Section 4
Health effects

Test Guideline No. 470
Mammalian Erythrocyte Pig-a Gene Mutation Assay



25 June 2025

OECD Guidelines for the Testing of Chemicals

OECD
BETTER POLICIES FOR BETTER LIVES

OECD/OCDE 470
Adopted:
25 June 2025

OECD GUIDELINES FOR THE TESTING OF CHEMICALS

Mammalian Erythrocyte *Pig-a* Gene Mutation Assay

INTRODUCTION

1. The OECD Test Guidelines for the testing of chemicals are periodically reviewed in light of scientific progress, changing regulatory needs and animal welfare considerations. Previously, this Test Guideline included three equal considerations when assessing whether a response is positive or negative. First, the test chemical response was assessed as to whether there was a statistically significant increase from the concurrent negative control. Second, the response was evaluated for a concentration/dose related effect. Third, historical negative control distributions were used to assess biological relevance. As noted above, the three considerations were given equal weight. However, recent analyses suggest more nuance and flexibility should be applied to the third consideration (i.e., comparison to historical controls). In particular, distributions can only serve as a proxy for normal biological variation when they are of sufficient quality for this purpose. The changes found herein address these issues by revising language in the "Evaluation and Interpretation of Results," and by updating an associated Annex.

2. The *in vivo* erythrocyte *Pig-a* gene mutation assay (hereafter called the *Pig-a* assay) uses an endogenous mammalian gene, the phosphatidylinositol glycan class A gene (*Pig-a*), as a reporter of somatic-cell gene mutation. *In vivo* gene mutation tests, such as the *Pig-a* assay, are especially relevant for assessing mutagenicity because physiological factors, such as absorption of the test chemical from the site of exposure, distribution of the test chemical throughout the test system *via* systemic circulation, and *in vivo* metabolism and DNA repair processes, all contribute to the mutagenic responses. For these reasons, the *in vivo Pig-a* assay can be used to further investigate genotoxic potential detected by *in vitro* systems.

3. The *Pig-a* assay can be performed with commonly used strains of rats or mice (1) (see paragraph 28) and the test can be conducted without euthanizing the animals. These properties facilitate integration of the *Pig-a* assay into many *in vivo* rodent testing protocols (see paragraphs 8-9 and Annex 2) (2) (3).

4. This Test Guideline (TG) is based on the recommendations included in a Detailed Review Paper (DRP), which describes the scientific principles, technical conduct, performance, and context of use of the *Pig-a* assay (3). In addition, the DRP compares the strengths and weaknesses of the *Pig-a* assay in relation to other *in vivo* genetic toxicology tests (3). The DRP was based on the review by the International Workshops on Genotoxicity Testing of several *Pig-a* protocols (2). The main components and procedures for the *Pig-a* assay are described below. Further guidance for conducting the test,

© OECD (2025)
You are free to use this material subject to the terms and conditions available at http://www.oecd.org/termsandconditions/

including detailed laboratory protocols that are compliant with the recommendations in this TG, can be found in these and other publications (4) (5) (6) (7).

5. An independent peer review (8) of the Retrospective Performance Assessment conducted on the *in vivo* erythrocyte *Pig-a* gene mutation assay (3) (9) concluded that the test was adequate for identifying *in vivo* mutagens with a high degree of accuracy. The review indicated that the test was sufficiently well validated for the development of a TG used to support regulatory decisions.

6. Definitions of key terms are set out in Annex 1.

INITIAL CONSIDERATIONS AND LIMITATIONS

7. *Pig-a* gene mutations have been observed to accumulate with repeat dosing of mutagens, and increased frequencies of mutant erythrocytes generally persist for months in peripheral blood (10) (11) (12). These observations suggest that the kinetics by which *Pig-a* mutant cells appear and disappear from blood circulation are similar to those of *Pig-a* wild-type cells, indicating that the mutations measured in the *Pig*-a assay act as if they have a relatively neutral phenotype, *i.e*., one that does not confer a cell growth advantage or disadvantage. Although different dosing regimens have been employed for the test, a protocol employing 28 consecutive days of dosing (or longer, see paragraph 44) takes advantage of mutant accumulation, provides sufficient time for mutant expression, and facilitates integrating the test into standard repeat-dose toxicity studies. The 28-day dosing protocol also has displayed slightly greater accuracy for detecting weak *in vivo* mutagens than shorter-term (*i.e.,* less than 14 consecutive days) dosing protocols (*e.g*., concordance of 93% and 89% using two sets of criteria *vs*. 88% using data from short-term treatments only: see (3) for full discussion). Finally, a 28-day dosing protocol detects some mutagens at lower (per day) dose levels compared to shorter dosing schedules (13).

8. 3Rs principles (*i.e*., replacement, reduction, and refinement) and experimental data support integrating the *Pig-a* assay with other test(s) whenever possible. However, there are several important factors that should be considered in order to do so effectively. The recommended procedure for conducting the *Pig-a* assay involves 28 consecutive days of dosing followed by testing for *Pig-a* gene mutation (paragraph 41). Thus, the test readily integrates into study designs that use such repeat-dosing schedules, *e.g*., the 28-day subacute general toxicology studies described in OECD TG 407 (14) and TG 412 (15), or the transgenic rodent gene mutation test described in OECD TG 488 (16). The *Pig-a* assay also can be combined with tests with extended exposure durations, such as the subchronic toxicity studies described in OECD TG 408 (17) and TG 413 (18). Combination of *Pig-a* assays with tests having short-term treatment and sampling schedules is discussed in Annex 2.

9. Under the conditions outlined in paragraph 8 and Annex 2, the *Pig-a* and erythrocyte micronucleus tests can be integrated into single experiments, since only small volumes of peripheral blood are required for each endpoint. Thus, both clastogenicity/aneugenicity (micronucleus test) and gene mutation (*Pig-a* assay) can be evaluated in the same study (19). The *Pig-a* assay also complements the *in vivo* comet test (so long as mutant manifestation time is taken into account), since the latter detects a different type of endpoint (DNA damage in the form of strand breaks) than the *Pig-a* assay. The comet test also shows low sensitivity for detecting test substances that induce mutation in the bone marrow compartment (20), while the *Pig-a* assay specifically measures mutations induced in bone marrow erythroid cells (paragraph 11).

10. The *Pig-a* gene is relatively large (1302 bp of protein-coding sequence in 5 exons). Mutations that produce the *Pig-a* mutant phenotype (see paragraph 16) have been detected throughout the coding region and flanking splice-site sequences, without any major hot-spots for mutation (21) (22) (23) (24).

© OECD (2025)

Most of the mutations detected by the test have been base-pair substitutions and frameshifts, including mutations resulting in exon deletions, without any known bias as to the nature of the mutations detected.

11. The erythrocyte *Pig-a* assay detects mutations induced in erythroid precursor cells, which in adult rodents are primarily found in the bone marrow. Because the bone marrow is a well-perfused tissue, many substances that induce mutations in other tissues, induce mutation in the bone marrow as well. The assay is sensitive to both test chemicals that are direct-acting mutagens and test chemicals that require metabolic activation to induce mutation (*i.e.*, promutagens). However, the *Pig-a* assay is not applicable for assessing the mutagenicity of test chemicals or their metabolites that do not reach the bone marrow (see paragraph 75). In general, the *Pig-a* assay described in this TG cannot investigate genotoxicity at the site of first contact.

12. Analyses of *Pig-a* gene mutation spectra have been conducted by DNA sequence analysis of *Pig-a* mutant phenotype erythroid precursor cells from the bone marrow of rodents dosed with known mutagens. Findings from these studies indicate that *Pig-a* mutant phenotype cells contain *Pig-a* mutations while *Pig-a* wild-type cells do not (21) (22) (23) (24).

13. Although this TG is applicable only to rodent blood erythrocyte-based *Pig-a* assays, *Pig-a* mutations can be detected *in vitro*, in cultured mammalian cell lines, and in various animal species, including humans (3) (25) (26) (57) (58). Analogous measurements in the peripheral blood of other mammalian species may be feasible and eventually valuable, provided adequate sensitivity and specificity to detect mutagenic and non-mutagenic substances have been demonstrated.

14. Most peer-reviewed *Pig-a* assay data are from rats. Mouse data collected to date (1) (25) indicate this species also can be used for the assay (3). However, several characteristics of mouse-based tests, including the low circulating blood volume in mice relative to rats, should be taken into consideration when designing studies in mice (*e.g.*, see paragraph 57).

15. The most appropriate *in vivo* methods for testing manufactured nanomaterials may be tests that can measure effects at the site of contact (27), making the *Pig-a* assay of lesser relevance for testing these types of substances. The *Pig-a* assay, however, may be appropriate for testing the mutagenicity of manufactured nanomaterials provided that bone marrow or systemic availability of the manufactured nanomaterials can be demonstrated (*e.g.,* following intravenous treatments) (28) (29).

PRINCIPLE OF THE TEST METHOD

16. The protein encoded by the *Pig-a* gene is necessary for the biosynthesis of glycosylphosphatidylinositol (GPI) cell membrane anchors (30). GPI anchors link approximately 150 unique proteins to the exoplasmic surface of mammalian cytoplasmic membranes. Of the approximately 30 genes involved in GPI synthesis, *Pig-a* is the only gene located on the X-chromosome. As there is only one X chromosome in males and one functional X chromosome in females, a single mutation has the potential to inactivate the *Pig-a* gene product and result in cells deficient in GPI anchors and their associated GPI-anchored proteins. Thus, male and female animals are equally sensitive to *Pig-a* mutant induction (3). Other genes involved in GPI biosynthesis only can be inactivated by mutations in both alleles, making them less likely targets for GPI inactivation and their mutation less likely to cause GPI anchor deficiency, which is the mutant phenotype measured by the test (31).

17. Microliter volumes of peripheral blood contain millions of red blood cells (RBC) [*e.g.*, (32)], a few percent of which are immature erythrocytes, also known as reticulocytes (RET). RET contain residual amounts of RNA and express surface markers, such as CD71, that are typical of erythroid precursor

© OECD (2025)

cells, whereas mature RBC do not. For the purpose of this guideline, RBC are used to signify all circulating erythrocytes, irrespective of age, while RET refers to the most immature fraction of circulating erythrocytes. RET can be discriminated from mature RBC based on their residual RNA and/or presence of the CD71 surface marker using appropriate fluorescent dyes or antibodies.

18. Peripheral blood RBC and RET can be analysed by flow cytometry for the *Pig-a* mutant phenotype (*i.e.*, loss of GPI anchors) by labelling with fluorochrome-conjugated antibodies against GPI-anchored proteins (33) (34) (35). Thus, *Pig-a* mutant cells will be non-fluorescent with regard to GPI-anchored epitope(s), and *Pig-a* wild-type cells will fluoresce. In this way, millions of labelled RBC and RET can be tested for the *Pig-a* mutant phenotype using flow cytometry. As noted in paragraphs 24-26 and 61, methods that concentrate RET and/or mutant cells have been developed to make evaluating million(s) of cells per sample practical.

19. Both the RET and total RBC populations should be evaluated for the mutant phenotype. Increased frequencies of mutant reticulocytes (MUT RET) appear sooner in peripheral blood circulation compared to mutant erythrocytes (MUT RBC). Even so, it is important to enumerate MUT RBC in addition to MUT RET because MUT RBC represent an important confirmatory, lagging indicator of mutation. The demonstration of mutant manifestation in both cohorts of erythrocytes represents the strongest evidence of test chemical-induced mutagenic activity (see paragraph 68), and the dosing/blood collection schedules described herein (see paragraphs 41-46) were designed to provide sufficient time for both MUT RET and MUT RBC to be increased by exposure to mutagenic substances. Indeed, statistically significant increases in only one population, MUT RET or MUT RBC, is a rare occurrence when rodents are exposed to known mutagenic substances and when a 28-day repeat-dose protocol is utilised. However, since mutations are assayed at only one or a few sampling times, it is conceivable that positive MUT RET, but not MUT RBC responses could occur for a late arising RBC mutant frequency or that negative responses are detected stochastically in one or the other erythrocyte cohort. Such results benefit from expert evaluation for establishing the mutagenicity of the test chemical.

20. In addition to collecting data on *Pig-a* mutant induction, the fraction of RET among total RBC can be used as a measure of toxicity to the erythroid cell compartment. Percent RET (%RET) data can be useful for establishing that the bone marrow has been exposed to biologically significant levels of the test chemical or its metabolites ((2) (7); see also paragraphs 50 and 56). When blood samples are collected within approximately 3 days following cessation of dosing, bone marrow toxicity often manifests as a reduction in %RET relative to %RET in concurrent vehicle/solvent control animals. However, it is important to note that when blood samples are collected later than approximately 3 days following cessation of dosing, bone marrow toxicity can at times be expressed as elevated %RET relative to concurrent vehicle/solvent control animals (36). Elevated %RET following dosing is colloquially referred to as a ‘rebound effect’ or ‘stress erythropoiesis’, and usually represents a compensatory response to a period of bone marrow toxicity, especially involving reduced erythropoiesis. Finally, dose-related changes in %RET may occur if the test chemical causes haemolysis (37). Whatever the cause, dose-related changes in %RET indicate systemic exposure at biologically significant levels.

VERIFICATION OF LABORATORY PROFICIENCY

Training and proficiency investigations

21. Before laboratories conduct the *Pig-a* assay for regulatory safety assessments, it is important for staff to gain experience with blood collection, blood labelling and flow cytometric procedures. To

© OECD (2025)

minimise the number of animals required for the preliminary training exercises, reconstruction experiments, also known as spiking experiments, can be conducted as described in Annex 3 (38).

22. Establishing laboratory proficiency requires reproducing the expected results from published data (1). This should be demonstrated for both MUT RBC and MUT RET frequencies using a minimum of two mutagens (including weak responses induced by low doses of mutagens; see Table 1) and with compatible vehicle/solvent controls (see paragraph 32). These experiments should use doses of mutagens that give reproducible and dose-related increases in mutant frequency and demonstrate the sensitivity and dynamic range of the test system using the methods employed in the laboratory. Note that by collecting and analysing pre-dosing blood samples (see also paragraph 47) along with vehicle/solvent control samples from rodents used in these proficiency experiments, laboratories can begin building up their historical negative control data base (described in more detail in paragraph 23 and Annex 4, section 4.1).

Table I. Recommended Mutagens for Training and Proficiency Testing.

Chemical[a]	Cas No.	Vehicle	Route	Highest Suggested Rat Dose[b] (mg/kg/day)	Highest Suggested Mouse Dose[b] (mg/kg/day)	Blood Collection Day[c]
N-Ethyl-*N*-Nitrosourea, also known as ENU[d]	759-73-9	Phosphate-buffered saline, pH 6.0	Oral gavage	20	40	15-30
Benzo[*a*]pyrene, also known as B[a]P[e]	50-32-8	Sesame oil or Corn oil	Oral gavage	125	125	15-30
7,12-Dimethylbenz[a]-anthracene, also known as DMBA[e]	57-97-6	Sesame oil or Corn oil	Oral gavage	30	75	15-30

[a]Other mutagens can be used if justified.
[b]Generally well-tolerated, highly mutagenic dose level assuming 3 consecutive days of administration (Days 1-3) (11) (39).
[c]Blood collection day describes a range of recommended days for collecting peripheral blood for the *Pig-a* assay with the days relative to the start of dosing (*i.e.*, Day 1).
[d]Potent direct-acting mutagen; pH 6 phosphate-buffered saline should be used to dissolve ENU in order to inhibit hydrolysis; the ENU dosing solution(s) should be prepared fresh each day of use; commercial supplies of ENU have large amounts of stabilisers and this needs to be accounted for when preparing dosing solution(s).
[e]Requires metabolic activation.

Historical negative control data

23. During the course of the proficiency investigations, each laboratory should establish historical negative control distributions for MUT RET and MUT RBC frequencies. Negative control data can be obtained from two sources. First, negative control data can be derived from animals tested as negative controls in conjunction with animals dosed with test chemicals (hereafter referred to as concurrent negative controls). These animals typically will be dosed with the solvent/vehicle alone (paragraph 32). Secondly, for laboratories that conduct pre-dosing *Pig-a* assays (paragraph 47), negative control data can be obtained from these untreated, naïve animals (referred to as pre-dose control data). Pre-dose animals have been shown to have *Pig-a* mutant frequencies equivalent to solvent/vehicle concurrent control animals when tests are conducted soon after a 28-day repeat-dose treatment protocol, as recommended in this TG (3). Whatever the source(s) of negative control data, animals that exhibit exceptionally high mutant cell frequencies (*e.g.*, identified by an outlier test) should not be used to

© OECD (2025)

construct negative control distributions, as they may have experienced a jackpot-type mutational event (as defined in paragraph 47 and Annex 1). Negative control data and the distribution of the negative control data are important for determining the minimum number of cells that should be evaluated for the mutant phenotype (paragraph 61), and assessing test acceptability (paragraph 65). Furthermore, depending on the quantity and quality of the underlying data, negative control data and the distribution of the negative control data may also be useful for evaluating and interpreting test chemical-related effects (paragraphs 67-72). More detailed recommendations for developing historical negative control datasets, and evaluating their quality, are given in Annex 4, sections 4.1 and 4.4, respectively.

DESCRIPTION OF THE METHOD

Intellectual property considerations

24. Some of the specific techniques used for assessing the mutagenic DNA-damaging potential of an agent and for evaluating the effects of an agent that may modify mutagenic DNA damage based on GPI-anchor-deficient erythrocytes are patented by Litron Laboratories in the U.S., Canada, and Europe. Note that this patent protection does not cover the principle of the assay or the general design of scoring GPI negative cells. Laboratories implementing the test method can obtain kits for cell analysis from Litron (commercially known as MutaFlow® kits) or ship blood samples to a proficient laboratory for testing with a Litron kit. MutaFlow kits come with a limited-use license, and a signature is not necessary. Alternatively, laboratories that do not purchase a kit from Litron or an authorised kit distributor must sign a license agreement with Litron Laboratories to perform the techniques covered by the patents in the U.S., Canada, and Europe.

25. The PIGRET method including the immunomagnetic separation for reticulocyte enrichment is patented by Teijin Limited in Japan only, and requires signing a license with Teijin to use this technique in Japan.

26. This TG describes the requirements and important characteristics of erythrocyte-based analysis of *Pig-a* mutation. In regions where patents apply, protocols using alternative techniques, that do not infringe upon the patents held by Litron Laboratories or Teijin Limited, may be conducted freely using the guidance provided by this TG. The use of Litron or Teijin patented techniques are not necessary to conduct a valid test as long as the acceptability criteria (paragraph 65) are fulfilled.

Preparations

Selection of animal species

27. Common laboratory strains of healthy young adult rats and mice are normally used (*i.e.*, 6-12 weeks old at the start of dosing). Slightly younger or older animals are also acceptable (*e.g.*, 4 weeks of age at the start of treatment). However, such deviations should be justified. Justification could include facilitating integration with other tests or to avoid killing animals that have been bred but not used in a procedure. Provision should be made for any alterations to the historical negative control data base such deviations in age may cause. Strains of rats and mice such as Han Wistar, Sprague-Dawley and Fisher 344 rats and CD-1 mice have been used extensively (1). The choice of rodent species and strain should be based on (i) the animals used in other toxicity studies (to be able to correlate data and to facilitate integration), (ii) animals that developed tumours in a carcinogenicity study (when investigating the mode of action for carcinogenesis), or (iii) animals with the most relevant metabolism for humans, if known.

© OECD (2025)

Animal housing and feeding conditions

28. The temperature in the experimental animal room ideally should be 22°C (±3°C). The relative humidity ideally should be 50-60%, being at least 30% and preferably not exceeding 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 *ad libitum* drinking water. The choice of diet may be influenced by the need to ensure a suitable mixture of a test chemical when administered by the diet. Animals should be group housed in small groups of the same sex; animals may be housed separately if scientifically justified. For group caging, no more than five animals should be housed per cage. Furthermore, cages should conform with animal welfare standards [*e.g.*, (40)]. Based on Animal Care and Use Committee (or equivalent) recommendations, solid floors should be used wherever possible and appropriate environmental enrichment should be provided.

Preparation of animals

29. Animals are randomly assigned to solvent/vehicle control and test chemical dosing groups. The animals are identified uniquely and acclimated to the laboratory conditions for a least five days (or as prescribed by the applicable Animal Care and Use Committee or equivalent) before the start of dosing. The least invasive method of uniquely identifying animals should be used. Appropriate methods include ringing, tagging, micro-chipping and biometric identification. Toe and ear clipping are not scientifically justified for these tests. Cages should be arranged in such a way that possible effects due to cage placement are minimised. At the commencement of the study, the weight variation for each sex should not exceed ±20% of the mean weight.

Preparation of doses

30. Solid test chemicals should be dissolved or suspended in appropriate solvents or vehicles (*e.g.*, water, corn oil; see (3) for examples) or mixed in diet or drinking water prior to dosing of the animals. Liquid test chemicals can be dosed directly, if necessary; however, dilution with an appropriate vehicle should be used to generate different doses and maintain a consistent dose volume among the treatment groups. For inhalation exposures, test chemicals can be administered as a gas, vapour, or solid/liquid aerosol, depending on their physicochemical properties (15) (18).

31. Fresh preparations of the test chemical should be employed unless stability data demonstrate the acceptability of the storage conditions used.

Test conditions

Solvent/vehicle

32. The solvent/vehicle should not produce toxic effects at the dose volumes used and should not be suspected of chemical reactions with the test chemicals. If other than well-known solvents or vehicles are used, their inclusion should be supported with reference data indicating their non-reactivity with the test chemical, and their compatibility in terms of test animals, route of administration, and endpoint. It is recommended that wherever possible, the use of aqueous solvents/vehicles should be considered first.

© OECD (2025)

Controls

Positive controls and flow cytometry standards

33. Laboratories will have established their proficiency for conducting the *Pig-a* assay as described in paragraphs 21-22. It is only in this context that the inclusion of concurrent positive control animals are not required for each test. Rather than relying on mutagen-exposed animals, biological flow cytometry standards are included in each analysis (paragraph 34) and represent a useful type of analytical control that is 3Rs-friendly and focuses quality assurance measures on important factors that demonstrate the reliability of a study (2) (7) (11). Specifically, the use of blood-based flow cytometry standards ensures that i) blood collection and cell staining/labelling were performed properly, ii) the reagents used to stain/label cells were effective, and iii) the flow cytometer adequately differentiated mutant phenotype cells from wild-type cells. Finally, in order to assure continued proficiency in detecting increases in mutant frequency, laboratories should occasionally (at least once per year) perform additional tests with mutagen-treated animals similar to the proficiency studies described in paragraph 22.

34. Given their importance, it is necessary that biological flow cytometry standards be made and used each time *Pig-a* mutant analyses are conducted. For this purpose, a “mutant mimic” (*i.e.*, a blood sample prepared to have the properties of mutant erythrocytes when tested) or comparable sample (see Annex 5) is prepared that demonstrates the light scatter and fluorescence characteristics of MUT RET and MUT RBC versus wild-type RET and wild-type RBC. Mutant mimics can be created by processing extra blood from a solvent/vehicle concurrent control animal; a portion of the blood is prepared normally for evaluating *Pig-a* mutant frequencies, but another portion is prepared by omitting the fluorescent antibody(s) for GPI-anchored proteins from the fluorescent antibody labelling protocol (11) (38). To make the standards, these two blood samples are then mixed in approximately equal proportion. Since standards created from mutant mimics (or the alternative described in Annex 5) are valuable for guiding instrumentation settings and software/data analysis parameters, as well as serving as a control on the adequacy on the analytical procedure (paragraph 33), they should be generated for every study, and used each day blood samples are analysed. Once established, instrument settings generally should remain unchanged for all samples processed for mutant analysis throughout that particular data acquisition session.

Negative controls

35. Concurrent negative control animals should be included at every sampling time and otherwise handled in the same way as animals in the dosing groups, except for not being dosed with the test chemical. If a solvent/vehicle is used in administering the test chemical, the control group should receive the same solvent/vehicle. The MUT RET and MUT RBC frequencies from concurrent negative control animals of a given study should be consistent with the historical negative control distribution. There are several valid approaches for calculating lower and upper bound limits that can be used to facilitate such comparisons, and they are described in Annex 4, section 4.2.

36. The requirement for analysing blood from concurrent solvent/vehicle control animals at every sampling time also applies to studies that include the collection and analysis of pre-dose blood samples (paragraph 47). That is, pre-dose blood samples are not a substitute for concurrent solvent/vehicle controls, irrespective of the statistical approach(es) utilised.

© OECD (2025)

PROCEDURE

Number and sex of animals

37. It is technically feasible to conduct the *Pig-a* gene mutation assay using blood from animals of either sex (3). Where human exposure to the test chemical may be sex specific, as for example with some pharmaceuticals, the test should be performed with the appropriate sex.

38. When performing a single-sex study, group sizes at study initiation (and during establishment of laboratory proficiency) should be established with the aim of providing a minimum of 6 analysable animals [for justification of group size, see statistical power calculation in (41)]. Thus, a study conducted with one sex according to the parameters established in paragraph 48, with three dose groups and a concurrent negative control, would require 24 animals at the start of dosing. The number of animals will be higher when pre-dose blood samples are used to eliminate outliers as described in paragraph 47, or when additional animals are included to accommodate unforeseen events that reduce the number of animals. Sound scientific judgement should be used in interpreting the data in situations where the number of animals in any group is less than 6 (*e.g*., due to excluding an animal with a jackpot mutation response).

39. When data demonstrate relevant differences for the test chemical between males and females (*e.g.,* differences in systemic toxicity, metabolism, and bioavailability, including in a range-finding study), it is preferable to perform a study using both sexes.

40. For studies that treat the two sexes with different dose levels, the number of animals required per sex is the same as indicated in paragraph 38, at least 6 males and 6 females per group. For studies that treat sexes with the same dose levels, it is often possible to reduce the number of animals per group by taking advantage of a factorial design (see Annex 4 section 2.1). The factorial statistical design is capable of maintaining adequate statistical power while using fewer animals of each sex.

Dosing schedule and sampling times

41. A 28-day repeat-dose protocol is recommended for conducting the *Pig-a* assay. Of the dosing schedules published and analysed to date, 28-day repeat-dose schedules offer the most compelling evidence that a negative test result for mutation is accurate [*e.g*., has the highest percent negative predictivity (3)]. Alternate dosing schedules can be utilised, if scientifically justified (see paragraph 46). Test chemicals may be administered as a single dose per day or as a split dose, *i.e*., two or more exposures on the same day, to facilitate administering a large volume.

42. When animals are exposed to a test chemical for 28 consecutive days, blood should be collected at least once, within hours to day(s) of cessation of exposure (*e.g*., Day[1] 28-31). Data collected from testing diverse genotoxicants suggest that sampling on Day 28-31 gives sufficient time for manifestation of MUT RET and MUT RBC responses and that the exact timing of sample collection is not critical. This schedule has the advantage of facilitating integration of the *Pig-a* assay with commonly utilised general toxicology studies and other genetic toxicology tests (paragraphs 8-9). Note that an extra administration of the test chemical on Day 29, followed by sample collection several hours later, is permissible for facilitating tissue harvest for conducting the *in vivo* comet test.

[1] Here and elsewhere, study day is relative to start of dosing, which is defined as Day 1.

© OECD (2025)

43. Optionally, and when logistically feasible, there may be merit to conducting *Pig-a* analyses on blood samples collected from a 28-day repeat-dose protocol at an additional, later time point. For instance, some toxicology experiments include satellite “recovery” or “withdrawal” groups to evaluate whether toxic effects are diminished or increased upon discontinuation of dosing. Such blood samples are typically collected between 2 to 4 weeks after cessation of dosing, and represent an opportunity to evaluate erythrocytes for the *Pig-a* mutant phenotype after additional time for mutant manifestation.

44. When the dosing period exceeds that of a 28-day repeat-dose study (*e.g.*, 90-day subchronic studies), blood samples should be collected at least once, after approximately 28 days of dosing (*e.g.*, at Day 28-31). As explained in paragraph 43, there may be merit to collecting blood later in the experimental protocol, (*e.g.*, within a few days of cessation of dosing), but recognizing the possibility that clonal expansions of erythroid precursor cells may disproportionately affect mutant frequencies at extended sampling times [*e.g.*, (42)]. In addition, sampling of animals at much later times (*e.g.*, sampling animals that are ≥8 months of age) may require developing an historical negative control data base for older animals (see Annex 4, section 4.1).

45. Dosing schedules that involve fewer than 28-consecutive days of exposure (*e.g.*, dosing on one or three consecutive days) should be scientifically justified as they exhibit lower sensitivity than repeat-dose, 28-day treatment protocols for detecting mutagens (paragraph 7). If a short-term treatment protocol is used, blood collection and analyses should occur at least two times: at approximately Day 15, and again at approximately Day 30. Test sensitivity for detecting mutagens is anticipated to be higher for short-term dosing schedules that involve two post-dosing blood collection/analysis times as opposed to one (13).

46. Other dosing and sampling schedules may be used when scientifically justified, especially in instances when they facilitate integration with other toxicology studies, or combination with other genotoxicity tests. For instance, protocols that do not involve daily dosing but do use several administrations of test chemical per week for several weeks have been shown to retain high *Pig-a* assay sensitivity while facilitating combination with other genotoxicity tests (43). As above, such studies need to take into consideration the different manifestation times of MUT RET and MUT RBC responses and the manifestation times for any other genotoxicity tests being conducted as part of the study (see discussion in Annex 2, section 2 regarding combining *Pig-a* assays with short-term treatment schedules with other short-term genetic toxicology assays).

47. Whatever the dosing schedule, it may be advantageous (but not required) to perform *Pig-a* analyses before the first administration of the test chemical. Pre-dosing samples are taken within one week prior to test chemical dosing (if adult mice are used, it is recommended that pre-dose sampling be conducted ~7 days prior to dosing to allow the haemopoietic system to re-establish homeostasis). Mutation ‘jackpots’, resulting in mutant frequencies that are unusually high and outside the historical negative control distribution (see Annex 4, section 4.2 for approaches for describing the distribution), may occur due to mutations in germ cells or in early erythroid progenitor cells. It may be useful to identify animals having such mutant frequencies so that they can be excluded from the main study and replaced with animals having mutant frequencies within the historical negative control distribution.

Dose groups and dose levels

48. A minimum of three test chemical dosing groups and a negative control group generally should be used. Additional dose groups may be necessary for a detailed analysis of dose-response (see paragraph 49); the use of fewer dose groups should be justified. Except for dosing with the test chemical, animals in the concurrent negative control group should be handled in an identical manner to

© OECD (2025)

animals in the test chemical groups. If a solvent/vehicle is used in administering the test chemical, the concurrent control group and all dosed groups should receive an equal volume of the vehicle or solvent.

49. Any existing toxicity and toxicokinetic data for the test chemical should be taken into consideration in setting dose levels. If a preliminary range-finding study is performed because there are insufficient suitable data already available to guide dose selection, it should be performed in the same laboratory, using the same strain, source, sex, and treatment route to be used in the main study. The highest dose level should be chosen with the aim of inducing toxic effects but not death or severe suffering [see (44)]. This definition is taken from TG 407 (14) and is used here to facilitate integration of the *Pig-a* endpoint with other repeat-dose studies, thereby maximising the toxicological information acquired. When limited by other factors, such as palatability for dietary or drinking water administration (and when dosing cannot be done by gavage) or explosiveness for test chemicals administered by inhalation, the highest dose level will be the maximum feasible dose. Thereafter, a descending sequence of dose levels should be selected with a view to demonstrating any dose-related response. Two- to four-fold intervals are frequently used for setting the descending dose levels when genotoxicity endpoints are evaluated. With most test chemicals, the dose levels used should cover a range from the maximum to little or no toxicity. When additional toxicity endpoints are integrated into the study, other dose spacing may be considered. The addition of a fourth test group is often preferable to using excessively large spacing between doses.

50. The highest dose also may be defined as a dose that produces toxicity to the bone marrow compartment (see paragraph 20). When this evidence is provided by a reduction in mean %RET (relative to the mean concurrent solvent/vehicle control, measured *e.g.*, with blood sampled 1-3 days after a 28-day dosing schedule), the reduction should be statistically significant, but ideally it should not exceed 80%. This is because new erythrocytes need to be generated in order to manifest a chemical-induced effect in this test.

51. For test chemicals that do not produce toxicity, or that are not limited by other factors (see paragraph 49), the highest dose for dosing protocols of $\geq$ 14 consecutive days is 1000 mg/kg/day, while the maximum for protocols involving shorter-term dosing is 2000 mg/kg/day.

Limit test

52. As noted above, studies normally should employ three test chemical dosing groups and a solvent/vehicle concurrent control group. However, if from assessment of other data, no observable toxic effects (including no depression of bone marrow proliferation or other evidence of target tissue cytotoxicity) would be expected at the highest test dose of 1000 mg/kg body weight/day in a study of $\geq$ 14 days of dosing or 2000 mg/kg/day in studies employing <14 days of dosing (paragraph 51), and if genotoxicity would not be expected based upon *in vitro* genotoxicity studies or data from structurally related substances, then this one dose level can be used. In the case of inhalation exposures, the limit test concentrations are 20 mg/L, 5 mg/L or 20,000 ppm for vapours, dusts/mists (aerosols) and gases, respectively. Limit test doses may not apply in cases where human exposure indicates the need to use a higher dose level (15). Also, when integrating or combining the *Pig-a* assay with other test(s), the applicability of using a single limit dose to these other test(s) should be ensured before conducting a limit test. For instance, the single limit dose provision does not include TG 489 for which the full 3-dose regimen applies (45).

Administration of doses

53. The primary concern in choosing a route of administration is ensuring exposure of the target tissue, bone marrow; however, the anticipated route of human exposure also should be considered.

© OECD (2025)

Therefore, routes of exposure such as oral (by gavage), dietary, drinking water, topical, subcutaneous, intravenous, inhalation, intratracheal, or implantation may be chosen as justified. Intraperitoneal injection is generally not recommended since it is not a usual route of human exposure and only should be used with specific scientific justification. If the test chemical is mixed in diet or drinking water, especially in the case of single or short-term dosing, care should be taken to ensure that the delay between food and water consumption and blood sampling provide time for mutant manifestation (see paragraphs 41-46).

54. 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 and all volumes should be consistent with applicable Animal Care and Use Committee (or equivalent) guidelines. 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.

Observations

55. General clinical observations of the test animals should be made, and clinical signs recorded at least once each day, preferably at the same time(s) and considering the peak period of anticipated effects after dosing (44). At least twice each day all animals should be observed for morbidity and mortality. All animals should be weighed at least once each week, and when they are euthanised. Quantitative measurements of food consumption should be made at least weekly when test chemicals are administered in the diet; general observations of food consumption should be made to ascertain animal health. If the test chemical is administered *via* the drinking water, water consumption should be measured at each change of water and at least weekly. Animal Care and Use Committee (or equivalent) guidelines should be observed; but in general, animals exhibiting non-lethal indicators of excess toxicity should be euthanised prior to completion of the test period.

Target tissue exposure

56. A blood sample should be taken at appropriate times(s) in order to permit measuring the plasma levels of the test chemical and/or its metabolite(s) for the purposes of demonstrating that exposure of the bone marrow occurred and where other exposure data do not exist (*e.g.*, significant perturbations to %RET, see paragraphs 20 and 50; see paragraph 75 for further discussion). The blood sample should be large enough to detect relevant levels of the test chemical and its metabolites without perturbing the biological processes being measured in the animals. The bone marrow is well perfused by the peripheral blood circulation, receiving up to 15% of the cardiac output, and measurement of plasma levels of the test chemical and its metabolites is considered qualitative evidence for exposure of bone marrow (46) (47) (48) (49).

Blood collection for conducting the *Pig-a* assay

57. Applying animal welfare standards, small volumes of peripheral blood (less than about 300 µl) are obtained either using a method that permits survival of the test animal, such as bleeding from the tail vein, jugular vein, or other appropriate blood vessel, or by cardiac puncture or sampling from a large blood vessel when the animal is humanely killed. The size of the animal and animal welfare

© OECD (2025)

standards may limit the options for the method used. As flow cytometric analysis requires single cell suspensions, care should be taken to avoid blood coagulation; this is normally accomplished using an anticoagulant, such as heparin and/or EDTA. Samples should be collected and processed blinded as to the treatment or by employing a blocked design (Annex 4, section1).

Storage of blood

58. In the presence of anticoagulant, blood samples can be stored for up to 5 days before they are processed for flow cytometric analysis as long as they are maintained cold, but not frozen (*e.g.*, in a 2-8°C refrigerator) (2). Furthermore, in the presence of anticoagulant, blood samples can be shipped to an analytical test site provided they are maintained cold, not frozen, throughout transportation and any subsequent storage, and as long as they are further processed and analysed within approximately 5 days of collection (2). Whatever storage and/or shipment method is utilised, each laboratory should develop their own data to support it.

59. Procedures have been described for freezing and later thawing blood samples for subsequent processing and flow cytometric analysis of *Pig-a* mutations (50). These procedures can be useful for delaying analysis for reasons that include but are not limited to instrument failure, deferring the decision to acquire *Pig-a* data, and storing blood from mutagen-treated animals for use as analytical standards (Annex 5). Furthermore, frozen blood samples can be transported from the animal experimental site to an analytical site as long as they are maintained frozen throughout transportation (*e.g.*, on dry ice) (50). Whatever freezing and thawing method is employed, it is important to demonstrate minimal lysis of RBCs, and that the freezing and thawing process and the length of storage have minimal impact on MUT RET, MUT RBC, and %RET frequencies.

Methods of measurement

60. Flow cytometric analysis is used for determining %RET, MUT RET and MUT RBC frequencies in peripheral blood samples.

61. Proficient laboratories have shown that for commonly used rodent models, mean baseline MUT RET and MUT RBC frequencies typically range from just under 1×10^{-6} up to approximately 3×10^{-6}. With this in mind, it is generally necessary for most laboratories to evaluate a minimum of 1 to 3×10^6 RBC and 1 to 3×10^6 RET per animal per time point. Note that given the rarity of RET in peripheral blood circulation, it is not practical to evaluate $\geq 1 \times 10^6$ RET directly from blood samples. In order to overcome this problem, immunomagnetic separation procedures were developed (51) (52) to increase the number of RET (and in some cases the number of RBC) investigated for mutant cell frequency measurements. These immunomagnetic separation techniques, or a validated alternative, can be used to facilitate conducting the *Pig-a* assay.

62. Each laboratory should set the number of cells analysed to minimise the occurrence of zero mutant cell frequency readings. Even so, it is possible that for a particular study, more than an occasional zero mutant cell frequency reading will be observed. In order to avoid this possibility, the number of cells evaluated per animal may be increased by processing left-over blood samples that have been stored properly (paragraphs 58-59). However, when circumstances do not permit such additional processing/scoring, it should be the case that the occurrence of zero MUT RBC values will usually be very rare, owing to the assay's ability to evaluate many times more MUT RBC compared to MUT RET. Thus, when conducting a study with 28 consecutive days of test chemical treatment, a low incidence of

© OECD (2025)

zero readings in the MUT RBC population should be considered sufficient to offset concerns with the number of zero MUT RET readings in a particular study.

63. The preferred antibodies for scoring the incidence of MUT RET and MUT RBC are anti-CD59 for rats and anti-CD24 for mice, which target the GPI-anchored proteins CD59 and CD24, respectively. Other GPI-anchored proteins exist on the surface of wild-type RBCs (*e.g.*, CD55) that potentially could be used in the test, and it also is possible to use combinations of antibodies for GPI-anchored proteins to distinguish wild-type and mutant phenotype cells (*e.g.*, anti-CD59 plus anti-CD55). Alternative antibodies or combinations of antibodies for identifying *Pig-a* mutant and wild-type erythrocytes can be used if they generate known *Pig-a* responses (1) and upon demonstration of proficiency by the laboratory (see paragraph 22).

DATA AND REPORTING

Treatment of results

64. Individual animal data, as well as group means and a measure of within-group variation (*e.g.*, standard deviation or standard error), should be presented in tabular form. The *number* of RET, RBC, MUT RET, and MUT RBC scored should be listed separately for each animal and time point analysed, and they should be accompanied by RET, MUT RET, and MUT RBC *frequencies* for each individual animal. Whereas RET frequency is usually expressed as a percentage, it is preferable to express MUT RET frequency as number of MUT RET per 1 million total RET, and MUT RBC frequency as the number of MUT RBC per 1 million total RBC. Data on animal toxicity and clinical signs also should be reported.

Statistical evaluation and interpretation of results

Acceptability criteria

65. For a valid test, data analysis is conducted only when the study fulfils all acceptability criteria in terms of animal number, dosing, sampling times, mutant analysis, etc. described in paragraphs 37-63. The following criteria determine the acceptability of the test:

- The concurrent negative control frequencies are consistent with the distribution described by the laboratory historical negative control data base. It is important that the historical negative control data base was constructed appropriately (for instance as recommended in paragraph 23 and Annex 4, section 4.1), and the historical negative control mutant frequencies and their distribution are consistent with literature values (1) (3) (7).
- When they are used, concurrent positive controls should induce a statistically significant increase compared to the concurrent negative control (see paragraphs 35-36).
- Mutant mimic controls or historical samples from animals treated with a positive control substance were used for setting flow cytometer gates for experimental sample evaluations as described in paragraph 34 and Annex 5.
- The number of dose levels, route of exposure, and the criteria for the selection of the highest dose are consistent with those described in paragraphs 48-52.
- The appropriate numbers of MUT RBC and MUT RET have been analysed (paragraph 61).

© OECD (2025)

Evaluation and interpretation of results

66. Data from *Pig-a* experiments are generally evaluated with the intention of determining whether the frequencies of MUT RET and MUT RBC in test chemical dosing groups are increased relative to those of the concurrent solvent/vehicle control group. Such analyses should be accomplished by appropriate statistical methods, using the animal as the experimental unit. The statistical method(s) employed should be indicated.

67. When evaluating the responses, all data should be taken into consideration and, in all cases, sound scientific judgement applied.

68. Providing that all acceptability criteria above are fulfilled, a test chemical is considered positive if both criteria a and b, below, are met in any of the experimental conditions examined:

a) At least one of the treatment groups exhibits a statistically significant increase in both MUT RET and MUT RBC frequency compared with the concurrent negative control; and

b) The mutant frequency responses are dose-related, for example when evaluated with an appropriate trend test (Annex 4) (not applicable to the limit test)

Criteria a- and b-type analyses, both of which are based on concurrent negative control data, are considered the statistical analyses of primary importance for interpreting study data. Historical negative control data and their distribution may also be useful for contextualising the results of the current study. However, this comparison carries less weight than criterion a or b. Furthermore, the degree to which historical control data are useful for this purpose is highly dependent on the quantity and quality of the underlying data. For instance, historical control data cannot serve as a useful comparator (e.g., as a proxy for normal biological variation), when inter-study variation represents the predominant source of variation. This and other concepts surrounding the appropriate use of historical control data are provided in Annex 4. Thus, only to the degree historical negative control data have been generated in sufficient quantity, and these data are deemed of sufficient quality, that in addition to criteria a and b, a third criterion, c, can be applied as follows:

c) At least one of the treatment groups exhibits MUT RET and MUT RBC frequencies exceeding the upper bound limit derived from the historical negative control data distribution for at least one post-treatment time point.

The upper bound limit derived from the historical negative control distribution should be determined using a method appropriate to the quantity and quality of the laboratory's data (e.g., control charts, tolerance intervals, or quantiles), with justification provided in the study report. As inter-laboratory variability may influence the control distribution, no fixed numerical value can be prescribed, and laboratories should define their upper bound limits in a scientifically sound and transparent manner, as described in Annex 4.

In many instances, the frequencies being considered will be the current study's mean values compared to a distribution derived from historical negative control mean values. However, there are situations that make historical negative control data and related distributions based on individual animal data useful; for example, when historical control values are low in number, or when it is of interest to estimate sources of variation. As explained in more detail in Annex 4, in these cases it is important to make appropriate comparisons. Specifically, a distribution based on individual animal data should not be used as a comparator for the current study's treatment group means. Rather, a distribution based on individual animal data should only be used as a comparator for the current study's individual animal data. This requirement is referred to as a "like to like" comparison in Annex 4.

© OECD (2025)

Positive results indicate that, under the test conditions, the test chemical induces *Pig-a* gene mutations in the erythropoietic cells of the test species.

69. Providing that all acceptability criteria above are fulfilled, and bone marrow exposure to the test chemical and/or its metabolites occurred (see paragraph 75), a test chemical is considered negative if both criteria a and b, below, are met in all experimental conditions examined:

a) No treatment group exhibits a statistically significant increase in either MUT RET or MUT RBC frequency compared with the concurrent negative control; and

b) None of the mutant frequency responses are dose-related, for example when evaluated by an appropriate trend test (Annex 4) (not applicable to the limit test).

Criteria a- and b-type analyses, both of which are based on *concurrent* negative control data, are considered the statistical analyses of primary importance for interpreting study data. *Historical* negative control data and their distribution may also be useful for contextualising the results of the current study. However, this comparison carries less weight than criterion a or b. Furthermore, the degree to which historical control data are useful for this purpose is highly dependent on the quantity and quality of the underlying data. For instance, historical control data cannot serve as a useful comparator (e.g., as a proxy for normal biological variation), when inter-study variation represents the predominant source of variation. This and other concepts surrounding the appropriate use of historical control data are provided in Annex 4. Thus, only to the degree historical negative control data have been generated in sufficient quantity, and these data are deemed of sufficient quality, that in addition to criteria a and b, a third criterion, c, can be applied as follows:

c) None of the treatment groups exhibits MUT RET and MUT RBC frequencies exceeding the upper bound limit derived from the historical negative control data distribution at any of the post-treatment time points.

The upper bound limit derived from the historical negative control distribution should be determined using a method appropriate to the quantity and quality of the laboratory's data (e.g., control charts, tolerance intervals, or quantiles), with justification provided in the study report. As inter-laboratory variability may influence the control distribution, no fixed numerical value can be prescribed, and laboratories should define their upper bound limits in a scientifically sound and transparent manner, as described in Annex 4.

In many instances, the frequencies being considered will be the current study's mean values compared to a distribution derived from historical negative control mean values. However, there are situations that make historical negative control data and related distributions based on individual animal data useful; for example, when historical control values are low in number, or when it is of interest to estimate sources of variation. As explained in more detail in Annex 4, in these cases it is important to make appropriate comparisons. Specifically, a distribution based on individual animal data should not be used as a comparator for the current study's treatment group means. Rather, a distribution based on individual animal data should only be used as a comparator for the current study's individual animal data. This requirement is referred to as a "like to like" comparison in Annex 4.

Negative results indicate that, under the test conditions, the test chemical does not induce *Pig-a* gene mutations in the erythropoietic cells of the test species.

70. Regarding the statistical approaches described in paragraphs 68 and 69 above, it is important to recognise there is no single correct method of conducting a statistical analysis. A practical approach is to suggest a particular set of statistical analyses as *an example* of the sort of evaluations that can be

© OECD (2025)

carried out. This is the context that three types of analyses (a, b, c) described above were presented. See Annex 4 for information that testing laboratories and regulatory reviewers may find useful when considering whether a particular statistical analysis is appropriate and whether the data interpretation is scientifically sound. Whatever statistical approach is used, it is important that it is described in advance of conducting a study.

71. Especially in those instances when the Criteria being considered (a and b; or a, b and c) are not all in alignment with respect to positive or negative call, sound scientific judgment should be applied in an effort to interpret the results as either positive or negative. The application of sound scientific judgement is consistent with recommendations made by an OECD genotoxicity working group that met in Ottawa, Canada in 2013 (53). It should also be noted that, in general, criterion c has less weight than criteria a and b.

72. If the application of sound scientific judgement is unable to resolve whether a response is either positive or negative, further investigations of the *existing experiments* may be necessary to establish the biological relevance of a result. For example, these investigations may include analysing more cells.

73. In rare cases, even after further investigation, if the data preclude concluding whether the test chemical produced a positive or negative result, the study is deemed equivocal.

74. In some cases, equivocal studies may prompt the need for a repeat experiment using modified experimental conditions, or changing the test system.

75. As noted in paragraph 69, a negative result is only considered reliable when evidence is provided that the bone marrow (and/or general circulation) has been exposed to the test chemical and/or its metabolites. With toxic substances, direct evidence of bone marrow exposure is provided by a statistically significant change in the fraction of RET among total RBC (%RET; see paragraph 20); in the absence of such evidence, measurement of the plasma or blood levels of the test chemical and its metabolites (*e.g.*, using the blood sample described in paragraph 56) can be useful for documenting exposure. A discussion of this topic is provided by the European Food Safety Agency (54). Alternatively, ADME (absorption, distribution, metabolism, and excretion) data, obtained in an independent study, using the same dosing route and same species (or a different species, with sound scientific justification), can be used to demonstrate bone marrow exposure. In general, plasma levels calculated from such studies can be used for this purpose. In some instances, modelling software programs can assist in bridging information on blood levels and tissue concentrations and help in extrapolating from one route to another or from one species to a different species (55). Finally, a weight of evidence approach can be used to demonstrate systemic and/or bone marrow exposure to the test chemical, *e.g.*, for coloured test chemicals, a matching urine colour could be an indicator of systemic exposure to the test chemical.

Test report

76. Consistent with Good Laboratory Practice, the test report should include the following information:

Evidence for laboratory proficiency

Test chemical:

- Source, lot number, and expiration/retest date, if available;
- Physical nature and purity;
- Physicochemical properties relevant to the study;
- Stability of the test chemical, if known;

© OECD (2025)

- For mono-constituent substances: chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, chemical identity of impurities as appropriate and practically feasible;
- For multi-constituent substances, substances of unknown or variable composition, complex reaction products or biological materials (UVBCs) and mixtures: characterization of chemical identity as far as possible (see above), quantitative occurrence and relevant physicochemical properties of the constituents;

Test chemical preparation:

- Justification for choice of solvent/vehicle;
- Solubility and stability of the test chemical in the solvent/vehicle, if known;
- Preparation of dietary, drinking water or inhalation formulation, if appropriate;
- Analytical determinations on formulations (*e.g.*, stability, homogeneity, nominal concentrations), when conducted;

Test animals:

- Species and strain and justification for choice;
- Number, age and sex of animals;
- Cage type, housing conditions, diet, enrichment, etc.;
- Method used to uniquely identify animals;
- Procedure used to form dosing groups;
- Individual weight of animals at the start of the test, including body weight range, mean and standard deviation for each dosing group;

Test conditions (reference can be made to established/published methods in lieu of detailed descriptions):

- Report whether (yes/no) blinding of the study (or parts of the study) was performed
- Protocol for dose range-finding study, if conducted;
- Rationale for dose level selection;
- Details of test chemical preparation;
- Details on the administration of the test chemical;
- Rationale for route and duration of administration (if other than recommended);
- Methods for analytical determinations of dose, when used;
- Details of food and water composition and quality;
- Methods for the measurement of animal toxicity, including %RET and, where available, histopathological or clinical pathological analysis, and frequency with which animal observations and body weights were obtained;
- Methods for the bioanalytical/toxicokinetic analysis (if conducted);
- Details and rationale for the peripheral blood sampling protocol, including pre-dose sampling (if conducted), and sampling within (if conducted) and after the dosing period, including recovery sampling (if conducted);
- Methods used for blood collection and storage, if used;
- Method used for humane killing, if part of the study;
- Methods for verifying that the test chemical reached the bone marrow, or general circulation, if negative results are obtained;

© OECD (2025)

- Methods used for immunomagnetic separation of cells;
- Materials used for quantification of mutant cells, including source and lot of antibody(s) and immunomagnetic separation reagents;
- Methods used for calculation of mutant frequencies; rationale used for the number of RET and RBC interrogated for mutation (paragraph 61);
- Methods used for a positive control, if included in the analysis;
- Criteria for acceptability of the study;

Results:

- Data from dose range-finding study, if conducted;
- Animal condition prior to and throughout the test period, including body weight observations and signs of toxicity for individual animals;
- For dietary or drinking water exposures, dose (mg/kg body weight/day) calculated from diet/drinking water test chemical concentrations and individual animal consumption;
- Body weight, organ weight, clinical pathology and histopathology for individual animals, if available, at the conclusion of the study;
- Instrument calibration approach, including descriptions of: gating strategy, how the MUT RET and MUT RBC scoring regions were constructed (*e.g.*, use of mutant mimics, etc.), and graphical output from the flow cytometer demonstrating the fluorescent resolution achieved between mutant erythrocytes (or mutant mimics) and wild-type erythrocytes;
- Pre-dosing data for each animal, if collected, including number of cells and cell equivalents evaluated, the number of mutant cells observed (MUT RET and MUT RBC), the frequency of mutant cells (reported as number per million), and %RET. In addition, group means should be reported and a measure of within group variation (standard deviation or standard error) should be reported for each proposed dosing group;
- If pre-dosing data were used to eliminate any animals from the main study, provide the rationale;
- From the main study, data from each negative (solvent/vehicle) concurrent control animal and each dosed animal at each sampling and analysis point, including number of cells evaluated, the number of mutant cells observed (MUT RET and MUT RBC), the frequency of mutant cells (reported as number per million) and %RET. In addition, group means and a measure of within group variation should be reported for each sampling and analysis point. In addition, graphical flow cytometric output from each sample analysis should be available, upon request;
- Concurrent positive control data, if any, at each analysis time, including number of cells evaluated, the number of mutant cells observed (MUT RET and MUT RBC), the frequency of mutant cells (reported as number per million), and %RET. In addition, group means and a measure of within group variation should be reported for each analysis point. Include any laboratory historical control data or literature values for comparison;
- Historical negative control data with number of rodent blood specimens clearly specified, statistical description of the distribution including type of calculation used to generate lower and/or upper bound limits used to aid data interpretation and justification for its use (see Annex 4), as well as the time period covered and evidence that the assay is under control during the period covered. A representative table that captures this and other useful information is provided in Annex 4, section 6;
- Statistical analyses performed to evaluate the effect of the test chemical on MUT RET, MUT RBC and %RET frequency;
- Data supporting that bone marrow exposure occurred for negative findings (see paragraph 75);

© OECD (2025)

- Criteria for supporting a positive or negative response that are met;

Discussion of the Results:

- Explain how scientific judgement was used in evaluating the results of the test; especially in those instances when some, but not all, of the criteria are met for a negative or a positive result.

Conclusions:

Provide a concise summary of the study findings, including:

- Whether the test chemical showed evidence of gene mutation under the test conditions;
- An assessment of the biological and statistical relevance of the results;
- Any limitations or uncertainties that may affect the interpretation of the findings.

© OECD (2025)

LITERATURE

1. Shemansky JM, LP McDaniel, C Klimas, SD Dertinger, VN Dobrovolsky, T Kimoto, K Horibata, JE Polli, RH Heflich (2019) *Pig-a* gene mutation database. Environ Mol Mutagenesis 60:759-762.

2. Gollapudi BB, AM Lynch, RH Heflich, SD Dertinger, VN Dobrovolsky, R Froetschl, K Horibata, MO Kenyon, T Kimoto, DP Lovell, LF Stankowski Jr, PA White, KL Witt, JY Tanir (2015) The *in vivo Pig-a* assay: A report of the International Workshop on Genotoxicity Testing (IWGT) workgroup. Mutat Res 783:23-35.

3. OECD (Organisation for Economic Cooperation and Development) (2020) The in vivo erythrocyte *Pig-a* gene mutation assay - Part 1: Detailed Review Paper and Retrospective Performance Assessment. Available as Publication 315 [ENV/JM/MONO(2020)6].

4. Chikura S, T Kimoto, S Itoh, H Sanada, S Muto, K Horibata (2019) Standard protocol for the total red blood cell *Pig-a* assay used in the interlaboratory trial organized by the Mammalian Mutagenicity Study Group of the Japanese Environmental Mutagen Society. Genes Environ 41:5. Available at: https://doi.org/10.1186/s41021-019-0121-z.

5. Bemis JC, NE Hall, SD Dertinger (2019) Erythrocyte-based *Pig-a* gene mutation assay. In: A Dhawan, M Bajpayee (eds), Genotoxicity Assessment: Methods and Protocols, Methods in Molecular Biology, vol 2031, Humana, NY, NY. pp 29-57. Available at: https://doi.org/10.1007/978-1-4939-9646-9_2.

6. Chikura S, T Kimoto, S Itoh, H Sanada, S Muto, K Horibata (2021) Standard protocol for the PIGRET assay, a high-throughput reticulocyte *Pig-a* assay with an immunomagnetic separation, used in the interlaboratory trial organized by the Mammalian Mutagenicity Study Group of the Japanese Environmental Mutagen and Genome Society. Genes Environ 43:10. Available at: https://doi.org/10.1186/s41021-021-00181-7.

7. Dertinger SD, JA Bhalli, DJ Roberts, LF. Stankowski Jr, BB Gollapudi, DP Lovell, L Recio, T Kimoto, D Miura, RH Heflich (2021) Recommendations for conducting the rodent erythrocyte *Pig-a* assay: A report from the HESI GTTC *Pig-a* Workgroup. Environ Mol Mutagenesis 62:227-237.

8. OECD (Organisation for Economic Cooperation and Development) (2020) Report of the peer review of the validation status of the in vivo erythrocyte Pig-a gene mutation assay. Available as Publication 319 [ENV/JM/MONO(2020)10].

9. OECD (Organisation for Economic Cooperation and Development) (2020) The in vivo erythrocyte Pig-a gene mutation assay - Part 2 - Validation report. Available as Publication 316 [ENV/JM/MONO(2020)7] at: https://one.oecd.org/document/ENV/JM/MONO(2020)7/en/pdf

10. Miura D, VN Dobrovolsky, T Kimoto, Y Kasahara, RH Heflich (2009) Accumulation and persistence of *Pig-A* mutant peripheral red blood cells following treatment of rats with single and split doses of *N*-ethyl-*N*-nitrosourea. Mutat Res 677:86-92.

11. Phonethepswath S, D Franklin, DK Torous, SM Bryce, JC Bemis, S Raja, S Avlasevich, P Weller, O Hyrien, J Palis, JT MacGregor, SD Dertinger (2010) *Pig-a* mutation: Kinetics in rat erythrocytes following exposure to five prototypical mutagens. Toxicol Sci 114:59-70.

12. Dertinger SD, SL Avlasevich, DK Torous, JC Bemis, S Phonethepswath, C Labash, K Carlson, J Mereness, J Cottom, J Palis, JT MacGregor (2014) Persistence of cisplatin-induced mutagenicity in hematopoietic stem cells: Implications for secondary cancer risk following chemotherapy. Toxicol Sci 140:307-314.

13. Elhajouji A, TT Hove, O O'Connell, H Martus, SD Dertinger (2020) *Pig-a* gene mutation assay study design: Critical assessment of 3- versus 28-day repeat-dose treatment schedules. Mutagenesis 35:349-

© OECD (2025)

358. Available at: https://doi.org/10.1093/mutage/geaa014.

14. OECD (Organisation for Economic Cooperation and Development) (2008) OECD Guidelines for the testing of chemicals: Repeated dose 28-day oral toxicity study in rodents. Test Guideline 407. Available at: https://www.oecd.org/en/publications/2008/10/test-no-407-repeated-dose-28-day-oral-toxicity-study-in-rodents_g1gh292f.html

15. OECD (Organisation for Economic Cooperation and Development) (2018) OECD Guidelines on the testing of chemicals: 28-day (subacute) inhalation toxicity study. Test Guideline 412. Available at: https://www.oecd.org/en/publications/2018/06/test-no-412-subacute-inhalation-toxicity-28-day-study_g1gh2939.html .

16. OECD (Organisation for Economic Cooperation and Development) (2020) OECD Guideline for testing of chemicals: Transgenic rodent somatic and germ cell gene mutation assays. Test Guideline 488. Available at: https://www.oecd.org/en/publications/2022/06/test-no-488-transgenic-rodent-somatic-and-germ-cell-gene-mutation-assays_g1g3404a.html .

17. OECD (Organisation for Economic Cooperation and Development) (2018) OECD Guideline for the testing of chemicals: Repeated dose 90-day oral toxicity study in rodents. Test Guideline 408. Available at: https://www.oecd.org/en/publications/2018/06/test-no-408-repeated-dose-90-day-oral-toxicity-study-in-rodents_g1gh2931.html .

18. OECD (Organisation for Economic Cooperation and Development) (2018) OECD Guidelines for the testing of chemicals: 90-day (subchronic) inhalation toxicity study. Test Guideline 413. Available at: https://www.oecd.org/en/publications/2018/06/test-no-413-subchronic-inhalation-toxicity-90-day-study_g1gh293b.html .

19. Dertinger SD, S Phonethepswath, D Franklin, P Weller, DK Torous, SM Bryce, S Avlasevich, JC Bemis, O Hyrien, J Palis, JT MacGregor (2010) Integration of mutation and chromosome damage endpoints into 28-day repeat dose toxicology studies. Toxicol Sci 115:401-411.

20. Kirkland D, DD Levy, MJ LeBaron, MJ Aardema, C Beevers, J Bhalli, GR Douglas, PA Escobar, CS Farabaugh, M Guerard, GE Johnson, R Kulkarni, F Le Curieux, AS Long, J Lott, DP Lovell, M Luijten, F Marchetti, JJ Nicolette, S Pfuhler, DJ Roberts, LF Stankowski Jr, V Thybaud, SK Weiner, A Williams, KL Witt, R Young (2019) A comparison of transgenic rodent mutation and *in vivo* comet assay responses for 91 chemicals. Mutat Res 839:21-35.

21. Kimoto T, K Suzuki, Xm Kobayashi, VN Dobrovolsky, RH Heflich, D Miura, Y Kasahara (2011) Manifestation of *Pig-a* mutant bone marrow erythroids and peripheral blood erythrocytes in mice treated with *N*-ethyl-*N*-nitrosourea: Direct sequencing of *Pig-a* cDNA from bone marrow cells negative for GPI-anchored protein expression. Mutat Res 723:36-42.

22. Revollo JR, MG Pearce, A Dad, DM Petibone, TW Robison, D Roberts, VN Dobrovolsky (2018) Analysis of mutation in the rat *Pig-a* assay: I) Studies with bone marrow erythroid cells. Environ Mol Mutagenesis 59:722-732.

23. Revollo JR, A Dad, MG Pearce, RA Mittelstaedt, TW Robison, VN Dobrovolsky (2019) *Pig-a* mutations in bone marrow erythroblasts of rats treated with 7,12-dimethyl-benz[*a*]anthracene. Mutat Res 848:503106.

24. Revollo JR, A Dad, MG Pearce, RA Mittelstaedt, A Casildo, RG Lapidus, TW Robison, VN Dobrovolsky (2020) CD59-deficient bone marrow erythroid cells from rats treated with procarbazine and propyl-nitrosourea have mutations in the *Pig-a* gene. Environ Mol Mutagenesis 61:797-806.

25. Olsen A-K, SD Dertinger. CT Krüger, DM Eide, C Instanes, G Brunborg, A Hartwig, A Graupner (2017) The *Pig-a* gene mutation assay in mice and human cells: A review. Basic Clin Pharmacol Toxicol

© OECD (2025)

121:78-92. Available at: https://doi.org/10.1111/bcpt.12806.

26. Bemis JC, RH Heflich (2019) *In vitro* mammalian cell mutation assays based on the *Pig-a* gene: A report of the 7th International Workshop on Genotoxicity Testing (IWGT) workgroup. Mutat Res 847:403028.

27. OECD (Organisation for Economic Cooperation and Development) (2009) Preliminary review of OECD Test Guidelines for their applicability to manufactured nanomaterials. Available at: https://www.oecd.org/en/publications/preliminary-review-of-oecd-test-guidelines-for-their-applicability-to-manufactured-nanomaterials_b1b0e314-en.html .

28. OECD (Organisation for Economic Cooperation and Development) (2014) Genotoxicity of manufactured nanomaterials: Report of the OECD Expert Meeting. Series on the Safety of Manufactured Nanomaterials, No. 43 [ENV/JM/MONO(2014)34]. Available at: https://one.oecd.org/document/ENV/JM/MONO(2014)34/en/pdf .

29. Elespuru R, S Pfuhler, MJ Aardema, T Chen, SH Doak, A Doherty, CS Farabaugh, J Kenny, M Manjanatha, B Mahadevan, MM Moore, G Ouédraogo, LF Stankowski Jr, JY Tanir (2018) Genotoxicity assessment of nanomaterials: Recommendations on best practices, assays, and methods. Toxicol Sci 164:391-416.

30. Kinoshita T (2014) Biosynthesis and deficiencies of glycosylphosphatidylinositol. Proc Jpn Acad, Ser B 90:130-143.

31. Dad A, VN Dobrovolsky, RH Heflich, JR Revollo (2020) Molecular analysis of GPI-anchor biosynthesis pathway genes in rat strains used for the *Pig-a* gene mutation assay. Mutat Res 858-860:503256.

32. Giknis MLA, CB Clifford (2006) Clinical laboratory parameters for Crl:CD(SD) rats. Charles River Laboratories. pp 1-18. Available at: https://www.criver.com/sites/default/files/resources/rm_rm_r_clinical_parameters_cd_rat_06.pdf.

33. Miura D, VN Dobrovolsky, Y Kasahara, Y Katsuura, RH Heflich (2008) Development of an in vivo gene mutation assay using the endogenous *Pig-A* gene: I. Flow cytometric detection of CD59-negative peripheral red blood cells and CD48-negative spleen T-cells from the rat. Environ Mol Mutagenesis 49:614-621.

34. Bryce SM, JC Bemis, SD Dertinger (2008) In vivo mutation assay based on the endogenous *Pig-a* locus. Environ Mol Mutagenesis 49:256-264.

35. Phonethepswath S, SM Bryce, JC Bemis, SD Dertinger (2008) Erythrocyte-based *Pig-a* gene mutation assay: Demonstration of cross-species potential. Mutat Res 657:122-126.

36. Dertinger SD, S Phonethepswath, SL Avlasevich, DK Torous, J Mereness, SM Bryce, JC Bemis, S Bell, P Weller, JT MacGregor (2012) Efficient monitoring of *in vivo Pig-a* gene mutation and chromosomal damage: Summary of 7 published studies and results from 11 new reference compounds. Toxicol Sci 130:328-348.

37. Kenyon MO, SL Coffing, JI Ackerman, WC Gunther, SD Dertinger, K Criswell, KL Dobo (2015) Compensatory erythropoiesis has no impact on the outcome of the *in vivo Pig-a* mutation assay in rats following treatment with the haemolytic agent 2-butoxyethanol. Mutagenesis 30:325-334.

38. Raschke M, B-W Igl, J Kenny, J Collins, SD Dertinger, C Labash, JA Bhalli, CCM Tebbe, KM McNeil, A Sutter (2016) In vivo *Pig-a* gene mutation assay: Guidance for 3Rs-friendly implementation. Environ Mol Mutagenesis 57:678-686.

39. Labash C, SL Avlasevich, K Carlson, A Berg, DK Torous, SM Bryce, JC Bemis, JT MacGregor, SD Dertinger (2016) Mouse *Pig-a* and micronucleus assays respond to *N*-ethyl-*N*-nitrosourea,

© OECD (2025)

benzo[*a*]pyrene, and ethyl carbamate, but not pyrene or methyl carbamate. Environ Mol Mutagenesis 57:28-40.

40. EU (European Union) (2010) Directive 2010/63/EU of the European Parliament and the Council of 22 September 2010 on the protection of animals used for scientific purposes. Available at: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:276:0033:0079:en:PDF .

41. Dertinger SD, S Phonethepswath, P Weller, S Avlasevich, DK Torous, JA Mereness, SM Bryce, JC Bemis, S Bell, S Portugal, M Aylott, JT MacGregor (2011b) Interlaboratory *Pig-a* gene mutation assay trial: Studies of 1,3-propane sultone with immunomagnetic enrichment of mutant erythrocytes. Environ Mol Mutagenesis 52:748-755.

42. Mittelstaedt RA, VN Dobrovolsky, JR Revollo, MG Pearce, Y Wang, A Dad, PB McKinzie, H Rosenfeldt, B Yucesoy, R Yeager, S-C Hu, Y Tang, S Min, H-K Kang, D-J Yang, M Basavarajappa, RH Heflich (2019) Evaluation of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) mutagenicity using *in vitro* and *in vivo Pig-a* assays. Mutat Res 837:65-72.

43. Dertinger SD, SL Avlasevich, DK Torous, P Singh, S Khanal, C Kirby, A Drake, JT MacGregor, JC Bemis (2019) 3Rs friendly study designs facilitate rat liver and blood micronucleus assays and *Pig-a* gene mutation assessment: Proof-of-concept with 13 reference chemicals. Environ Mol Mutagenesis 60:704-739.

44. OECD (Organisation for Economic Cooperation and Development) (2000) Guidance document on the recognition, assessment, and use of clinical signs as humane endpoints for experimental animals used in safety evaluation. Available at: https://www.oecd.org/en/publications/guidance-document-on-the-recognition-assessment-and-use-of-clinical-signs-as-human-endpoints-for-experimental-animals-used-in-safety-evaluation_9789264078376-en.html .

45. OECD (Organisation for Economic Cooperation and Development) (2016b) OECD Guideline for the testing of chemicals: *In vivo* mammalian alkaline comet assay. Test Guideline 489. Available at: https://www.oecd.org/en/publications/2016/07/test-no-489-in-vivo-mammalian-alkaline-comet-assay_g1g6fb2c.html.

46. US Federal Register (1996) International Conference on Harmonisation; Guidance on specific aspects of regulatory genotoxicity tests for pharmaceuticals. Vol 61 (no. 80):18198-18202. Department of Health and Human Services, Food and Drug administration. Available at: https://www.govinfo.gov/content/pkg/FR-1996-04-24/html/96-10021.htm.

47. Marenzana M, TR Arnett (2013) The key role of the blood supply to bone. Bone Research 3:203-215.

48. Grüneboom A, I Hawwari, D Weidner, S Culemann et al. (2019) A network of trans-cortical capillaries as mainstay for blood circulation in long bones. Nature Metabolism 1:236-250.

49. Kirkland D, Y Uno, M Luijten, C Beevers, J van Benthem, B Burlinson, S Dertinger, GR Douglas, S Hamada, K Horibata, DP Lovell, M Manjanatha, H-J Martus, N Mei, T Morita, W Ohyama, A Williams (2019) *In vivo* genotoxicity testing strategies: Report from the 7th International Workshop on Genotoxicity Testing (IWGT). Mutat Res 847:403035.

50. Avlasevich SL, DK Torous, JC Bemis, JA Bhalli, CC Tebbe, J Noteboom, D Thomas, DJ Roberts, M Barragato, B Schneider, J Prattico, M Richardson, BB Gollapudi, SD Dertinger (2019) Suitability of long-term frozen rat blood samples for the interrogation of *Pig-a* gene mutation by flow cytometry. Environ Mol Mutagenesis 60:47-55.

51. Dertinger SD, SM Bryce, S Phonethepswath, SL Avlasevich (2011) When pigs fly: Immunomagnetic separation facilitates rapid determination of *Pig-a* mutant frequency by flow cytometric analysis. Mutat Res 721:163-170.

© OECD (2025)

52. Kimoto T, S Chikura, K Suzuki, Xm Kobayashi, Y Itano, K Horibata, M Honma, VN Dobrovolsky, RH Heflich, D Miura, Y Kasahara (2011) Further development of the rat *Pig-a* mutation assay: Measuring rat *Pig-a* mutant bone marrow erythroids and a high throughput assay for mutant peripheral blood reticulocytes. Environ Mol Mutagenesis 52:774-783.

53. OECD (Organisation for Economic Cooperation and Development). (2017) Overview of the set of OECD Genetic Toxicology Test Guidelines and updates performed in 2014-2015. Series on Testing and Assessment, No. 238 – 2nd edition. Available at: https://www.oecd.org/en/publications/overview-of-the-set-of-oecd-genetic-toxicology-test-guidelines-and-updates-performed-in-2014-2015-second-edition_61eca5cd-en.html .

54. EFSA (European Food Safety Authority) Scientific Committee, A Hardy, D Benford, T Halldorsson, M Jeger, HK Knutsen, S More, H Naegeli, H Noteborn, C Ockleford, A Ricci, G Rychen, V Silano, R Solecki, D Turck, M Younes, G Aquilina, R Crebelli, R Gürtler, KI Hirsch-Ernst, P Mosesso, E Nielsen, J van Benthem, M Carfi, N Georgiadis, D Maurici, J Parra Morte, J Schlatter (2017) Scientific Opinion: Clarification of some aspects related to genotoxicity assessment. EFSA Journal 15:5113. Available at: https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2017.5113 .

55. Shin HK, YM Kang, KT No (2017) Predicting ADME Properties of Chemicals. In: Leszczynski J, A Kaczmarek-Kedziera, T Puzyn, MG Papadopoulos, H Reis, MK Shukla (eds). Handbook of Computational Chemistry. Springer, Cham. Available at: https://doi.org/10.1007/978-3-319-27282-5_59.

56. OECD (Organisation for Economic Cooperation and Development) (2016) OECD Guideline for the testing of chemicals: Mammalian erythrocyte micronucleus test. Test Guideline 474. Available at: https://www.oecd.org/en/publications/2016/07/test-no-474-mammalian-erythrocyte-micronucleus-test_g1g6fb16.html .

57. Torous DK, SL Avlasevich, MG Khattab, A Baig, LJ Saubermann, Y Chen, JC Bemis, DP Lovell, VE Walker, JT MacGregor, SD Dertinger (2020) Human blood *PIG-A* mutation and micronucleated reticulocyte flow cytometric assay: Method optimization and evaluation of intra- and inter-subject variation. Environ Mol Mutagen 61:807-819.

58. Lawrence R, K Munn, H Naser, L Thomas, H Haboubi, L Williams, S Doak, G Jenkins (2023) The *PIG-A* gene mutation assay in human biomonitoring and disease. Environ Mol Mutagen 64:480-493.

© OECD (2025)

Annex 1 - Definitions and acronyms

3Rs = Replacement, reduction, and refinement; these principles provide a framework for performing more humane animal research

ADME = absorption, distribution, metabolism, and excretion

Clonal expansion = over-representation of the progeny of a cell in a population, usually due to the accelerated growth of that cell and its progeny relative to other cells in the population

Confidence interval (CI) = a range of values that is likely to include a population value with a certain degree of confidence

Control limit = horizontal line(s) drawn on a statistical control chart; these are investigator-defined and use-case-dependent values; in the field of Quality Control, these values are typically the sample mean ± 3 standard deviations, but some other multiples of the standard deviation can be useful *e.g.*, 2x or 1.96x

Erythrocytes = red blood cells, irrespective of age; for the purposes of this document, this term is interchangeable with RBC

Erythroid mutagen = a chemical capable of altering the DNA sequence of erythroid cells

Flow cytometry = the determination of the characteristics of a population of cells using a flow cytometer

GPI = glycosylphosphatidylinositol, a type of cell membrane anchor that links approximately 150 unique proteins to the exoplasmic surface of mammalian cytoplasmic membranes; *Pig-a* is among the genes necessary to form GPI anchors

Jackpot mutation = a germ cell mutation or more likely a mutation in a bone marrow erythroid precursor cell that results in a mutant frequency that is unusually high and outside the historical negative control distribution

Mutant mimic = erythrocytes that are processed in such a manner that they exhibit the same fluorescence characteristics as *bona fide Pig-a* mutant erythrocytes; these cells are often generated by processing extra blood from a vehicle/solvent control animal and omitting the fluorescent GPI-anchored antibody(s) from the fluorescent antibody labelling protocol; mutant mimics are valuable for guiding instrumentation settings and software/data analysis parameters

MUT RBC = mutant red blood cells; for the purposes of this document, the term describes circulating erythrocytes, irrespective of age, that exhibit the *Pig-a* mutant phenotype (*i.e.*, loss of GPI anchors)

MUT RET = mutant reticulocytes; for the purposes of this document, the term describes the most immature fraction of circulating erythrocytes, identified by their residual RNA and/or presence of the CD71 surface marker, that exhibit the *Pig-a* mutant phenotype (*i.e.*, loss of GPI anchors)

Negative control = in most instances solvent- or vehicle-dosed groups of animals used to identify significant substance-related effects in the *Pig-a* assay. In some instances, *e.g.*, establishing a negative control data base, pre-dose data may substitute for data from solvent- or vehicle-treated animals

Neutral phenotype = a term used to describe mutant cells that are not highly selected for or against *in vivo*, *e.g.*, cells that grow at the same rate as cells from the predominant, wild-type cell population

***Pig-a* gene** = phosphatidylinositol glycan class A gene

© OECD (2025)

Prediction interval = an estimate of an interval in which a future observation will fall, with a certain probability, given what has already been observed

Promutagen = a chemical requiring structural alteration, usually by enzymatic metabolism, to induce mutations

Quantile = divisions of a probability distribution or frequency distribution into equal, ordered subgroups, for example quartiles or percentiles

RBC = red blood cells; for the purposes of this document, it encompasses all circulating erythrocytes, irrespective of age

RET = reticulocytes (in some contexts called polychromatic erythrocytes), an immature erythrocyte developed from bone marrow erythroblasts that is identified by its residual RNA and/or presence of the CD71 surface marker

Residual distribution = residuals are estimates of experimental error obtained by subtracting the observed responses from the predicted responses; the predicted response is calculated from the chosen model, after all the unknown model parameters have been estimated from the experimental data; examining residual distribution is important because it provides information about model assumptions and appropriateness

Tolerance interval = statistical interval within which, with some confidence level, a specified proportion of a sampled population falls

UVBC = substances of unknown or variable composition, biological materials or complex reaction products

Variance homogeneity = this is an assumption of most parametric tests including the independent samples t-test and ANOVA, and states that all comparison groups have the same variance; the independent samples t-test and ANOVA utilise the *t* and *F* statistics respectively, which are generally robust to violations of the assumption as long as group sizes are equal

© OECD (2025)

Annex 2 - Combination of micronucleus and comet tests with *Pig-a* assays

1. Combination of the micronucleus and comet assays with Pig-a assays employing a 28-day dosing protocol

Micronucleus and comet tests often are conducted hours to day(s) following short-term dosing protocols (*e.g.*, two or three consecutive days of exposure). However, TG 474 for the *in vivo* micronucleus test [OECD, 2016] and TG 489 for the *in vivo* alkaline comet test [OECD, 2016b] also describe integration of these genotoxicity tests into longer, repeat-dosing protocols. Thus, these two genotoxicity tests can be combined with *Pig-a* assays that use the recommended 28-day repeat-dosing protocol.

Blood sampling for blood-based assays, such as the *in vivo* micronucleus assay, can be performed early in the *Pig-a* repeat-dose treatment schedule (*e.g.*, on Day 4). In addition, the blood-based micronucleus assay and the comet assay on multiple tissues can be conducted after the 28-day treatment protocol is complete, often at Day 29, with the addition of a dose on Day 29 prior to tissue harvest for conducting the comet assay.

2. Combination of micronucleus and comet assays with Pig-a assays employing a short-term dosing protocol

While it is technically possible to combine the *Pig-a* assay with the micronucleus and comet assay while using a single, short-term treatment protocol for the tests, two important considerations should be addressed. First, when acute dosing schedules are used to generate *Pig-a* data, such studies should be accompanied by scientific justification for measuring the *Pig-a* endpoint as they generally result in the animals being tested with higher daily doses, but a lower total dose of test chemical than with a 28-day repeat-dosing protocol. Such treatments generally result in lower *Pig-a* mutant frequencies than generated by longer-term treatment schedules. Second, the *Pig-a* assay requires an extended period for mutant manifestation - usually a minimum of 2 to 4 weeks [OECD, 2020], which is much longer than the duration of a short-term treatment comet or micronucleus test.

The following are three possible designs for combining micronucleus and comet assays with *Pig-a* assays employing short-term treatments:

1. Use a single set of animals and a single short-term treatment (*e.g.*, 2 or 3 consecutive days of treatment). Using this protocol, combining the *Pig-a* assay with micronucleus and/or comet tests limits the comet and micronucleus tests to blood-based analyses that can be accomplished without euthanizing the animal. Thus, blood-based micronucleus and comet tests are conducted following the short-term treatment, and the animals are held for an additional period (*e.g.*, a total of 28 days) for manifestation of *Pig-a* mutants. Limiting the analysis to blood is problematic for the comet test, which has displayed reduced accuracy for detecting DNA damage in blood [Kirkland et al., 2019]. Thus, using this protocol, the *Pig-a* assay combines much more readily with short-term blood-based micronucleus tests than it does with short-term comet tests.
2. Incorporate a second round of short-term treatment(s) during the final days of the *Pig-a* mutant manifestation period. The comet and micronucleus tests, as well as the *Pig-a* assay, then can be performed after this second round of short-term exposure(s). Following this protocol, the *Pig-a* assay would mainly be measuring mutations induced by the first round of treatment(s),

© OECD (2025)

with the micronucleus and comet tests measuring genotoxicity induced by the second round of treatment(s). Note that if desired, a blood-based micronucleus test also can be performed following the first round of short-term treatment(s).

3. A less 3Rs-friendly alternative consists of a single short-term treatment, with half the animals euthanised shortly thereafter to perform the comet and/or micronucleus tests as described in TG474 and TG 489. In order to perform the *Pig-a* assay, the remaining animals are maintained until later blood collection times, *e.g.*, Day 15 and Day 30 as described in paragraph 45. This ensures that animals are available that have an adequate time for manifestation of the *Pig-a* mutant phenotype.

References for Annex 2, in order of appearance

OECD (Organisation for Economic Cooperation and Development) (2016) OECD Guideline for the testing of chemicals: Mammalian erythrocyte micronucleus test. Test Guideline 474. Available at: https://www.oecd.org/en/publications/2016/07/test-no-474-mammalian-erythrocyte-micronucleus-test_g1g6fb16.html.

OECD (Organisation for Economic Cooperation and Development) (2016b) OECD Guideline for the testing of chemicals: *In vivo* mammalian alkaline comet assay. Test Guideline 489. Available at: https://www.oecd.org/en/publications/2016/07/test-no-489-in-vivo-mammalian-alkaline-comet-assay_g1g6fb2c.html.

OECD (Organisation for Economic Cooperation and Development) (2020) The in vivo erythrocyte *Pig-a* gene mutation assay - Part 1: Detailed Review Paper and Retrospective Performance Assessment. Available as Publication 315 at: http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm.

Kirkland D, DD Levy, MJ LeBaron, MJ Aardema, C Beevers, J Bhalli, GR Douglas, PA Escobar, CS Farabaugh, M Guerard, GE Johnson, R Kulkarni, F Le Curieux, AS Long, J Lott, DP Lovell, M Luijten, F Marchetti, JJ Nicolette, S Pfuhler, DJ Roberts, LF Stankowski Jr, V Thybaud, SK Weiner, A Williams, KL Witt, R Young (2019) A comparison of transgenic rodent mutation and *in vivo* comet assay responses for 91 chemicals. Mutat Res 839:21-35.

© OECD (2025)

Annex 3 - Animal-friendly training exercises for learning techniques used in the *Pig-a* assay

A 3Rs-friendly strategy for learning animal-treatment, blood-collection, blood-processing, and flow-cytometric analysis techniques used in the *Pig-a* assay has been described by [Raschke et al., 2016]. For this strategy, a single rodent is exposed to a known, potent mutagenic chemical, *e.g.*, one listed in Table 1. After an appropriate phenotypic expression time that allows elevated MUT RET and MUT RBC frequencies to appear in the peripheral blood circulation, blood from the exposed rodent can be collected along with blood from a sex- and age-matched untreated or vehicle/solvent control animal. A series of mixtures of the two blood samples is made that contains various amounts of blood from the mutagen-exposed animal (*i.e*., spiked samples). After determining MUT RET and MUT RBC frequencies for the rodent dosed with a mutagen and the negative control rodent, expected (intermediate) frequencies can be calculated for the spiked samples based on the proportion of blood from the rodent dosed with a mutagen added to negative control blood. By conducting reconstruction experiments on 2 or more separate occasions, and with several replicates per spiked sample, staff can be trained on important elements of the test and the proficiency of staff members can be established by demonstrating good agreement between observed and expected mutant cell frequencies. Successful completion of spiking experiments represents a useful, 3Rs-friendly gateway to the laboratory proficiency investigations described in paragraph 22.

Reference for Annex 3

Raschke M, B-W Igl, J Kenny, J Collins, SD Dertinger, C Labash, JA Bhalli, CCM Tebbe, KM McNeil, A Sutter (2016) In vivo *Pig-a* gene mutation assay: Guidance for 3Rs-friendly implementation. Environ Mol Mutagenesis 57:678-686.

© OECD (2025)

Annex 4 - Statistical analysis and data interpretation

1. Preface

1. This Annex draws partly from an International Workshop on Genotoxicity Testing (IWGT) report [Dertinger et al., 2023]. For illustrative datasets and detailed method descriptions, readers are encouraged to review this publication.

2. Statistical analysis methods used in this assay can vary, and this Test Guideline provides three examples in paragraphs 68 - 69 as practical, non-prescriptive approaches. Alternative methods are acceptable if justified with sound statistical reasoning. Investigators should document their chosen methods in a written study or validation plan before starting the study and be prepared to defend their approach.

3. The statistical approaches below are representative tools for evaluating genotoxicity, but sound scientific judgement is essential, particularly when statistical analyses exhibit conflicting results. The application of sound scientific judgement is consistent with recommendations made by an OECD genotoxicity working group that met in Ottawa, Canada in 2013 [OECD, 2017].

4. Considering the language in Test Guideline paragraphs 68 - 69 and the data evaluation tools described in this Annex, the following overarching principles should be kept in mind:

i. When concurrent negative control data fulfil study acceptability criteria, they represent the most important comparator for judging whether a particular test chemical induced a genotoxic effect.
ii. Historical control data can provide useful context for interpreting study results, but this requires supporting evidence including: a) relevant historical control data that have been generated in sufficient quantity, and; b) data that accurately describes inter-animal variability.
iii. Historical control data should be visualized before any study comparisons take place; graph(s) that show the degree to which the data are stable over time are particularly useful.
iv. Qualitative and semi-quantitative assessments of historical control data should be supplemented with quantitative evaluations including: a) the stability of the data over time, and b) the degree to which inter-study variation explains the total variability observed.
v. When inter-animal variation is the predominant source of variability, the relationship between study responses and a historical control-derived interval or upper bounds value (*i.e.,* criterion c) can be used with a strong degree of confidence in contextualizing the results.
vi. When inter-study variation is the major source of variability, comparisons between study data and the historical control-derived bounds are less useful, and consequentially, less emphasis should be placed on using historical control data to contextualize a particular study's results.
vii. Most statistical tests, including those in this Annex, assume random sampling or blocking to mitigate uncontrolled factors on study results.

2. Pairwise comparisons

5. One type of statistical test that may be implemented are pairwise comparisons that determine if the concurrent vehicle/solvent (negative) control group is significantly different from the test chemical dose groups. Parametric analyses that use analysis of variance (ANOVA) with an appropriate multiple comparisons test are commonly used, but other methodologies are equally acceptable so long as assumptions underlying the analyses are met. For example, if data transformation does not result in homoscedasticity, weighted (variance-corrected) ANOVA and/or t-tests or appropriate non-parametric methods can be considered.

© OECD (2025)

2.1 Factorial design

6. If more than one sex is used in a study, factorial design approaches are generally advantageous as it evaluates the effects of sex and test chemical dose simultaneously and examines how these factors interact. Standard statistical software can be used to analyse the data generated from this design.

7. The analysis first inspects the sex-by-dose interaction term in the ANOVA table. (Note that statisticians who take a modelling approach such as using General Linear Models may approach the analysis in a different but comparable way but will not necessarily derive the traditional ANOVA table). In the absence of a significant interaction term, the combined values across sexes or across dose levels provide valid statistical tests between the levels based upon the pooled within-group variability term of the ANOVA.

8. The analysis continues by partitioning the estimate of the between-dose-level variability into contrasts, which test for linear and quadratic contrasts of the responses across the vehicle/solvent concurrent control and test chemical dose levels. When there is a significant sex x dose interaction, this term can be partitioned into linear x sex and quadratic x sex interaction contrasts. These terms test whether the dose level responses are parallel for the two sexes or there is a differential response between the two sexes.

9. The pooled within-group variability estimate allows pairwise tests between means, such as between sexes or dose levels (*e.g.,* against negative (solvent/vehicle) control levels). Where there is a significant interaction, comparisons can be made between the means of different dose levels within a sex or sexes within a dose level.

3. Trend test

10. A second type of analysis that may be used is the trend test to identify a dose-response relationship. For this test, data should be available from the negative control group and all experimental dose groups. However, trend tests are normally not applicable for Limit Tests, which involve only a single test chemical dose level as described in paragraph 52). When employing these analyses, care is needed in interpreting the results. For example, a simple linear trend test may fail to detect a trend if the dose-response is non-monotonic. In such cases, more advanced tests, like the downturn protection test proposed by Bretz and Hothorn [2003] may be useful. Creating a graph of the dose response can aid in choosing the appropriate test or determining if a transformation is needed.

4. Historical negative control data distribution

11. Distributions of historical negative control data are important for determining the minimum number of cells that should be evaluated and for assessing a particular study's acceptability and also when determining whether the results of any test chemical-exposed treatment group exceeds the upper bounds of the historical negative control data distribution ("criterion c"). Given their importance, guidance for building historical negative control datasets is provided below.

4.1 Building historical negative control datasets: Group means versus individual animal data

12. There is value in compiling historical negative control data based on both individual animal data and as study group means. First, individual animal data facilitate certain statistical analyses that cannot

© OECD (2025)

be accomplished with group means. One example is Variance Components Analysis aimed at estimating the extent to which inter-animal differences versus study-dependent nuisance factors explain the variation observed in an assay. This important consideration cannot be addressed when analyses are based on group mean data. Second, initially focusing on individual animal data represents a practical recommendation because, for some *in vivo* test systems, it may be difficult to achieve adequately sized databases with group means due to limitations on animal testing.

13. As a laboratory gains experience and conducts more studies, the historical control databases can expand to include both individual animal data and group mean data. Individual animal data will remain essential for evaluating the quality and stability of historical negative control data. However, the group mean historical negative control data distribution may replace the individual animal data distribution in the context of criterion c-type assessments.

14. While individual animal- and group mean-based databases are interrelated, intervals for historical negative control data based on individual animals are wider than those based on group means. Therefore, a guiding principle is *like should be compared to like* meaning individual animal results should be compared with intervals based on individual animal data, while group mean results are most appropriately compared to intervals based on group means.

4.2 Database size

15. A historical negative control database (HCD) should consist of data from at least ≥ 20 independent studies and ≥ 100 individual animals. These values are consistent with recommendations by Hayashi et al. [2011] and Igl and colleagues [2019]. A database with < 20 independent studies and < 100 individuals may still be useful but should carry less weight compared to later versions derived from ≥ 20 independent studies and ≥ 100 individual animals.

16. The goal of achieving 20 independent studies and upwards of 100 animals is complicated by the fact that some endpoints may be affected by certain experimental factors such as species, strain, age, sex, vehicle composition, etc. For the TGR mutation assay, the tissue being analysed for mutations or DNA strand breaks should also be considered. When factor(s) are found to appreciably affect the endpoint in question, separate historical control databases may become necessary, which is an unrealistically large burden that may result in many small databases that never reach the goal of 100 observations. In these instances, a practical solution may become necessary, for example conducting a relatively small number of independent experiments, each with 10 animals, and each with a different, commonly used vehicle. Once these pilot studies are complete, and so long as the data are found to be "under control" (see Section 4.4, "Evaluating the Quality of Historical Control Data"), an appropriately calculated historical control interval based on individual animal data may be used for assay acceptability and addressing criterion c. However, as explained previously, analyses conducted with such data should carry less weight relative to later databases that are more appropriately sized.

17. In most instances, the historical negative control databases will be generated with data from animals dosed with the solvent/vehicle alone. For some endpoints, historical negative control databases can benefit from the inclusion of naïve animal data. This assumes solvent/vehicle effects are not appreciable. This example stresses the importance of supplementing the historical control genotoxicity data with the reporting of metadata that includes experimental factors such as animal model, strain, sex, solvent/vehicle, age at time of tissue harvest, and analytical equipment configuration.

18. Related to this, assuming supporting literature or other data are available, it may be appropriate to initially consider factor(s) such as sex and/or solvent/vehicle to have no effect on a genotoxicity endpoint. This helps one more efficiently construct an adequately sized database. However, as more

© OECD (2025)

data accumulate, these assumptions should be re-evaluated, and in this manner decisions about pooling or parsing of databases can be made on sound scientific reasoning.

4.3 Maintaining historical control databases

19. Historical control data can be from studies conducted under Good Laboratory Practices (GLP) guidelines, non-GLP studies, or from both, so long as the experimental conditions are similar, and quality control measures confirm accuracy of data not covered by a GLP-compliance statement. An established historical control database should be maintained and reissued at some consistent interval, or else by some other defined triggering event, unless no new data have been generated in the previous 12 months. While historical control databases may be updated on a predetermined schedule, assay acceptability criteria must be evaluated on a per study basis allowing for real time feedback on assay performance so as to be useful for modifying a fixed database update schedule.

20. For example, if a study does not meet assay acceptance criteria, this should trigger work aimed at identifying the underlying cause(s). If an underlying cause *is identified* (*e.g.,* an explainable technical error, reagent issue, or equipment failure), these assay control data should *not* be used in the next iteration of the historical control database. For the sake of transparency, such data should be added to the database, but omitted from the calculation of interval(s) that form the basis of assay acceptability criteria or for addressing criterion c.

21. If an assay fails assay acceptance criteria and the underlying cause *is not identified*, then a decision must be made as to whether the assay control data should be included in a future historical control database. To maximize transparency, the decision process for excluding control data from a database should most ideally be articulated *a priori*. If the unusual control data point(s) are deemed acceptable for inclusion in a future historical control database because no technical error, etc. was identified, this should ideally prompt a reissue of the database and associated analyses (*e.g.,* generation of a control chart, new interval calculation, etc.) ahead of schedule, so that the historical control database remains representative of upcoming studies.

22. Finally, a timeframe should be selected to analyse the historical control data for stability, such as 5 years as an upper bound. Accordingly, assay-specific factors will often need to be considered. For instance, if the assay is conducted infrequently, or a particular tissue is rarely studied, older data may be used to obtain approximately 20 studies and 100 independent observations, provided there have been no changes in methodology over the time period.

4.4 Evaluating the quality of historical negative control data

23. To evaluate the quality of HCD, a qualitative and semi-quantitative approach should be taken [OECD, 2017]. Initial assessments should evaluate whether (i) the genotoxicity data are consistent with published results from proficient laboratories (here, it can be useful to seek out interlaboratory validation/transferability-centric papers); (ii) the level of variation across samples within a study and across studies compare to published results from proficient laboratories; and (iii) there is an obvious drift to higher or lower frequencies with respect to time. Quantitative analyses should also be applied and usually begin by characterizing the distribution of the data, which can be useful as residuals that do not approximate a normal distribution should not support some of the analyses described below.

24. When data transformation is deemed necessary, computational tools may be utilized to select an appropriate transformation. For instance, the Box-Cox approach available in many statistical software [Box and Cox, 1964] calculates an estimate of the lambda power parameter, which can be used to identify an optimal transformation. However, simply rounding lambda to the nearest 0.5 value helps

© OECD (2025)

select from common transformations. For instance, a lambda of -1 = reciprocal, 0 = logarithmic, 0.5 = square root, 1 = untransformed, and 2 = squared.

25. The next analysis should use statistical tools widely used in the fields of manufacturing and process control. Control charts are a useful for studying historical control data [Hayashi et al., Mutat. 2011] and are recommended in several genetic toxicology OECD TGs [OECD, 2017]. A control chart is a graph used to study how a process changes over time. Data are plotted in a sequential order that reflects when they were collected over time. A control chart has a central line plotted parallel to the X-axis describing the mean, an upper line for the upper control limit (UCL; mean plus 3 standard deviations, *i.e*., sigma +3), and a lower line for the lower control limit (LCL; sigma -3). Graphs can show sigma ±2 lines, which are known as "warning limits." (Note that standard deviation values plotted on control charts are not the same standard deviation values from summary statistics of the sample; rather, control chart standard deviation values are estimated by the average range divided by a normalizing constant, d_2, see Dertinger et al., 2023). By comparing current data to these lines, conclusions can be drawn about whether a process is consistent ("under control") or variable ("out-of-control").

26. Different types of control charts with different objectives may be useful for genetic toxicology data are briefly described here. I-charts can be used with either continuous or discrete (count) data and are time-ordered sequence charts that plot individual values on the y-axis and the order of the individuals on the x-axis. C-charts also plot individual values where no attempt is made to take into consideration any grouping that may exist in the data. The primary difference between I- and C-charts is that the latter are restricted to counts (numbers of events) and assume an underlying Poisson distribution. X-bar charts are plots of the means derived from a series of experiments, each of which is made up of several individuals (*e.g.,* the mean of 5 animals in a negative control group).

27. The X-bar chart is generally more informative when there are a reasonable number of studies (*e.g.,* 20 or more), while the I-chart is useful for a smaller number of studies. However, caution is needed when using I-charts since the data may not be truly independent because, for instance, observations from the same study may have correlated responses. If a group falls outside the limits of an X-bar chart, it is probable that several individual values from the group will also fall outside the limits for the I-chart. Ultimately, these control charts should be seen as complementary tools rather than one being superior to the other.

28. Decision rules can be used with control charts to highlight when a process is out-of-control. Codified rules are used to ensure that various control charts are interpreted in a consistent manner. Western Electric rules and Nelson rules are two widely used examples that can highlight an out-of-control process based on non-random characteristics of the data [Nelson, 1984]. For instance, Nelson rule number 3 is "violated" when six (or more) data points in a row are continually increasing or decreasing.

29. Some statistical software programs supplement control charts with metrics developed for monitoring manufacturing processes, which quantify the variability of a process. One example is the Stability Index (SI) [Jensen et al., 2019] for non-grouped data that calculates by dividing the long-term standard deviation (*i.e.,* the total standard deviation of the database) by the short-term standard deviation (*i.e.,* the average of the moving range divided by d_2). SI values close to 1.0 are considered stable with low variability. Thus SI, together with Nelson or Western Electric rules (or subsets), provide warning signals that aspect(s) of the assay have drifted to, or are in the process of drifting to, an out-of-control status.

30. Variance Components Analysis (VCA) is another category of tests that can provide insights into the quality of historical control data. There are a variety of related methods that include hierarchical or nested ANOVA, Restricted Maximum Likelihood (REML) approaches, and Bayesian modelling methods [Gelman and Hill, 2006]. Another strategy, called Evaluating the Measurement Process, has been

© OECD (2025)

described by Wheeler [2006]. These methods partition and estimate the proportion of total variation accounted for by various experimental factors providing an estimate of standard deviation for the source.

4.5 Calculating intervals that describe historical negative control distributions

31. There are several valid approaches for characterizing the distribution of historical negative control data, and each laboratory should use an appropriate method for describing their data. This typically involves calculating upper and/or lower bounds to define the range where most negative control animal values are expected to fall. Key considerations include sample size and whether the data are normally distributed. A brief overview of methods for calculating an upper and/or lower bound limit is provided below with further details available in the literature [Vardeman, 1992; InfinityQS International, 2014; and Kluxen et al., 2021].

4.5.1. Inappropriate methods

- Range: The range is typically based on all negative control animal values and is the difference between the minimum and maximum observed value. Range *does not adequately* describe the historical negative control distribution for the purpose of establishing useful lower and upper bound limits. This is because the range will widen as the number of samples increases and may depend on one or two extreme (unusual/outlier) values. A wide range may "reward" poorly performing laboratories by masking potentially adverse study outcomes.

- Confidence interval: A confidence interval is a range of estimates likely to include a population value (parameter), such as the mean with a defined degree of confidence (*e.g.,* 95%). While formal definitions involve considerations of repeat sampling from hypothetical populations, these are often impractical to apply in real-world contexts. Importantly, confidence intervals are *not useful* for adequately describing historical negative control distribution when establishing an upper bound limit for criterion c. This is because confidence intervals become narrower as sample size increases, reflecting the increased precision in estimating the population value (parameter), which makes them *inappropriate* for OECD criterion c data interpretation.

4.5.2. Appropriate methods

32. Methods for calculating an interval that describe the distribution of historical negative control data are listed below. A database of ≥ 100 individual rodents/data points from ≥ 20 independent experiments is preferred. However, smaller intervals may suffice, particularly if the data are normally distributed or can be normalized, and the assay is well-control. Intervals from smaller datasets should be weighted less heavily than those from larger databases, as they are less likely to cover the desired proportion.

33. Although historical control data may fall within these intervals, this does not mean that the data are appropriate for criterion c-type assessments. Excessive variability can result in limit/interval values that are too wide. This would be indicative of poor performance by the laboratory and should not be mistaken for the "normal, biological variability" of the assay. This cautionary statement applies to all intervals described below and underscores the recommendation to use methods such as Nelsen or Western Electric rules, SI, and/or VCA to evaluate the quality of historical control data before calculating distribution intervals.

- Control limit: In the field of Quality Control, multiples of the standard deviation, usually the mean ± 3 standard deviations, are used as control limits. These values are lines plotted on a control chart and may be accompanied by other standard deviation multiples, for instance the mean ± 2 standard deviations, which are referred to as warning limits. In conjunction with control charts, control and warning limits are valuable tools for assessing the degree to which a repeated process or test is under

© OECD (2025)

control.

When setting up control charts, one should consider that the standard deviation used is not the same as the standard deviation from summary sample statistics. Multiplying this standard deviation by 2 or 3 yields values different from those used to create the warning and control limits. This discrepancy arises because statistical software often estimates the standard deviation using the "average moving range," as the true population standard deviation is unknown and needs to be estimated from the data. The sample estimate of the standard deviation is adjusted accordingly to minimize biases or errors.

Assuming a normal distribution and an "under control" process, 99.73% of the negative control data should fall within 3 standard deviations of the mean, and approximately 95% of the data should fall within two standard deviations. Control and/or warning limits represent a useful resource for evaluating historical negative control data and for providing upper and/or lower bound limits that aid in data interpretation.

When considering study data in the context of criterion c-type assessments, warning limits that describe where approximately 95% of normally distributed data fall (± 2 standard deviations) are generally more appropriate than control limits that are based on 3 standard deviations. The former is consistent with many OECD Test Guidelines (*e.g.,* OECD TG474, 2016), while the latter characterizes exceptionally rare/unusual data points, thereby generating intervals that are too wide for the criterion c use case. However, control limits based on 3 standard deviations may be useful for data acceptability criteria.

- Prediction intervals: Prediction intervals are designed to predict one, or several, future observation(s) based on existing data. For example, with a 95% prediction interval, a new result (or specified number of future observations) from a negative test article would be expected to fall in the prediction interval with 95% probability. Non-normal data should be transformed as necessary prior to calculating the interval, which can be back transformed to original units for use and reporting. Alternately, some computer software programs enable a non-parametric calculation that does not assume a normal distribution.

  Prediction intervals typically address a specified *number* of future observations, usually one or several, rather than a percentage of future observations. This may pose challenges for using a prediction interval to judge study validity for criterion c-type purposes, unless one accompanies the calculation with rationale for choosing a particular number of future observations.

- Tolerance intervals: Tolerance intervals, like prediction intervals, are forward-looking but are designed to predict future observations with specified coverage and confidence levels. Coverage refers to the percentage of future observations defined by the calculated lower and upper bounds (*e.g.,* 95% or 99%), modeled with a specified confidence level. While tolerance intervals are often calculated using 95% confidence, this parameter can be user defined.

  Considerations for normality and optional non-parametric approaches discussed for prediction intervals also apply to tolerance intervals. In general, tolerance intervals are wider than prediction intervals since they are designed to predict a high percentage of future values, while prediction intervals typically focus on estimating one or a few new observations.

- Quantiles: Quantiles are used for summarizing the rank of data points based on size without assuming any specific probability distribution. Quantiles are widely used in many biomedical applications where non-normality because of outliers and/or skewness is common. For instance, intervals based on percentiles are often used to help interpret test results. Confidence intervals for quantiles can be calculated to provide estimates of uncertainty around the quantile measurement. These can help evaluate the quality of the underlying data set. Quantile confidence intervals will be especially wide for the tails of the distribution unless the sample size is large.

© OECD (2025)

4.6 Using historical control data to interpret study data

34. Low quality historical control data are not suitable for criterion c-type assessments. With this in mind, this section describes the use of historical control data to interpret study data and is based on the assumption that the database is large enough, and of an acceptable quality, for this purpose.

35. A common approach for fulfilling criterion c involves conducting a large number of studies, then using the study mean values to calculate a historical control distribution-based interval (or else just the upper bound limit value). In this case, the aforementioned “like to like” comparison translates to a straight-forward assessment. That is, for any particular study, the treatment group mean values are compared to the upper bound limit value (derived from historical negative control study means) that separates “expected” from “elevated” values. This is a valid and useful approach, with the caveat that such a strategy requires many separate studies and consequentially large numbers of rodents, and the use of laboratory animals for this and other purposes is becoming more challenging in many regions of the world.

36. As discussed previously, it is also possible to use *individual* animal data to evaluate the quality of historical control data, and to calculate a useful interval (or else just an upper bound limit value) that can be used for criterion c purposes (at least until enough studies have been performed to justify a study mean-based interval). However, after arriving at an upper bound that separates expected from elevated values, it may be less obvious how one should proceed to evaluate study results on a per animal basis. In this situation, make use of well documented, sound scientific reasoning.

37. When sound scientific reasoning is used to assess individual animal data against an appropriately derived upper bound limit value, at least four factors should be considered. First, how many animals and numbers of studies were used to calculate the interval or upper bound limit? For example, a database and upper limit based on 10 studies and 50 animals should carry more weight relative to 5 studies and 25 animals. Second, how many test chemical treated animal responses were above the upper bound value? That is, the greater the number, the higher the concern. Third, which treatment group(s) exhibited elevated values? For example, all the elevated value(s) appearing in the low dose group may be indicative of a technical issue as opposed to a treatment-related effect. Fourth, what was the magnitude of the elevated response(s)? That is, greatly elevated values would be of higher concern than values that barely exceeded the upper bound limit.

38. This is not an exhaustive list of factors that should be considered. Rather, this subset focuses on the interpretation of elevated individual animal responses against an upper bound limit, and how they may contribute to the contextualization of statistical analyses (*i.e.,* criteria a and b). Other important considerations, such as the distribution and metabolism of a test substance, kinetics of appearance and disappearance of the endpoint, etc., are additional important elements that must be considered when assessing biological plausibility.

5. Interpretation of study results

39. The statistical approaches described in the preceding sections are regarded as representative tools for evaluating whether the test chemical induces genotoxicity. That being said, sound scientific judgement must take a leading role when considering any data set. This is especially obvious when statistical analyses exhibit, what on the surface, may appear to be conflicting results. This paradigm was reinforced by an expert OECD genotoxicity working group that met in Ottawa, Canada, 2013 [OECD, 2017].

40. In some cases, even after applying scientific judgement, appropriate statistical tools, and possibly evaluating additional data, it will not be possible to classify a response as positive or negative.

© OECD (2025)

In these cases, the response is equivocal and further testing may be required to resolve the genotoxicity of the test chemical.

6. Representative historical negative control table

41. As indicated in the Test Report section, study reports should provide information regarding the manner in which historical negative control data were collected, evaluated, and utilised. The following table is offered as a guideline.

Annex 4, Table 1. Representative historical negative control table*

	Information	**Description and Comments**
1. General description	Date range	Period over which the HCD were collected.
	Test system	The endpoint investigated and the model used. For an in vivo study, this may include the species, strain, sex, and age range of the animals at the time of dose initiation and tissue harvest, and the randomization or blocking procedures employed.
	Analytical system	Method of data collection.
2. Basic metrics	Number of studies and observations	Number of studies the data were collected from, and the total number of observations used in the HCD, including a statement on whether the observations are individual animals or group means.
	Median, mean** and Std. Dev.	Median and mean of the HCD, with its standard deviation. **Alternative measures of central tendency may be used.
	Observed range	Minimum and maximum values observed in the HCD.
	Group structures	Describe the rationale for combining or separating historical control databases based on experimental factors such as sex, vehicles, dose routes, age groups, amount of available data, and so forth.
3. Quality metrics	Transformation	A description of the analysis performed to select the appropriate transformation and justification for the final transformation selected. Visualization of the transformed HCD is recommended.
	Control chart	A longitudinal graphical display of the HCD, that includes the type of control chart presented and displays both warning and control limits.
	Variability over time	An assessment of the stability of HCD over time, including a brief description of the methods used.
	Sources of variation	An assessment of the predominant source of variability in the HCD, including a brief description of the methods used.
	Distribution interval	Interval that describes the distribution of the HCD, together with a brief description of

© OECD (2025)

		the method used to calculate the interval and a justification for why the specific interval was selected.

*Given the amount and type of information requested above, one practical suggestion would be for sponsors to supply these data in an appendix or annex to a regulatory study report. It would remain consistent for some period of over time, until the historical negative control database has been updated with new data/distributions.

7. References for Annex 4, in order of appearance

Dertinger SD, D Li, C Beevers, GR Douglas, RH Heflich, DP Lovell, DJ Roberts, R Smith, Y Uno, A Williams, KL Witt, A Zeller, C Zhou (2023) Assessing the quality and making appropriate use of historical negative control data: A report of the International Workshop on Genotoxicity Testing (IWGT). Environ Mol Mutagen, in press, doi: 10.1002/em.22541.

OECD (Organisation for Economic Cooperation and Development). (2017) Overview of the set of OECD Genetic Toxicology Test Guidelines and updates performed in 2014-2015. Series on Testing and Assessment, No. 238 – 2nd edition. Available at: https://www.oecd.org/en/publications/overview-of-the-set-of-oecd-genetic-toxicology-test-guidelines-and-updates-performed-in-2014-2015-second-edition_61eca5cd-en.html.

Bretz F, LA Hothorn (2003) Statistical analysis of monotone or non-monotone dose-response data from in vitro toxicological assays. ATLA-Altern Lab Animal 31(Suppl 1):81-96.

Hayashi M, K Dearfield, P Kasper, D Lovell, H-J Martus, V Thybaud (2011) Compilation and use of genetic toxicity historical control data. Mutat Res 723:87-90.

Igl B-W, A Bitsch, F Bringezu et al. (2019) The rat bone marrow micronucleus test: Statistical considerations on historical negative control data. Regul Toxicol Pharmacol 102:13-22.

Box GEP, DR Cox (1964) An Analysis of Transformations, Journal of the Royal Statistical Society: Series B (Methodological), 26(2):211-243.

LS Nelson (1984) The Shewhart Control Chart—Tests for Special Causes. Journal of Quality Technology, 16(4):237-239.

WA Jensen, J Szarka III, J., K White (2019). Stability assessment with the stability index. Quality Engineering, 31(2):289-301.

A Gelman, J Hill (2006) Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge: Cambridge University Press.

DJ Wheeler (2006) EMP III (Evaluating the Measurement Process) Using Imperfect Data. SPC Press, Knoxville, TN.

Vardeman SB (1992) What about the other intervals? The American Statistician 46:193-197.

InfinityQS International (2014) A practical guide to selecting the right control chart. pp 1-21. Available at: http://www.infinityqs.com/sites/infinityqs.com/files/files/PDFs/InfinityQS_Practical_Guide_to_Selecting_the_Right_Control_Chart_Oct2013.pdf.

Kluxen FM, K Weber, C Strupp, SM Jensen, LA Hothorn, J-C Garcin, T Hofmann (2021) Using historical control data in bioassays for regulatory toxicology. Reg Toxicol Pharmacol 125:105024.

© OECD (2025)

OECD (Organisation for Economic Cooperation and Development) (2016) OECD Guideline for the testing of chemicals: Mammalian erythrocyte micronucleus test. Test Guideline 474. Available at: https://www.oecd.org/en/publications/2016/07/test-no-474-mammalian-erythrocyte-micronucleus-test_g1g6fb16.html.

© OECD (2025)

Annex 5 - An alternative to mutant mimics for establishing flow cytometer settings for conducting the *Pig-a* assay

An alternative to mutant mimics is to use blood samples previously collected from animals with elevated *Pig-a* mutant frequencies (such as from positive controls, if used). Such samples can be stored and used as flow cytometry standards to identify the light scatter and fluorescence characteristics of mutant versus wild-type RET and RBC. In these cases, it is usually sufficient to include 1 to 3 such blood samples at each time point studied. Examples of test chemicals that will produce appropriate increased mutant frequencies are listed in Table 1. Note that it is not necessary to administer the substance using the same route of administration as the test chemical, and the dosing and blood sampling schedule can be different, as long as samples with elevated MUT RBC and MUT RET frequencies are available at the time when experimental analyses are conducted. When blood sample(s) with elevated mutant frequencies come from previously treated rodents, blood should be stored appropriately (paragraphs 58-59). When these samples are being used in place of mutant mimics, the blood sample(s) should demonstrate levels of MUT RET and MUT RBC that are elevated sufficiently to establish the fluorescence characteristics of mutant phenotype cells. For this purpose, it is ideal for the mutant frequency in these two cell populations be at least 100 mutant cells per million erythrocytes.

© OECD (2025)