The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) appreciates the opportunity to provide comment on revised draft OECD Test Guideline (TG) 487 "*In Vitro* Micronucleus Test." Accordingly, ICCVAM would likely to offer the following general and technical comments for consideration by the U.S. National Coordinator to the OECD and the Technical Lead as the unified U.S. position on this matter. We especially want to draw attention to the first two points.

1. First and foremost, the OECD needs to fully recognize the importance of having TGs based on adequately considered and evaluated draft TGs and that providing critical background information with only a few days to consider is entirely inappropriate. It is critical to the success of the TG program and the acceptance of data under MAD that all supporting materials should be made available with sufficient time for consideration. Thus, the review process should be delayed to allow for: (1) the significant issues raised by the ESAC Peer Review to be addressed, and (2) careful consideration of the total data package by member countries. We note that the final and critically important ESAC Peer Review document was received only on 30 January, allowing only one day for consideration, in order for comments from representatives of 15 U.S. Federal agencies to be collated, reviewed, and submitted to the U.S. National Coordinator in time to meet the OECD mandated deadline. Thus, although we provide comments, we request that the OECD delay the 15 February due date to allow for due consideration of all of the supporting documents. We also request that, in the future, the OECD take into account the critical importance of the TG review process and provide sufficient time (at least two months if not three) to ensure an adequate review. Otherwise, it appears as if the OECD cares more about schedules than making sure that a TG meets the needs of GD 34 and the regulatory and scientific community.

2. The purpose of validation is to determine the usefulness *and* limitations of a test method for a specific purpose. The data used to support validation of the *in vitro* micronucleus (MN) assay leaves several critically important questions unaddressed. These questions include how and whether it is appropriate to use cytochalasin B (CB) for cell lines, the method(s) by which cytotoxicity should be measured when CB is not used, and the maximum level of cytotoxicity appropriate for a valid test. The validation data sets, while more extensive than those used to validate some of the older test methods when their guidelines were first approved, are less extensive than those available for several recently validated test methods and do not cover all product categories (e.g., food additives) or functional classes (e.g., a sufficient number of aneugens, chemicals that require metabolic activation). Use of the protocol described in this guideline is clearly appropriate in certain circumstances (e.g., as a preliminary screen or as a follow-up test in the case of an ambiguous result in another assay or battery). However, the available published data do not support the substitution of the *in vitro* micronucleus assay for *all* current uses of the *in vitro* chromosome aberrations assay. In particular, we do not agree that it is, at this time, appropriate to substitute this test guideline for TG 473 in standard batteries designed to detect agents that interact with DNA to cause genetic damage. In fact, because the two different assays each provide unique information, the *in vitro* MN test, even

when adequately validated, should not be considered a replacement for the *in vitro* chromosomal aberration test but rather as another test that might be used to evaluate the mammalian cell genotoxicity of a test compound. In any case, it is not the role of an OECD TG to determine how the results of the test should be used within an overall safety evaluation. Paragraphs 3 and 5 contain discussions unlike that found in other OECD TGs in that they address the use of the assay with respect to other assays. These comments should be eliminated or altered to make it clear that the TG does not include a recommendation of how the results of the test are to be interpreted beyond the finding that the test article does or does not induce chromosomal damage under the conditions of the test. However, we do agree that it is useful to describe the context for why this test might be conducted. We also agree with the comment from Canada that the purpose and intended use of the test should be clarified.

General Comments

1. The latest ESAC Peer Review document raises several issues/concerns that should be addressed before a TG on this assay is finalized. Specifically:
   - The data sets on indirect-acting chemicals requiring metabolic activation and aneugens should be increased to allow a careful evaluation of the performance of the test method for these classes of chemicals
   - An optimize test method protocol and corresponding performance standards should be included in this TG and made available for consideration by the scientific community prior to the TG being finalized.

2. In reading the comments on the previous version of this TG, there were a number of times where the *in vitro* chromosomal aberration TG was referred to as if it was the gold standard. That TG was published 10 years ago; surely, with the increased knowledge we have in this area, we should be writing the most appropriate TG possible and not assuming that what was considered adequate a decade ago must still be considered adequate now. In fact, issues raised with this TG may indicate the need to update other *in vitro* genetic toxicity TGs. Also, considering that these assays are used in a weight-of-evidence approach to predict carcinogenicity or germ cell damage, an evaluation should have been conducted comparing the two *in vitro* assays against those endpoints to demonstrate whether or not detecting aneugens improves the performance of the assay.

3. The justification for not using CB in all cell systems is inadequate and unless there is adequate data to support a conclusion that CB interferes with the MN assay, it should be present throughout the exposure period to ensure that all cells at risk for MN formation divide in the presence of CB. Furthermore, the TG reiterates throughout the importance of knowing the proliferative history of the cells being scored for MN and there is no direct way to accomplish that goal other than by using CB. The statement that comparable results have been obtained in the presence and the absence of CB for a certain group of substances is not sufficient in and of itself to support a conclusion that CB should not be required.

4. As different measures of cytotoxicity can result in different maximum concentrations being considered acceptable for testing in these *in vitro* assays, the appropriate measure should be the same for all MN assays. In addition, a more careful evaluation of what constitutes an acceptable level of cytotoxicity should be identified and used throughout all in vitro MN assays, taking into account the need to avoid mimicking the high false positive rate associated with the *in vitro* Chromosomal Aberration Assay. The required level of cytotoxicity should also be the same across the assays.

5. Considering the number of significant concerns and issues we have, this TG clearly requires at least another round of review before being considered for final approval by the WNT. In addition, we recommended that this TG require the use of CB during and, for short exposure periods, after the exposure period. The TG could also state that methodologies where CB is not used are acceptable with adequate justification and as long as the cytotoxicity data are comparable.

6. The ICCVAM and its Genetic Toxicity Working Group would be glad to assist with revision of the draft TG in order to ensure that it adequately addresses all of the concerns of member countries.

Specific Comments

pp. 1, ¶ 1 The second sentence is not completely accurate as some chromosomes migrate during anaphase but fail to reach the poles. The term “whole chromosomes” is also inaccurate as micronuclei (MN) may contain centric fragments that fail to migrate to the poles and are indicative of aneugenic damage. This statement applies to the use of the word “whole” throughout the TG. A more accurate statement would be “or chromosomes that fail to reach the poles during anaphase.”

pp. 1, ¶ 1 A more correct third sentence would say, “The assay detects the activity of clastogenic and/or aneugenic chemicals…” The reason is that several clastogens have also been reported to induce aneuploidy by destroying the kinetochore (e.g., mitomycin C) while there have been reports that some aneugens also induce chromosome aberrations.

pp. 1, ¶ 1 Last sentence; the success of this assay depends on the fact that the cells being evaluated for MN have undergone cell division during and post exposure. Saying that it is sufficient for cells to have "likely” undergone cell division is inadequate and will potentially lead to false negative studies. The statement “during and/or after exposure to the test substance” should be added to the end of the sentence for improved clarity.

pp. 1, ¶ 2 The statement that “The use of cytokinesis block facilitates the acquisition of the additional mechanistic information (e.g., chromosome non-disjunction) that can be obtained by FISH-techniques (6-15).” is inaccurate. Using the

cytokinesis-block method has no impact on the ability of the FISH technique (or any other technique for that purpose) to identify the presence of a centromere in a MN. The studies reported used the cytokinesis block technique to enrich for cells at risk for the presence of MN, nothing more. Also, if this were true, it would be sufficient grounds to support the use of the cytokinesis block method in all studies.

pp. 1, ¶ 3 In the fifth sentence, the word "reliability" should be changed to "power."

pp. 1, ¶ 4 In previous comments in response to the first draft, the term "mutagen" (see line 3) was considered inappropriate for this assay. We agree with that comment that the MN assay is a genotoxic assay rather than a mutagenic assay and "genotoxicity" should be consistently used throughout the document.

pp. 1, ¶ 4 If this test is of particular value because it detects clastogens and/or aneugens, then it is critically important that discrimination between clastogens and aneugens be required. Otherwise, regulatory agencies will need to take the conservative approach and classify a test substance that may be an aneugen (i.e., having a threshold) as a clastogen (i.e., not having a threshold). As pointed out by the ESAC Peer Review Committee, the database for testing aneugens is inadequate.

pp. 1, ¶ 4 In line 3, insert "initially" between "for" and "investigating" because the current statement seems to imply that the Test 474 does not detect inducers of aneuploidy. Also, substitute "genotoxic" for "mutagenic."

pp. 1, ¶ 5 The identity of the international *in vitro* MN assay working group needs to be clarified. It is the "International Workshops on Genotoxicity Testing *in vitro* MN assay working group".

pp. 1, ¶ 5 The supporting documents for the ESAC statement need to be cited and publicly available (consistent with GD 34 and the need for transparency).

pp. 1, ¶ 5 At the end of line 4, change "has" to "have" for subject/verb agreement.

pp. 1, ¶ 5 The current data do not justify substitution of the *in vitro* micronucleus assay for the *in vitro* chromosome aberrations assay in all instances. For example, the ECVAM ESAC evaluation did not find that the data they evaluated supported the assessment of some product classes (e.g., agrochemicals and pharmaceuticals) (31). In addition, an evaluation of the ability of the test to correctly predict the results of a rodent carcinogenicity test found that while few rodent non-carcinogens have been evaluated, among those few the specificity of the test was quite poor compared to other *in vitro* tests. (Kirkland et al., Mutat Res. 584:1-256, 2005).

pp. 2, Initial Considerations Section	This section seems to support the standard use of the cytokinesis-block technique in all protocols.
pp. 2, ¶ 7	Is it known which conditions give rise to false positives in the various *in vitro* MN assays? The papers cited deal with clastogenicity and mutations. What are the conditions associated with aneuploidy? If these are not known, then how can it be stated that the limits of the assay are known, a requirement for an adequately validated test method?
pp. 2, ¶ 7	In the second to last line, the word "artifactual" is misspelled.
pp. 2, ¶ 8	This paragraph states, "…it is essential that nuclear division has occurred in both treated and untreated cultures." This statement again supports the use of cytokinesis-block (CB) to make sure that the cells being scored have divided.
pp. 2, ¶ 9	Consider changing "toxic" to "cytotoxic".
pp. 2, ¶ 10	While it is true that primary cells with metabolizing capability would eliminate the need to add an exogenous source of metabolic activation, it would not allow for test compounds to be evaluated for genotoxicity in the absence of metabolic activation. This is critical because a genotoxic compound may be metabolized to an inactive form by the primary cells, leading to a false negative call (i.e., a search of the genetox database would reveal a number of genotoxic compounds that are positive only in the absence of metabolic activation). Thus, only using primary cells with metabolizing capability or any similar approach should not be permitted.
pp. 2, ¶ 10	The statement that "After exposure to the test substance, cell cultures are grown for a period sufficient to allow chromosome or spindle damage to lead to the formation of micronuclei in interphase cells and to trigger the aneuploidy sensitive cell stage (G2/M)." is internally inconsistent as the first part of the sentence negates the need for the second part.
pp. 2, ¶ 10	There is a continual reference to the importance of scoring interphase cells at risk for MN formation (i.e., those that have divided following exposure to the test substance). Other than the CB method, there is no other method capable of ensuring that only cells at risk are scored for MN. Thus, unless it can be demonstrated that CB adversely affects the sensitivity of the assay, this method should be required. In Paragraph 11, it is stated that there is some concern with the use of CB in mouse lymphoma cells but it is not clear what the concern is. If the concern is significant, this cell line may not be suitable for this assay.
pp. 2, ¶ 11	Suggest rewording "Cultured cells from human peripheral blood lymphocytes or from Syrian Hamster Embryo (SHE) may be used." to "Cultured human

peripheral blood lymphocytes or Syrian hamster embryo (SHE) cells may be used."

pp. 2, ¶ 11 The statement "The frequency of micronuclei in the negative control cultures should be within the historic negative control range for the laboratory." does not belong in this section since it is a QC issue and implies that an appropriate cell line has already been selected.

pp. 2, ¶ 11 In line 3, the term "interactions" is not clear in this context, consider changing to "artifactual effects".

pp. 2, ¶ 11 Are there stable human cell lines that can be used in this assay?

pp. 3, ¶ 12 The acceptable donor age range should be defined; otherwise different individuals/organizations may define "acceptable" differently. If the concern is that females show an increase in MN frequency as they age, then there must be publications showing at what age this increase becomes significant. It is not clear what is meant by "pooling of samples." Presumably, this section refers to donors and not to data so pooling is referring to pooling of blood samples. What is the rationale for pooling samples (saying that a lab has done this successfully is not a scientific rationale for putting it in a TG)? The ECVAM Validation Management Team on the Micronucleus Test *in vitro* should have evaluated the number of donors needed and whether or not blood could be pooled. The recommendation is that, if variability in response could occur among donors, then each donor should be tested separately. Furthermore, more than one donor should be tested, either concurrently or sequentially. The number of donors and their gender needed should be defined here.

pp. 3, ¶ 12 The TG should include a precautionary statement on the safe handling of human tissue.

pp. 3, ¶ 12 In the sentence "Established cell lines and strains should be checked routinely for the stability of the modal chromosome number and the absence of mycoplasma contamination and cultures should not be used if contaminated," it may be more correct to state that "cell lines/strains" rather than "cultures" should not be used if contaminated.

pp. 3, ¶ 16 There is a "." missing after (47).

pp. 3, ¶ 16 Examples of the cases where it is appropriate to use more than one concentration of S9 should be provided. Otherwise, it is a generic statement that offers no guidance to those doing the testing or receiving the results (regardless of the fact that this statement is in other *in vitro* genetox TGs).

pp. 3, ¶ 17 Again, only using a cell type that metabolizes to evaluate the genotoxicity of a test substance (except for a specific reason unrelated to general testing) should

not be considered or recommended. In addition, genetically engineered cell lines that do not express the entire suite of activating enzymes would give incomplete results. Such cell lines would seem to be more appropriate in a research, rather than regulatory, setting.

pp. 4, ¶ 19 This states that "If other than well-known solvent/vehicles are used, their inclusion should be supported by data indicating their compatibility with the test." which implies that there are well-known solvents that do not need to be tested for compatibility. These solvents should be identified to avoid possible differences of opinion as to what are examples of well-known solvents.

pp. 4, ¶ 20 In regard to the statement that "This is not the case with cell lines and cytochalasin B need not be used provided that cell proliferation is demonstrated to ensure that the majority of cells scored have progressed through mitosis," as cytotoxicity can also be a result of a cytostatic effect, how can it be assured that a majority (50+ %) of cells scored have progressed through mitosis, and why would 50% be considered adequate to identify a weak MN-inducing agent as opposed to knowing absolutely that every cell scored was at risk for MN formation?

pp. 4, ¶ 21 Why must the test substance be removed prior to adding CB? Is there evidence that CB, which inhibits actin assembly, interacts with genotoxic chemicals and/or alters the sensitivity and specificity of this test method when administered at the same time as the test compound? Administering CB only after the test compound has been removed will, of course, allow cells to divide and produce MN in the absence of CB. This must greatly decreases the sensitivity of the assay, especially when long exposure durations are used.

pp. 4, ¶ 22 As CB is required for PBL cultures, why would there be parallel cultures with CB to evaluate effects on cell cycling? What about SHE cells? If parallel cultures with CB are being run, they should be scored for MN in binucleate cells.

pp. 4, ¶ 23 The wording here suggest that any change in pH or osmolality should be avoided as opposed to changes that would result in false positive responses only. This is not correct as small changes in pH or osmolality are of no consequence. Furthermore, the conditions for excessive pH or osmolality for MN induction have not been evaluated. It may be that this assay is less sensitive than the chromosomal aberration test to one or the other.

pp. 4, ¶ 24 Clarification is needed as to what method for measuring cytotoxicity should be used?

pp. 4, ¶ 25 The sentence "In the case of studies without cytochalasin B, cell proliferation should be measured by the cell counts or the population doubling, combined with an assessment of cytotoxicity." is confusing. First, the word "time" or

"duration" is missing after "population doubling". Second, the term "assessment of cytotoxicity" needs to be clarified by providing examples of ways to assess cytotoxicity that are meaningful in the context of this test method.

pp. 4, ¶ 25 A better example should be provided to indicate how the formula works rather than the statement that "Thus, a CBPI of 1 (all cells are mononucleate) is equivalent to 100% cytostasis." In this case and assuming that in the control cells, 100% were binucleates, would not the equation %Cytostasis = 100-100{($CBPI_T$ - 1)/($CBPI_C$-1)} be equal to 100-100{(1-1)/(2-1)} = 100-100 (0/1) which is indeterminate?

pp. 5, ¶ 28 In regard to "For poorly soluble compounds that are not cytotoxic at concentrations lower than the insoluble concentration, the highest concentration should produce a precipitate visible by the unaided eye or with the aid of an inverted microscope at the end of the treatment", there is considerable difference in what can be seen with the unaided eye versus that with an inverted microscope (and at what magnification?). One or the other but not both methods should be recommended. Furthermore, why must the precipitate be present only at the end of the culture period? Would it not be sufficient to state, "For poorly soluble compounds that are not cytotoxic at concentrations lower than the insoluble concentration, the highest concentration should produce a precipitate visible (by the unaided eye) (with the aid of an inverted microscope) in the cultures."?

pp. 5, ¶ 30 In the statement "This provides a control for the activity of the metabolising system, whether endogenous or exogenous." the last phrase is unnecessary and should be deleted.

pp. 5, ¶ 30 The term "mutagenic" is used here, as opposed to "genotoxic".

pp. 5, ¶ 28 The current wording seems to imply that two positive controls (one control that is directly acting and a second control that requires metabolic transformation to produce a response) are required per experiment but the term experiment is not defined (perhaps an experiment only tests for MN activity with metabolic activation). Suggested alternative language is that "A concurrent positive control that is appropriate for the presence or absence of metabolic activation (e.g., S9) in the culture should be included in each experiment."

pp. 6, ¶ 32 Examples of cell lines where cyclophosphamide should not be used as a positive control, or a concentration that would be deemed unacceptable should be provided.

pp. 6, ¶ 34 The statement "In addition, untreated control (lacking solvent) should also be used unless there are historical control data demonstrating that no deleterious

or mutagenic effects are induced by the chosen solvent." could be interpreted to indicate that each lab needs to generate its own historical control for each solvent. However, paragraph 19 states "If other than well-known solvent/vehicles are used, their inclusion should be supported by data indicating their compatibility with the test." which implies that well-known solvents need not be independently tested by a lab. The apparent discordance needs to be reconciled.

pp. 6, Treatment Schedule — This section needs more clarity. Specifically what is needed is a table providing information on the exposure durations and the sampling times as it relates to the exposure duration, along with information on what sets of experiments need to be conducted to adequately demonstrate that a compound is negative for MN induction. Of course, using CB in all protocols and basing cytotoxicity on the BN index would greatly decrease the complexity of this section.

pp. 6, ¶ 35 — In line 2, the statement "over time of the cell cycle" is unclear. Is the decrease in synchrony seen over a single cell cycle or over a period of time (multiple cell cycles)? Furthermore, the synchrony is only partial as cells can enter their first mitosis any time between ~40 and ~96 hrs after PHA stimulation

pp. 6, ¶ 35 — In the penultimate sentence, add "of the cell cycle" after "…at all stages."

pp. 6, ¶ 36 — In the statements "Theoretical considerations based on the non-synchronised cycling of cell lines in culture, together with data (ref) indicate that most aneugens and clastogens will be detected by a short term treatment (3 – 6 hours) in the presence and absence of S9 followed by a recovery period, if required (5). Cells are sampled at a time equivalent to about 2 times the normal (i.e. untreated) cell cycle lengths after the beginning of treatment. In some instances a longer recovery period employing sampling times of about 3 cell cycles) may be appropriate," there are missing ref, and it is not clear what is meant by "most". Since most aneugens affect mitosis, the longer the exposure duration and the shorter the cell cycle, the more an aneugen is likely to be detected. However, if CB is only added after treatment, the cells most at risk for forming a MN by this mechanism will have divided in the absence of the aneugen. The statement that a recovery period may not be required is confusing given that the next sentence states that cells are sampled 2 cell cycles after the beginning of treatment. Examples should be provided to indicate what is meant by "In some instances a longer recovery period employing sampling times of about 3 cell cycles" may be appropriate." When are such conditions needed? If the conditions cannot be defined then this extended recovery time is needed in all situations where a negative result is obtained in both the presence of S9 mix for 3-6 hours and in the absence of S9 mix for 3-6 hours and 20 hours using a 2-cycle sample time.

pp. 6, ¶ 37 The extended exposure duration described here is incompatible with the use of CB as described in paragraph 38.

pp. 7, ¶ 39 To make the cell cycle effect data comparable, should not the protocol with and without CB be the same? Also, if this approach were recommended, it would be more efficient to collect the MN data on the CB-treated cultures.

pp. 7, ¶ 40 Why is it required to include a prolonged (20 h) exposure to the test substance in the absence of S9 with the short-term exposure studies for human PBL, while for cell lines, the prolonged exposure is only conducted if the short-term exposure is negative (Paragraph 37)?

pp. 7, ¶ 40 Based on the information provided, some cultures would be treated starting at 48 hours for 20 hours, with CB added for another 28 hours, so that the cultures would be sampled at 96 hours. Especially in the case of aneugens, this protocol seems problematic as the interphase cells with MN resulting from cell division during the extended exposure duration may not be able to divide again and give rise to binucleate cells during the CB period.

pp. 7, ¶ 41 What is the data supporting the statement that "If the protocols give negative or equivocal results, confirmation should be considered by varying the conditions, such as commencing exposure at 24 hours after PHA stimulation…." Why would that time make a difference? Also, if PBL are tested in the absence of S9 for 3-6 and 20 hours (similar to what is required for cell lines) and the data are negative under both conditions why is more testing needed and does this mean a lab would need to test 3-6 and 20 hours at both 48 hours and at 24 hours for multiple donors? Clarity is needed as to how many different protocols need to be conducted to demonstrate a test compound is negative for MN induction.

pp. 7, ¶ 42 This paragraph states that "If it is known or suspected that the test substance acts at a specific, identified phase of the cell cycle, the protocol should be modified to target exposure to this phase." Why does this apply only to PBLs and what examples/citations can be provided to support this statement and how such testing should occur?

pp. 7, ¶ 43 The statements that "Duplicate cultures should be performed at each concentration and are strongly recommended for negative/solvent control cultures. Where minimal variation between duplicate cultures can be demonstrated, from historical data, it may be acceptable for single cultures to be used at each concentration." Even if historical data existed for some compounds that minimal variation existed between duplicate cultures, it is difficult to appreciate how historical data can be assumed to pertain to all compounds that might be tested. Duplicate cultures should be required at all concentrations, including negative and positive controls. Otherwise, the criteria by when duplicate cultures would not be needed would need to be

provided. Furthermore, Paragraph 47 states that 1000 cells should be scored per culture with 2000 cells total, which indicates that only duplicate cultures are being used.

pp. 7, ¶ 45 There is an extra period in line 6.

pp. 7, ¶ 45 Examples of other methods for differentiating between clastogens and aneugens should be provided.

pp. 8, ¶ 50 This, of course, would not be an issue if CB was present throughout the exposure period, as the only way MN could be increased in mononuclear cells is if they represent cells that had divided during the exposure period.

pp. 8, ¶ 51 It is not evident why cells with one, two, or more MN need to be tabulated, as there is no discussion on how to analyze such data. For example, are there data indicating that compounds can increase the frequency of cells with multiple MN without increasing the frequency of cells with MN? This requirement should be omitted unless adequately justified and the method of analysis and interpretation discussed.

pp. 8, ¶ 54 "**Evaluation and interpretation of results**" should be on a new line.

pp. 8, ¶ 55 The statements in paragraph 55, describing the criteria for a positive call, and in paragraph 54 appear to be in contradiction. In paragraph 54, positive responses are said not to require verification. In paragraph 55, one criterion is "a reproducible increase." How can something be demonstrated to be reproducible (in the classic sense) without conducting an independent repeat experiment? If this refers to data within the same experiment, how is it different for saying a "concentration-related" increase is needed? This apparent conflict should be resolved with more specific wording.