Majority ICCVAM Comments on the Draft OECD Guideline for the Testing of Chemicals: The Uterotrophic Bioassay in Rodents

The following comments regarding the OECD Draft Test Guideline (TG) on the Uterotrophic Bioassay in Rodents are based on provisions in OECD Series on Testing and Assessment Number 34: Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment (OECD GD 34). OECD GD 34 provides for the use of general international TGs to allow for flexibility in fulfilling testing requirements for regulatory purposes rather than the use of a specific test method protocol. Nevertheless, OECD GD 34 specifically recommends that TGs are to be based on test method protocols that have been optimized, standardized, and validated. Thus, based on the lack of an appropriately validated test method protocol, development of this TG at the present time is premature. Comments below reflect this concern, as well as identifying issues related specifically to the TG.

Comments

1. OECD GD 34 requirements for validation includes the demonstration of intra- and inter-laboratory reproducibility using relevant classes and numbers of reference chemicals. Validation studies for the uterotrophic assay were conducted across 19 laboratories, using four different protocols and involving over 6000 animals under a variety of different conditions (e.g., animal strain and age, diet, housing, bedding, vehicle). The TG claims that intra- and inter-laboratory reproducibility was supported by the above validation studies. However, considering the limited protocol standardization, we question the adequacy of the number of reference chemicals (one strong agonist, five weak agonists, one pure antagonist, and one negative) that were used to evaluate the reliability of this test method.

2. The proposed uterotrophic TG currently includes an adult ovariectomized rat protocol, which involves potential pain and distress associated with the required surgical procedure, as well as a non-surgical immature rat protocol. The U.S.

Government Principles for the Utilization and Care of Vertebrate Animals in Research, Testing, and Education states that the avoidance or minimization of discomfort, distress, and pain, consistent with sound scientific principles, is imperative. Furthermore, it states that U.S. Government agencies should consider these principles whenever they develop requirements for testing procedures involving animals. It is our understanding that the ovariectomized rat protocol is the preferred protocol at the current time because of difficulties in predicting when the immature animals are going to reach puberty. Thus, based on the number of chemicals tested, we don't believe that the validation effort was sufficiently robust to demonstrate that the immature rat and the ovariectomized rat versions of the uterotrophic assay are equally reliable and accurate. If the two protocols are equal in performance and since minimizing pain and suffering is a critically important goal, it would seem that any use of the ovariectomized rat protocol instead of the immature rat protocol must be scientifically justified.

3. The extent to which the test method can accurately identify negative substances has not been adequately determined with the use of only one negative substance in the OECD study. Even if proposed as a screen, the extent to which false positives occur are needed to define the limitations of the assay. This is especially true given that non-estrogen receptor dependent effects can result in increased uterine weight. Efforts should be made to increase the number and expand the types of negative compounds evaluated in order to strengthen conclusions regarding assay specificity and the likelihood of false positives.

4. The extent to which the test method can generate relevant data based on the recommendations for dosing stated in the TG is unclear. The TG indicates that the maximum dose should represent the standard limit dose of 1000 mg/kg bw/dy, the maximum tolerated dose (MTD), or a dose inducing uterotrophic effects. The TG should state, consistent with other OECD TGs, that the maximum dose should represent the standard limit dose of 1000 mg/kg bw/dy, the maximum tolerated dose (MTD), or the maximum dose below 1000 mg/kg bw/dy that can be administered

given the physico-chemical properties of the test substance and the route of exposure. As there is no way, a priori, to identify a dose of an unknown test substance that would be capable of inducing uterotrophic effects, including that phrase in this context is inappropriate. No rationale is provided for the upper limit and since the regulatory needs may vary, the upper limit should be an option. The TG also states that one or more reduced dose levels should be selected with a view to demonstrating dose-response relationships and identifying a no-observable-effect-level (NOEL). It is unclear how relevant dose-response information (including a NOEL) can be determined using this approach.

5. The data generated in the OECD validation program demonstrated the ability of the test method to reproducibly detect a small number of estrogenic substances when laboratories were instructed to use specific doses for each non-coded test article. However, the ability of laboratories to test coded substances, to select appropriate doses, and to obtain reproducible and accurate results using the complete test method protocol has not been demonstrated. Dose-setting procedures must be included in test method protocols, and these procedures must be evaluated to determine if they can reliably identify appropriate test substance doses for testing and especially doses that will allow for the detection of weakly active substances.

6. The route of administration is stated in the TG as being either subcutaneous (s.c.) or by oral gavage. Rather than just stating the test substance can be administered either s.c. or oral, the TG should state that the study protocol should include administration of the test substance by the same route as humans are likely to be exposed (which is more likely to be oral or dermal than s.c.). Should additional routes be judged more appropriate for one reason or another, clear justification of those routes of exposure in the test species in relation to those anticipated for the target species is necessary.

7. The TG states that the uterotrophic response is not entirely of estrogenic origin (i.e., factors and compounds other than estrogen agonists can cause an increase in uterine weight) and that all positive outcomes should initiate actions for further clarification.

However, such methods are not specifically stated in the TG. Further clarification of methods to evaluate whether a positive uterotrophic response constitutes a relevant estrogenic response is essential if the uterotrophic assay is to be used as a screen for estrogenicity. Controlling for specificity of response is of particular importance given potentially confounding issues that are not sufficiently addressed in the TG.

8. Strain specific differences exist in the onset of puberty, metabolic competency, and uterine response to estrogenic compounds. The TG states that the laboratory should be able to demonstrate the sensitivity of the strain used but no guidance is provided as to how this is to be done. It would be more appropriate to give guidance as to the characteristics a strain should have in order for it to be used in the assay and to require justification (or additional validation) if another strain is used.

9. The issue of phytoestrogen content in food is addressed by stating that dietary levels of phytoestrogens should not exceed 350 µg of total genistein equivalents (TGE)/gram of laboratory diet for immature female rats. This statement is contradicted by current literature (Thigpen et al. 2004), which specifically states that diets containing less than 325 to 350 µg/g TGE still have the potential to alter the results of vaginal opening and uterotrophic assays. This is clearly a critical and unresolved issue, especially for the detection of weak acting estrogenic compounds. Thigpen et al. recommends the use of feed containing no more than 20 µg/g TGE. Considering the importance of diet in this test method, only low phytoestrogen content, certified diets (or diets demonstrated by testing to be appropriately low in phytoestrogens) should be used.

10. As a related comment, the TG states that, in case of unexpected results, an analysis of the diet for estrogenic compounds should be considered. The TG does not state what constitutes an unexpected result or what should be done to address the findings of any feed analysis conducted. If the feed is analyzed and found to have high phytoestrogen content, are the experiments to be discarded, or repeated using a different feed formulation? This seems to be an approach that could lead to the

unnecessary use of animals and therefore is to be avoided by using a low phytoestrogen content, certified diet.

11. In regards to potential estrogenic exposures resulting from animal bedding, the TG states that in the case of unexpected results an analysis of the bedding for estrogenic compounds should be considered. The TG does not state what constitutes an unexpected result or what should be done to address the findings of any bedding analysis conducted. If the bedding is analyzed and found to have high phytoestrogen content, are the experiments to be discarded, or repeated using different bedding? This seems to be an approach that could lead to the unnecessary use of animals and therefore is to be avoided by only using bedding known to be free of phytoestrogens.

12. Differences in consumption of feed can lead to differences in the magnitude of response to test substance. This issue is addressed in the TG by stating that measurement of daily food consumption is an option. However, no guidance is provided for determining individual food consumption for animals that are communally housed.

13. A vehicle for administration of a test substance is not specified. The TG recommends the use of several different oils, which have different densities and different caloric and fat contents. This is an issue of concern, since the vehicle may affect total metabolizable energy (ME) intake, thereby potentially altering measured endpoints such as uterine weights (Thigpen et al. 2004).

14. Paragraph 52 of the TG states that the uterus weight at termination can be used to assure that the appropriate age in the immature intact rat was not exceeded. As a guide, the mean blotted uterus weight should be "around 30 mg at postnatal day 23. However, the historical data of the rat strain used by the laboratory are decisive in this respect." Using the uterine weight of the control group to decide that animals of the appropriate age were used is inadequate because rats (even those from the same litter, housed in identical situations, and with equal access to food) do not all enter puberty

at the same age. A more adequate control for pubertal status would be measurement of serum estradiol concentrations, which can be ascertained by several simple, readily available methods.

15. The TG states that laboratories carrying out this assay on a routine basis should periodically verify assay performance using a positive control; for example, once per year by the response to a reference dose of 17ß-ethinyl estradiol (CAS No. 57-63-6). If such testing is only done sporadically then a positive control group treated with 17ß-ethinyl estradiol should be considered for inclusion within each assay. The justification for concluding that once a year is sufficient to demonstrate laboratory proficiency needs to be provided.

16. As a related positive control issue, no guidance on the magnitude of the increase in uterine weight from the positive control, 17ß-ethinyl estradiol, is provided other than that the increase should be statistically significant (part 29). As the number of animals per group is not specified, it may be possible to achieve a statistically significant increase with a relatively small absolute increase in uterine weight if the number of animals tested is relatively large. In such cases, the sensitivity of the test method may be less than optimal. It may be important to consider criteria for a positive control based on an absolute or percent increase in weight and/or a more stringent statistical criterion than $p<0.05$.

Thigpen JE, Setchell KD, Saunders HE, Haseman JK, Grant MG, Forsythe DB. 2004. Selecting the appropriate rodent diet for endocrine disruptor research and testing studies. Ilar J 45(4):401-416.