OECD/OCDE 475
Adopted:
29 July 2016

OECD GUIDELINE FOR THE TESTING OF CHEMICALS

Mammalian Bone Marrow Chromosomal Aberration Test

INTRODUCTION

1. The OECD Guidelines for the Testing of Chemicals are periodically reviewed in the light of scientific progress, changing regulatory needs and animal welfare considerations. The original Test Guideline 475 was adopted in 1984. In 1997, a revised version was adopted, based on scientific progress made to that date. This modified version of the Test Guideline reflects scientific knowledge from more than thirty years of experience with this assay and the interpretation of the data. This Test Guideline is part of a series of Test Guidelines on genetic toxicology. A document that provides succinct information on genetic toxicology testing and an overview of the recent changes that were made to these Test Guidelines has been developed (1).

2. The mammalian *in vivo* bone marrow chromosomal aberration test is especially relevant for assessing genotoxicity because, although they may vary among species, factors of *in vivo* metabolism, pharmacokinetics and DNA-repair processes are active and contribute to the responses. An *in vivo* assay is also useful for further investigation of genotoxicity detected by an *in vitro* system.

3. The mammalian *in vivo* chromosomal aberration test is used for the detection of structural chromosome aberrations induced by test chemicals in bone marrow cells of animals, usually rodents (2) (3) (4) (5). Structural chromosomal aberrations may be of two types, chromosome or chromatid. While the majority of genotoxic chemical-induced aberrations are of the chromatid-type, chromosome-type aberrations also occur. Chromosomal damage and related events are the cause of many human genetic diseases and there is substantial evidence that, when these lesions and related events cause alterations in oncogenes and tumour suppressor genes, they are involved in cancer in humans and experimental systems. Polyploidy (including endoreduplication) could arise in chromosome aberration assays *in vivo*. However, an increase in polyploidy per se does not indicate aneugenic potential and can simply indicate cell cycle perturbation or cytotoxicity. This test is not designed to measure aneuploidy. An *in vivo* mammalian erythrocyte micronucleus test (Test Guideline 474) or the *in vitro* mammalian cell micronucleus test (Test Guideline 487) would be the *in vivo* and *in vitro* tests, respectively, recommended for the detection of aneuploidy.

4. Definitions of terminology used are set out in Annex 1.

This Guideline was adopted by the OECD Council by written procedure on 29 July 2016 [C(2016)103].

INITIAL CONSIDERATIONS

5. Rodents are routinely used in this test, but other species may in some cases be appropriate if scientifically justified. Bone marrow is the target tissue in this test since it is a highly vascularised tissue and it contains a population of rapidly cycling cells that can be readily isolated and processed. The scientific justification for using species other than rats and mice should be provided in the report. If species other than rodents are used, it is recommended that the measurement of bone marrow chromosomal aberration be integrated into another appropriate toxicity test.

6. If there is evidence that the test substance(s), or its metabolite(s), will not reach the target tissue, it may not be appropriate to use this test.

7. Before use of the Test Guideline on a mixture for generating data for an intended regulatory purpose, it should be considered whether, and if so why, it may provide adequate results for that purpose. Such considerations are not needed, when there is a regulatory requirement for testing of the mixture.

PRINCIPLE OF THE TEST METHOD

8. Animals are exposed to the test chemical by an appropriate route of exposure and are humanely euthanised at an appropriate time after treatment. Prior to euthanasia, animals are treated with a metaphase-arresting agent (e.g. colchicine or Colcemid®). Chromosome preparations are then made from the bone marrow cells and stained, and metaphase cells are analysed for chromosomal aberrations.

VERIFICATION OF LABORATORY PROFICIENCY

Proficiency Investigations

9. In order to establish sufficient experience with the conduct of the assay prior to using it for routine testing, the laboratory should have demonstrated the ability to reproduce expected results from published data (e.g. (6)) for chromosomal aberration frequencies with a minimum of two positive control substances (including weak responses induced by low doses of positive controls), such as those listed in Table 1 and with compatible vehicle/solvent controls (see paragraph 22). These experiments should use doses that give reproducible and dose related increases and demonstrate the sensitivity and dynamic range of the test system in the tissue of interest (bone marrow) and using the scoring method to be employed within the laboratory. This requirement is not applicable to laboratories that have experience, i.e. that have a historical database available as defined in paragraphs 10-14.

Historical Control Data

10. During the course of the proficiency investigations, the laboratory should establish:

- A historical positive control range and distribution, and
- A historical negative control range and distribution.

11. When first acquiring data for a historical negative control distribution, concurrent negative controls should be consistent with published control data, where they exist. As more experimental data are added to the historical control distribution, concurrent negative controls should ideally be within the 95% control

limits of that distribution. The laboratory's historical negative control database should be statistically robust to ensure the ability of the laboratory to assess the distribution of their negative control data. The literature suggests that a minimum of 10 experiments may be necessary but would preferably consist of at least 20 experiments conducted under comparable experimental conditions. Laboratories should use quality control methods, such as control charts (e.g. C-charts or X-bar charts (7)), to identify how variable their data are, and to show that the methodology is 'under control' in their laboratory. Further recommendations on how to build and use the historical data (i.e. criteria for inclusion and exclusion of data in historical data and the acceptability criteria for a given experiment) can be found in the literature (8).

12. Where the laboratory does not complete a sufficient number of experiments to establish a statistically robust negative control distribution (see paragraph 11) during the proficiency investigations (described in paragraph 9), it is acceptable that the distribution can be built during the first routine tests. This approach should follow the recommendations set out in the literature (8) and the negative control results obtained in these experiments should remain consistent with published negative control data.

13. Any changes to the experimental protocol should be considered in terms of their impact on the resulting data remaining consistent with the laboratory's existing historical control database. Only major inconsistencies should result in the establishment of a new historical control database, where expert judgement determines that it differs from the previous distribution (see paragraph 11). During the re-establishment, a full negative control database may not be needed to permit the conduct of an actual test, provided that the laboratory can demonstrate that their concurrent negative control values remain either consistent with their previous database or with the corresponding published data.

14. Negative control data should consist of the incidence of structural chromosomal aberration (excluding gaps) in each animal. Concurrent negative controls should ideally be within the 95% control limits of the distribution of the laboratory's historical negative control database. Where concurrent negative control data fall outside the 95% control limits, they may be acceptable for inclusion in the historical control distribution as long as these data are not extreme outliers and there is evidence that the test system is 'under control' (see paragraph 11) and no evidence of technical or human failure.

DESCRIPTION OF THE METHOD

Preparations

Selection of animal species

15. Commonly used laboratory strains of healthy young adult animals should be employed. Rats are commonly used, although mice may also be appropriate. Any other appropriate mammalian species may be used, if scientific justification is provided in the report.

Animal housing and feeding conditions

16. For rodents, the temperature in the animal room should be 22°C (±3°C). Although the relative humidity ideally should be 50-60%, it should be at least 40% and preferably not exceed 70% other than during room cleaning. Lighting should be artificial, the sequence being 12 hours light, 12 hours dark. For feeding, conventional laboratory diets may be used with an unlimited supply of drinking water. The choice of diet may be influenced by the need to ensure a suitable admixture of a test chemical when administered

by this route. Rodents should be housed in small groups (no more than five per cage) of the same sex and treatment group if no aggressive behaviour is expected, preferably in solid floor cages with appropriate environmental enrichment. Animals may be housed individually only if scientifically justified.

Preparation of the animals

17. Healthy young adult animals (for rodents, ideally 6-10 weeks old at start of treatment, though slightly older animals are also acceptable) are normally used, and are randomly assigned to the control and treatment groups. The individual animals are identified uniquely using a humane, minimally invasive method (e.g. by ringing, tagging, micro-chipping or biometric identification, but not ear or toe clipping) and acclimated to the laboratory conditions for at least five days. Cages should be arranged in such a way that possible effects due to cage placement are minimised. Cross contamination by the positive control and the test chemical should be avoided. At the commencement of the study, the weight variation of animals should be minimal and not exceed ± 20% of the mean weight of each sex.

Preparation of doses

18. Solid test chemicals should be dissolved or suspended in appropriate solvents or vehicles or admixed in diet or drinking water prior to dosing the animals. Liquid test chemicals may be dosed directly or diluted prior to dosing. For inhalation exposures, test chemicals can be administered as a gas, vapour, or a solid/liquid aerosol, depending on their physicochemical properties. Fresh preparations of the test chemical should be employed unless stability data demonstrate the acceptability of storage and define the appropriate storage conditions.

Solvent/vehicle

19. The solvent/vehicle should not produce toxic effects at the dose levels used, and should not be suspected of chemical reaction with the test chemicals. If other than well-known solvents/vehicles are used, their inclusion should be supported with reference data indicating their compatibility. It is recommended that wherever possible, the use of an aqueous solvent/vehicle should be considered first. Examples of commonly used compatible solvents/vehicles include water, physiological saline, methylcellulose solution, carboxymethyl cellulose sodium salt solution, olive oil and corn oil. In the absence of historical or published control data showing that no structural aberrations or other deleterious effects are induced by a chosen atypical solvent/vehicle, an initial study should be conducted in order to establish the acceptability of the solvent/vehicle control.

Controls

Positive controls

20. A group of animals treated with a positive control substance should normally be included with each test. This may be waived when the testing laboratory has demonstrated proficiency in the conduct of the test and has established a historical positive control range. When a concurrent positive control group is not included, scoring controls (fixed and unstained slides) should be included in each experiment. These can be obtained by including within the scoring of the study appropriate reference samples that have been obtained and stored from a separate positive control experiment conducted periodically (e.g. every 6-18 months) in the laboratory where the test is performed; for example, during proficiency testing and on a regular basis thereafter, where necessary.

21. Positive control substances should reliably produce a detectable increase in the frequency of cells with structural chromosomal aberrations over the spontaneous level. Positive control doses should be chosen so that the effects are clear but do not immediately reveal the identity of the coded samples to the scorer. It is acceptable that the positive control be administered by a route different from the test chemical, using a different treatment schedule, and for sampling to occur only at a single time point. In addition, the use of chemical class-related positive control substances may be considered, when appropriate. Examples of positive control substances are included in Table 1.

Table 1. Examples of positive control substances

Substance and CASRN
Ethyl methanesulphonate [CASRN 62-50-0]
Methyl methanesulphonate [CASRN 66-27-3]
Ethyl nitrosourea [CASRN 759-73-9]
Mitomycin C [CASRN 50-07-7]
Cyclophosphamide (monohydrate) [CASRN 50-18-0; (CASRN 6055-19-2)]
Triethylenemelamine [CASRN 51-18-3]

Negative controls

22. Negative control group animals should be included at every sampling time and otherwise handled in the same way as the treatment groups, except for not receiving treatment with the test chemical. If a solvent/vehicle is used in administering the test chemical, the control group should receive this solvent/vehicle. However, if consistent inter-animal variability and frequencies of cells with structural aberrations are demonstrated by historical negative control data at each sampling time for the testing laboratory, only a single sampling for the negative control may be necessary. Where a single sampling is used for negative controls, it should be the first sampling time used in the study.

PROCEDURE

Number and sex of animals

23. In general, the micronucleus response is similar between male and female animals (9) and it is expected that this will be true also for structural chromosomal aberrations; therefore, most studies could be performed in either sex. Data demonstrating relevant differences between males and females (e.g. differences in systemic toxicity, metabolism, bioavailability, bone marrow toxicity, etc. including e.g. a range-finding study) would encourage the use of both sexes. In this case, it may be appropriate to perform a study in both sexes, e.g. as part of a repeated dose toxicity study. It might be appropriate to use the factorial design in case both sexes are used. Details on how to analyse the data using this design are given in Annex 2.

24. Group sizes at study initiation should be established with the aim of providing a minimum of 5 analysable animals of one sex, or of each sex if both are used, per group. Where human exposure to chemicals may be sex-specific, as for example with some pharmaceuticals, the test should be performed with the appropriate sex. As a guide to maximum typical animal requirements, a study in bone marrow at two sampling times with three dose groups and a concurrent negative control group, plus a positive control group (each group composed of five animals of a single sex), would require 45 animals.

Dose levels

25. If a preliminary range-finding study is performed because there are no suitable data already available to aid in dose selection, it should be performed in the same laboratory, using the same species, strain, sex, and treatment regimen to be used in the main study (10). The study should aim to identify the maximum tolerated dose (MTD), defined as the highest dose that will be tolerated without evidence of study-limiting toxicity, relative to the duration of the study period (for example, by inducing body weight depression or hematopoietic system cytotoxicity), but not death or evidence of pain, suffering or distress necessitating humane euthanasia (11).

26. The highest dose may also be defined as a dose that produces some indication of toxicity to the bone marrow.

27. Substances that exhibit saturation of toxicokinetic properties, or induce detoxification processes that may lead to a decrease in exposure after long-term treatment may be exceptions to the dose-setting criteria and should be evaluated on a case-by-case basis.

28. In order to obtain dose response information, a complete study should include a negative control group and a minimum of three dose levels generally separated by a factor of 2, but by no greater than 4. If the test chemical does not produce toxicity in a range-finding study or based on existing data, the highest dose for a single administration should be 2000 mg/kg body weight. However, if the test chemical does cause toxicity, the MTD should be the highest dose administered and the dose levels used should preferably cover a range from the maximum to a dose producing little or no toxicity. When target tissue (bone marrow) toxicity is observed at all dose levels tested, further study at non-toxic doses is advisable. Studies intending to more fully characterise the quantitative dose-response information may require additional dose groups. For certain types of test chemicals (e.g. human pharmaceuticals) covered by specific requirements, these limits may vary.

Limit test

29. If dose range-finding experiments, or existing data from related animal strains, indicate that a treatment regime of at least the limit dose (described below) produces no observable toxic effects, (including no depression of bone marrow proliferation or other evidence of target tissue cytotoxicity), and if genotoxicity would not be expected based upon *in vitro* genotoxicity studies or data from structurally related substances, then a full study using three dose levels may not be considered necessary, provided it has been demonstrated that the test chemical(s) reach(es) the target tissue (bone marrow). In such cases, a single dose level, at the limit dose, may be sufficient. For an administration period of >14 days, the limit dose is 1000 mg/kg body weight/day. For administration periods of 14 days or less, the limit dose is 2000 mg/kg/body weight/day.

Administration of doses

30. The anticipated route of human exposure should be considered when designing an assay. Therefore, routes of exposure such as dietary, drinking water, topical subcutaneous, intravenous, oral (by gavage), inhalation, intratracheal, or implantation may be chosen as justified. In any case, the route should be chosen to ensure adequate exposure of the target tissue(s). Intraperitoneal injection is generally not recommended since it is not an intended route of human exposure, and should only be used with specific scientific justification. If the test chemical is admixed in diet or drinking water, especially in case of single

dosing, care should be taken that the delay between food and water consumption and sampling should be sufficient to allow detection of the effects (see paragraphs 33-34). The maximum volume of liquid that can be administered by gavage or injection at one time depends on the size of the test animal. The volume should not normally exceed 1 mL/100 g body weight except in the case of aqueous solutions where a maximum of 2 mL/100 g may be used. The use of volumes greater than this should be justified. Except for irritating or corrosive test chemicals, which will normally produce exacerbated effects at higher concentrations, variability in test volume should be minimised by adjusting the concentration to ensure administration of a constant volume in relation to body weight at all dose levels.

Treatment schedule

31. Test chemicals are normally administered as a single treatment, but may be administered as a split dose (i.e. two or more treatments on the same day separated by no more than 2-3 hours) to facilitate administering a large volume. Under these circumstances, or when administering the test chemical by inhalation, the sampling time should be scheduled based on the time of the last dosing or the end of exposure.

32. There are little data available on the suitability of a repeated-dose protocol for this test. However, in circumstances where it is desirable to integrate this test with a repeated-dose toxicity test, care should be taken to avoid loss of chromosomally damaged mitotic cells as may occur with toxic doses. Such integration is acceptable when the highest dose is greater or equal to the limit dose (see paragraph 29) and a dose group is administered the limit dose for the duration of the treatment period. The micronucleus test (Test Guideline 474) should be viewed as the *in vivo* test of choice for chromosomal aberrations when integration with other studies is desired.

33. Bone marrow samples should be taken at two separate times following single treatments. For rodents, the first sampling interval should be the time necessary to complete 1.5 normal cell cycle lengths (the latter being normally 12-18 hours following the treatment period). Since the time required for uptake and metabolism of the test chemical(s) as well as its effect on cell cycle kinetics can affect the optimum time for chromosomal aberration detection, a later sample collection 24 hours after the first sampling time is recommended. At the first sampling time, all dose groups should be treated and samples collected for analysis; however, at the later sampling time(s), only the highest dose needs to be administered. If dose regimens of more than one day are used based on scientific justification, one sampling time at up to approximately 1.5 normal cell cycle lengths after the final treatment should generally be used.

34. Following treatment and prior to sample collection, animals are injected intraperitoneally with an appropriate dose of a metaphase-arresting agent (e.g. Colcemid® or colchicine), and samples are collected at an appropriate interval thereafter. For mice this interval is approximately 3-5 hours prior to collection and for rats it is 2-5 hours. Cells are harvested from the bone marrow, swollen, fixed and stained, and analysed for chromosomal aberrations (12).

Observations

35. General clinical observations of the test animals should be made and clinical signs recorded at least once a day, preferably at the same time(s) each day and considering the peak period of anticipated effects after dosing. At least twice daily during the dosing period, all animals should be observed for morbidity and mortality. All animals should be weighed at study initiation, at least once a week during repeated-dose studies, and at euthanasia. In studies of at least one-week duration, measurements of food consumption should be made at least weekly. If the test chemical is administered via the drinking water, water

consumption should be measured at each change of water and at least weekly. Animals exhibiting non-lethal indicators of excessive toxicity should be humanely euthanised prior to completion of the test period (11).

Target tissue exposure

36. A blood sample should be taken at appropriate time(s) in order to permit investigation of the plasma levels of the test chemicals for the purposes of demonstrating that exposure of the bone marrow occurred, where warranted and where other exposure data do not exist (see paragraph 44).

Bone marrow and chromosome preparations

37. Immediately after humane euthanasia, bone marrow cells are obtained from the femurs or tibias of the animals, exposed to hypotonic solution and fixed. The metaphase cells are then spread on slides and stained using established methods (see (3) (12)).

Analysis

38. All slides, including those of positive and negative controls, should be independently coded before analysis and should be randomised so the scorer is unaware of the treatment condition.

39. The mitotic index should be determined as a measure of cytotoxicity in at least 1000 cells per animal for all treated animals (including positive controls), untreated or vehicle/solvent negative control animals.

40. At least 200 metaphases should be analysed for each animal for structural chromosomal aberrations including and excluding gaps (6). However, if the historical negative control database indicates the mean background structural chromosomal aberration frequency is <1% in the testing laboratory, consideration should be given to scoring additional cells. Chromatid and chromosome-type aberrations should be recorded separately and classified by sub-types (breaks, exchanges). Procedures in use in the laboratory should ensure that analysis of chromosomal aberrations is performed by well-trained scorers and peer-reviewed if appropriate. Recognising that slide preparation procedures often result in the breakage of a proportion of metaphases with a resulting loss of chromosomes, the cells scored should, therefore, contain a number of centromeres not less than $2n\pm2$, where n is the haploid number of chromosomes for that species.

DATA AND REPORTING

Treatment of Results

41. Individual animal data should be presented in tabular form. The mitotic index, the number of metaphase cells scored, the number of aberrations per metaphase cell and the percentage of cells with structural chromosomal aberration(s) should be evaluated for each animal. Different types of structural chromosomal aberrations should be listed with their numbers and frequencies for treated and control groups. Gaps, as well as polyploid cells and cells with endoreduplicated chromosomes are recorded separately. The frequency of gaps is reported but generally not included in the analysis of the total structural aberration frequency. If there is no evidence for a difference in response between the sexes, the data may be combined for statistical analysis. Data on animal toxicity and clinical signs should also be reported.

Acceptability Criteria

42. The following criteria determine the acceptability of the test:

a) The concurrent negative control data are considered acceptable for addition to the laboratory historical control database (see paragraphs 11-14);

b) The concurrent positive controls or scoring controls should induce responses that are compatible with those generated in the historical positive control database and produce a statistically significant increase compared with the negative control (see paragraphs 20-21);

c) The appropriate number of doses and cells has been analysed;

d) The criteria for the selection of highest dose are consistent with those described in paragraphs 25-28.

Evaluation and Interpretation of Results

43. Providing that all acceptability criteria are fulfilled, a test chemical is considered clearly positive if:

a) At least one of the treatment groups exhibits a statistically significant increase in the frequency of cells with structural chromosomal aberrations (excluding gaps) compared with the concurrent negative control,

b) This increase is dose-related at least at one sampling time when evaluated with an appropriate trend test, and

c) Any of these results are outside the distribution of the historical negative control data (e.g. Poisson-based 95% control limits).

If only the highest dose is examined at a particular sampling time, a test chemical is considered clearly positive if there is a statistically significant increase compared with the concurrent negative control and the results are outside the distribution of the historical negative control data (e.g. Poisson-based 95% control limits). Recommendations for appropriate statistical methods can be found in the literature (13). When conducting a dose-response analysis, at least three treated dose groups should be analysed. Statistical tests should use the animal as the experimental unit. Positive results in the chromosomal aberration test indicate that a test chemical induces structural chromosomal aberrations in the bone marrow of the species tested.

44. Providing that all acceptability criteria are fulfilled, a test chemical is considered clearly negative if in all experimental conditions examined:

a) None of the treatment groups exhibits a statistically significant increase in the frequency of cells with structural chromosomal aberrations (excluding gaps) compared with the concurrent negative control,

b) There is no dose-related increase at any sampling time when evaluated by an appropriate trend test,

c) All results are inside the distribution of the historical negative control data (e.g. Poisson-based 95% control limits), and

d) Bone marrow exposure to the test substance(s) occurred.

Recommendations for the most appropriate statistical methods can be found in the literature (13). Evidence of exposure to the bone marrow to a test substance may include a depression of the mitotic index or measurement of the plasma or blood levels of the test substance(s). In the case of intravenous administration, evidence of exposure is not needed. Alternatively, ADME data, obtained in an independent study using the same route and same species can be used to demonstrate bone marrow exposure. Negative results indicate that, under the test conditions, the test chemical does not induce structural chromosomal aberrations in the bone marrow of the species tested.

45. There is no requirement for verification of a clear positive or clear negative response.

46. In cases where the response is not clearly negative or positive and in order to assist in establishing the biological relevance of a result (e.g. a weak or borderline increase), the data should be evaluated by expert judgement and/or further investigations of the existing experiments completed. In some cases, analysing more cells or performing a repeat experiment using modified experimental conditions could be useful.

47. In rare cases, even after further investigations, the data will preclude making a conclusion that the test chemical produces either positive or negative results, and the study will therefore be concluded as equivocal.

48. The frequencies of polyploid and endoreduplicated metaphases among total metaphases should be recorded separately. An increase in the number of polyploid/endoreduplicated cells may indicate that the test chemical has the potential to inhibit mitotic processes or cell cycle progression (see paragraph 3).

Test Report

49. The test report should include the following information:

Summary

Test chemical:

- source, lot number, limit date for use if available;
- stability of the test chemical, if known.

Mono-constituent substance:

- physical appearance, water solubility, and additional relevant physicochemical properties;
- chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, purity, chemical identity of impurities as appropriate and practically feasible, etc.

Multi-constituent substance, UVBCs and mixtures:

- characterised as far as possible by chemical identity (see above), quantitative occurrence and relevant physicochemical properties of the constituents.

Test chemical preparation:

- justification for choice of vehicle;

- solubility and stability of the test chemical in solvent/vehicle, if known;
- preparation of dietary, drinking water or inhalation formulations;
- analytical determinations on formulations (e.g. stability, homogeneity, nominal concentrations), when conducted.

Test animals:

- species/strain used and justification for use;
- number, age and sex of animals;
- source, housing conditions, diet, etc.;
- method for uniquely identifying the animals;
- for short-term studies: individual weight of the animals at the start and end of the test; for studies longer than one week: individual body weights during the study and food consumption. Body weight range, mean and standard deviation for each group should be included.

Test conditions:

- positive and negative (vehicle/solvent) controls;
- data from range-finding study, if conducted;
- rationale for dose level selection;
- details of test chemical preparation;
- details of the administration of the test chemical;
- rationale for route and duration of administration;
- methods for verifying that the test substance(s) reached the general circulation or bone marrow;
- actual dose (mg/kg body weight/day) calculated from diet/drinking water test chemical concentration (ppm) and consumption, if applicable;
- details of food and water quality;
- method of euthanasia;
- method of analgesia (where used);
- detailed description of treatment and sampling schedules and justifications for the choices;
- methods of slide preparation;
- methods for measurement of toxicity;
- identity of metaphase arresting chemical, its concentration, dose and time of administration before sampling;
- procedures for isolating and preserving samples;
- criteria for scoring aberrations;
- number of metaphase cells analysed per animal and the number of cells analysed for mitotic index determination;
- criteria for acceptability of the study;
- criteria for considering studies as positive, negative or inconclusive.

Results:

- animal condition prior to and throughout the test period, including signs of toxicity;
- mitotic index, given separately for each animal;

- type and number of aberrations and of aberrant cells, given separately for each animal;
- total number of aberrations per group with means and standard deviations;
- number of cells with aberrations per group with means and standard deviations;
- changes in ploidy, if seen, including frequencies of polyploid and/or endoreduplicated cells;
- dose-response relationship, where possible;
- statistical analyses and method applied;
- data supporting that exposure of the bone marrow occurred;
- concurrent negative control and positive control data with ranges, means and standard deviations;
- historical negative and positive control data with ranges, means, standard deviations, and 95% control limits for the distribution, as well as the time period covered and number of observations;
- criteria met for a positive or negative response.

Discussion of the results.

Conclusion.

References.

LITERATURE

(1) OECD (2016). Overview of the set of OECD Genetic Toxicology Test Guidelines and updates performed in 2014-2015. ENV Publications. Series on Testing and Assessment, No. 234, OECD, Paris.

(2) Adler, I.D. (1984), "Cytogenetic Tests in Mammals", in *Mutagenicity Testing: A Practical Approach*, Venittand, S., J.M. Parry (eds.), IRL Press, Washington, DC, pp. 275-306.

(3) Preston, R.J. et al. (1987), Mammalian in vivo cytogenetic assays. Analysis of chromosome aberrations in bone marrow cells, *Mutation Research*, Vol. 189/2, pp. 157-165.

(4) Richold, M. et al. (1990), "*In Vivo* Cytogenetics Assays", in *Basic Mutagenicity Tests, UKEMS Recommended Procedures. UKEMS Subcommittee on Guidelines for Mutagenicity Testing. Report. Part I revised*, Kirkland, D.J. (ed.), Cambridge University Press, Cambridge, pp. 115-141.

(5) Tice, R.R. et al. (1994), Report from the working group on the in vivo mammalian bone marrow chromosomal aberration test, *Mutation Research*, Vol. 312/3, pp. 305-312.

(6) Adler, I.D. et al. (1998), Recommendations for statistical designs of in vivo mutagenicity tests with regard to subsequent statistical analysis, *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, Vol. 417/1, pp. 19-30.

(7) Ryan, T.P. (2000), *Statistical Methods for Quality Improvement*, 2nd ed., John Wiley and Sons, New York.

(8) Hayashi, M. et al. (2011), Compilation and use of genetic toxicity historical control data, *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, Vol. 723/2, pp. 87-90.

(9) Hayashi, M. et al. (1994), In vivo rodent erythrocyte micronucleus assay, *Mutation Research/Environmental Mutagenesis and Related Subjects*, Vol. 312/3, pp. 293-304.

(10) Fielder, R.J. et al. (1992), Report of British Toxicology Society/UK Environmental Mutagen Society Working Group. Dose setting in in vivo mutagenicity assays, *Mutagenesis*, Vol. 7/5, pp. 313-319.

(11) OECD (2000), "Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation", OECD Environment, Health and Safety Publications (EHS), Series on Testing and Assessment, N°19, OECD Publishing, Paris.

(12) Pacchierotti, F., V. Stocchi (2013), Analysis of chromosome aberrations in somatic and germ cells of the mouse, *Methods in Molecular Biology*, Vol. 1044, pp. 147-163.

(13) Lovell, D.P. et al. (1989), "Statistical Analysis of *In Vivo* Cytogenetic Assays", in *Statistical Evaluation of Mutagenicity Test Data. UKEMS SubCommittee on Guidelines for Mutagenicity Testing, Report, Part III*, Kirkland, D.J. (ed.), Cambridge University Press, Cambridge, pp. 184-232.

ANNEX 1

DEFINITIONS

Aneuploidy: Any deviation from the normal diploid (or haploid) number of chromosomes by one or more chromosomes, but not by multiples of entire set(s) of chromosomes (*cf.* polyploidy).

Centromere: Region(s) of a chromosome with which spindle fibers are associated during cell division, allowing orderly movement of daughter chromosomes to the poles of the daughter cells.

Chromatid-type aberration: Structural chromosome damage expressed as breakage of single chromatids or breakage and reunion between chromatids.

Chromosome-type aberration: Structural chromosome damage expressed as breakage, or breakage and reunion, of both chromatids at an identical site.

Endoreduplication: A process in which after an S period of DNA replication, the nucleus does not go into mitosis but starts another S period. The result is chromosomes with 4,8,16...chromatids.

Gap: An achromatic lesion smaller than the width of one chromatid, and with minimum misalignment of the chromatids.

Mitotic index: The ratio between the number of cells in mitosis and the total number of cells in a population, which is a measure of the proliferation status of that cell population.

Numerical aberration: A change in the number of chromosomes from the normal number characteristic of the animals utilised (aneuploidy).

Polyploidy: A numerical chromosomal aberration involving a change in the number of the entire set of chromosomes, as opposed to a numerical change in part of the chromosome set (*cf.* aneuploidy).

Structural chromosomal aberration: A change in chromosome structure detectable by microscopic examination of the metaphase stage of cell division, observed as deletions and fragments, intrachanges or interchanges.

ANNEX 2

THE FACTORIAL DESIGN FOR IDENTIFYING SEX DIFFERENCES IN THE *IN VIVO* CHROMOSOMAL ABERRATION ASSAY

The factorial design and its analysis

In this design, a minimum of 5 males and 5 females are tested at each concentration level resulting in a design using a minimum of 40 animals (20 males and 20 females, plus relevant positive controls).

The design, which is one of the simpler factorial designs, is equivalent to a two-way analysis of variance with sex and concentration level as the main effects. The data can be analysed using many standard statistical software packages such as SPSS, SAS, STATA, Genstat as well as using R.

The analysis partitions the variability in the dataset into that between the sexes, that between the concentrations and that related to the interaction between the sexes and the concentrations. Each of the terms is tested against an estimate of the variability between the replicate animals within the groups of animals of the same sex given the same concentration. Full details of the underlying methodology are available in many standard statistical textbooks (see references) and in the 'help' facilities provided with statistical packages.

The analysis proceeds by inspecting the sex x concentration interaction term in the ANOVA table[1]. In the absence of a significant interaction term the combined values across sexes or across concentration levels provide valid statistical tests between the levels based upon the pooled within group variability term of the ANOVA.

The analysis continues by partitioning the estimate of the between concentrations variability into contrasts which provide for a test for linear and quadratic contrasts of the responses across the concentration levels. When there is a significant sex x concentration interaction this term can also be partitioned into linear x sex and quadratic x sex interaction contrasts. These terms provide tests of whether the concentration responses are parallel for the two sexes or whether there is a differential response between the two sexes.

The estimate of the pooled within group variability can be used to provide pair-wise tests of the difference between means. These comparisons could be made between the means for the two sexes and between the means for the different concentration level such as for comparisons with the negative control levels. In those cases where there is a significant interaction comparisons can be made between the means of different concentrations within a sex or between the means of the sexes at the same concentration.

References

There are many statistical textbooks which discuss the theory, design, methodology, analysis and interpretation of factorial designs ranging from the simplest two factor analyses to the more complex forms used in Design of Experiment methodology. The following is a non-exhaustive list. Some books

[1] Statisticians who take a modelling approach such as using General Linear Models (GLMs) may approach the analysis in a different but comparable way but will not necessarily derive the traditional anova table, which dates back to algorithmic approaches to calculating the statistics developed in a pre-computer age.

provide worked examples of comparable designs, in some cases with code for running the analyses using various software packages.

Box, G.E.P, Hunter, W.G. and Hunter, J.S. (1978). Statistics for Experimenters. An Introduction to Design, Data Analysis, and Model Building. New York: John Wiley & Sons.

Box G.E.P. & Draper, N.R. (1987) Empirical model-building and response surfaces. John Wiley & Sons Inc.

Doncaster, C.P. & Davey, A.J.H. (2007) Analysis of Variance and Covariance: How to choose and Construct Models for the Life Sciences. Cambridge University Press.

Mead, R. (1990) The Design of Experiments. Statistical principles for practical application. Cambridge University Press.

Montgomery D.C. (1997) Design and Analysis of Experiments. John Wiley & Sons Inc.
Winer, B.J. (1971) Statistical Principles in Experimental Design. McGraw Hill.

Wu, C.F.J & Hamada, M.S. (2009) Experiments: Planning, Analysis and Optimization. John Wiley & Sons Inc.