ICCVAM comments and recommendations made after the review of previous versions of the Test Guideline that have not been specifically addressed in the latest revision of the Test Guideline:

Test Guideline (TG) Page (pp) and Paragraph (¶)	Comments
General	As stated in previous comments submitted by ICCVAM to the U.S. National Coordinator on this test method and the proposed test guideline, ICCVAM and its Endocrine Disruptor Working Group have concluded, with only one agency in disagreement, that the uterotrophic bioassay has not been adequately validated for its intended purpose. Thus, ICCVAM's recommendation is that this material be placed in an OECD Guidance Document, which could then be used as the basis for further studies that could lead to an adequate demonstration of validation. This is identical with the approach proposed by the OECD for the uterotrophic bioassay to detect estrogen antagonists. The previous sets of comments relating this position are attached and it is requested that they be addressed before proceeding with further consideration of this draft TG. In terms of the revised TG, there are several key technical points, which when each is considered alone, appear adequate. However, when considered together, they raise serious concerns about the adequacy of this revised TG and the ability of scientists to produce useful data without the inappropriate use of animals.
pp. 1, ¶ 1	This paragraph states that "extensive" intra- and interlaboratory studies were conducted. However, considering the number of chemicals tested, the word "extensive" is inappropriate or needs to be defined (the validation study may have been extensive in the number of laboratories participating but it certainly was not in the number of substances tested or in the extent to which the complete protocol, including range finding tests, was evaluated).

Test Guideline (TG) Page (pp) and Paragraph (¶)	Comments
pp. 2, ¶ 7	This paragraph states that the immature rat has an intact HPG axis, making the test system less specific but covering a larger scope of investigation because (in contrast to the OVX female method), it can respond to substances that interact with the HPG axis rather than just the ER. This additional information indicates that either the immature system should not be used because of the increased likelihood of obtaining a false ER response or should be used because the test method also detects substances that interact with the HPG axis, which is also important information. Regardless, this statement means, on face value, that the two model systems must differ in sensitivity, a factor which is inconsistent with an earlier statement in this paragraph that the two methods have comparable sensitivity. The TG should more clearly state the scientific advantages and limitations of the two methods, as a screening assay for detecting substances with estrogenic activity.
pp. 2, ¶ 9	This paragraph states that the uterotrophic response is not entirely of estrogenic origin, and that certain non-estrogenic steroids and synthetic progestins may also lead to a stimulative response. The TG states that "Any response may be analyzed histologically for keratinization and cornification of the vagina." Considering that the purpose of this test is to identify substances with estrogenic activity, it would seem that this step is necessary to make the test method more accurate. This paragraph states that any positive outcome should normally initiate actions for further clarification by the use of *in vitro* and other *in vivo* assays. Why would that occur if, as Paragraph 4 states, this test method "is intended to be included within a battery of *in vivo* and *in vitro* tests…". Furthermore, if a different i*n vivo* test is more specific, should not these tests have preference over the uterotrophic test?
pp. 3, ¶ 16	The TG states that Sprague-Dawley and Wistar rat strains were used during validation but also states that other commonly used laboratory rodent strains may be used, unless known or suspected to be less responsive. The TG goes on to state that the laboratory should demonstrate the sensitivity of the strain as described in paragraphs 26 and 27. Considering that the validation study did not evaluate the utility of other strains, ICCVAM believes that the TG should specifically recommend that these two strains should be used and that the use of other strains must be adequately justified (e.g., by providing comparative sensitivity data). This approach will minimize an unnecessary variable in the test method and will help to clarify what constitutes an acceptable phytoestrogen level in the diet. The TG should include, as examples, appropriate mouse strains.

Test Guideline (TG) Page (pp) and Paragraph (¶)	Comments
pp. 4, ¶ 22	ICCVAM again questions the use of 350 microgram genistein equivalents/gram as the limit for laboratory diet. This is contradicted by current literature (Thigpen et al. 2004), which specifically states that diets containing less than 350 micrograms/g TGE still have the potential to alter the results of vaginal opening and uterotrophic assays, depending on the rat strain. The citations (i.e., 6, 9) refer only to Sprague-Dawley or Wistar rats. Thus, specifying the strains used as being Sprague-Dawley or Wistar would eliminate this concern and could be accomplished by including in the second sentence "…laboratory diet for immature female Sprague-Dawley or Wistar rats (6)(9)." The TG, for the statement that "if adult ovariectomized mice are to be used, proportional reduction in dietary phytoestrogen levels must be considered (2)" should provide guidance on what is meant by a proportional reduction.
pp. 5, ¶ 23-24	The TG should include a recommendation that only diets and bedding that have been determined (either analytically or biologically) to be suitable for this test should be used. Otherwise, there is always the possibility that either there will be animal wastage and/or unsatisfactory test results.
pp. 5, ¶ 27	While there is appreciation that the use of a concurrent positive control increases the numbers of animals per study, given the potential variations in this test method associated with factor such as diet, bedding, the presence of estrogenic tissue in OVX females, the onset of puberty among some immature females, and the fact that a failed quality control test conducted periodically would necessitate all studies conducted after the last qualifying test be discarded, it still appears critical that a positive control be included in each study. The OECD should consider also that the uterine weight of 40-45 mg for control animals, as a quality control (QC) measure for responsive animals, is for Sprague-Dawley and Wistar rats only. If the acceptable limit for mice and other strains of commonly used rats is unknown, the QC measure is not generally applicable and should be applied to Sprague-Dawley and Wistar rats only. Therefore, a statement that the QC step negates the need for a concurrent positive control is fallacious. Furthermore, and consistent with other OECD TGs (in this case for *in vivo* genotoxicity), the positive control (whether periodic or concurrent) should be a weak acting dose of EE. Otherwise, it will not be possible to identify laboratory procedures or husbandry conditions that may affect the ability of the test method to detect weak-acting estrogens.

Test Guideline (TG) Page (pp) and Paragraph (¶)	Comments
pp. 6, ¶ 28	The TG states that each control and treated group should include at least 6 animals. Later (e.g., ¶ 30), it is stated that data from some animals may be excluded after necropsy because of residual ovarian tissue. Therefore, the TG needs to state how many animals with acceptable data are needed per dose group.
pp. 7, ¶ 34	The limit dose should be defined in this paragraph (see comment for paragraph 36).
pp. 7, ¶ 36	It would be useful if guidance could be provided for "a preliminary range finding study" as to what would be an efficient design and what represents a few animals. This paragraph states that the maximum limit dose should be 1000 mg/kg/dy. Again, as stated in a previous set of comments, other OECD short-term *in vivo* test guidelines mandate 2000 mg/kg/dy for studies of less than 14 days. It is not clear why the dose level for these studies are different, especially considering that the avowed purpose of this assay is be a sensitive method for detecting compounds with estrogenic activity, and based on the likelihood that infants and children may be especially sensitive to endocrine disruption. Having a limit dose of 2000 mg/kg/day will not increase the extent of animal pain and suffering as the criteria that the highest dose tested should not cause pain and suffering remains in effect.
pp. 7, ¶ 37	This section should instruct experimenters to provide a rationale to justify the route of administration.
pp. 8, ¶ 39	This section states that dosing of OVX rats for up to seven days may increase the sensitivity of the assay. If accurate, a seven-day exposure protocol should be required.
pp. 9, ¶ 43	Unless justified and the purpose explained, the optional weighing of feeders to measure food consumption should be deleted. It should be noted in the TG that immature animals are group housed which affects the reliability of this measurement. Also, the word "rat" should be replaced with "animal" unless this measurement is only attended for rats. Throughout this TG there are references to rats where the term "animal" is likely more appropriate (see para 44 – why would only rats be humanely killed?, or para 48 where uterine weight can be used to assure that the appropriate age in the immature intact rat was not exceeded)
pp. 10, ¶ 49	The purpose for the optional investigations (histopathology on the uterus and/or vagina) should be provided, as well as how such data are to be used.
pp. 12, ¶ 53	Under dosing, the number of dose groups and the limit dose should be provided.

Test Guideline (TG) Page (pp) and Paragraph (¶)	Comments
pp. 13, ¶ 56	The QC data for blotted uterine weights refers to immature Sprague-Dawley and Wistar rats and not to OVX rats or to other strains of rats, or to mice. This paragraph should include the phrase "As a guide, for immature Sprague-Dawley or Wistar rats, ". Furthermore, the statement that this "needs to be considered on a case-by-case basis, indicates that using this QC criteria as justification for not needing a concurrent positive control is without scientific merit.
pp. 13, ¶ 59	This paragraph now states that "If divergent results are obtained by the blotted versus the wet uterine weights, the blotted weights should be given preference for the final interpretation" But the next sentence states "However, a significant response in either measure would indicate that the test substance is positive for estrogenic activity." If the "However, …" sentence is accurate then the preceding sentence "If…" is contradictory and should be deleted.