups (24). It can thus be argued that the comet assay is a suitable method for the detection of biological differences between groups of subjects but that at the moment assay variation precludes firm conclusions about the validity or significance of values at the individual level. One can therefore not extrapolate from single measurements of the comet assay end points to draw conclusions concerning possible risk of disease, although the validity of the results can be increased by repeated measurements of samples from the same subject. The precision can be increased by analysing more images per gel and/or gels per sample for both sample and calibration curve (reference standard) cells. However, this approach always has the disadvantage that the adjustment of data is based on DNA damage analysed in different gels. In any adjustment of data by reference standards, the SD of the adjusted data is equal to the sum of the SDs in the sample and reference standard. This means that the adjustment of data is mainly successful if there is a large day-to-day variation in the assay. In our study, we found that the day-to-day variation was relatively small indicating that the variation between experiments was not the most important variable determining the overall variation. This indicates that the variation arises because of variation in DNA damage between gels; it should be possible to reduce variation between gels by including a true internal standard in the same gel as sample cells. Recent studies have shown that the variation in DNA damage is indeed decreased using such an internal standard in comet assay (25). The results reported here encourage further progress towards establishing reference standards, including internal standards, in the comet assay.

Funding

Environmental Cancer Risk, Nutrition and Individual Susceptibility, a Network of Excellence operating within the European

Union 6th Framework Program, Priority 5: 'Food Quality and Safety' (contract no. 513943); the Danish Research Council (to L.F., S.L. and P.M.); the Italian Association for Cancer Research (to A.A. and G.M.); FP6 COMICS Project (contract no. 037575 to A.R.C. and A.A.).

Conflict of interest statement: None declared.

References

- Collins, A. R. (2004) The comet assay for DNA damage and repair: principles, applications, and limitations. *Mol. Biotechnol.*, **26**, 249–261.
- Olive, P. L. (2002) The comet assay. An overview of techniques. *Methods Mol. Biol.*, **203**, 179–194.
- Tice, R. R., Agurell, E., Anderson, D. *et al.* (2000) Single cell gel/comet assay: guidelines for in vitro and in vivo genetic toxicology testing. *Environ. Mol. Mutagen.*, **35**, 206–221.
- Albertini, R. J., Anderson, D., Douglas, G. R. *et al.* (2000) IPCS guidelines for the monitoring of genotoxic effects of carcinogens in humans. *Mutat. Res.*, **463**, 111–172.
- Hartmann, A., Agurell, E., Beevers, C. *et al.* (2003) Recommendations for conducting the *in vivo* alkaline comet assay. *Mutagenesis*, **18**, 45–51.
- Burlinson, B., Tice, R. R., Speit, G. *et al.* (2007) Fourth International Workgroup on Genotoxicity testing: results of the in vivo Comet assay workgroup. *Mutat. Res.*, **627**, 31–35.
- Collins, A. R., Oscoz, A. A., Brunborg, G., Gaivao, I., Giovannelli, L., Kruszewski, M., Smith, C. C. and Stetina, R. (2008) The comet assay: topical issues. *Mutagenesis*, **23**, 143–151.
- Møller, P., Friis, G., Christensen, P. H. *et al.* (2004) Intra-laboratory comet assay sample scoring exercise for determination of formamidopyrimidine DNA glycosylase sites in human mononuclear blood cell DNA. *Free Radic. Res.*, **38**, 1207–1214.
- Forchhammer, L., Bräuner, E. V., Folkmann, J. K., Danielsen, P. H., Nielsen, C., Jensen, A., Loft, S., Friis, G. and Møller, P. (2008) Variation in assessment of oxidatively damaged DNA in mononuclear blood cells by the comet assay with visual scoring. *Mutagenesis*, **23**, 223–231.
- European Standards Committee on Oxidative DNA Damage (2003) Comparative analysis of baseline 8-oxo-7,8-dihydroguanine in mammalian cell DNA, by different methods in different laboratories: an approach to consensus. *Carcinogenesis*, **23**, 2129–2133.
- European Standards Committee on Oxidative DNA Damage (2003) Measurement of DNA oxidation in human cells by chromatographic and enzymic methods. *Free Radic. Biol. Med.*, **34**, 1089–1099.
- European Standards Committee on Oxidative DNA Damage (2005) Gedik, C. M. and Collins, A. (2005) Establishing the background level of base oxidation in human lymphocyte DNA: results of an inter-laboratory validation study. *FASEB J.*, **19**, 82–84.
- Garcia, O., Mandina, T., Lamadrid, A. I. *et al.* (2004) Sensitivity and variability of visual scoring in the comet assay results of an inter-laboratory scoring exercise with the use of silver staining. *Mutat. Res.*, **556**, 25–34.
- Kohn, K. W., Erickson, L. C., Ewig, R. A. G. and Friedman, C. A. (1976) Fractionation of DNA from mammalian cells by alkaline elution. *Biochemistry*, **15**, 4629–4637.
- Ahnström, G. and Erixon, K. (1981) Measurement of strand breaks by alkaline denaturation and hydroxyapatite chromatography. In Friedberg, E. C. and Hanawalt, P. C. (eds), *DNA Repair: A Laboratory Manual Research Procedures*. Marcel Dekker, New York, NY, USA, pp. 403–418.
- International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature*, **431**, 931–945.
- Zar, J. H. (1984) *Biostatistical Analysis*. 2nd edn. Prentice Hall, Englewood Cliffs, NJ, USA.
- Olive, P. L., Banáth, J. P. and Durand, R. E. (1990) Heterogeneity in radiation-induced DNA damage and repair in tumor and normal cells measured using the "comet" assay. *Radiat. Res.*, **122**, 86–94.
- Moneef, M. A., Sherwood, B. T., Bowman, K. J., Kockelbergh, R. C., Symonds, R. P., Steward, W. P., Mellon, J. K. and Jones, G. D. (2003) Measurements using the alkaline comet assay predict bladder cancer cell radiosensitivity. *Br. J. Cancer*, **89**, 2271–2276.
- Kumaravel, T. S. and Jha, A. N. (2006) Reliable Comet assay measurements for detecting DNA damage induced by ionising radiation and chemicals. *Mutat. Res.*, **605**, 7–16.
- Møller, P., Wallin, H., Holst, E. and Knudsen, L. E. (2002) Sunlight induced DNA damage in human mononuclear cells. *FASEB J.*, **16**, 45–53.
- Lovell, D. P., Thomas, G. and Dubrow, R. (1999) Issues related to the experimental design and subsequent statistical analysis of in vivo and in vitro comet studies. *Teratog. Carcinog. Mutagen.*, **19**, 109–119.

23. Møller, P. (2006) Assessment of reference values for DNA damage detected by the comet assay in humans blood cell DNA. *Mutat. Res.*, **612**, 84–104.
24. Holz, O., Jörres, R., Kästner, A., Krause, T. and Magnussen, H. (1995) Reproducibility of basal and induced DNA single-strand breaks detected by the single-cell gel electrophoresis assay in human peripheral mononuclear leukocytes. *Int. Arch. Occup. Environ. Health*, **67**, 305–310.
25. Zainol, M., Stoute, J., Almeida, G. M., Rapp, A., Bowman, K. J., Jones, G. D. D. and ECVAG (2009) Introducing a true internal standard for the Comet assay to minimise intra- and inter-experiment variability in measures of DNA damage and repair. *Nucleic Acids Res.* Advance Access published on October 14, 2009; doi:10.1093/nar/gkp826.

Received on June 18, 2009; accepted on October 8, 2009