Mutat Res Gen Tox En 837 (2019) 52–59

Contents lists available at ScienceDirect

Mutat Res Gen Tox En

journal homepage: www.elsevier.com/locate/gentox

Analysis of historical negative control group data from the *in vitro* micronucleus assay using human lymphocytes

David P. Lovell[a,*], Mick Fellow[b], Azeddine Elhajouji[c], Christopher S. Farabaugh[d], Ben G. Gilby[e], Kiyohiro Hashimoto[f], Yan Li[g], Shambhu Roy[h], Maik Schuler[i], James Whitwell[j], Jennifer Y. Tanir[k]

[a] St George's Medical School, University of London, London, SW17 0RE, UK
[b] Astra Zeneca, Drug Safety and Metabolism, Cambridge, CB4 0WG, UK
[c] Novartis Institutes for Biomedical Research, Preclinical Safety, Genetic Toxicology and photosafety, Basel, Switzerland
[d] Charles River Laboratories, Skokie, IN, USA
[e] Envigo CRS Limited, Cambs, UK
[f] Takeda Pharmaceutical Company Limited, Fujisawa, Kanagawa, Japan
[g] Covance Greenfield, Indianapolis, IN, USA
[h] Bioreliance, Rockville, MD, USA
[i] Pfizer, Groton, CT, USA
[j] Covance Laboratories, Harrogate, North Yorkshire, HG3 1PY, UK
[k] ILSI Health and Environmental Sciences Institute (HESI), 1156 15th Street NW, 2nd Floor, Washington, DC 20005, USA

ARTICLE INFO

Keywords:
Human lymphocytes
In vitro micronucleus assay
Historical negative control data
Quality control statistics

ABSTRACT

A database of the micronuclei counts was built up for historical negative control data from human lymphocyte *in vitro* micronuclei tests (MnVit) carried out in 8 laboratories with experience of the method. The mean incidence of micronucleated cells (mnt)/1000 cells ranged from 2.2/1000 to 15.9/1000. There were no large differences in incidence between the presence or absence of S9 mix or between different treatment lengths. There was also little evidence that different solvents affected the numbers of micronuclei appreciably. A number of laboratories did show significant inter-experiment variability, indicating that there remained unidentified factors affecting frequencies. Donor variance may be one such factor. Inter-individual variability may explain some of these differences. The approximate 7.5-fold difference in mnt/1000 scores in a relatively small group of experienced laboratories illustrates the potential complications that can arise if a metric like a fold increase was considered the only biologically important finding. Although there is inherent variability between experiments, it was evident that within a laboratory the overall laboratory mean remains constant over time. It is believed that these findings will provide help to laboratories conducting studies using human lymphocytes in the MnVit and to those involved in the assessment of MnVit results.

1. Introduction

The revised Organisation for Economic Cooperation and Development (OECD) guidelines [1] emphasised a more important role for historical negative control data in the analysis and interpretation of *in vitro* micronucleus assays. The guidance has indicated that comparison of individual study results with a laboratory's historical negative control data should form part of result evaluation and laboratories are encouraged to build up a data set of negative control data to provide evidence of their competence with the test [2]. Negative controls in this context are samples treated with just the solvent or vehicle used with the test article.

The ILSI/HESI Genetic Toxicology Technical Committee (GTTC) Data Interpretation Workgroup initiated a project to collect and collate sets of data from established and experienced laboratories into a standardized database. This analysis and publication are intended to provide high quality negative control data from genotoxicity tests carried out under OECD guidelines. This group has previously reported on the negative control data from a set of *in vitro* micronuclei studies from a series of experienced laboratories using TK6 cells; this paper also provides additional background to the GTTC Data Interpretation project [3].

The analysis of TK6 historical micronuclei control data showed that there was appreciable variability between laboratories, concluded that

* Corresponding author.
E-mail address: dlovell@sgul.ac.uk (D.P. Lovell).

https://doi.org/10.1016/j.mrgentox.2018.08.009
Received 7 June 2018; Received in revised form 23 August 2018; Accepted 24 August 2018
Available online 18 September 2018

the information on the variables that differed between studies did not explain this variability and provided support to the OECD proposal to increase the numbers of cells scored to 2000 in part to reduce the possibility of zero counts.

In the current paper, a similar analysis has been carried out using a data set created from data provided by a set of 8 laboratories with experience in the use of human lymphocytes for the *in vitro* micronucleus test.

The human lymphocyte micronucleus test is an established method used both in epidemiological investigations [4] and in the *in vitro* micronucleus test [5]. Human lymphocytes have the advantage that they are primary human cells which are relatively easy to culture in suspension. The use of isolated lymphocytes means that the ratio of erythrocytes to lymphocytes is very different compared with the use of whole blood with a consequent reduction on hemoglobin as potential target for chemicals. Detailed protocols for the *in vitro* micronucleus test have been produced by the HUMN consortium [6] which has provided a vehicle for collaboration in the use of the assay in human epidemiological studies [4]. OECD test guideline 487 has also provided recommendations for the conduct of the assay [1].

2. Material and methods

Data were collected from laboratories recruited by the GTTC following requests for relevant data sent to laboratories that were considered to have experience in using human lymphocytes in their assays. Eight laboratories (4 from the USA, 1 from Japan and 3 from Europe) participated. All the laboratories used manual scoring methods, in addition one laboratory (Lab D) used an image analysis system for a separate set of 52 replicates. One company provided data from two laboratories (one sited in Europe, the other in the USA). This company provided its data from each site as separate datasets for male and female donors. In all, 10 data sets (identified as A–J) were collected from 8 laboratories representing 7 independent organisations (with 4 datasets from one company).

All the laboratories had carried out their experiments using three different S9 conditions (−S9 short (approx. 4 h treatment + 20 h recovery), +S9 short (approx. 4 + 20 h) and –S9 long (approx. 24 + 0 h)) except for Laboratory A which did not carry out any experiments using the –S9 short combination.

Experiments were mainly carried out between 2012 and 2015. One laboratory (A), however, provided data from a series for 1997–2000. Lab A was included because they had considerable experience of using the test and a large historical database.

Most laboratories completed a short questionnaire on laboratory conditions. The data provided showed that staining was by acridine orange (5 labortories), DifQuik (1) or Giemsa (2). All the laboratories used cytochalasin B (CytoB).

One company provided separate data sets from blood pooled from 2 male or 2 female donors 18–35 years old (UK site) or reported that each study used a single blood sample from an 18–35 year old donor (US site). The other laboratories reported a variety of different sources for the donors:

- < 45 years, both males and females
- Age 18–35 years, male and/or female, blood pooled from two donors
- Either pooled or single blood samples from 18 year or older male non-smoker
- Single blood samples from male donors age 21–35

Where a range of donors was used these could not be identified with the specific study results. All laboratories reported the use of phytohaemagglutinin (PHA) stimulations.

S9 used in most studies was from rats with the inducer used being Aroclor (6 laboratories), phenobarbital/beta-naphthoflavone (PB/BNF) (2 laboratories).

In general, each study consisted of a set of three experiments, one for each of the three S9 duration conditions. Each experiment consisted of two replicates with either 1000 or 2000 cells per replicate. Data were subsequently converted to micronucleated cells (mnt)/1000 cells scored to allow comparisons between laboratories. One laboratory (Lab D) reported the results derived from scoring 2000 cells for two replicates as a single value of mnt/1000 cells.

Solvents employed were mainly dimethyl sulphoxide (DMSO) and water. One laboratory (J) used a large range of solvents, including: DMSO (25), Ethanol (5), 'Untreated' (5), Saline (1), phosphate buffered saline (PBS) (2), Water (5), and dimethylformamide (DMF) (1).

2.1. A description of data collected

Data were collected using a specially designed Excel spreadsheet similar to that used previously [3].

Results were obtained and examined for three main types of experiments: a short treatment in the absence of S9 followed by a recovery period (−S9 short), a short treatment in the presence of S9 followed by a recovery period (+S9 short) and a long treatment in the absence of S9 without recovery (−S9 long). All the laboratories reported that the duration of treatment was 3–4 h for the short treatment with a recovery period of 20–21 h. For the long treatment, they reported a 24 h treatment with either a 0 h or 24 h recovery period.

Some studies were described as experiments using all three S9/duration conditions. In others, it was not clear whether the experiments were from a study with only one condition or more. This limited formal statistical analyses across experimental conditions within a single study.

Most of the data provided was in the form of the number of micronucleated cells together with the number of cells scored for each replicate. Most of the laboratories, with the exception of Lab D, provided data on replicate samples within an experiment. Many laboratories reported counts on 1000 cells. In the case of some laboratories (Labs C and E) the number of cells scored were either 1000 or 2000 cells/replicate. In these cases, the data were converted to a frequency of micronucleated cells/1000 cells. The total number of replicates across all conditions scored by the laboratories was 1630 and ranged from 42 for Lab I to 373 for Lab D. There were a small number of cases where more than two replicates were scored in an experiment. Lab C reported many experiments with 4 replicates and with 10 replicates in at least one case.

No restriction was placed upon the amount of data that could be submitted other than that laboratories were expected to provide data from a minimum of 20 experiments. One complication was that in some cases there were very few experiments performed with a particular solvent (Lab J, for example, included some studies on a specific solvent with just 2 replicates). These numbers were considered acceptable because they could be combined. However, this limited a determination of whether there were appreciable effects on frequencies from different solvents.

Lab D carried out two separate series of studies: one using manual and the other using an image analysis system, both with scores based upon 2000 cells combined over two replicates and the results expressed as micronucleated cells/1000 cell.

2.2. Statistical methods

The methods used in this analysis were broadly based upon those reported in the previous paper [3]. Data were analysed using the statistical procedures available in Minitab (Minitab 16 Statistical Software. Minitab, Inc., State College, PA) and the R statistical programming language [7].

Data were analysed as counts if the data were from replicates of exactly 1000 cells. Alternatively, if the number of cells scored were not

exactly 1000 then analyses were done on the number of micronucleated cells/1000 cells scored.

The procedures used were: one-way and nested analyses of variances, tests for extra-binomial variation (*i.e.* Goodness of fit to a Poisson distribution), correlations between S9 conditions within the same study (where applicable) and the calculation of tolerance intervals.

A number of QC methods were used: C-Charts for Poisson counts, I-Charts for individual replicate values of counts or proportions of micronucleated cells and X-bar Charts when there were a number of replicates per experiment. (See Lovell et al. [3] for details on the interpretation of these charts). Detailed results of many of these analyses can be found in the Supplementary material.

The anovas were carried out on untransformed 'counts' or proportions. These gave satisfactory results for the types of exploratory analyses carried out.

3. Results

Data from all the participating laboratories were considered acceptable for inclusion in the analyses reported. The quality of the data provided was sufficient and the laboratories were diligent in their presentation of the data in a standard format for all the sets of treatment conditions.

The data sets from the 8 laboratories (anonymized as A to J) were broken down into 35 combinations of the various ± S9 and treatment times. Table 1 shows the mean and standard deviations (SD) for each of these 35 combinations. The summary results in different formats are presented in the Supplementary material section to help make further comparisons within and between the laboratories.

Table 1
Summary table of results from 35 conditions from the different datasets provided by 10 laboratories.

Row	Lab	Combination	N	Mean	SD	Vehicle	S9	Time
1	A	AS9 + Sn_	14	11.57	3.11	DMSO	+ S9	Short
2	A	AS9-Ln_	35	15.94	4.70	DMSO	− S9	Long
3	B	B1S9-Sn_	23	5.22	1.98	Water	− S9	Short
4	B	B1S9 + Sn_	24	4.96	2.60	Water	+ S9	Short
5	B	B1S9-Ln_	24	5.71	3.62	Water	− S9	Long
6	B	B2S9-Sn_	80	6.17	2.62	DMSO	− S9	Short
7	B	B2S9 + Sn_	80	6.33	2.73	DMSO	+ S9	Short
8	B	B2S9-Ln_	80	7.28	2.49	DMSO	− S9	Long
9	C	C1S9 + Sn_	98	4.62	2.69	DMSO	+ S9	Short
10	C	C2S9-Sn_	66	4.70	3.17	DMSO	− S9	Short
11	C	C3S9-Ln_	74	4.45	2.72	DMSO	− S9	Long
12	D	D1S9-Sn_	36	9.58	3.45	DMSO	− S9	Short
13	D	D1S9 + Sn_	94	9.13	3.32	DMSO	+ S9	Short
14	D	D1S9-Ln_	117	8.11	3.27	DMSO	− S9	Long
15	D	D2S9-Sn_	39	5.46	2.52	DMSO	− S9	Short
16	D	D2S9 + Sn_	43	5.64	2.75	DMSO	+ S9	Short
17	D	D2S9-Ln_	44	6.95	2.94	DMSO	− S9	Long
18	E	E1S9-Sn_	24	4.94	2.00	DMSO	− S9	Short
19	E	E2S9 + Sn_	46	4.59	2.37	DMSO	+ S9	Short
20	E	E3S9-Ln_	42	5.32	2.19	DMSO	− S9	Long
21	F	FS9-Sn_	46	4.58	2.38	DMSO	− S9	Short
22	F	FS9 + Sn_	46	4.26	2.27	DMSO	+ S9	Short
23	F	FS9-Ln_	46	4.98	2.87	DMSO	− S9	Long
24	G	GS9-Sn_	34	8.47	3.57	DMSO	− S9	Short
25	G	GS9 + Sn_	34	7.26	2.76	DMSO	+ S9	Short
26	G	GS9-Ln_	34	8.29	3.49	DMSO	− S9	Long
27	H	HS9-Sn_	17	13.24	8.13	DMSO	− S9	Short
28	H	HS9 + Sn_	17	13.32	7.54	DMSO	+ S9	Short
29	H	HS9-Ln_	17	11.65	8.01	DMSO	− S9	Long
30	I	IS9-Sn_	14	4.57	3.50	DMSO	− S9	Short
31	I	IS9 + Sn_	14	2.57	1.16	DMSO	+ S9	Short
32	I	IS9-Ln_	14	4.21	1.97	DMSO	− S9	Long
33	J	JS9-Sn_	72	2.36	1.18	DMSO +	− S9	Short
34	J	JS9 + Sn_	78	2.15	1.27	DMSO +	+ S9	Short
35	J	JS9-Ln_	64	2.41	1.33	DMSO +	− S9	Long

(DMSO + signifies a range of different vehicles).

Fig. 1 shows the mean and standard deviation for each of the 35 combinations and provides an indication of the relative variability between each set of replicates. Fig. 2 gives the mean and its associated 95% confidence interval (CI) for each set. This figure shows the precision associated with the mean values and the widths of the CIs reflect in part the numbers of replicates. Combinations where the n is small will generally have wide CIs. Table 1 shows that 7 combinations had less than 20 replicates.

The means for cells scored ranged from 2.2 to 15.9 /1000. In general, the variability within the laboratories was appreciably less than between laboratories which can be seen in Fig. 3. There were significant ($P < 0.001$) differences between laboratories compared with intra-laboratory differences in a one-way anova of the 35 combination means. The mean values for the 35 combinations were 6.6/1000 (SD 3.31/1000). Figs. 4–8 show the distribution of the means and individual values. There is some indication of a skew to the right with a small number of outlying results (especially from laboratories A and H which have high mean scores (11.5–15.9/1000) while Lab J has low mean scores (between 2–3/1000). Therefore, there is a 5 to 7-fold difference between the results from experienced laboratories.

Data from Labs A and H also seem to be quite variable, *i.e.* they have large SDs. However, this may, in part, reflect that the mean and variance of Poisson distributed data will, theoretically, be identical, *i.e.* the SD will increase with the mean. None of the results can be ruled out as unacceptable. The mean counts from Lab A which came from samples which were collected around 2000 are high but have been included in the analysis as they are similar to those from Lab H. Lab A stopped using human lymphocytes in 2000 and moved on to using CHO WBL and TK6 cells It is possible that the higher values reported by them are consistent with reports that frequencies are now lower because of improved culture condition or media.

Fig. 1 also shows a distribution of the means by the length of treatment with 'short' being shown in black and 'long' in grey. Similarly, Fig. 2 shows a distribution of the means based upon the presence or absence of S9 (+S9 or −S9) with presence of S9 being shown in grey and absence of S9 in black. There are no striking differences apparent in these figures between the three conditions indicating that there is little evidence that the presence or absence of S9 or short or long treatment time affected the incidence of micronucleated cells. Analyses show that differences between the conditions can be identified as statistically significant in an anova within Lab A. However, the effects are small and probably below the level of effect that might be considered biologically important and more a reflection of the statistical power possible with the number of samples analysed. In the case of some laboratories it was possible to compare frequencies across the different S9 conditions within the same experiment. In these case there were large and highly significant correlations between the scores from the different S9 conditions (see Supplementary material).

The majority of the data from the laboratories were based upon 1000 cells scored from each of 2 replicates cultures; or, in some cases 2000 cells. Most results were from duplicate cultures. (Lab D did not report the individual counts of the duplicate cultures). In all 35 of the 1630 counts across the treatment combinations, were zero. All the laboratories had a small number of zeros: (A 2; B 6; C 3; D 6; E 3; F 3; G 3; H 3; I 3; J 3)

The effects of other laboratory-associated variables is somewhat difficult to assess because of the small number of participants.

Only one laboratory (Lab D) used a non-manual method. There were appreciable differences between the scores using the manual analysis (8.1–9.6/1000) for three conditions for manual compared with 5.4–7.0/1000 for the image analysis (Table 2). Counts were made using the different methods on different samples. The image analysis was carried out more recently and although there was a large reduction (approx. 30%) in counts using image analysis there may be other factors/confounders affecting these results.

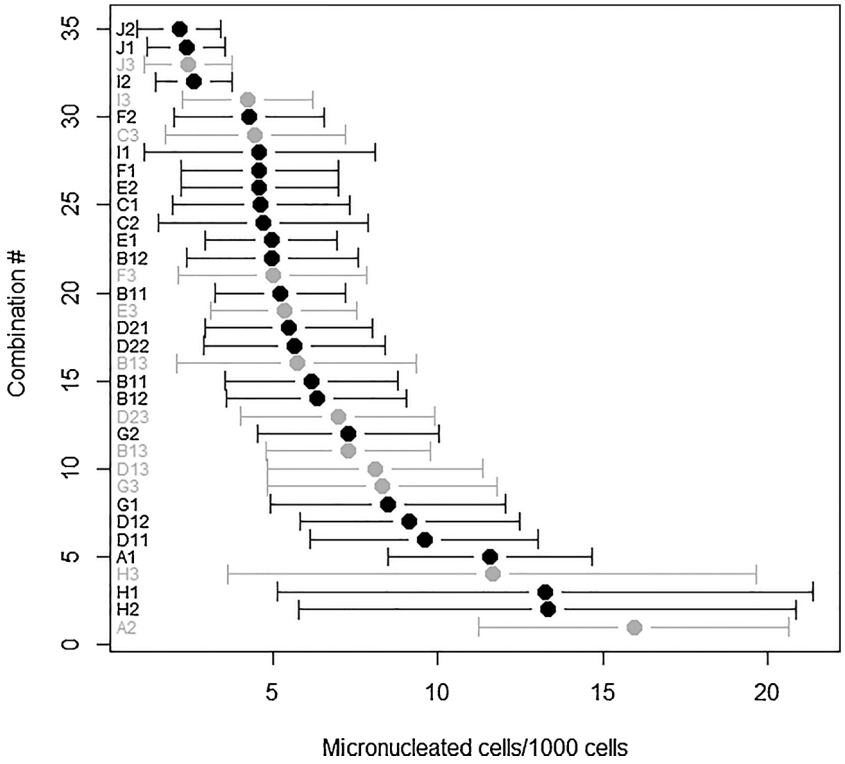

Fig. 1. Means and SD of 35 combinations from 8 laboratories: Black: Short; Grey: Long.

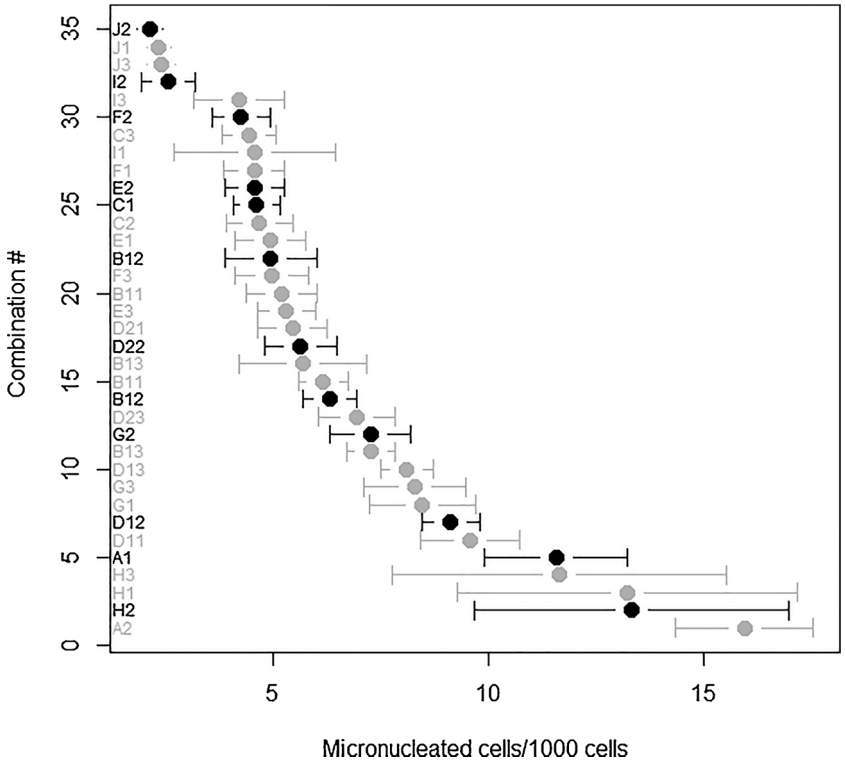

Fig. 2. Means and 95% CI of 35 combinations from 8 laboratories: Black: −S9; Grey: +S9.

One laboratory (Lab B) carried out a series of studies: (B1) with water and another (B2) with DMSO. The means of micronucleated cells of the B1 series were slightly smaller than the B2 (approximately 5.5/1000–6.4/1000) but series B1 was small (only 6 studies) and there are probably many other confounding effects. However, the results are similar enough that the two series could be combined.

Another laboratory (J) carried out studies using a series of different solvents (Table 3). A one-way anova showed significant differences in micronucleus incidence between the samples with different solvents. However, the number of studies was small and Lab J had some of the lowest scores. Conclusions as to whether the solvent employed is an important factor in introducing variability into the assay are difficult to

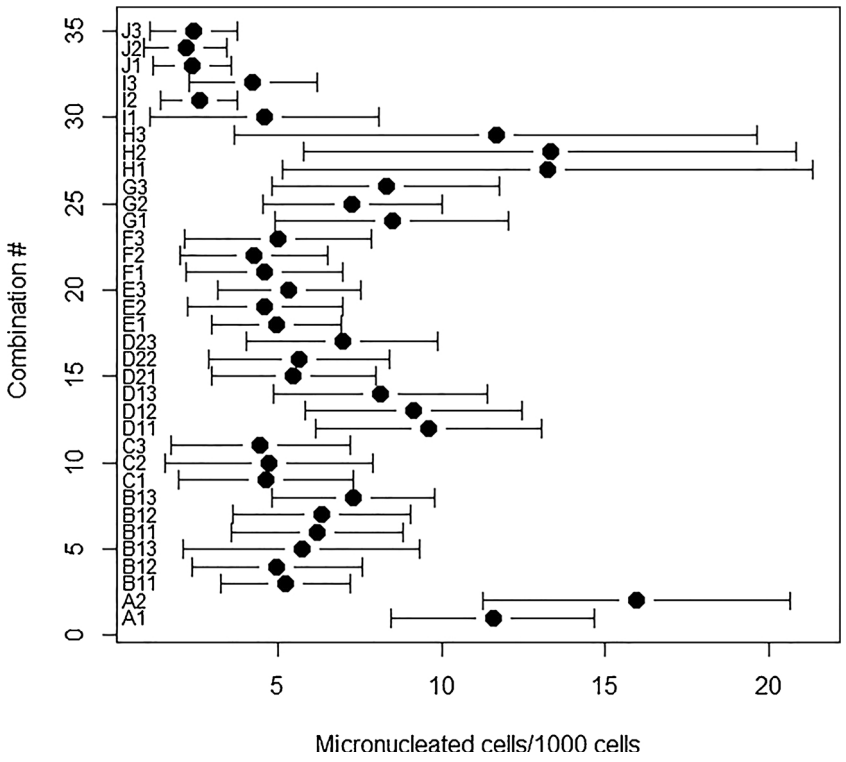

Fig. 3. Plot of the mean micronuclei/1000 cells from the 35 combinations from the 8 participating laboratories (including separate male and female samples).

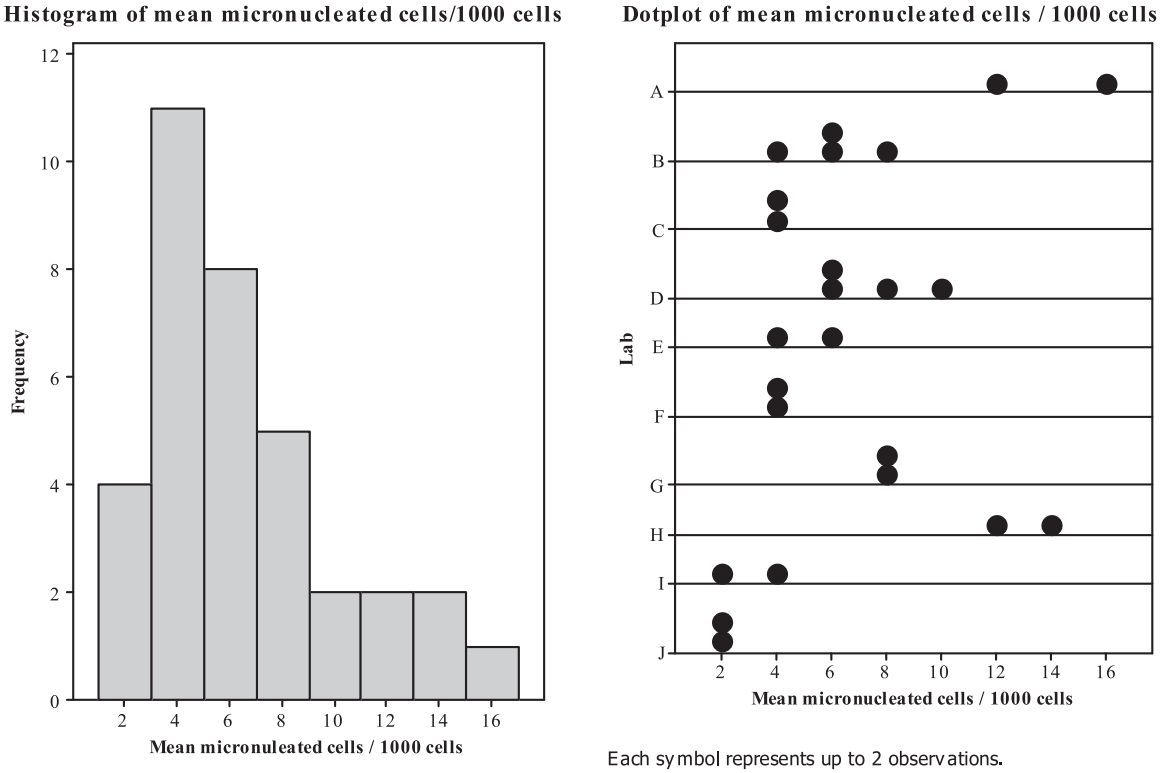

Fig. 4. Distribution of the mean micronuclei/1000 cells from the 35 combinations from the 8 participating laboratories.

draw from this data set.

Little of the variability in the results could be explained by other protocol variables. All of the laboratories used CytoB and PHA. Different stains and different P450 inducers were used and the treatment and recovery times differed slightly between laboratories but none of these variables appear to influence variability appreciably.

Several laboratories showed significant between experiment (study) variability indicating that there remained sources of variability that have yet to be identified. There is substantial between experiment variability (when compared with the within experiment variability in one-way anovas) which was detected, where an analysis was possible. A number of laboratories (*i.e.* Labs B, C, E, F, G, H, and J) showed

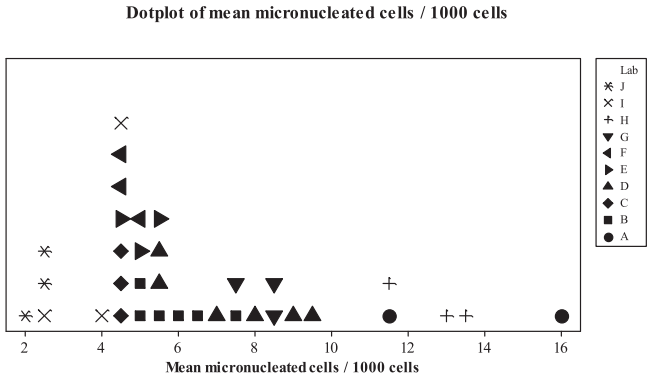

Fig. 5. Spread of the mean micronuclei/1000 cells from the 8 participating laboratories.

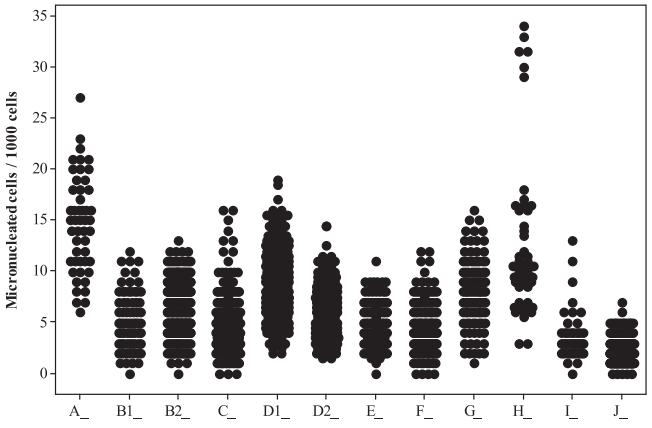

Fig. 6. Distribution of individual values of the 8 participating laboratories.

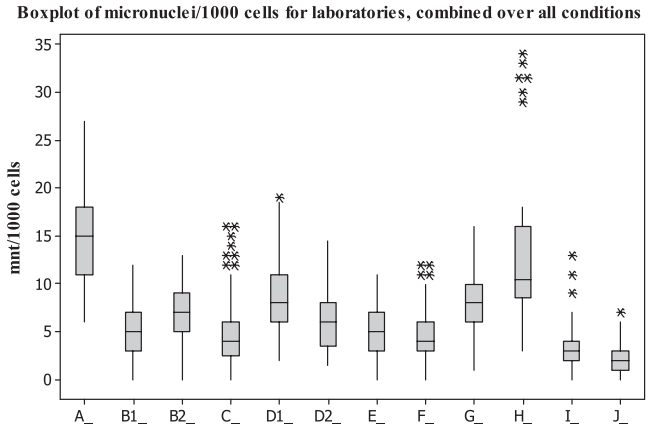

Fig. 7. Box plots of the distribution of the micronuclei counts from the 8 participating laboratories.

appreciable and statistically significant between experiment variability while Labs D and I showed borderline significance. This between experiment variability can also be seen in the QC charts for these laboratories (See Supplementary material).

One factor that did seem to differ between laboratories and could be an important source of variability was the method of donor selection for the lymphocytes. The laboratories' descriptions of the donor criteria were rather general and imprecise. It was not clear how many donors were in the pooled samples, and details of their sexes, ages and social habits were not given. The specific source of the lymphocytes used for a particular individual study was also unclear.

The exception was the laboratory with two independent sites and

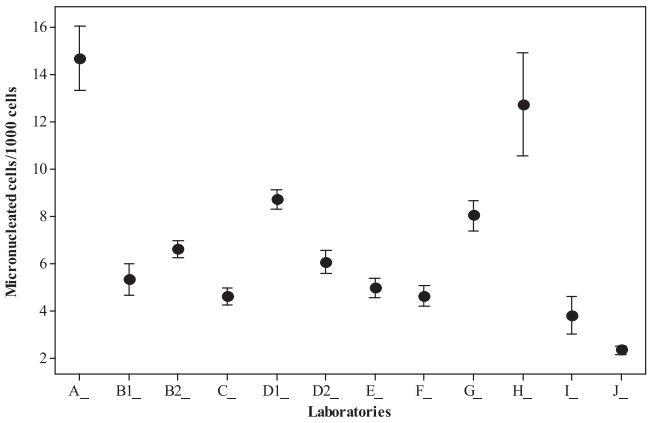

Fig. 8. Mean and 95% CI for the mean for the micronuclei counts from the 8 participating laboratories.

Table 2
Results for Lab D using manual and image analysis.

Combinatio	n	Mean	SD
Manual Analysis			
D1S9-Sm	36	9.58	3.45
D1S9 + Sm	94	9.13	3.32
D1S9-Lm	117	8.11	3.26
Image Analysis			
D2S9-Sm	39	5.46	2.52
D2S9 + Sm	43	5.64	2.75
D2S9-Lm	44	6.95	2.94

Table 3
Results from Lab J using different vehicles /solvents.

Solvent	n	Mean	SD
DMSO	120	2.29	1.13
Ethanol	20	1.55	0.69
Untreated	20	2.00	0.86
Saline	6	3.67	1.03
MF	12	1.83	0.84
Water	30	2.83	1.88

series of studies done using male and female donors. In this case there seemed to be only a slight difference between the experiments using male donors in the two series but in the series using female donors the counts from the US laboratory were appreciably higher (about 3/1000). There may, of course, be many other factors and confounders in these studies which could explain this difference.

4. Discussion

The results here will be useful in the assessment of the quality of negative control data for both an assessment of the competence of a laboratory to perform the assay and as a help to those who have to interpret the results of studies based upon the OECD Guidelines.

This study showed appreciable between laboratory variability ranging from 2.2/1000 to 15.9 /1000; approximately a 7.5-fold difference. This is a wider range than was found in the survey of the TK cell line where the range was from 3.5/1000 to 13.8/1000 (a fold-difference of just less than 4) and there were also more participating laboratories in the TK cell investigation [3]. This increased variability could reflect a greater inherent variability seen with human donor cells than may be expected from use of a well-defined cell line.

A number of laboratories had appreciable within laboratory (or between experiment) variability which was seen both with the results of

an anova and by examining the relevant QC plots. However, the variability seemed clustered around a specific background level which was associated with that laboratory rather than covering the full range across all the laboratories. This was so even though some of the QC plots for some of the laboratories show variability with some experiments falling outside the control limits based upon QC acceptance criteria (see Supplementary material).

From the single laboratory from which data were available, an appreciable difference was seen between manual scoring and the use of image analysis. The majority of samples were scored by manual techniques. There was not enough data to assess whether the use of automated scoring methods would improve the scoring by reducing potential subjective influences and through the opportunity to increase the number of cells that could be scored.

There were statistically significant differences between micronucleus frequencies between studies using different solvents but the sample sizes and the effects were small. In neither case could these be considered strong evidence for the effect of these factors. There was no clear evidence that other factors such as presence or absence of S9 or length of treatment contributed appreciably to the observed variability. The OECD Guidelines state that CytoB should be used as a cytokinesis blocker in human lymphocytes studies to account for variability in cell cycle times and because PHA stimulation will not occur amongst all lymphocytes. All the laboratories stated that they had used CytoB.

As mentioned earlier one potential source of variability is the donors. Inter-individual variability has been investigated in samples processed directly after been taken from individuals in the HUMN project [8] (rather than, as here, after culturing for use in the *in vitro* assay). The HUMN project identified considerable inter-individual variability in baseline micronucleus frequencies and identified a number of factors which influenced these frequencies. Kirsch-Volders et al. [5] also pointed to the increased variability in the assay because of the variability introduced using donors. The results from the HUMN project support the possibility that variability between donors could be one of the important factors contributing to inter-laboratory variability.

The OECD Guidelines suggest that lymphocytes should be obtained from young(ish) donors (approx. 16–35 years old) with no known illnesses or recent exposures to genotoxic agents and that details regarding the number of donors should be given if cells are pooled. In this study, a direct comparison between male and female donors was only possible for samples from the UK and US laboratory of one participating organization. In that organization, there was a pronounced difference in micronuclei between the scores with female donors from US studies compared with UK studies. Again, this represents just one study but it does suggest that the donor may be an important factor and more specific information on the characteristics of the donor are needed to identify if inter-donor variability is a major explanatory factor to the variability between and within laboratories. The investigation of between donor variability should be a topic for further research.

There were an appreciable number (2.1%) of zero counts which was probably because the number of cells counted per replicate was 1000 and the background level was low. This provides some support for the OECD Guidance recommendation to increase the number of cells counted per replicate to 2000. Although not a major constraint to the interpretation of the data with low background levels it is difficult to assess (for instance, using QC plots) whether zero scores are representative of an unusual (and possibly unacceptable) result in an experiment. There is some suggestion that the human lymphocyte assay is expected to have low frequencies [5] which, if generally so, could have an implication on the number of cells to score per replicate.

There are only a small number of published sets of negative control data for the micronucleus assay using human lymphocytes. One source, however, is the SFTG papers [9,10]. As part of this study nine coded substances were assessed by "…11 laboratories using human lymphocytes" [10]. The authors noted that "… the donor variability, inherent to primary cultures of human lymphocytes, added to the inter-laboratory variability, may have affected the power of the experimental system."

Lorge et al. [9] reported that "(I)n the presence of cytochalasin B, the lowest backgrounds of micronucleated cells were found in human lymphocytes in primary cultures (approx. 7–9 per 1000 cells)." Table 4 of their paper reported the following incidences for the three different treatment regimens: 7.4 ± 7.5 (3 + 26), 7.1 ± 5.3 (3 + 45) and 6.9 ± 4.9 (20 + 28). (Treatment hours + recovery hours in brackets).

Reanalysis of the data from the 99 samples in Appendix A of Clare et al. [10] gave slightly different values and showed that there was appreciable variability with individual culture values ranging values from 0.3 to 32.6/1000 cells. There were significant differences in the negative control values between the laboratories carrying out the assay ($P = 0.001$) which 'explained' about 24% of the total variability. The laboratory means ranged from 0.3 to 13.8 mnt/1000 cells. There were also significant lower counts in the cultures using 'Medium' compared with the other two solvents (ethanol and DMSO). Distinguishing the factors that affect the scores is complicated by the confounding of variability between the individual donors, the different solvents and any other factors that may differ in the conduct of the assays by the 11 participating laboratories. The variability found in this study could be investigated by specific studies investigating the effect of different donors, solvents and other variables across a number of laboratories.

One surprising aspect of the study was the relatively small number of laboratories that responded to the GTTC's multiple calls for data. The laboratories that volunteered data were considered experienced and it was not considered that any major laboratory conducting this assay had been overlooked. However, it is pertinent to ask why there are so few laboratories and what the pool of expertise/experience means for the application of this assay. It may be a reflection that whilst the assay has been developed over several decades, international acceptance as a regulatory compliant assay, with its own OECD guidance, is relatively new. Hence, we might expect the adoption of the assay by more laboratories in the future; alternatively, it may reflect a more general shift to the use of the TK6 assay.

An important consideration is whether the range of values obtained here provides an acceptable range of negative control data. The range is clearly wide especially given the relative small number of laboratories who were known experts. Clearly such a range raises questions over what would be an acceptable biologically important effect. A doubling effect would clearly be very different if it was from 15 to 30/1000 compared with 2.5 to 5.0/1000. The finding that while a laboratory may show appreciable intra-laboratory (or between experiment) variability that this is relative small compared to the range across laboratories, suggests that, provided a laboratory can show consistency in the QC chart, the position of its mean value in the between laboratory range is less critical. It might, however, be of concern if a new laboratory's mean value fell outside the 2–15 range. Again, further consideration of whether inter-donor variability is an important factor in the study characteristics might be of importance in reducing variability in the results.

In conclusion, the results of this study have identified the ranges of human lymphocyte solvent control micronuclei data collected from a group of experienced laboratories. These can provide guidance for other laboratories developing historical control databases. The findings suggest that, although there is variability between laboratories, the intra-laboratory means remain stable providing evidence to justify the OECD guidelines on optimal cell numbers for scoring. The analyses carried out also provide possible approaches for assessing factors influencing variability.

Acknowledgments

The authors wish to thank the participating laboratories and their staff who provided data. The authors also thank the HESI GTTC and the Data Interpretation Workgroup for intellectual and financial support.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.mrgentox.2018.08.009.

References

[1] OECD, Test No. 487: In Vitro Mammalian Cell Micronucleus Test, OECD Guidelines for Testing of Chemicals, Section 4, (2014).
[2] OECD, Overview of the Set of OECD Genetic Toxicology Test Guidelines and Updates Performed in 2014-2015 Series on Testing & Assessment No. 238, OECD Publishing, Paris, 2016http://www.oecd-ilibrary.org/docserver/download/9717131e.pdf?expires=1520263341&id=id&accname=guest&checksum=1CBCB4880A409F8F46AD7994AFDB13EE.
[3] D.P. Lovell, M. Fellows, F. Marchetti, J. Christiansen, A. Elhajouji, K. Hashimoto, S. Kasamoto, Y. Li, O. Masayasu, M.M. Moore, M. Schuler, R. Smith, L.F. Stankowski Jr., J. Tanaka, J.Y. Tanir, V. Thybaud, F. Van Goethem, J. Whitwell, Analysis of negative historical control group data from the *in vitro* micronucleus assay using TK6 cells, Mutat. Res. Toxicol. Environ. Mutagen. 825 (2018) 40–50.
[4] S. Bonassi, M. Fenech, C. Lando, Y.P. Lin, M. Ceppi, W.P. Chang, N. Holland, M. Kirsch-Volders, E. Zeiger, S. Ban, R. Barale, M.P. Bigatti, C. Bolognesi, C. Jia, M. Di Giorgio, L.R. Ferguson, A. Fucic, O.G. Lima, P. Hrelia, A.P. Krishnaja, T.K. Lee, L. Migliore, L. Mikhalevich, E. Mirkova, P. Mosesso, W.U. Muller, Y. Odagiri, M.R. Scarffi, E. Szabova, I. Vorobtsova, A. Vral, A. Zijno, HUman MicroNucleus Project: international database comparison for results with the cytokinesis-block micronucleus assay in human lymphocytes, I. Effect of laboratory protocol, scoring criteria and host factors on the frequency of micronuclei, Environ. Mol. Mutagen. 37 (2001) 31–45.
[5] M. Kirsch-Volders, I. Decordier, A. Elhajouji, G. Plas, M.J. Aardema, M. Fenech, In vitro genotoxicity testing using the micronucleus assay in cell lines, human lymphocytes and 3D human skin models, Mutagenesis 26 (2011) 177–184.
[6] M. Fenech, Cytokinesis-block micronucleus cytome assay, Nat. Protoc. 2 (2007) 1084–1104.
[7] R Development Core Team, R: a Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2011.
[8] M. Fenech, N. Holland, E. Zeiger, W.P. Chang, S. Burgaz, P. Thomas, C. Bolognesi, S. Knasmueller, M. Kirsch-Volders, S. Bonassi, The HUMN and HUMNxL international collaboration projects on human micronucleus assays in lymphocytes and buccal cells—past, present and future, Mutagenesis 26 (2011) 239–245.
[9] E. Lorge, V. Thybaud, M. Aardema, J. Oliver, A. Wakata, G. Lorenzon, D. Marzin, SFTG International Collaborative Study on in vitro micronucleus test. Part I. General conditions of the study, Mutat. Res. 607 (2006) 37–60.
[10] M.G. Clare, G. Lorenzon, L.C. Akhurst, D. Marzin, J. van Delft, R. Montero, A. Botta, A. Bertens, S. Cinelli, V. Thybaud, E. Lorge, SFTG international collaborative study on in vitro micronucleus test. II. Using human lymphocytes, Mutat. Res. 607 (2006) 37–60.