Mutagenesis vol. 26 no. 3 pp. 393–399, 2011
Advance Access Publication 11 January 2011
doi:10.1093/mutage/geq105

The influence of scoring method on variability in results obtained with the comet assay

Amaya Azqueta[1,*], Silja Meier[1,3], Catherine Priestley[2], Kristine Bjerve Gutzkow[3], Gunnar Brunborg[3], Jérôme Sallette[4], Françoise Soussaline[4] and Andrew Collins[1]

[1]Department of Nutrition, Faculty of Medicine, University of Oslo, PO Box 1046 Blindern, 0316 Oslo, Norway, [2]Safety Assessment, AstraZeneca Research & Development, Alderley Park, Macclesfield, Cheshire SK104TG, UK, [3]Department of Chemical Toxicology, Division of Environmental Medicine, Norwegian Institute of Public Health, PO Box 4404 Nydalen, 0403 Oslo, Norway and [4]IMSTAR SA, 60 rue ND des Champs, 75006, Paris, France.

[*]To whom correspondence should be addressed. Tel: +47 22851329; Fax: +47 22851341; Email: o.a.azqueta@medisin.uio.no

As part of a project to develop high throughput versions of the comet assay (single cell gel electrophoresis), with a consequent need for more efficient scoring, we have compared the performance of visual scoring, automated and semi-automated image analysis when assessing comets in the same set of gels from dose-response experiments with typical DNA-damaging agents. Human lymphoblastoid TK-6 cells were treated with concentrations of methylmethanesulphonate between 0.04 and 0.6 mM, and peripheral human lymphocytes were incubated, after embedding in agarose, with H_2O_2 concentrations from 2.5 to 160 μM. All three scoring methods proved capable of detecting a significant level of damage at the lowest concentration of each agent. Visual scoring systematically overestimates low levels of damage compared with computerised image analysis; on the other hand, heavily damaged comets are less efficiently detected with image analysis. Overall, the degree of agreement between the scoring methods is within acceptable limits according to a Bland–Altman analysis.

Introduction

The comet assay (single cell gel electrophoresis) is widely used in genotoxicity testing, human biomonitoring, ecogenotoxicology and basic research into mechanisms of DNA damage and repair. In its simplest form, with an alkaline treatment and electrophoresis at high pH, it detects DNA strand breaks (either single or double stranded) and also alkali-labile sites, notably the apurinic/apyrimidinic sites (AP sites) that are left when a base is lost from the DNA either spontaneously or during DNA repair. When combined with lesion-specific enzymes, it detects also oxidised or alkylated bases and ultraviolet-induced pyrimidine dimers.

The assay is sensitive, versatile and—after calibration—quantitative, but it suffers from the perception that it is less precise than more analytical techniques based on chromatography [even though the comet assay has been shown to be more accurate than high performance liquid chromatography or gas chromatography-mass spectrometry in estimating oxidative damage in DNA (1)]. A recent trial, in which identical samples of cells treated with ionising radiation to induce different levels of DNA breakage were distributed to 12 laboratories for analysis (2), showed substantial variation between results from the different laboratories. Discrepancies were clearly the result of variations in technique, including, possibly, the method of scoring.

Two kinds of comet scoring are in common use, both depending on fluorescence microscopy after staining with a suitable DNA-binding dye. With visual scoring, comets are assigned to one of five classes according to the perceived intensity of the comet tail; giving values 0–4 according to the damage class, the overall score for 100 comets will be between 0 and 400 arbitrary units. The alternative method is computer-based image analysis, which gives a value of % DNA in the tail (among other parameters) for each comet. The image analysis can be either semi-automated, where a gel is scanned and comets selected for analysis by the operator, or automated, with no involvement of the operator in selecting comets. In terms of objectivity, automated image analysis would be expected to be superior to semi-automated image analysis, which should be superior to visual scoring. However, there are potential problems with image analysis; for example, overlapping comets are not scorable by image analysis but can be resolved visually, and since they are more likely to occur when there are many breaks and large comet tails, image analysis may underestimate damage.

In this paper, we examine the results obtained in one laboratory applying three scoring methods—visual scoring, semi-automated image analysis and automated image analysis—to gels prepared with methylmethanesulphonate (MMS)-treated TK-6 cells or with freshly isolated human lymphocytes treated with H_2O_2. These chemicals and cell types were selected because they give different response patterns. TK-6 cells treated with MMS behave as a homogeneous population, i.e. as the concentration of MMS increases, the level of damage in all cells increases. In contrast, when lymphocytes are treated with H_2O_2, some cells show an increase in damage from a low concentration of the chemical, while a substantial fraction of the cells are resistant to damage until a relatively high concentration is reached, leading to a heterogeneous distribution of comets.

An important aspect of precision is intra-experimental repeatability; in other words, how similar are the results when the same sample (in this case, the same gel containing comets) is analysed several times? As well as carrying out repeated scoring, we look at the human factor—the additional variability introduced (in visual scoring and semi-automated image analysis) by different operators.

Materials and methods

Cells

TK-6 cells derived from human lymphoblast cell line were grown in RPMI (Roswell Park Memorial Institute) medium (Sigma) supplemented with 10% fetal calf serum (Sigma), 200 μg/ml sodium pyruvate, 2 mM L-glutamine,

100 U/ml penicillin and 100 μg/ml streptomycin. Cells were maintained as suspension culture at 37°C in a humidified atmosphere with 5% CO_2.

Human lymphocytes were isolated by venepuncture and centrifugation over Lymphoprep (AXIS-SHIELD PoC AS).

Treatment of TK-6 cells with MMS

Five millilitres of TK-6 cells at 2.5×10^5 cells/ml were treated with 0, 0.04, 0.06, 0.08, 0.1, 0.2, 0.4 or 0.6 mM MMS for 3 h at 37°C. After the treatment, cells were centrifuged and washed with fresh medium and phosphate-buffered saline (PBS) and then suspended at 4.5×10^5 cells/ml in PBS.

Treatment of human lymphocytes with H_2O_2

Lymphocytes were treated with 0, 2.5, 5, 10, 20, 40, 60, 80, 100, 120, 140 or 160 μM H_2O_2 after setting the gels as described below. Twelve 5 μl gels were set on each slide, one gel per concentration of H_2O_2. A 12-well silicone gasket clamped on the slide allowed us to incubate each gel independently without mixing (3), and so 100 μl of different concentrations of H_2O_2 was added to each well. After 5 min of treatment on ice, each gel was washed twice in PBS, slides were immersed in lysis solution and the comet assay protocol was followed.

The comet assay

Fifteen microlitres of cell suspension (4.5×10^5 cells/ml in PBS) was mixed with 135 μl of 0.5% low melting point agarose at 37°C and 5 μl aliquots were dropped onto a microscope slide precoated with 1% normal melting point agarose or onto a GelBond film of similar size. For the experiments with MMS, the gels were set on glass slides and on GelBond film; with H_2O_2, only glass slides were used. Each slide/GelBond film had either 8 gels from TK-6 cells treated with the different concentration of MMS or 12 gels from untreated lymphocytes for subsequent H_2O_2 treatment as above. Three identical slides were prepared in each case so that each concentration was represented by three gels. Gels were then left to set at 4°C for 5 min. TK-6 cells treated with MMS were lysed overnight by immersion of the slides in 2.5 M NaCl, 0.1 M Na_2EDTA, 0.1 M Tris base, pH 10 and 1% Triton X-100 (lysis solution) at 4°C. Lymphocytes were similarly lysed after H_2O_2 treatment. Slides were then placed in a horizontal gel electrophoresis tank and DNA was allowed to unwind for 20 min in 0.3 M NaOH and 1 mM Na_2EDTA, pH $>$ 13 before the electrophoresis was carried out for 20 min at 0.8 V/cm and ~300 mA in a cold room. The slides were neutralised by washing them three times for 5 min in 0.4 M Tris base (pH 7.5) at 4°C and then rinsed in distilled water. Then, slides were fixed in absolute ethanol for 1.5 h and air-dried overnight. DNA was stained by immersing the slides in SYBR Gold (Invitrogen) in TE buffer (10 mM Tris base, 1 mM ethylenediaminetetraacetic acid, pH 8.0) for 30 min at 4°C, followed by washing in water, drying, placing a drop of water on each gel and covering with a coverslip. A Nikon Eclipse TS-100 fluorescence microscope was used to evaluate the nuclei visually or using a semi-automated image analysis system (Comet Assay IV; Perceptive Instruments). A Pathfinder™ Cellscan Comet system (IMSTAR) was used for the automated image analysis.

With the visual scoring system, a total of 50 comets on each gel were classified as belonging to one of five categories according to the tail and head intensity. Each category was given a value between 0 and 4: (0) = undamaged and (4) = maximum damage (i.e. almost all DNA in tail). An overall score was calculated for each gel by applying the following formula: (percentage of cells in class 0×0) + (percentage of cells in class 1×1) + (percentage of cells in class 2×2) + (percentage of cells in class 3×3) + (percentage of cells in class 4×4). Consequently, the total score was in the range from 0 to 400 (4). The number of comets present in each gel was recorded for comparison with the number actually detected by automated image analysis. With the semi-automated image analysis system, the images of 50 comets on each gel were analysed. In this system, the operator selects the comets to be analysed; the programme defines the area of the comet, the beginning of the head, the middle of the head and the end of the tail, and calculates, on the basis of fluorescence intensity, the percentage of DNA in the tail (along with other descriptive parameters). The semi-automated image analysis system allows the operator to change the set parameters (area of comet, beginning of head, middle of head and end of tail) if the operator does not agree with the programme but this option was not used in these experiments and we only took into account the comets that the programme can detect. The automated image analysis system detects comets without limit on number and with no manual intervention. Thus, the number of comets scored varies from gel to gel. It analyses all the non-overlapping comets according to a number of characteristics (intensity, shape, size and symmetry), thus enabling parameter distribution analysis and records images of all comets detected in the scanned area. These comet images are analysed to give percentage of DNA in tail and a range of other morphological characteristics, such as shape and area of individual head and tail, as well as different inertia moments.

In order to check for experimental variation, one slide per treatment was scored repeatedly by the same operator. Also, with visual and semi-automated scoring, three different operators scored the same slides.

Statistics

Visual scoring gives a single value per gel as the unit of measurement. With image analysis, the unit of measurement from the MMS experiment was the median percentage of DNA in the tail per gel; this is recommended for most comet results because it does not give undue weight to extreme values. However, in the case of comets from lymphocytes treated with H_2O_2, the distribution tends to be concentrated at either high or low levels of damage (see Figure 1B); in this case, median values would show an abrupt change at a certain concentration, while the mean value gives a better impression of overall damage. Generally, in figures, we show the mean value of medians or means from three gels. *T*-test was used to compare percentage of DNA in tail of comets from cells treated with the solvent and cells treated with the lowest dose of MMS or H_2O_2 and also to compare results with different scoring methods. Coefficients of variation (CVs) were calculated for each treatment dose under all conditions and with all three scoring methods; we also calculated CVs for repeated scoring and for scores of different operators. Comparisons of different methods of scoring employed the Bland–Altman approach (5). For this purpose, visual scores were divided by four so that the range was 0–100, comparable with the range of 0–100 for percentage of DNA in tail.

Results

Different patterns of response to MMS and H_2O_2

As explained in the Introduction, the pattern of DNA damage depends both on the cell type and on the damaging agent.

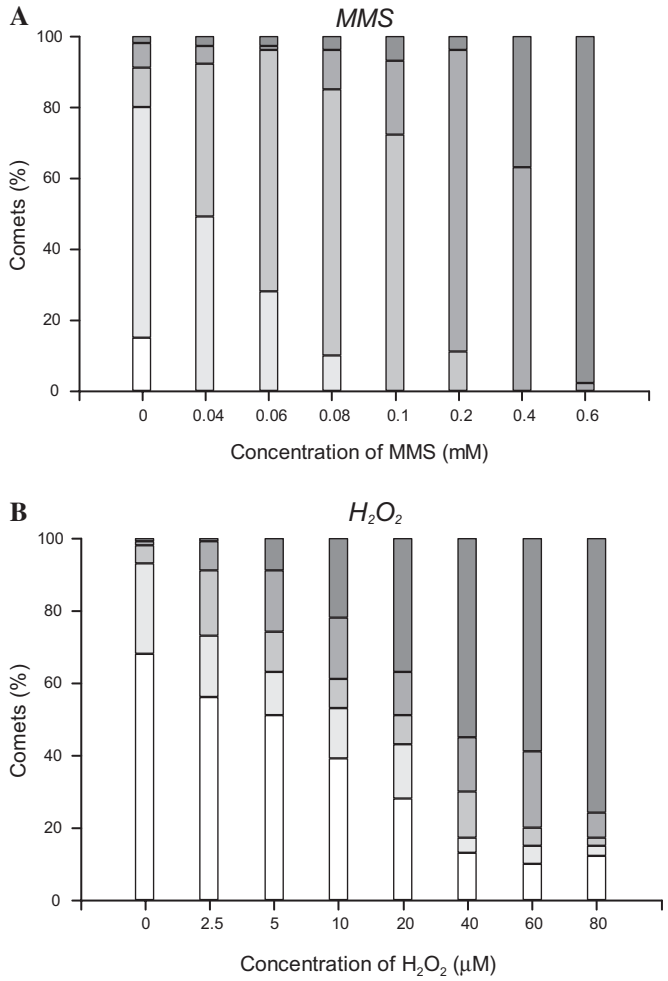

Fig. 1. Distribution of comets in different classes by visual scoring, across the range of concentrations of (**A**) MMS (TK-6 cells) and (**B**) H_2O_2 (lymphocytes). Classes are indicated by shading, from white (class 0) to dark grey (class 4).

Figure 1A illustrates the distribution of comets (classified by visual scoring) as the concentration of MMS increases: comets move homogeneously through the damage classes, until at 0.6 mM virtually all are in Class 4. In contrast, as Figure 1B shows, H_2O_2 seems to have an effect on a subpopulation of cells at low concentrations, leaving other cells virtually undamaged until much higher concentrations are reached. There are few comets in the intermediate damage classes, at any concentration of H_2O_2. {This pattern is typical of lymphocytes; other cell types treated with H_2O_2 tend to behave more homogeneously [(6) and common observations by the authors].} Thus these two distinct patterns of behaviour present different challenges in terms of image analysis.

DNA breaks induced by alkylating agent MMS

MMS causes alkylation damage mainly in the form of methylated guanine. TK-6 cells were treated with a range of concentrations of MMS for 3 h; during this time, processing of the initial damage leads to the presence of DNA strand breaks and AP sites. The cells were mixed with agarose to make gels on triplicate glass slides or GelBond films. The same slides were scored by visual scoring, semi-automated image analysis and automated image analysis. Figure 2 shows the results obtained with visual scoring expressed as arbitrary units. There is a clear dose-response; the lowest concentration, 0.04 mM, produces breaks significantly above background, while at the highest concentration, 0.6 mM, the saturation limit of the assay is reached (virtually all comets being designated as Class 4). Results with glass slides and with GelBond are indistinguishable. It is possible to scan these small gels visually and to count the comets present; Figure 2 shows that gels had similar densities of cells, except for two doses (0.1 and 0.6 mM) where numbers were lower.

Semi-automated image analysis of the same gels similarly detected the effect of the lowest concentration and reached a near-maximum percentage of DNA in tail at the highest concentration. Results were very similar for glass slides and GelBond (Figure 3). Routinely, 50 comets were picked (at random) for analysis; but in a few gels, the density of comets was low and that number could not be reached.

The dose-response curves obtained using automated image analysis were virtually identical for glass slides and GelBond (Figure 4). The automated system analyses all the comets it recognises in a gel, and so it is instructive to look at the numbers of comets analysed. Numbers vary from gel to gel, the maximum being ~100, at 0.06 and 0.08 mM MMS. There is a tendency for fewer comets to be detected at high MMS concentrations.

In order to discover whether the variable recovery was the result of a failure of the automatic focusing, we took one slide that had undergone fully automated scoring and located and focused the gels by eye before repeating the automatic analysis; we found that the results—both tail DNA percentages and numbers of comets scored—were identical to those obtained when the system was working fully automatically (data not shown).

Figure 5A plots the DNA damage results obtained using the three distinct scoring methods. All three methods show a strong association between DNA damage and MMS concentration. Dose-response curves for semi-automated and automated image analysis are very similar, while visual scoring apparently shows a higher level of background DNA breaks. With all three scoring methods, there was a significant difference between scores from

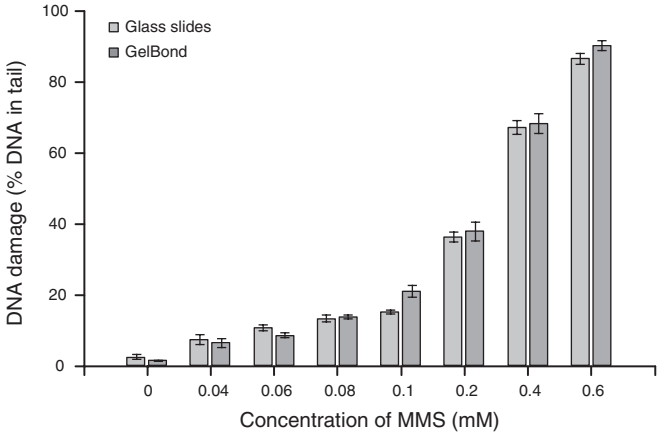

Fig. 3. DNA damage induced by MMS, assessed by semi-automated image analysis of gels on glass slides or on GelBond film; mean values from three gels, with SD.

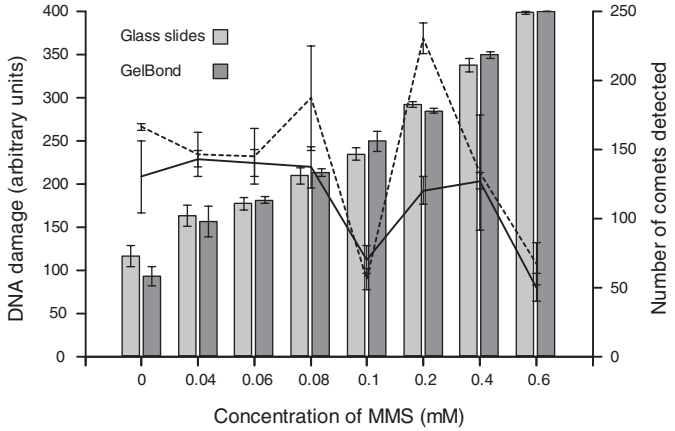

Fig. 2. DNA damage induced by MMS, assessed by visual scoring of gels on glass slides or on GelBond film (bars), with manual counts of the total number of detectable comets per gel shown as solid line, for glass slides and broken line, for GelBond. Mean values from three gels are shown, with SD.

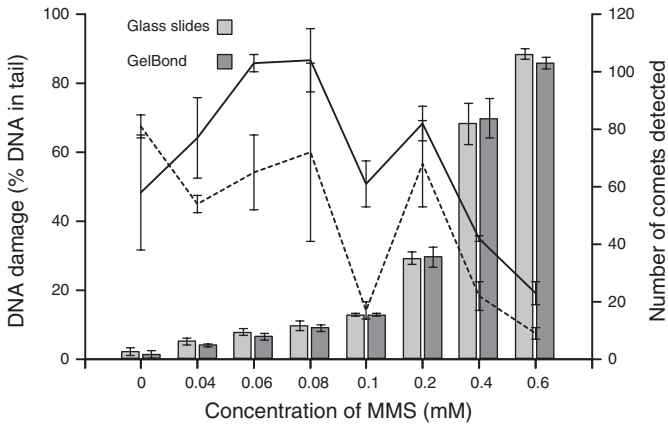

Fig. 4. DNA damage induced by MMS, assessed by automated image analysis of gels on glass slides or on GelBond film (bars), with numbers of comets per gel detected by the automated system shown as solid line, for glass slides and broken line, for GelBond. Mean values from three gels are shown, with SD.

untreated comets and the lowest concentration of MMS ($P \leq 0.023$). The recommended method for assessing agreement between two methods of measurement is the Bland–Altman approach; the difference between values obtained with two scoring methods is plotted against the mean of the values. Figure 6 (left-hand panels) shows that all data fall within limits of agreement of ±2 SDs, with no outliers.

Mean and median CVs for each condition and scoring method are shown in Table I.

DNA breaks induced by H_2O_2

We were interested to know how the different scoring systems would cope with a population of comets from lymphocytes showing a bimodal distribution of damage.

Cells were treated with H_2O_2 after setting in agarose in the 12-gel format on glass slides (GelBond films were not used in this trial as they had given the same results as glass slides in the MMS experiments). Visual scoring (Figure 7) and semi-automated image analysis (Figure 8) gave very similar dose-responses, with detection of the lowest dose (2.5 μM) and a maximum (i.e. saturation of the assay) at ~100 μM. Automated image analysis (Figure 9) showed a similar response, but levels of damage recorded were significantly lower at most doses of H_2O_2. The maximum level of damage reached was ~60% DNA in tail by automated image analysis compared with 75% with semi-automated image analysis. The number of comets present in each gel was constant at ~90, as estimated by visual examination (Figure 7); only at the highest concentration, 160 μM, was there an indication that some cells might have been lost (or comets were so highly damaged that DNA was fragmented and dispersed and comets invisible). However, the number of comets detected by the automated image analysis fell dramatically as the concentration of H_2O_2 increased (Figure 9)—suggesting that the programme was missing some of the more heavily damaged comets. This is essentially because the more damaged comets, without densely staining heads, are more difficult to detect as they are not considered as corresponding to the comet multi-parameter model. We also compared normal completely automated analysis, with manual location and focusing of comets followed by automated image capture and analysis. Except for one high concentration, dose-responses were similar, and in both cases, the number of comets detected showed the same decline with increasing H_2O_2 concentration (data not shown). As seen in Figure 5B, the dose-response for H_2O_2 is clearly non-linear, whichever scoring method is used, reflecting the fact that the highest doses tested produced damage above the saturation level of the assay. With all three scoring methods, there was a significant difference between scores from untreated comets and the lowest concentration (2.5 μM) of H_2O_2 ($P \leq 0.05$).

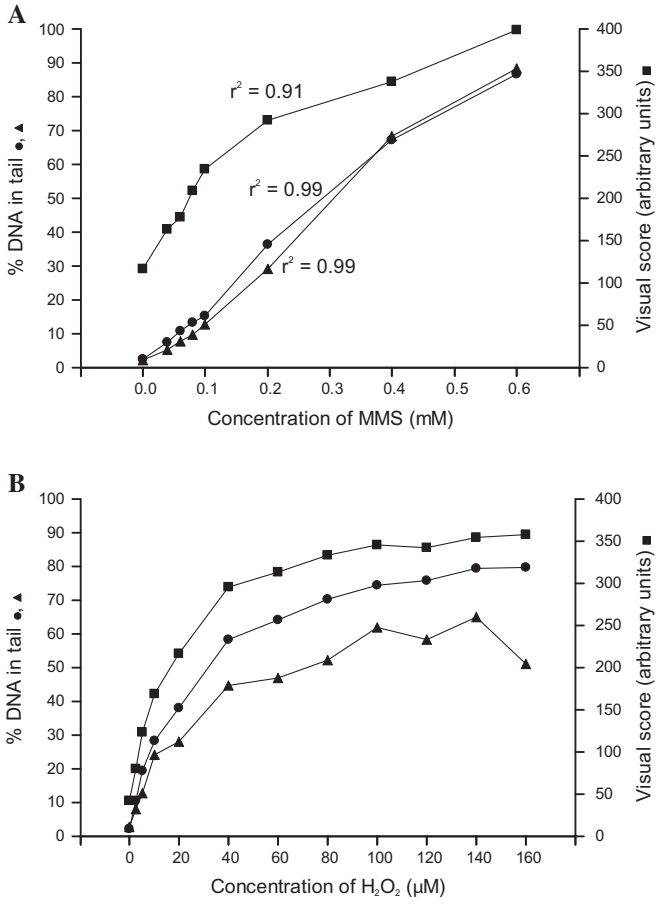

Fig. 5. DNA damage induced by different concentrations of MMS (**A**) or H_2O_2 (**B**), plotted on a linear scale: data from 12-gel glass slides, scored visually (squares), by semi-automated image analysis (circles) or by automated image analysis (triangles). Regression analysis of the image analysis data in (A) provide the r^2 values shown.

Figure 6 (right-hand panels) gives Bland–Altman plots for the H_2O_2-damaged comets and shows good agreement between different scoring methods but also a clear indication that the difference between scores depends on the concentration of H_2O_2 (i.e. the level of damage, indicated on the *x*-axis by the mean comet score).

Intra-experimental variation

An important aspect of validation is reproducibility. How much variation is there when the same slide is repeatedly scored, with the three scoring systems used here? And does this variation increase as different people are involved in the analysis? Table I gives CVs for visual scoring and semi-automated image analysis, based on triplicate scoring of the same slide and also on the use of three different operators for the scoring.

Visual scoring gave near identical results when the same slide of MMS-treated cells was scored three times; results were similar also for H_2O_2-damaged cells. Not surprisingly, there was more variation when different scorers were involved. Semi-automated image analysis gave similar results from triplicate scores of MMS-treated cells; variation was greater when H_2O_2-treated cells were analysed. The use of different scorers increased variability with this method, too.

In the case of automated image analysis, the operator does not have any influence on the results, and so we examined only the reproducibility of repeated scoring by one operator. Whether damage was inflicted by MMS or by H_2O_2, the damage scores were acceptably close (CVs <20% and in three of four cases <10%).

Discussion

Results with the three scoring systems on MMS-treated cells were qualitatively similar, with detection of breaks at the lowest concentration of MMS (0.04 mM) and saturation at ~0.6 mM. The dose-response curves with MMS (Figure 5A) are very close for both image analysis systems, while visual scoring clearly starts from a higher baseline, reflecting the

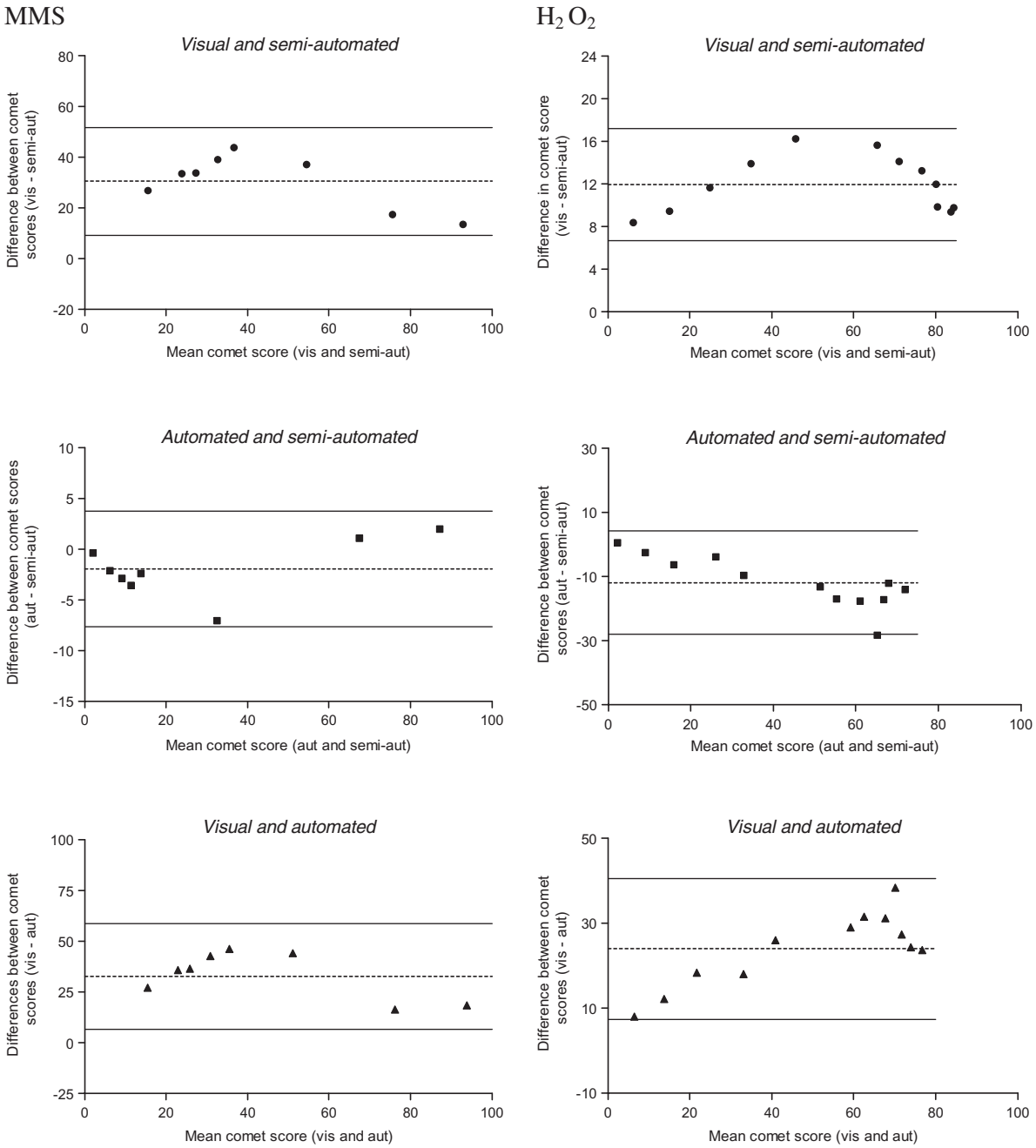

Fig. 6. Bland–Altman plots to assess the degree of agreement between different scoring methods. For each pairwise comparison, differences between comet scores by the two methods are plotted against the mean comet score by the two methods. The broken line indicates the mean difference, and the solid lines represent ±2 SDs from this mean. Left-hand panels: comparisons for MMS-treated cells. Right-hand panels: comparisons for H_2O_2-treated cells.

Table I. Reproducibility, as indicated by the CV (%) when the same slide was scored three times by the same operator and (in the case of visual scoring and manual semi-automated image analysis) when scored by three different operators (For automated image analysis, obviously, different scorers are not an issue.)

		Visual scoring: CV (%)		Manual semi-automated image analysis: CV (%)		Automated image analysis: CV (%)	
		Mean	Median	Mean	Median	Mean	Median
MMS	Repeat scores	2.6	1.6	7.9	6.2	9.0	4.3
	Different scorers	11.1	11.1	16.3	13.8	—	—
H_2O_2	Repeat scores	5.9	4.2	17.4	8.3	19.2	7.5
	Different scorers	16.3	9.9	25.3	17.0	—	—

CVs were calculated for each set of three scores; mean and median CVs were then calculated across the range of concentrations of MMS or H_2O_2 for each scoring method.

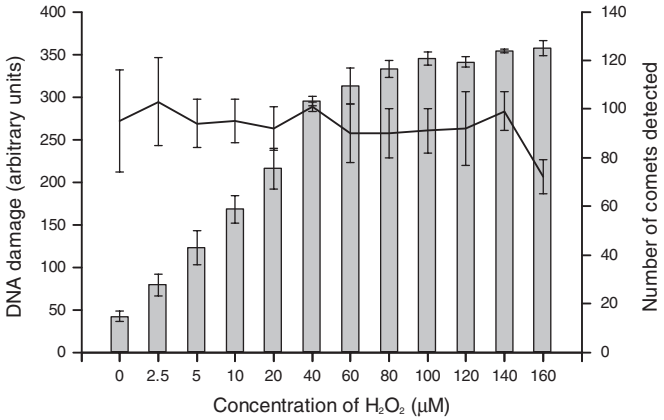

Fig. 7. DNA damage induced by H_2O_2, assessed by visual scoring of gels on glass slides (bars), with manual counts of detectable comets per gel shown as a solid line. Mean values from three gels are shown, with SD.

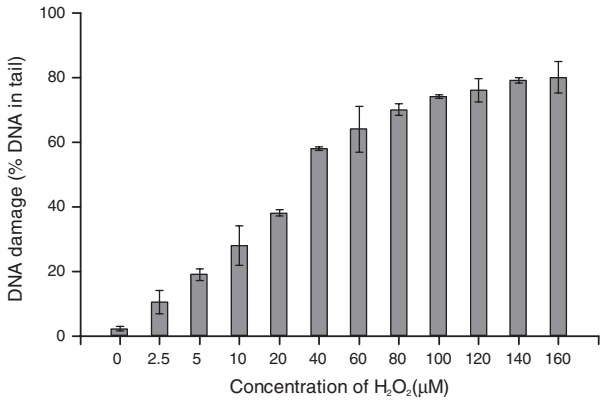

Fig. 8. DNA damage induced by H_2O_2, assessed by semi-automated image analysis of gels on glass slides; mean values from three gels, with SD.

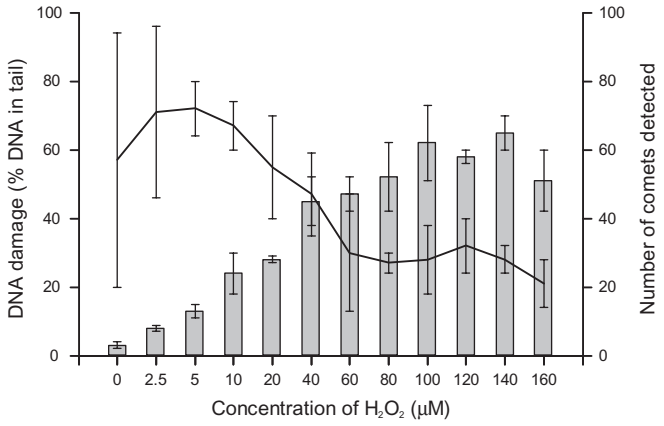

Fig. 9. DNA damage induced by H_2O_2, assessed by automated image analysis of gels on glass slides (bars), with manual counts of detectable comets per gel shown as a solid line. Numbers of comets per gel are detected by the automated system. Mean values from three gels are shown, with SD.

coarseness of the grading system; if a comet has only a faint, but detectable tail, it is given a value of 1, even if only a very small percentage of the DNA is in the tail. Automated image analysis does not detect all the comets—especially at higher levels of damage (Figure 4). This is not a result of poor focusing. Nor is it likely to be because of overlapping of the heavily damaged cells; the density in these gels was such that few comets overlapped. However, it is worth emphasising that, even when counting few cells, the amount of damage is measured accurately (agreeing with semi-automated image analysis) and precisely.

Relatively low numbers of comets were detected at MMS concentrations 0.1 and 0.6 mM—with both visual and automated scoring (Figures 2 and 4). At the highest dose, it is likely that cytotoxicity caused a decrease in detectable comets; the low number at 0.1 mM is presumably the result of a technical error.

Also in the case of H_2O_2-treated lymphocytes, as shown in Figure 5B, automated image analysis gives lower values, relative to the other two methods; this is apparently because more heavily damaged comets (which are less easily analysed) are present even at low concentrations of H_2O_2. The level of agreement between the two versions of image analysis is illustrated by the Bland–Altman plot (Figure 6).

There are reasonable doubts about scoring consistency and reproducibility of the comet assay, i.e. its precision. These doubts were addressed in our experiments with repeated scoring of the same slide by the same or different operators. CVs tend to be high at very low levels of damage, and so we calculated both mean and median CVs across the range of concentrations of damaging agent (Table I). The median value minimises the influence of the very high values. Mean CVs (for repeat counts with the same operator) ranged between 2.6 and 19.2% while median values were between 1.6 and 8.3%, depending on damaging agent and scoring method. Automated and semi-automated analysis resulted in very similar CVs. Comet score CVs were invariably higher when different scorers were involved; but agreement was remarkably good, when compared with the disparate results obtained by different operators (scoring the same slides) in the study of Forchhammer *et al*. (7).

In spite of its subjective nature, visual scoring is capable of giving reliable, quantitative results in the hands—and eyes—of an experienced microscopist. Image analysis is more rigorously quantitative than visual scoring and in theory less prone to operator bias—though it should be remembered that in the semi-automated version, comets are selected by the operator for analysis. (In fact, as Table I shows, trained operators agree more closely with visual scoring than with semi-automated image analysis). Obviously, as the number of samples processed per experiment increases, the need for automated scoring becomes paramount. Automated image analysis has the definite advantage that operator time is saved for less tedious tasks than gazing at comets through the microscope or on the screen. It requires, however, a significant outlay on equipment and dedicated software. Automated image analysis does not register all the more highly damaged comets although this does not affect the end result in terms of sensitivity (detection of low levels of damage) or the shape of the dose-response curve.

Bland–Altman plots, as in Figure 6, are a standard method to compare different ways of measuring the same samples. Differences between the pairs of measurements are plotted against the mean value; outliers are readily identified as falling outside the range of ±2 SDs. All comparisons were free of outliers. Clearly, the pattern is not random; it highlights those regions of the dose-response where there are greater or lesser differences in scoring by the various methods.

The method chosen for scoring comets will depend on resources available as well as on considerations of accuracy

or precision. Automated scoring is most appropriate for high throughput analyses; visual scoring is simple and cheap but shows some deviation from linearity with respect to break frequency. However, the important message from this work is that results from all three approaches can be regarded as trustworthy and—to a large extent—interchangeable.

Funding

European Commission (Contract LSHB-CT-2006-037575)—project 'Comet assay and cell array for fast and efficient genotoxicity testing' (COMICS).

Acknowledgements

Conflict of interest statement: None declared.

References

1. ESCODD, Gedik, C. M. and Collins, A. Establishing the background level of base oxidation in human lymphocyte DNA: results of an interlaboratory validation study. *FASEB J.*, **19**, 82–84.
2. Forchhammer, L., Johanson, C., Loft, S. *et al.* (2010) Variation in the measurement of DNA damage by comet assay measured by the ECVAG inter-laboratory validation trial. *Mutagenesis*, **25**, 113–123.
3. Shaposhnikov, S., Azqueta, A., Henriksson, S. *et al.* (2010) Twelve-gel slide format optimised for comet assay and fluorescent *in situ* hybridisation. *Toxicol. Lett.*, **195**, 31–34.
4. Collins, A., Dusinská, M., Franklin, M. *et al.* (1997) Comet assay in human biomonitoring studies: reliability, validation, and applications. *Environ. Mol. Mutagen.*, **30**, 139–146.
5. Bland, J. M. and Altman, D. G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, **327**, 307–310.
6. Collins, A. R., Ma, A. and Duthie, S. J. (1995) The kinetics of repair of oxidative DNA damage (strand breaks and oxidised pyrimidines) in human cells. *Mutat. Res.*, **336**, 69–77.
7. 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.