**Table 1 Response to comments on FEMALE PUBERTAL ASSAY**

| Comment ID Number | Reviewer | Comment | Response |
|---|---|---|---|
| 1. Clarity of purpose of the assay | | | |
| 1.1 | JB | The purpose is clear.<br>…<br>The statement of purpose and applicability is quite clear. | The statement of purpose will be revised to address comments by other reviewers. See response to comment 1.8. |
| 1.2 | BD | The statement of purpose of the assay is reasonably clear. A revision to the initial sentence to state more specifically what will actually be accomplished with this assay might be appropriate. For example: "The purpose of this protocol is to identify chemicals that affect, after oral administration, pubertal development and thyroid function in the intact juvenile/peripubertal female rat." | See response to comment 1.8. |
| 1.3 | DF | The assay's stated purpose is to test the effect of xenobiotics on the endocrine system of the prepubertal female rat, specifically with regard to estrogenic/antiestrogenic and antithyroid activity or disrupted hypothalamic and pituitary function as relates to the onset of puberty. This appears straightforward enough; however, in the peer review charges, the effect of mixtures is mentioned as an application of the assay. No instruction of how best to perform mixture analysis is provided (an increasingly | The assay can be used to test mixtures of related substances such as polybrominated diphenyl ether mixtures or mixed nonylphenol isomers. The problem of how widely to generalize the results, or how to test mixtures of widely different chemicals, is not an assay-specific issue and is not appropriate to address in the protocol. |

| | | | |
|---|---|---|---|
| | | important issue in toxicology), nor is it addressed in the ISR. | |
| 1.4 | HP | The purpose of the assay is clearly stated and well justified. The protocol is intended to detect alterations in sexual maturation and thyroid function by exogenous chemical exposures during the prepubertal period. | The statement of purpose will be revised to address comments by other reviewers. See response to comment 1.8. |
| 1.5 | DR | Though the intent of pubertal assay described in both ISR and the Appendix 1 female pubertal protocol is the same, the description of the purpose of the assay is not exactly the same. For example,<br><br>(i) In the Appendix 1 protocol, Section I. Purpose and Applicability, the first sentence states that “the purpose of this protocol is to quantify the effects of chemicals on pubertal development and thyroid function in the intact juvenile/peripubertal female rat.”<br>On page 8, para 3 of ISR, Section III: Purpose of the assay, first sentence: It states that “The purpose of the female pubertal assay is to provide information obtained from an in vivo mammalian system that will be useful in assessing the potential of a chemical substance or mixture to interact with endocrine system.”<br><br>(ii) The Appendix 1 protocol in the section I. Purpose and Applicability, second sentence: It states that “this assay detects chemicals | See response to comment 1.8. |

| | | | |
|---|---|---|---|
| | | that display antithyroid, estrogenic, or antiestrogenic activity (e.g., alterations in receptor binding or steroidogenesis), or alter hypothalamic function or gonadotropin or prolactin secretion".<br>On page 8, para 3 of ISR, section III Purpose of the assay, second sentence: It states that "This assay is capable of detecting chemicals with antithyroid, estrogenic, or antiestrogenic activity or agents which alter pubertal development via changes in luteinizing hormone, follicle stimulating hormone, prolactin or growth hormone levels or via alterations in hypothalamic function.<br>Page 3 and the Table 1 recommended by the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) of ISR: Pubertal female (rat): "An assay to detect chemicals that act on estrogen or through the hypothalamus-pituitary gonadal (HPG) axis that controls the estrogen and androgen systems. It is also enhanced to detect chemicals that interfere with thyroid system". | |
| 1.6 | DR | (iii) On a similar smaller note, the ISR title is "Validation of a Test Method for Assessment of on Pubertal Development and Thyroid Function in Juvenile Female Rats as a Potential Screen in The Endocrine Disruptor Screening Program Tier 1 Battery". In this | The title of the protocol will be changed to add "intact" and "peripubertal". The title will therefore read "Test Method for Assessment of Pubertal Development and Thyroid Function in Intact Juvenile/Peripubertal Female Rats, for Use in the Endocrine Disruptor Screening Program". |

| | | | |
|---|---|---|---|
| | | title the context of use of the report is very clear, i.e., validation. However, in the pubertal assay protocol of the Appendix 1, the title starts from “Test Method----.” The context in which this pubertal protocol would be used was not clear in the title. Moreover, the title of both ISR and appendix 1 should be“------Intact Juvenile/Peripubertal Female Rats” instead of “----Juvenile Female Rats”. This will be consistent with the description in the purpose of the assay and also removes the ambiguity of intact versus ovariectomized. | |
| 1.7 | DR | The above different descriptions in the female pubertal protocol and in the ISR report of the purpose of female pubertal rat assay clearly create ambiguity and confusion, and it does not take into the account new knowledge in the field of endocrine disruptors.<br><br>For example, patterns of gonadotropin secretion during puberty in girls have become clearer as measurement techniques have improved. It is now widely recognized that endocrine or paracrine factors different from gonadotropins may play a relevant role as modulators of estrogen (E2) secretion early in the process of ovarian maturation that leads to premature sexual development in girls. A variety of growth factors, including | See response to comment 1.8 concerning the purpose statement, and comment 11.15 concerning new endpoints. |

| | | | |
|---|---|---|---|
| | | IGF-I, are considered to have a synergistic effect on gonadotropin-induced stimulation of ovarian steroid synthesis or aromatization and breast development. | |
| 1.8 | DR | The purpose of the assay in ISR needs to be re-worded to remove the ambiguity and to make it comprehensive and clear. Here is a draft of an attempt to re-word it: "The purpose of the female pubertal assay is to assess the potential of a chemical substance or mixture to interact with endocrine system which influences pubertal development and thyroid function in the intact juvenile/peripubertal female rat. This assay measures indices of pubertal development and is capable of detecting chemicals interacting with the estrogen, androgen, and thyroid hormonal systems, or agents which alter pubertal development via changes in luteinizing hormone, follicle stimulating hormone, prolactin or growth hormone levels or via alterations in hypothalamic function." | The purpose statement will be modified to read as shown in the next paragraph. The modified statement retains the idea that the pubertal female assay should not be evaluated in isolation from other assays in the screening battery when determining the potential of a chemical or mixture to interact with the endocrine system.<br><br>"The purpose of the female pubertal assay is to help identify chemicals or mixtures that have the potential to interact with the endocrine system, by identifying effects on pubertal development and thyroid function in the intact juvenile/peripubertal female rat. This assay is capable of detecting anti-thyroid, estrogenic or anti-estrogenic chemicals (including agents which act via alterations in receptor binding or steroidogenesis), or agents which alter pubertal development via changes in luteinizing hormone, follicle stimulating hormone, prolactin or growth hormone levels or via alterations in hypothalamic function." |
| 1.9 | DR | The descriptive text in the protocol needs some further improvement for the clarity. As described above, the first sentence of the protocol states that "the purpose of this protocol is to quantify the effects of chemicals on pubertal development and | See response to comment 1.8. |

| | | | |
|---|---|---|---|
| | | thyroid function in the intact juvenile/peripubertal female rat." The word "quantify the effects---" is misleading because some of the data, particularly histology of thyroid section are qualititative and semi-quantitative in nature. It would be appropriate to use "determine or investigate the effects----" instead of "quantify the effects-----." | |
| 1.10 | DR | The second sentence of this section states that "this assay detects chemicals that display antithyroid, estrogenic, or antiestrogenic activity (e.g., alterations in receptor binding or steroidogenesis), or alter hypothalamic function or gonadotropin or prolactin secretion". Are these measures or effects of chemicals described in the second sentence indices of pubertal development is not clear? The purpose of the protocol should clearly match with assay objectives with multiple endpoints. The impaired pubertal development includes early or delayed onset of puberty, impaired gonadal maturation (steroidogenesis) or ovarian function, shown by decreased or increased estrogen levels, impaired secretion of gonadotropin, thyroid hormones or prolactin. The clear connection between first two sentences is missing. The second sentence should be re-worded as "The assay is expected to identify the endocrine-mediated | See response to comment 1.8. |

| | | | |
|---|---|---|---|
| | | effects on female pubertal development by measuring puberty indices following exposure to chemicals with estrogenic or anti-estrogenic activity, inhibitors of steroid and thyroid hormone synthesis,-----." | |
| | | | |
| | | | |
| 2. Relevance of the assay to its purpose | | | |
| 2.1 | JB | If the purpose of the assay is to quantify the effects of chemicals on pubertal development and thyroid function, then the procedures should optimize the chance of success and minimize confounds that would obscure the results. This reviewer sees a number of serious problems with the protocol that present confounds. *[The specific comments listed by the reviewer are presented and addressed elsewhere in this response-to-comments document.]* | Responses to comments on specific confounding factors are addressed elsewhere in this response-to-comments document. In addition, the phrase "to quantify the effects of chemicals on pubertal development and thyroid function" will be removed from the statement of purpose. The emphasis of this assay is on *detection*, not quantification. |
| 2.2 | JB | General Statement about stressors: The influence of potential sources of stressors really needs serious consideration by an expert in developmental influences of stress on physiology, and not just the influence of stress on stress-related hormones. | The protocol will be changed to direct that care should be taken to minimize stress from all sources, including noise, other species housed nearby, or other disturbances. See the response to comment 12.i.8. The interlaboratory validation study showed that despite the lack of such direction in that study, the assay yielded correct results for the identification of interaction of the chemicals with the endocrine system. |
| 2.3 | JB | In this reviewer's opinion, this assay will add very little to a battery. For the cost involved, I would think that more direct tests of, for | The female pubertal assay provides valuable information from a mammalian system for detecting agents which affect the hypothalamic/pituitary/gonadal |

| | | | |
|---|---|---|---|
| | | example, estrogenicity could be used. For example, a uterotropic assay in ovariectomized or prepubertal rats or mice eliminates many of the confounds described in this review. | (HPG) axis and/or which interfere with steroidogenesis in the female. The Agency believes it important to screen *in vivo* for effects that may arise from modes of endocrine action other than receptor binding, as the Endocrine Disruptor Screening and Testing Advisory Committee recommended. The uterotrophic assay, which is included in the proposed battery, responds almost exclusively to agents that interfere with estrogen receptor binding by estradiol and certainly cannot detect agents that affect the HPG axis. |
| 2.4 | JB | Similarly, vaginal opening is an estrogen-dependent process, so this is really predominantly a bioassay for estrogenicity. There are, however, better, straightforward bioassays for estrogenicity (*e.g.,* uterotropic assays). | The proximal cause of changes in VO may be estrogen levels, but these levels may be affected by changes in the HPG axis or in steroidogenesis. Other assays for estrogenicity such as the uterotrophic assay are generally limited to detection of agents which interfere with the estrogen receptor. |
| 2.5 | JB | Although not the task I was assigned, I will make an unsolicited statement. This protocol is very disappointing from an endocrinological point of view, and although it addresses endocrinological questions, it appears to have been developed primarily by toxicologists without sufficient input from experts in endocrinology. My opinion is that the protocol could have benefited from the inclusion of at least reproductive and developmental endocrinologists in its development. My personal assessment is that, in its current form, it will provide scant information relative to the amount of work that will go into the experiments. As I have | Responses to the specific scientific issues that may have led to this comment are addressed elsewhere in this response-to-comments document. The Agency continues to believe that the vaginal opening endpoint provides a remarkably sensitive indicator of interaction with the endocrine system, and that the additional endpoints maximize use of animals which otherwise might have been discarded. (This assumes that the male pubertal assay and the female pubertal assay are run simultaneously.) |

| | | | |
|---|---|---|---|
| | | indicated, much of the work would not be publishable in a reputable endocrine journal. While probably not my place, I recommend that a group of scientists with diverse expertise (from toxicology to reproductive and thyroid physiology and endocrinology) be convened in the style of a scientific network to discuss this protocol from a wide range of perspectives. To do it serially, as is being done, slows down the process. | |
| 2.6 | BD | EPA has provided ample background and discussion on the biological and toxicological relevance of the assay and its ability to multiple mechanisms of interference with estrogen and thyroid hormone activities. The endpoints specified are amenable to a large scale screening program. Concern over a lack of a clear demonstration of assay specificity to this point is an issue that has been recognized and discussed. | Agree. No change in protocol. |
| 2.7 | DF | Measurements of are deemed appropriate for measuring interference with estrogen and thyroid hormone endocrine systems. Biologically relevant endpoints for thyroid hormone such as T4 and TSH measurements and thyroid histology are appropriate, as are time to vaginal opening, parameters of the onset and length of the estrous cycle, and uterine histology for alterations in the function of the | Agree. No change in protocol. |

| | | | |
|---|---|---|---|
| | | hypothalamic-pituitary- gonadal axis. | |
| 2.8 | HP | The need for an effective pubertal assay is unequivocal. There is sufficient evidence to suggest that the proposed protocol will be able to detect a large cohort of compounds that are capable of disrupting the endocrine system and thus alter the timing of puberty. Disruption of thyroid activity will also be detected in this assay although the inclusion of these endpoints complicates the design of the assay and introduces a number of caveats as discussed in detail below. Although the assay is generally well constructed, there are a number of critical issues that diminish the biological and toxicological relevance of the assay. | Agree. No change in protocol. Specific issues raised by the reviewer are addressed elsewhere in this document. |
| 2.9 | DR | The pubertal period is a very sensitive age for exposure to agents which alter the endocrine system. Therefore, this assay when validated should be able to detect chemical disruptors of estrogen, androgen, and thyroid action. The female pubertal rat assays can also identify compounds that alter hypothalamic-pituitary control of the gonads or thyroids. | Agree. No change in protocol. |
| | | | |
| | | | |
| 3. Transferability of the protocol | | | |
| 3.1 | JB | I do not agree with the conclusion that “The current study demonstrates that the female pubertal protocol is transferable and | Reproducibility of every endpoint across labs is not required in order to show transferability. Indeed, as EPA recognizes and several commenters have |

| | | | |
|---|---|---|---|
| | | reproducible in contract laboratories." As indicated above, day of vaginal opening was reasonably transferable, but many of the other parameters were not. Not being a toxicologist, I do not know what level of replication from lab-to-lab is expected. From an endocrinological point of view, a well-controlled study should be entirely (or at least nearly entirely) repeatable from lab-to-lab. | pointed out, reproducibility of ovarian and uterine weight measurements is difficult given that the animals are cycling. However, the assay is considered transferable when the overall weight of evidence for a chemical is considered. The overall weight of evidence relies heavily on vaginal opening as an endpoint, and this endpoint was consistent across laboratories. |
| | | | |
| | | | |
| 4. Repeatability and reproducibility of the assay | | | |
| a. General comments | | | |
| 4.a.1 | JB | The ISR (page 68) states in reference to the methoxychlor data that "…all three laboratories did identify a similar pattern of response for this weak estrogen and this response was positive for interaction with the endocrine system at the same dose level." While this was true for vaginal opening, it was not true for ovaries, pituitary, liver, adrenal, but most importantly, the prototypical estrogen-dependent tissue, the uterus. As stated elsewhere, this indicates that either the protocol cannot be followed reliably or more likely that many of the parameters do not show reliable responses in animals treated in the way that they are in this protocol (e.g., offspring of mothers shipped during mid-gestation, ovaries intact | Age and body weight at vaginal opening are the primary endpoints in this assay; the other estrogen-related endpoints are secondary. See the response to comment 3.1. |

| | | | |
|---|---|---|---|
| | | so possibly cycling, tissues taken at a time that they are still potentially responding directly to the compounds). | |
| 4.a.2 | JB | The ISR (page 73) indicates that "for the chemical which had never been tested before, results were consistent across all three laboratories. This reviewer does not see data to support this statement. Just examining yellow highlights on table 30, one can see that, while this is reasonably true for vaginal opening and liver weight, there were mismatches in ovarian weight, pituitary weight, uterine dry weight and adrenal weight. This reviewer's interpretation is that either the end-points measured are not meaningful in tests of these compounds, or the protocol cannot be reliably followed with sufficient sensitivity of outcome measurements to be meaningful. | This assay is intended to be used for detection of interaction of a test chemical with the endocrine system. Vaginal opening is the endpoint of greatest sensitivity in this assay and is consistent across laboratories. Consistency in detection of a test chemical as interacting with the endocrine system has been shown through the VO endpoint. Lack of response in other endpoints is not surprising given that the animals are cycling, but a positive response in such secondary endpoints may provide additional evidence of an interaction with the endocrine system. |
| 4.a.3 | JB | Repeatability is unlikely in many cases, because of the design of the experiments. There are simply too many confounds, ranging from differential exposure to stress conditions of some animals, food with potentially high levels of xenoestrogens, possible exposure to xenoestrogens from caging to tremendous variability because these are females, some of whom are cycling (and killed at random times in their very volatile cycle), some of whom are not. These have been covered extensively in | Data show that vaginal opening is sensitive and consistent despite the confounding factors cited by the commenter. The secondary endpoints are not as consistent, but provide useful information when positive. |

| | | | |
|---|---|---|---|
| | | previous sections | |
| 4.a.4 | JB | It was not clear to this reviewer, if TherImmune was supposed to be following the “test method” supplied, or a completely different protocol. There were many differences, which could compromise repeatability.<br><br>In the Therimmune study, 1143-103 and 1143-101, page 12, timed pregnant rats were received on GD 12. The protocol states protocol states that the animals could be rec’d on GD 7, 8, 9, or 10. I do not know if this was a change in protocol, or lack of attention by Therimmune.<br><br>In response to my question to the EPA re: what the difference was between 1143-101 and 1143-103, we were told that the dosing period of 1143-103 was one week later than 1143-101. Since the animals were timed-pregnant and all animals were received (according to the protocols) on the same day, and the protocol called for starting the experiment on the same postnatal day, how can that be? If the dosing started one week later, then the animals would be one week older than the protocol required. If this is the case, and I have no way of determining this, it would suggest to me that the laboratory does not understand the | The first TherImmune study was done in the early stages of prevalidation and the protocol was improved in later versions. The main outline of the assay as performed in that study has been retained, but several details have been changed. The primary test of repeatability was the interlaboratory validation study, which was done in three laboratories using the same written protocol.<br><br>Concerning the dosing periods of studies 1143-101 and 1143-103: Study 101 animals were received on Dec. 7 (see page 12 of Study 101 report), and Study 103 animals were received on Dec. 14 (see page 12 of Study 103 report). Both sets of animals were on their respective gestation day 12. These reports were available to the reviewers as Appendix 3. |

| | | | |
|---|---|---|---|
| | | importance of sticking to the protocol. | |
| 4.a.5 | JB | On Page 27, it is indicated that the experiment showed transferability of the protocol for methoxychlor. While it is true that the transferability of the protocol for vaginal opening was shown, they failed to find significant effects on two other parameters which would have been expected (uterine and ovary weight). My conclusion from the data would be that they did not replicate two previous findings | See the responses to comments 3.1, 4.a.1, 4.a.2, and 4.a.3. |
| 4.a.6 | BD | In general, the results reported in the validation studies, particularly the interlaboratory study conducted with three chemicals, indicate that the protocol generates data that lead to similar ultimate conclusions. Some issues that need to be clarified in interpreting the data have been raised above under responses to the earlier questions. | Agree. No change in protocol. |
| 4.a.7 | DF | The female prepubertal rat assay seeks to develop a reproducible and sensitive screening assay for the influence of xenobiotics on endocrine related endpoints, within an intact animal. The assay is sensitive and reproducible for the panel of compounds tested, and the endopoints measured are appropriate as a Tier 1 screen in conjunction with a battery of other in vitro and in vivo tests. | Agree. No change in protocol. |
| 4.a.8 | DF | In general, repeatability and reproducibility | Agree. No change in protocol. |

| | | | |
|---|---|---|---|
| | | was satisfactory, with the exception of the T4 measurements discussed under charge question 4. Furthermore, it is not unexpected that the wet weights of tissues would have so much variance and thus be of limited value since they cannot obviously be compared to wet weights prior to treatment. | |
| 4.a.9 | HP | Overall the repeatability and reproducibility of the assay is good and has been sufficiently demonstrated. However, the report identifies a number of areas where the data are inconsistent. In nearly all cases, the inconsistencies are noted in endpoints that are not needed to identify the disruption of puberty. Ovarian and uterine weights are unreliable measures of endocrine disruption, particularly when collected without regard to cycle. More guidance as to how cycle should be considered when collecting organ weight is needed. | Agree. No change in protocol.<br><br>The Agency considered standardizing the day of kill at a specific stage of the cycle in order to minimize variability in uterine and ovarian weight. However, this would have made it difficult to compare thyroid weights since the animals are still growing. Also, dose setting would have been more difficult inasmuch as body weights of both controls and treated animals would have a greater range of values. Therefore the Agency decided to retain the day of kill at a specific age despite the difficulties it presents for ovarian and uterine weights. |
| 4.a.10 | DR | It is assumed that all studies were conducted in professional contract laboratories with GLP facility. On minor points, it was not clear from the document whether each contractor purchased the animal, chemicals, and kits from the same source and the protocol of each laboratory assay was the same. This would have reduced some of the variability. | The goal of the interlaboratory study was to examine reproducibility of the assay when minor variations likely to be encountered in the “real world” during the Endocrine Disruptor Screening Program were allowed. The Agency recognizes that variability could have been limited even further by specifying exclusive sources of materials but did not feel that that would be an appropriate test of robustness of the assay in the Screening Program. |

| | | | |
|---|---|---|---|
| 4.a.11 | DR | Based on the ISR document, it appears that results obtained with the pubertal assay in the different contract laboratories are repeatable and reproducible, because all three laboratories data showed that the female pubertal assay may be useful for identifying chemicals that operate through a variety of mechanisms. This was true for both estrogenic and thyroid system interacting chemicals. For example, the TherImmune 1 study used a single dose of six different compounds in three different laboratories. Three different laboratories identified expected endocrine effects from exposure to chemicals with estrogenic, anti-estrogenic, androgenic or anti-androgenic activity, inhibitors of steroid and thyroid hormone synthesis, and a dopamine antagonist. ethynyl estradiol, tamoxifen (e.g, antagonist and partial estrogen agonist), and methoxychlor advanced the onset of vaginal opening. Propylthiouracil (e.g., an inhibitor of thyroid hormone synthesis), ketoconazole (e.g., an inhibitor of steroid synthesis) or pimozide (e.g., a dopamine antagonist) delayed the age of vaginal opening. The sensitivity of the protocol was assessed through multi-chemical study. Two different doses of six compounds were used for this study. The low doses of all six compounds showed expected changes in | Agree. No change in protocol. |

| | | | |
|---|---|---|---|
| | | the estrogen-related and thyroid system-related endpoints. The multi-dose study (TherImmune1 2) used three compounds, ethynyl estradiol, methoxychlor and phenobarbital and showed similar sensitivity to estrogenic compounds. Thus, the EPA in this ISR document has very correctly concluded that the female pubertal protocol is transferable, sensitive and reproducible | |
| | | | |
| | | | |
| b. Variability in endpoint values – uterine and ovarian weights | | | |
| 4.b.1 | JB | Problems with uterine weights, etc. Test Method, Page 17, para 4. Uterine weights and ovarian weights are not meaningful. The changes over the estrous cycle are likely to outweigh any effects of treatment. Likewise body weight has the same problem. An alternative would be to have parallel groups that are euthanized prior to puberty, so that the effects of the xenoestrogens on these variables can be assessed in the absence of ovarian hormones. | The Agency recognizes that uterine and ovarian weights are not likely to be particularly sensitive indicators of interaction with the endocrine system due to the normal variation in these weights in cycling animals. It considered having a parallel group that is killed prior to puberty (i.e., prior to onset of cycling), but this would significantly increase animal use. Also, the onset of puberty cannot be predicted for test chemicals; some chemicals may advance puberty and it would not be clear when to take measurements. The Agency decided to optimize the in-life measures (age and body weight at vaginal opening) rather than rely on uterine and ovarian weight. Nevertheless, uterine and ovarian weight may provide useful information when positive results are seen and so these endpoints are retained as part of the protocol. |
| 4.b.2 | JB | Test Method, Section X, para 10. If animal is not cycling, uterus and ovary will be either | The Agency agrees that age at vaginal opening and the percent of animals that are cycling regularly are |

| | | | |
|---|---|---|---|
| | | heavy or low, depending on stage (actually depending on the steroid hormone profile of the animal in the preceding day, so this is not meaningful. All that is important is age at vaginal opening, and perhaps whether the animals are cycling. | the most important endpoints for detection of estrogenic effects in this assay. However, it believes that uterine and ovarian weights are reasonably inexpensive to obtain and may provide additional weight of evidence to consider when a test chemical strongly interacts with the endocrine system. |
| 4.b.3 | JB | Problems with uterine weights, etc. Test Method, P 17. Uterine weights, ovarian weights and body weights are rather meaningless in the context of this particular protocol, evidenced by many of the "acceptable ranges" given in the protocol's performance criteria. Normal physiological fluctuation over the estrous cycle in response to cyclic changes in ovarian hormones is likely to outweigh many effects of treatment. Although the protocol states that regularity of cycling should be given more weight than lack of statistical significance for the difference in weight of ovary or uterus in treated animals compared to controls, this begs the question of what a positive result is likely to be attributable to. Uterine and body weights are very informative within the context of an ovariectomized animal, but not in an ovary-intact. As pointed out in the Integrated Summary Report, "It is also important to note that the variation in the uterine weighs was expected, since uterine weights | See response to comments 4.b.1 and 4.b.2.<br><br>Concerning the last two sentences of the second paragraph of this comment, and the last paragraph: The purpose of this assay is not to determine a mode of endocrine activity, but only to detect the potential for interaction with the endocrine system. The Agency agrees that an apical assay such as this one is unlikely to provide sufficient information to differentiate modes of action, even though in some cases it may provide information useful in hypothesizing a mode. |

| | | | |
|---|---|---|---|
| | | fluctuate during the estrous cycle and these females were killed on various days of their cycles."<br><br>Equally important, since animals are killed on PND 42, the last day of treatment, weights of any estrogen-responsive tissue or end-point will be a hodge-podge of direct estrogenic or antiestrogenic effects of the compound, indirect effects of the compound on the estrous cycle (Does it result in increased estradiol secretion? Does it result in long periods of anestrus?), and stage of the estrous cycle if the animals are indeed cycling. How can one know if a particular compound causes an increase in for example, uterine weight or a decrease in body weight via the compound, which is still available, having its uterotropic effect or body weight reducing effect, or by influencing these outcomes secondarily to perturbation of the hypothalamo-pituitary-gonadal axis (or by some other system)? There is a reason why the vast majority of endocrine studies on sex hormones is done in gonadectomized animals, and this is it. Considering the methods used, all that is important is whether the rats are cycling and age of vaginal opening, etc. But even with these variables, it must be remembered that advancing the age of vaginal opening can | |

| | | | |
|---|---|---|---|
| | | occur by a direct effect of an estrogenic compound on the vagina, or it can be secondary to HPG dysregulation.<br><br>An estrogenic xenoestrogen will advance puberty, as will estradiol, but it will also increase weight of uterus, and it will decrease estrous cyclicity. Therefore, an animal can develop a high uterine weight in a number of ways, for example, by an estrogenic xenoestrogens acting directly on the uterus (e.g., methoxychlor) or by lack of cyclicity with the ovary stalled in stage of follicular development (high endogenous estradiol). | |
| 4.b.4 | JB | On page 23, line 20, it is stated that large variations in weights were expected because the animals were killed on different stages of the estrous cycle. This is correct, but if so, why bother going to the trouble and expense of collecting all of these weights? | See the response to comment 4.b.2. The animals need to be killed in order to obtain thyroid weights, and obtaining ovarian and uterine weights is then a relatively minor additional expense. |
| 4.b.5 | DF | The assay protocol includes uterine wet weight as an endpoint (p. 7 line 7), yet in the ISR p. 59 line 8 it states that this is deleted from the protocol due to variability. This is unfortunate, since in ovariectomized rats, uterine wet weight is an excellent predictor of estrogenicity of a compound although the EPA peer review web site lists a rat | The protocol will be changed where indicated by the reviewer to remove uterine wet weight as an endpoint.<br><br>The use of ovariectomized animals as in the uterotrophic assay would not be appropriate in the female pubertal assay inasmuch as the female pubertal assay is intended to detect agents that interfere with the HPG axis and steroidogenesis as |

| | | | |
|---|---|---|---|
| | | uterotrophic assay as one of the battery of tests to be evaluated separately. | well as agents that interfere with estrogen-receptor binding. |
| | | | |
| | | | |
| c. Variability in endpoint values – thyroid hormones | | | |
| 4.c.1 | BD | The issue of variability of hormone measurements was discussed in the ISR, and it was noted that performance criteria for TSH had not yet been established. The mean TSH levels in control animals vary widely across the various studies discussed in the ISR. The multichemical study used carbon dioxide as a method of kill (Appendix 5, page 13) and had a high control TSH level, but similar levels were seen in one of the laboratories in the interlaboratory studies where brief carbon dioxide followed by decapitation was the method of kill (Appendix 15). The two other laboratories in the interlaboratory study used the same method of kill as the latter study and both of these studies reported control levels of TSH considerably lower than those reported in Appendices 5 and 15. The Office of Research and Development (ORD) studies reported in Table 22 (page 51) of the ISR had the lowest control levels of TSH, which no doubt reflects the extensive experience of this laboratory with this assay. Perhaps strict adherence to the ORD protocol should be specified. However, while the variability | The lower levels of TSH which were reported in the ORD studies are attributable to use of measurement materials from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDKD), which are known to report lower values than the Amersham kit materials used by the contractors. The method of kill is not important for TSH levels.<br><br>The protocol will be changed to ensure that laboratories meet the quality control standards established by the manufacturer of the specific TSH kit used. If the kit does not provide or specify a standard control, then the lab should use its own historical quality control samples. |

| | | | |
|---|---|---|---|
| | | of this endpoint is problematic, it appears that high control levels did not necessarily interfere with the ability of a laboratory to detect treatment effects. | |
| 4.c.2 | DF | The lack of a standard "recommended" assay for hormone measurements (p. 7 line 28), (several options are deemed acceptable as long as quality control samples are run) may have to be revisited due to the outlying or inconsistent T4 and TSH values reported in the interlaboratory exercise (Argus laboratory values; Figure 2 page 67 and Table 30 page 73 of the ISR). This may include having a separate lab coordinate all the hormone assays, or settling on a recommended assay kit and vendor to improve reproducibility. Nevertheless, the T4 and TSH values at least changed in the same direction in all three laboratories. | The values cited are regarded as outside the performance criteria and it does not appear to be appropriate to revise the performance criteria to cover this single set of data.<br><br>Although the Agency considered serving as the repository for quality control standards for hormone assays for the Endocrine Disruptor Screening Program, this was considered outside the scope of EPA's responsibilities. |
| 4.c.3 | DR | It is not clear how much blood is needed for hormone assay and at what speed it should be centrifuged. For the methodology of hormone measurement, four methods are described and it appears that the choice to choose was left to the contract laboratory. If available, the preferred choice of assay should have been time-resolved immunofluorometric assays (IFMA) particularly for measurement of gonadotropin concentrations. This is more | The assay will be changed to add that the blood should be centrifuged at 3000 g for 30 minutes. The amount of blood needed for the assays is specified by each kit's manufacturer.<br><br>The Agency has found through experience that IFMA does not work for rat TSH and provides poor results for rat $T_4$. |

| | | | |
|---|---|---|---|
| | | sensitive than radioimmunoassay or immunoradiometric assays. This would have helped in reducing intra-laboratory variability | |
| | | | |
| | | | |
| d. Variability in endpoint values – other | | | |
| 4.d.1 | JB | Without going into great detail, my general sense from scanning the spreadsheet and Table 30 was that, although there was reasonable consistency among labs in day of vaginal opening, there was not a great deal of consistency among labs in many of the other parameters. But even in "age at vaginal opening," Lab 1 saw no effect of DE-71 or 2-chloronitrobenzene, but saw an effect of methoxychlor, Lab 2 saw effects of all three test compounds, and Lab 3 saw effects of 2-chlornitrobenzene and methoxychlor, but not DE-71. Some reasons for discrepancies have been given in the foregoing discussion. This does not appear to this reviewer like a high degree of replicability/transferability. In addition, the lack of reproducibility of the other parameters raises the question of why all of these other parameters were measured, and then, why the pubertal assay is being developed instead of more straightforward tests. | As shown in Table 30 of the ISR, all three labs detected a delay in VO from 2-CNB, and all three labs detected an acceleration of VO from methoxychlor. There were some discrepancies between labs at the low doses, but the important point is that these two chemicals were consistently detected by all three labs as interacting with the endocrine system. Only one lab detected a change in VO from DE-71, but this chemical was included in the study because of its effects on the thyroid system, which all three labs correctly detected.<br><br>The Agency agrees that the inclusion of some of the other endpoints, such as ovarian and uterine weights, is a matter of judgment. The intent was to maximize the amount of information that is obtained from animals that otherwise might have been discarded unused. |
| 4.d.2 | BD | The issue of the variability of tissue weights of the pituitary and adrenal glands has been | In the case of the thyroid, fixing is needed in order to be able to separate the thyroid from the trachea. |

| | | | |
|---|---|---|---|
| | | addressed by indicating that steps need to be taken to avoid drying out prior to weighing (page 7). Weighing the organs after fixation (tissues placed in fixative immediately after removal) might also be considered as an approach to remedy the variability in these organ weights. | While fixation of the pituitary and adrenals could help reduce variability due to drying, the Agency believes that other, simpler methods will suffice. See the response to comment 12.n.3. |
| | | | |
| | | | |
| 5. Clarity of the protocol | | | |
| 5.1 | JB | Replace phrase "*in extremis"* by « at the point of death, » since this Latin term is not understood by all. It is also in the wrong place in the sentence, since as written, it states that they will be euthanized to the point of death, and it should probably state that the animals are found in the cage near the point of death should be euthanized. | The sentence will be changed to read "Animals which are found dead or which must be euthanized because they are near the point of death are removed from the cage." |
| 5.2 | JB | There are a number of problems in the description of the protocol that are likely to incorrect procedure. In general, I do not think it is sufficiently proscriptive and leaves too much to the judgment of the experimenter. Why should any aspects of the protocol be left to the discretion of the experimenter. This would be likely to result in variability in results. *[The specific comments listed by the reviewer are presented and addressed elsewhere in this response-to-comments document.]* | The Agency has attempted to standardize variables which are known to have an effect on the outcome of the assay while maintaining reasonable flexibility for variables which are thought to be less important. In general, flexibility in less-important variables helps to ensure that studies are implemented in a timely manner. |
| 5.3 | BD | For the most part, the methodology is clearly | Specific issues raised by the reviewer are addressed |

| | | | |
|---|---|---|---|
| | | described. I would propose the following for consideration. *[The specific comments listed by the reviewer are presented and addressed elsewhere in this response-to-comments document.]* | elsewhere in this document. |
| 5.4 | DF | In general, the protocol is well written, easy to follow, and instructions on data collection and reporting are clear. *[The specific comments listed by the reviewer are presented and addressed elsewhere in this response-to-comments document.]* | Specific issues raised by the reviewer are addressed elsewhere in this document. |
| 5.5 | HP | If the assay is to serve as a reliable and predictive tool for the identification of endocrine-disrupting compounds in females, it should be straightforward, clearly written, and easily performed in laboratories with reasonable testing experience. It should also be easily transferable and reproducible. There are a number of issues that impair these goals in the current protocol.<br><br>a. Objective – The objective is clearly stated in the protocol.<br><br>b. Methods – There are a few places within the methods where clarification is needed. *[The specific comments listed by the reviewer are presented and addressed elsewhere in this response-to-comments document.]* | Specific issues raised by the reviewer are addressed elsewhere in this document. |
| 5.6 | DR | The protocol is well written. Examples of | Agree. No change in protocol. |

| | | | |
|---|---|---|---|
| | | Tables 1-5 must be very helpful for the contract laboratories in measuring endpoints and preparing reports. | |
| | | | |
| | | | |
| 6. Clarity of reporting format | | | |
| 6.1 | JB | Reporting of results. In order to facilitate an understanding of the effects, standardized bar graphs should be required with a specific format. Generally, control group with mean +/- some indicator of variance, followed by the same for each of the treatments. Standard indicators of significance as would be found in a journal article should be placed above and below the bars. Although the TherImmune report had bar graphs, they were poorly set up. RTI had only very complex tables to wade through. In addition, it is standard to convert values like 0.0087 grams to 8.7 milligrams to facilitate digestion of the numbers. | The Agency believes that the reporting format included in the protocol, which was developed after the TherImmune transferability report and the RTI multi-chemical study, are adequate to understand the results.<br><br>The protocol will be changed to recommend reporting values for small organ weights in milligrams rather than grams. |
| 6.2 | BD | Page 12, second paragraph: This is somewhat confusing given the guidance on interpreting pituitary, liver, and kidney weights given on page 17 of the protocol. There it is indicated that organ weight changes for these organs should be considered only if they change significantly relative to terminal body weight. As mentioned previously, it wasn't clear in the ISR or lab study report whether the ratios | The inconsistency in the protocol will be resolved by changing the paragraph on page 12 to read as follows (additions highlighted in italics): "Report the mean, standard deviation, coefficient of variation, number of animals (N), and p-value for liver, kidneys, pituitary, adrenals, ovaries, uterus, and thyroid weights, for each treatment group, both unadjusted (U) and adjusted (A) for body weight on PND 21. *Report the mean, standard deviation, and p-value of the organ-weight-to-body-weight ratio for liver, kidney, and* |

| | | | |
|---|---|---|---|
| | | were being taken into account in the discussion of changes in the weights of these organs. It seems that for these three organs, the ratios to terminal body weight should also be reported in the table. | *pituitary only. For ovaries, uterus, and thyroid weights, d*o not use relative organ to body weight ratios, and do not adjust for body weight at necropsy."<br><br>Table 2 will be altered to add the relative weights for liver, kidneys, and pituitary. |
| 6.3 | HP | The data sheet for observations and measurements is sufficient and specific details about reservations regarding data collection were given above. | See response to comment 6.2. |
| 6.4 | HP | Data reporting is thorough and acceptable with the exception of the histological data. | See response to comment 6.2. |
| | | | |
| 7. Performance criteria | | | |
| 7.1 | JB | Positive controls compounds. A positive control with results that are known with certainty should be used. This is essential to demonstrate that the laboratory has the expertise and laboratory conditions sufficient to support replicating a previous result. Although a high dose can be used as a secondary control, a low dose, positive control to demonstrate reliability of the laboratory should be included in the protocol. | The Agency agrees that concurrent positive controls (preferably weak positives) are scientifically the best means of ensuring that the assay is being run acceptably. However, since this is an apical assay that responds to several different endocrine modes of action, positive controls would be necessary not only for estrogenicity through modulation of receptor binding but also through changes in the HPG axis and alterations in steroidogenesis – as well as for different thyroid-related mechanisms. The Agency discussed this with the Endocrine Disruptor Methods Validation Subcommittee and concluded that it would be unreasonable to require concurrent positive controls for all of these mechanisms when testing a single chemical. A certification program for laboratories was also briefly considered but would not have provided enough assurance that a study on a particular test |

| | | | |
|---|---|---|---|
| | | | chemicals was performed correctly. The Agency has chosen performance criteria on the control animals as the means to provide at least minimal assurance that the assay is being performed correctly. |
| 7.2 | HP | The lack of a positive control is a serious concern. Within the Integrated Summary Report, this omission is justified by the argument that it is highly unlikely that a single compound that will generate a positive result for all endpoints in the assay. This problem results from the inclusion of experimental endpoints designed to address two different and largely unrelated questions. By lumping pubertal endpoints, which assess estrogen action, together with thyroid endpoints, the choice of an appropriate positive control becomes complicated. It is readily apparent that the most salient and critical goal of this assay is to identify compounds that affect puberty. As such, a positive control that reliably and consistently advances puberty should be included, regardless of whether or not any thyroid endpoints are altered. Estradiol, DES, or estradiol benzoate would all be appropriate positive controls and at least one should be used by all labs for this purpose. Any labs not observing an effect with the positive control would then immediately know that they have a problem | See the response to comment 7.1. The issue is not whether to include positive controls for thyroid activity *vs.* estrogenicity, but rather inclusion of positive controls for all modes of endocrine activity such as the HPG axis and steroidogenesis. |

| | | | |
|---|---|---|---|
| | | executing the assay properly. | |
| | | | |
| | | | |
| 8. Data interpretation | | | |
| 8.1 | JB | It is unclear which data interpretation this refers to, since data interpretation is done at various stages… by the contract lab and by EPA. Having said that, clarity of the expected data interpretation may not be optimal. There a number of statistical issues to be considered in data interpretation. Many of the statements referred to, for example, an increase that was not significant. Statistically, an increase that is not significant is not an increase, and should not be phrased as though it is. Similarly, in the Summary Pubertal Interlab Results document, blue cells highlight "apparent," not statistical, dose-response relationships. Why? In biological research, all that counts is statistically significant effects. More statistical issues are discussed in section 4.d.<br>To be truly objective about the value of the work and to be statistically correct, the interpretation should rely on good statistical practices. The ISR in places has an appearance of wanting to "prove" the hypothesis/conclusion that this protocol is transferable. | The Agency agrees that interpretation of studies should be based on good statistical practices. |
| 8.2 | JB | The ISR indicates that listing vaginal cycles | The protocol includes a definition of "regular cycling" |

| | | | |
|---|---|---|---|
| | | as “regular” or not offered an informed summary of the data. I could not find any indication on how “regular” cycles were determined, nor what the definition was for “cycling.” Although page 9 of the protocol indicates how cycle length was to be computed, it does not indicate how many cycles are needed to qualify as having cycles, nor what constitutes regular, nor how to deal with the first days, which are usually acyclic. These are complex issues. It was very surprising that daily treatment with a fairly potent estradiol did not lead to disruption of estrous cyclicity.<br>On page 557 of the Therimmune final report 7244-600, for example, on page 557, animal number 9186 shows ten straight days of a diestrous vaginal smear, yet on page 558 she is referred to as cyclic. She actually exemplifies acyclicity. I did not go through all of the data for examples like this, because it is clear that the instructions did not indicate precise definitions. | on page 6. However, the Agency agrees that the observation period in the assay is short and there may not be sufficient time to determine whether cycling is regular over several cycles, particularly if vaginal opening is delayed and cycling does not begin until just a few days before the end of the study. The protocol will be clarified to read as follows (additions highlighted in italics):<br><br>“At the end of the study, the overall pattern of each female is characterized as regularly cycling (having recurring 4- to 5-day cycles), irregularly cycling (having cycles with a period of diestrus longer than 3 days or a period of cornification longer than 2 days), or not cycling (having prolonged periods of either vaginal cornification or leukocytic smears). *In cases where there are too few days between vaginal opening and the end of the study to observe more than one cycle, classification will have to be based on the available data with the default assumption that animals are cycling regularly if the partial data fit the definition, and are irregularly cycling if the study ends without being able to distinguish between irregular cycling and not-cycling.”*<br><br>The Agency also notes that if VO is delayed by many days, the delay itself strongly indicates interaction with the endocrine system and the regularity of subsequent cycling is of less importance than if VO had not been delayed. |
| 8.3 | JB | Statistical issues. Statements like (ISR, | The Agency agrees that interpretation of studies |

| | | | |
|---|---|---|---|
| | | page 23, line 8) “The ovarian weight was reduced, although this effect was present at p=.06, marginally beyond the p<0.05 cut-off for statistical significance” are inappropriate. Either a result is statistically significant (< .05), or it is not. Results that are not statistically significant cannot be used as support. Likewise, statements like “There was a nonsignificant increase in uterine weight (p = 0.08 adjusted for weaning), and a non-significant decrease in pituitary weight (P = 0.10)” (ISR, page 26, lin 16) are inappropriate. If the result is not statistically significant, and it is therefore due to chance, it really does not matter if it is an increase or a decrease; stating as the report does, that the non-significant increase is consistent with another paper’s non-significant increase is inappropriate. | should be based on good statistical practices. When interpreting data on specific chemicals, such statistical standards will apply. For validation of the assay, too, the main criterion was whether an assay provided the overall correct answer for a chemical given the weight of evidence. However, when considering individual endpoints during validation, it seemed reasonable to consider whether the endpoints were behaving similarly to expectations even though the arbitrary cutoff point of p < 0.05 might not have been achieved. |
| 8.4 | JB | *[Regarding transferability study]* It is stated that they showed transferability, because all of the drugs had expected results on advancing or delaying puberty. However, tamoxifen advanced, so it is unclear why this is considered success. While it was fine for day of vaginal opening, what would be considered positive results for the other parameters? | Tamoxifen is a selective estrogen receptor modulator and induced pseudoprecocious puberty. The age at vaginal opening decreased, while age at first estrus increased. This was expected behavior. It should also be noted that the assay is being used only to detect interaction with the endocrine system, not to determine mechanism. Thus with either of the endpoints alone being significant (and assuming no further information was available from other endpoints or other assays in the battery), tamoxifen would have correctly been identified as interacting with the endocrine system. |

| | | | |
|---|---|---|---|
| 8.5 | JB | On page 42, it is stated that ethynyl estradiol-treated rats had normal cycles. This is quite unexpected, since chronic estradiol treatment should stop the cycling by negative feedback. So, why would ethynyl estradiol result in normal estrous cycles? | The Integrated Summary Report (page 42) says that 100% of the animals were cycling but that only 20% were cycling normally. |
| 8.6 | BD | In general, the interpretations of the data generated in the validation studies of the female pubertal protocol that are presented in the Integrated Summary Report (ISR) were clear and thorough. There were a few points that caused some confusion.<br>1) In the ISR discussion of the results of the validation study that included bisphenol A (ISR, page 28, lines 21-25) it is basically concluded that the assay did not detect BPA activity with the possible exception of the body weight depression. In later discussion of the effects of estrogens on body weight (ISR, page 53, paragraph starting on line 22) it appears that, based on results with ethynyl estradiol and methoxychlor, a body weight depression in the absence of an effect on other endpoints such as vaginal opening would not be interpreted as an estrogenic activity. It is not clear how the BPA data would have been interpreted for a compound that did not have the extensive body of published data that exists for BPA. If possible, a more | The multi-dose-level study (TherImmune 2) showed that a significant decrease in age and weight at vaginal opening can be observed even at dose levels of estrogens which do not affect body weight. (See Tables 15 and 16 on methoxychlor and ethynyl estradiol, respectively, on page 39 of the ISR.) That is, VO appears to be more sensitive than body weight to the effects of estrogens. Thus, body weight depression in the absence of an effect on other endpoints such as vaginal opening would not generally be interpreted as evidence of estrogenicity. BPA would probably not be considered estrogenic given only the data from this assay, as stated on page 35 of the ISR. |

| | | | |
|---|---|---|---|
| | | definitive statement should be made about the sensitivity of body weight relative to the other endpoints with regard to estrogenic test compounds | |
| 8.7 | BD | Another instance where the ISR was in disagreement with the laboratory study report was the case of 2-chloronitrobenzene (2-CNB). The Argus report (Appendix 13) indicated that the thyroid showed histological changes consistent with a hypothyroid state, but that there were no significant changes in thyroid hormone levels. The ISR (Table 30, page 71) indicates that there was a significant decrease in T4 in the Argus study. The ISR does not mention histopathological results from the 2-CNB study thyroid component and there seems to be only sporadic use of histopathology results in interpreting the various validation studies throughout the ISR. In the case of 2-CNB, it is interesting to note that the laboratory that did not report changes in thyroid hormone changes did report treatment-related histological changes while the other laboratories reported thyroid hormone changes in the absence of histological changes. None of the laboratories reported effects on thyroid weight for this compound. One of the laboratory study reports (Appendix 15) indicated that the meaning of an isolated | Interpretation of the thyroid endpoints must be done in the context of results from other relevant assays in the Tier 1 battery. Consistent results from the male pubertal and the amphibian metamorphosis assay, for example, will allow a more confident interpretation of thyroid-related results from the female pubertal assay. In particular, a small change in $T_4$ alone, in the absence of any confirmatory evidence from other thyroid-related endpoints in this assay or in other assays, should generally not be interpreted as evidence of interaction with the thyroid system. |

| | | | |
|---|---|---|---|
| | | rise in TSH in the absence of other significant thyroid effects was unclear. The variability of hormone measurements in the thyroid assay was discussed in the ISR, but further discussion of what will constitute a positive call when results from only one laboratory are available might be helpful. | |
| 8.8 | BD | There are several instances in the ISR (for example, pages 26, 30, 31, 50, and 70) where changes in means that are not statistically significant are discussed when the changes are consistent with what was expected based on previous studies or are consistent with other observations in the study. Comment on how such results would be used in reaching a decision on the activity of a compound might be appropriate in Section XVI (Data Interpretation) of the protocol. | See the response to comment 8.3. |
| 8.9 | BD | Page 5, Section VIII, Vaginal Opening: It is indicated that documentation of vaginal threads or the appearance of pin holes are important and that it is critical that the "initiation" of vaginal opening be recorded. However, guidance on the interpretation and reporting of these endpoints is not provided. | The protocol will be modified to clarify that a "pinhole" is not considered initiation of vaginal opening, even though it must be recorded when observed. A photograph of a "pinhole" will be included in the protocol to assist the investigators in proper identification. |
| 8.10 | BD | Page 17, last paragraph: I don't believe that this paragraph is necessary. Presumably, the fact that an isolated endpoint showed significance at the low dose, but not at the high dose would be taken into account in | There seems to be nothing in the paragraph in question that conflicts with the comment, and the Agency believes the paragraph to be important. |

| | | | |
|---|---|---|---|
| | | reaching a conclusion about the substance based on the results of this assay and when considered with results obtained in the other battery assays. | |
| 8.11 | BD | Page 18, Table 6: Consider adding a column describing anticipated changes for an estrogen antagonist. | The Agency believes it would be imprudent to add such a column until data from more anti-estrogens are available. |
| 8.12 | BD | Page 9, to paragraph: "When statistically significant effects are observed ($p < 0.05$), treatment means are examined further using appropriate pairwise comparison tests to compare the control with each dose group." This approach may be overly conservative for a screening assay for certain multiple comparison procedures. In many of the validation study procedures, Dunnett's test was used for comparisons of treatment groups to controls. The following statement is taken from Haseman et al., Statistical issues in the analysis of low-dose endocrine disruptor data" Toxicol. Sci. 61: 201-210, 2001: "For example, Dunnett's test is a standalone test that does not require statistical significance of an overall ANOVA to be valid. However, many investigators who used Dunnett's test required statistical significance of an overall ANOVA before making pairwise comparisons. Because the critical values for Dunnett's test were derived without consideration of an overall ANOVA, requiring this additional | The Agency will take this comment under advisement and may change the protocol. Dunnett's test is acknowledged to be independent of ANOVA. Because circumstances vary and one statistical test may be appropriate in one circumstance but not another, the Agency hesitates to specify that Dunnett's test must be used for pairwise comparison of a treated group to control, in place of ANOVA and ANCOVA followed by a different pairwise comparison test. |

| | | | |
|---|---|---|---|
| | | significance may result in a somewhat conservative test. Specifically, there were a few instances in which our reanalysis found significant pairwise differences by Dunnett's test that were not reported as significant by the study investigators who themselves also used Dunnett's test. Such differences were apparently due to the extra requirement of a significant overall ANOVA imposed on Dunnett's test by the study investigator." Revision of this recommendation should be considered, depending on the multiple comparisons test to be applied. While this statistical guidance may not have affected the interpretation of the validation studies, it could potentially affect conclusions in future screens. | |
| 8.13 | DF | Clarity, comprehensiveness and consistency of the data interpretation with the stated purpose of the assay.<br>These issues all appear appropriate; some caveats are discussed in other sections below. One item to note here is that changes in liver enzyme profile are an indicator of the presence of a potential thyrotoxicant (p. 18 Table 6), but liver enzymes are not measured in the protocol. | Liver enzymes are relatively difficult to measure accurately and were not judged to be appropriate for a screen. An increase in liver weight is a surrogate for liver enzyme induction in this assay. |
| 8.14 | HP | The assay would benefit from simplification. A few salient endpoints, collected well and with careful controls, would be far superior to the broad spread of largely unrelated | Age and weight at vaginal opening are the simple and sensitive endpoints that form the core of this assay. Other estrogen-related endpoints were included, even though they are not as consistent, in order to |

| | | | |
|---|---|---|---|
| | | endpoints currently proposed. Inclusion of both estrogen and thyroid related endpoints complicate the interpretation of the data and as such, in many cases, the results of the data are listed as "difficult to interpret." For the assay to be effective and reliable a data set that is "difficult to interpret" should be rare rather than the norm. A screening assay should yield straightforward results that clearly and quickly identify compounds that require further testing. | maximize the amount of potentially useful data obtained from this assay. Similarly, thyroid-related data are obtained so that when data from the battery of Tier 1 assays is considered together the various pieces, which may not be convincing on their own, may together provide a higher level of confidence. |
| 8.15 | HP | It should be noted that the majority of endpoints within the assay are related to puberty rather than thyroid function. The thyroid endpoints, although interesting in their own right, feel out of place in the context of the assay. They also complicate the interpretation of the results (as discussed more in depth below) and preclude the inclusion of a positive control for the estrogenic endpoints. A "female pubertal assay" should be specific, simple, and straightforward. If possible, it would be advisable to develop a separate assay to assess thyroid function. | The amphibian metamorphosis assay was developed as a separate assay that focuses on thyroid function. Nevertheless, the Agency believes it inappropriate to rely solely on one assay to identify the potential for interaction with the thyroid system (unless the test chemical is not testable in other systems). Information from the mammalian pubertal assays may in some cases be important in determining the weight of evidence for a chemical's interaction with the thyroid system. |
| 8.16 | HP | In many cases, the results of the data are listed as "difficult to interpret." This is a significant concern as an EDSP Tier 1 Screening Assay should generate a data set that is relatively simple to interpret and reliably identifies compounds that require | The female pubertal assay is intended to be interpreted in the context of data from all of the assays in the Tier 1 battery. Thus it is inappropriate to conclude from this assay alone whether a chemical would move to Tier 2. |

| | | | |
|---|---|---|---|
| | | further testing. A laboratory running this assay should easily be able to conclude that a compound either produces or fails to produce an effect. The interpretation of the results should be as unequivocal as possible. The results of the study using 2-chloronitrobenzene (beginning on page 70 of the Integrated Summary Report) best illustrates how results from a compound, for which little about potential endocrine activity is known, might be interpreted. Two of the three laboratories reported significantly delayed vaginal opening. Significant changes in weight at vaginal opening, liver weight, adrenal weight and uterine weight were also observed in at least two of the laboratories. This was an unexpected finding but the authors ultimately conclude (albeit tenuously) that 2-chloronitrobenzene interacts with the endocrine system though in indeterminate mechanism. No information is given as to how this compound would then be classified. Would it move to Tier 2 screening? Would the results be questioned? How would this data be received by the EPA? This compound was selected for screening because it was hypothesized to have no effect on the endocrine system. The data do not support the hypothesis. How would that data be used? Would it be questioned? | It should be noted for the record that the text on page 70 of the ISR is incorrect. Data from all three laboratories showed an increase in both age and weight at vaginal opening from 2-CNB, as correctly shown in Table 30 (page 71), when analyzed for covariance with body weight at weaning as required by the protocol. |

| | | | |
|---|---|---|---|
| 8.17 | HP | A citation for the grading scale associated with the thyroid sections is given, but no citations are listed for the evaluation of follicular development or uterine histology. Collecting uterine and ovarian weight has consistently proven to be problematic and it is not clear how these endpoints are informative. Time of vaginal opening, time of first estrus, and regularity of the estrus cycle 10-12 weeks post puberty are far more valid and reliable endpoints of pubertal alteration and are sufficient to draw conclusions about whether or not a compound should undergo Tier 2 testing. Therefore the organ weights are a time consuming and problematic component that are not needed and could be eliminated. Uterine and ovarian weights are also confounded by cycle. Although the protocol states that the estrous cycle at the time of necropsy should be taken into account, guidance as to how to do this is insufficient. | The Agency agrees that age and weight at vaginal opening are the primary endpoints in this assay, with time of first estrus and regularity of cycling also providing valuable information. In EPA's judgment, the uterine and ovarian weights are not difficult to obtain and may also yield useful information in some cases. |
| 8.18 | HP | Section XII of the protocol states that uterus, thyroid, ovary and kidney are to be evaluated for pathologic abnormalities but no guidance is given as to how the histological findings are to be quantitatively assessed. It is unclear how the histological data will contribute to the interpretation of the data. | The protocol will be changed to add reference to the rapid screening method for ovarian toxicity described by Smith BJ et al. (Reprod Toxicol 5:379, 1991) as discussed in Plowchalk DR et al. (Chapter 5 in Tyson CA, Witschi H, Methods in Toxicology, Vol 3B, Female reproductive toxicology, Heindel JJ, Chapin RE eds.,1993). The following paragraph will be added to the Data Interpretation section of the protocol: |

| | | | |
|---|---|---|---|
| | | | "The judgment of the histopathologist as to whether an effect on ovary, uterus, and/or thyroid is associated with exposure to the test chemical must be considered when evaluating the organ weights and hormone levels measured in this assay. Severity and incidence of effect(s), and dose-response relationship may also be important information to consider."<br><br>(Histopathology of the kidney was already mentioned in the second paragraph of the Data Interpretation section.) |
| 8.19 | DR | In the pubertal protocol (Appendix 1), Table 6 entitled "Potential changes indicative of different mode of action that may be observed in female pubertal protocol" provides very clear comprehensive summary of expected different effects that may be observed from different modes of action. Using this table, chemical substances that that exert effects via various mechanisms or different modes of interaction with the endocrine system can be identified. The description of the data interpretation is very consistent with the stated purpose of the assay. The guidelines described in the text given for data interpretation for doses level tested, explanation of negative results in the context of interaction with endocrine system, performance criteria, and evaluations of endpoints are very clear. The same is true | Agree. No change in protocol. |

| | | | |
|---|---|---|---|
| | | for ISR report which describes this in the data interpretation section (page 74-77). | |
| | | | |
| | | | |
| 9. Appropriateness and completeness of validation | | | |
| a. Number and kinds of chemicals tested in validation | | | |
| 9.a.1 | BD | With regard to the discussion of BPA in the ISR, it is indicated on lines 18-19 of page 28 that body weight at vaginal opening was increased at the high dose. However, the data in Table 12 and the data presented in the laboratory study report seem to indicate that the body weight at vaginal opening was decreased at the high dose. Consideration of this result along with the ovarian weight data would likely change the conclusion reached on line 19 of page 29 that the expected estrogenicity of BPA was not detected. | The text of the ISR on page 28 lines 18-19 is incorrect and was based on an incorrect value for bodyweight at vaginal opening. A correction was eventually made to Table 12 but inadvertently not in the text. |
| 9.a.2 | BD | The selected test substances covered a variety of mechanisms of action purported to be detected by this assay and, for the most part, expected results were obtained. The selection of fenarimol, which apparently has a mixed mode of action that was not known at the time of selection, was described in the ISR as unfortunate, but it is likely that some unknown test compounds will interact with multiple targets and the data obtained with that compound did have utility. Inclusion of a broader range of compounds in the | Agree. The Agency notes that early in the validation effort it had discussed with the Endocrine Disruptor Methods Validation Subcommittee its plan to test a wide range of chemicals in individual laboratories during prevalidation, but to focus on only a few chemicals across several laboratories in the interlaboratory study. As noted, this approach was taken to minimize costs while still obtaining relevant information on the applicability of the assay. |

| | | | |
|---|---|---|---|
| | | interlaboratory study would have provided added confidence in the performance of the assay, although it is recognized that the additional cost may have made that impractical. | |
| 9.a.3 | DF | The choice of test substances for the testing the performance of the assay are appropriate and span a spectrum of agents with distinct suspected modes of action. A few comments are included for future consideration however. For a test substance to detect thyroid hormone disruption, a more relevant choice than PTU may have been ammonium perchlorate since this compound is a known inhibitor of iodine uptake by the thyroid and is found in ground water, particularly near air force bases. Also, a more specific aromatase inhibitor such as fadrazole rather a general steroidogenesis inhibitor such as ketoconazole than may have been useful for assessing the specificity of the female rat assay in particular. Interestingly, it is apparently difficult to find a test substance known to be generally toxic but does not in some way impact any of the endocrine related endpoints in this assay. One possibility is that high doses of generally toxic chemicals induce hepatic phase I and II metabolizing enzymes and phase III transporters as a by-product of exposure. This possibility was | The Agency originally intended to test perchlorate in the interlaboratory validation study but was discouraged from doing so because of concerns in the contract lab about preparing a potentially explosive solution of this substance in an oil vehicle. The obvious alternative (preparing the solution in water) was not acceptable because the pubertal study which was used as the reference for this chemical had actually tested using corn oil as solvent (summary data shown on page 51 of the ISR). DE-71 was substituted as a weak positive thyroid-active agent. PTU had been tested in the very first pubertal study when a strong agent was needed to determine whether the pubertal assay responded at all to thyroid agents, and was repeated at lower doses to try to determine the sensitivity of the assay.<br><br>The Agency also made persistent efforts to obtain fadrozole for the multi-chemical study but was unable to do so for reasons unrelated to the study.<br><br>Screening for mRNA expression may be appropriate to consider for future improvement of this assay. |

| | | | |
|---|---|---|---|
| | | noted in the ISR (p 70 line 19)In this regard, screening the activity and/or mRNA expression of a panel of known steroid and thyroid hormone metabolizing enzymes in liver biopsies should be an important endpoint to consider. | |
| 9.a.4 | HP | Given that no compound to date has tested negative in this assay it is difficult to evaluate the potential effects of test substance on outcome. | The Agency notes that while no substance has tested negative for both estrogenicity and thyroid-related effects, methoxychlor and atrazine gave positive results only on estrogenic endpoints, while DE-71 was specific in two out of three labs for thyroid effects. This suggests that the assay is specific, and may be the best indication of specificity that is possible in the absence of a reference compound that is known to be negative for both estrogenicity and thyroid-related effects. |
| 9.a.5 | DR | The choice of test substances and analytical methods is well thought and is well described in this document. It has been identified in the report that the cost associated with animal experiments did not allow to use appropriate ranges of positive and negative test substances, particularly weak positive controls. However, the use of weak positive compound was critical for validation of this study because most of the unknown test compounds are hormonally weak substances. | Agree. No change in protocol. The Agency deliberately tested weak compounds in the interlaboratory validation study as an intentional challenge to the repeatability of the assay in different laboratories. |
| | | | |
| | | | |
| b. Statistical methods used in validation | | | |

| | | | |
|---|---|---|---|
| 9.b.1 | JB | For reasons discussed above, the statistical analysis has problems. *[The specific comments listed by the reviewer are presented and addressed elsewhere in this response-to-comments document.]* | Specific issues raised by the reviewer are addressed elsewhere in this document. |
| 9.b.2 | HP | Analysis of variance is to be used with body weight at weaning as the covariate. Day of the estrus cycle is not taken into consideration in the analysis of organ weights but should be for the ovary and uterus as it likely has a greater effect on the weights of these organs than overall body weight. | This approach has been attempted in certain in-house studies but there are usually too few animals per group (based on day of estrus cycle) to allow acceptable analysis. |
| | | | |
| | | | |
| 10. Strengths of the assay | | | |
| 10.1 | JB | completely agree with the “strengths” section of the ISR (page 73) that an *in vivo* screen is desirable. I do not agree that this assay has as much strength as is discussed in the ISR. | General comment. No response needed. |
| 10.2 | BD | EPA has provided a good discussion of the strengths and limitations of this assay. Strengths include the coverage of multiple mechanisms of action that are not covered in other proposed assays for the battery and exposure during a sensitive time period, although it could be argued that the perinatal period would be more sensitive for some endpoints. | Agree. No change in protocol. |

| | | | |
|---|---|---|---|
| 10.3 | DF | Strengths of the assay: The major strength of the assay is that it is able to detect the activity of chemicals with regard to the endocrine system in an intact mammalian system. Cell culture and in vitro biochemical screening assays are useful to provide information on what effect a chemical can have on the function of the endocrine system but does not account for cell specific uptake, metabolism, effect of other circulating growth factors/hormones etc. On the other hand, in vitro studies provide more mechanistic information than are inferred from the in vivo system. Therefore as a component of a battery of assays, the prepubertal female rat assay appears relatively sensitive (see below for caveat) and standardizable to detect interactions between xenobiotic chemicals and estrogen and thyroid hormone physiology. It is especially noteworthy that a so-called "weak" estrogen, methoxychlor, and a "weak" thyroid disrupting compound DE-17 produced quantifiable and generally reproducible effects on the expected endpoints. | Agree. No change in protocol. |
| 10.4 | DR | The IRS report very clearly describes various studies of low and high doses of chemical substances conducted in the different laboratories. Each contract laboratory showed that the pubertal assays | Agree. No change in protocol. |

| | | | |
|---|---|---|---|
| | | can identify compounds that alter hypothalamic-pituitary control of the gonads. All studies using thyroid-active agents showed that the female pubertal assays detect alterations in thyroid function following exposure to compounds interacting to thyroid system. I concur with EPA conclusion in regards to the strengths and weakness of various assays. The one of the major strengths of this study is that it is an in vivo assay, and it can measure the effects of both parent compounds and their metabolites. This assay estimates the interaction with the endocrine system. Additionally, this assay measures the effects of the endocrine disruptors at one of the critical time period of the development of the animal, which is highly sensitive to changes in the endocrine system. This would help in identifying weak endocrine disruptors. Use of the redundant multiple endpoints increased the credence of the assay. This was further strengthened by the use of very well though performance criteria. | |
| | | | |
| | | | |
| 11. Limitations of the assay | | | |
| 11.1 | BD | A major current limitation, as pointed out by the EPA, is the lack of demonstrated specificity of the assay. | See response to comment 9.a.4. |
| 11.2 | DF | The specificity of the assay remains a | See response to comment 9.a.4. |

| | | | |
|---|---|---|---|
| | | question due to the effect of 2-CNB. A potential negative control, 2-CNB, delayed vaginal opening and increased TSH levels. This issue is discussed in more detail below, but may be the result of increased metabolism of estradiol and/or thyroid hormones by the 2-CNB exposed liver. | |
| 11.3 | BD | Another limitation is the impracticality of conducting the screen in more than a single strain to provide confidence in studies where endocrine activity is not detected. | The impracticality of conducting assays in multiple strains is a situation common to almost all *in vivo* assays and is not a "limitation" specific to the pubertal assays. See also section 12.d of this document. |
| 11.4 | DF | The major drawbacks include continued concern about rat strains chosen for the study, the window of exposure to the compounds, and the lack of behavior endpoints in the assay. | Strain of rat: see section 12.d.<br>Window of exposure: see response to comment 11.5.<br>Behavior endpoints: see response to comment 11.6. |
| 11.5 | DF | Exposure duration is confined to PND 22-PND 42. The effect of fetal or perinatal exposure to chemicals is not accounted for in the current protocol. There is growing evidence that early exposure to chemicals or maternal stress can "program" adult physiology of the offspring (e.g. Newbold et al. Reproductive Toxicology 23:290–296 (2007)), in addition to sexually dimorphic behaviors (see c). | The need for an assay that would cover exposure *in utero* through lactation was recognized by the Endocrine Disruptor Screening and Testing Advisory Committee and acknowledged by the Agency in 1998. The Agency attempted an *in utero* through lactation assay and discussed the results with the FIFRA Scientific Advisory Panel in February 2007. The study results were disappointing, possibly due to the chemical chosen (methoxychlor). (See http://www.epa.gov/scipoly/sap/meetings/2007/february/sap-2007-04-report-f.pdf ) |
| 11.6 | DF | The assay does not account for behavioral effects. In many cases, low doses of estrogenic compounds can affect sexually dimorphic brain organization, at lower or | Sexually dimorphic brain organization is affected primarily from prenatal exposure. See response to comment 11.5 for discussion of *in-utero*-through-lactation assay attempted for the screening program. |

| | | | |
|---|---|---|---|
| | | similar doses to those required to elicit morphological changes in reproductive tissues. Therefore, I was somewhat surprised that behavioral endpoints were not included in the assay, such as lordosis, for example. | Behavioral endpoints would require extending the assay for a significantly longer period of time. |
| 11.7 | HP | A number of elements within the protocol diminish the functional utility of the assay. Ovarian and uterine weights are generally uninformative and complicated by cycle. Inclusion of the thyroid endpoints precludes the use of a needed positive control group for the pubertal measures. The duration of estrus monitoring is too short and the use of daily lavage will likely induce pseudopregnancy in some animals, potentially confounding the data. Failure to eliminate phytoestrogens introduces an unnecessary confound and increases the risk of inter-laboratory variability. Finally, only two doses are to be used, both of which are based on body weight and neither of which will approximate a "typical" human or wildlife exposure. The failure to include a dose within a reasonably physiological range is a considerable concern. | Ovarian and uterine weights: See section 4.b of this document.<br><br>Inclusion of thyroid endpoints: See responses to comments 7.1, 8.14, and 8.15.<br><br>Duration of monitoring of estrous cycles: See response to comment 11.9.<br><br>Daily lavage: See response to comment 12.l.1.<br><br>Phytoestrogens: See section 12.j of this document.<br><br>Dose at physiological range: See response to comment 12.b.5. |
| 11.8 | HP | Daily lavage frequently induces pseudopregnancy in rats and can skew data regarding regularity of the estrus cycle and cycle length (Marcondes et al., 2002; Yener et al., 2007). Pseudopregnancy lasts 1-2 | See response to comment 12.l.1. |

| | | | |
|---|---|---|---|
| | | weeks, during which time the vaginal smear will have the appearance of persistent diestrus. No discussion as to how to deal with this frequent phenomenon is given in the assay. | |
| 11.9 | HP | Animals are to be killed on PND 42. This means that estrus cyclicity will be assessed for less than two weeks and animals will be killed before they have completed the transition to full adulthood. This is an insufficient amount of time to evaluate the regularity of the estrus cycle and will likely fail to detect a number of compounds that ultimately impact endocrine action in females. The impact of endocrine disrupting compounds on the estrus cycle can be delayed by several weeks (Gallo et al., 1999; Patisaul and Polston, 2007; Rubin et al., 2001; Whitten et al., 1993). Observation of the estrus cycle for at least six weeks post puberty is strongly recommended. | Extending the observation of the estrus cycle from PND 42 to PND 75 (i.e., six weeks past the typical day of VO in normal SD rats) would add over a month to the in-life portion of this assay. The references cited seem to indicate that differences in estrus cycling between treated groups and controls were seen at various extended times chosen by the investigators, but with one possible exception (Whitten et al. 1993, in which fewer *control* than treated animals were cycling after 99-108 days of daily exposure) there seems to be little or no examination of whether an additional month of exposure would add a significant capability to detect changes in cyclicity, particularly when using the pubertal regimen in which exposure does not begin until PND 22 (as opposed to the perinatal exposure used in several of the studies cited). There does not appear to be adequate support for extending the length of an assay that is intended to be a screening assay by 6 weeks.<br><br>See also the response to comment 8.2. |
| 11.10 | HP | As discussed in detail above, the assay maximizes breadth at the expense of a robust experimental design to answer a specific question. Removal of the thyroid endpoints and the organ weights would | See the responses to comments 7.1, 7.2, and 8.14. |

| | | | |
|---|---|---|---|
| | | allow for the inclusion of a needed and critical positive control group and make the data vastly easier to interpret. Simplification would also make the assay easier to replicate across laboratories. To assess pubertal disruption, only age at vaginal opening, day of first estrus, and regularity of the estrus cycle (at least 6 weeks post-puberty) is required. All other measures are extraneous and unnecessary. All compounds testing positive in this screening could then be advanced to a Tier-2 screening protocol. If the thyroid elements remain in the protocol, the use of a positive control for the pubertal endpoints is strongly recommended. | |
| 11.11 | DR | Table 3 on page 9 in the Section III: Purpose of the assay lists the end points for the female pubertal protocol: It is not clear why one of the highly sensitive hormone dependent organs, i.e., mammary gland is not included for the analysis of its weights and histopathology. In the various strain of rats, it has been shown that the treatment of 14-21 days with endocrine disruptors, particularly estrogenic in nature produces profound changes in the mammary gland. | Age and weight at vaginal opening is a sensitive indicator of interaction with the endocrine system. Studies on whether mammary gland endpoints are more sensitive in the pubertal assay context may be appropriate. The Agency does not feel that it would be appropriate to delay the Screening Program further to investigate this, but may consider this for potential improvement of the assay in the future. |
| 11.12 | DR | The central nervous system is one of other system should have been included for analysis of its weight and histopathology, because we now that endocrine disruptors | Changes in brain weight and histopathology are more relevant to assessing effects from perinatal exposure than from exposure during the juvenile/pubertal period. Changes in neurotransmitter levels may be |

| | | | |
|---|---|---|---|
| | | influences its development and functions. | helpful in assessing neuroendocrine effects from peripubertal exposure, but the myriad of neurotransmitters that control the secretion of GnRH from the hypothalamus, which are most relevant to the initiation of cycling and VO, would be too technically challenging to measure for a screen. |
| 11.13 | DR | Why the levels of estrogen, androgen and progesterone were not proposed to be measured is not clear. It is the ratio of androgen and estrogen or estrogen and progesterone which determines their effects on the target organs. | Androgens are not detectable in females at this age. Also, progesterone and estrogen are variable due to cycling (similar to ovarian and uterine weights). These hormones have been measured in some pubertal studies, but were judged not to add significant value. |
| 11.14 | DR | Minor weakness<br>It has been shown that the rodent pubertal female assay is useful for identifying potential endocrine disruptors having not only estrogenic/antiestrogenic but also androgenic/antiandrogenic activities, therefore it is not clear why androgenic/antiandrogenic activities were not monitored. | The female pubertal protocol has not been tested with androgens or anti-androgens. |
| 11.15 | DR | We now know the non-receptor-mediated mechanisms exist by which unknown disruptors can affect the embryo/fetus without showing positive effects on the proposed classical multiple endpoints. In situ biochemical and gene activation measurements or biomarkers for assessing pubertal development could have really helped to detect subtle changes in the endocrine systems which would be not | While the information from such recent methods looks promising, these methods were not considered validated during the time the pubertal assays were being validated across laboratories. They may be appropriate for future improvements to the assay. |

| | | | |
|---|---|---|---|
| | | detected otherwise by proposed multiple endpoints in this assay. ChIP on ChIP assay would have been more sensitive for screening effects of endocrine disruptors by studying changes in genes involved in androgen, estrogen, or thyroid systems. | |
| 11.16 | DR | The methodologies for measuring indices of puberty are not very modern, and they may not be very sensitive in detecting the initiation and progression of molecular changes that ultimately impair the pubertal development. There is a concern that all these functional assays may not be able to detect subtle changes in the animals exposed to weak endocrine disruptors. | See the response to comment 11.15. |
| | | | |
| | | | |
| 12. Details of the protocol | | | |
| a. Test compound | | | |
| 12.a.1 | BD | Some guidance on acceptable purity might be provided here. The methoxychlor (MXC) used in some of the validation studies had a purity of less than 90%. Earlier literature indicated that technical grade MXC was more potent than highly purified MXC due to presence of active metabolites in the technical grade material. For unknown compounds going through the screen, will use of technical grade material be acceptable or is highly purified compound preferred? In any case, some statement | Designation of the substance to test is not part of the protocol *per se* and was not part of validation. The testing orders issued by the Agency will identify the substance to be tested (e.g., technical grade vs. analytical grade). |

| | | | |
|---|---|---|---|
| | | might be provided. | |
| | | | |
| | | | |
| b. Dose selection | | | |
| 12.b.1 | JB | Test Method, Section XVI., para 2. It is stated that "body weight loss that does not exceed approximately 10 % is an indication that MTD was approached but not exceeded." It cannot be over-emphasized that since reduction in body weight, and in particular body fat, is an outcome of estradiol treatment, body weight loss is not useful as an indicator of MTD.<br>…<br><br>Dose of compounds. Test Method, Section V, para 1. Two dose levels are used with the maximum tolerated dose (MTD) chosen based on a reduction of no greater than 10 % of the mean body weight for controls. Since compounds with estrogenic activity are being investigated, and since estrogens have anorectic and body weight suppressive effects, a different factor than MTD needs to be used.<br>…<br><br>Maximum tolerated dose: The ISR makes a number of statements such as the following (ISR, page 68): "Laboratories 2 and 3 reported that terminal body weight was | As indicated in the response to comment 8.6 and discussed in Section VI.C.3 (pages 53-54), data show that vaginal opening is more sensitive than body weight gain to the effects of estrogens. Thus, while a test chemical that is estrogenic may cause a decrease in body weight gain (compared to controls) that is not easily distinguishable from a decrease that is due to general toxicity, lowering the dose level to the point where the decrease is significant but not exceeding 10% will still result in changes in age and weight at vaginal opening.<br><br>While the reviewer makes a valid point that a decrease in body weight gain that is ultimately attributable to estrogenicity is not related to the concept of "maximum tolerated dose", the two causes cannot be distinguished at the time of dose-setting (in the absence of other indications of toxicity). For simplicity of wording in the protocol, the term "maximum tolerated dose" will be retained. |

| | | | |
|---|---|---|---|
| | | significantly reduced in the high dose methoxychlor rats, both finding that this group weighed 94.1 percent of the controls. This indicates that the Maximum Tolerated Dose was reached, but that the body weight decrease compared t controls was not so severe as to interfere with endocrine endpoints." The use of body weight data like these to draw conclusions of maximum tolerated dose are not meaningful within the context of an experiment examining the effects of a potential estrogen, since treatment with estradiol itself can suppress body weight by 20 %, and is responsible for quite dramatic shifts over the four day estrous cycle. The animals, who have lost weight, are not ill, and their body weight loss has nothing to do with maximum tolerated dose. They are not losing weight because of toxicity; this is just a normal physiological response to estradiol, so would seem to be unrelated to the toxicological term, MTD. | |
| 12.b.2 | JB | In addition, these choices of doses are odd for endocrinological work. Since effects can differ at various points on the dose-response curve, usually a pilot dose-response curve is run. While the meaning of the term that is used in endocrinology… physiological levels… might not have meaning, can dosages be chosen based on reasonable expectation of level of exposure. The | The purpose of Tier 1 screening is to identify the potential of a chemical to interact with the endocrine system, not to characterize risk. The assay is not intended to provide enough information to derive a dose-response curve. Thus use of two dose levels appears appropriate for this assay.<br><br>There are many difficulties in estimating a level of exposure that might adequately represent human |

| | | | |
|---|---|---|---|
| | | choice of two doses leaves a great deal to be desired, since most physiological work requires a dose-response curve. | exposure. Rarely are there sufficient monitoring data; often there are none at all. Exposure models require many assumptions about the behavior of the chemical in the environment as well as about human behavior. Such models may allow a more reasonable estimation of upper-end exposure than piling worst-case assumption upon worst-case assumption, but are not adequate to identify levels of exposure which should be tested in order to ensure that areas of potential low-dose non-monotonicity in the dose-response curve that might be of concern for risk assessment have been covered. Finally, exposure conditions are subject to change (depending, for example, on uses) and a dose-setting scheme based on such a changeable basis would require constant updating. It is not clear that Congress intended the Screening Program to be of such magnitude.<br><br>In addition, the Endocrine Disruptor Screening Program is concerned with effects on wildlife as well as humans. Identifying "expected exposure levels" for wildlife is even more challenging in most cases than it is for humans.<br><br>Thus while testing at "expected exposure levels" would be ideal, it does not appear feasible at this time. |
| 12.b.3 | JB | In discussion of specificity, the ISR mentions (page 82) that "a good faith effort was made to identify a chemical that was both toxic to other systems but without endocrine effects." It is not surprising that one could | The ISR addresses the issue of decreased feed intake on the pubertal endpoints in some detail (section VI.C.3, pages 53-55). As noted there, the feed restriction study performed by the Agency provides reason to believe that decreased feed intake (i.e., a |

| | | | |
|---|---|---|---|
| | | not be found, because a toxic compound will decrease body weight, and this seems to be required to demonstrate that the dose has exceeded the MTD. However, body weight loss due to toxicity would likely be accompanied by a decrease in nutrient intake. From a physiological point of view, food deprivation causes reproductive dysfunction. Unfortunately, approaching a problem like this from a toxicological view-point with little regard to the underlying endocrinology/physiology has problems. In short, any compound that compromises nutrition would be expected to have endocrine effects. | decrease in nutrient intake) resulting in a decrease in body weight gain of less than approximately 10% (the indicator of approaching but not exceeding the MTD) will not affect the endpoints measured in the female pubertal assay. |
| 12.b.4 | BD | One issue that was not directly addressed in the Integrated Summary Report or the protocol itself was the question of whether confining testing to the Maximum Tolerated Dose and one half of the Maximum Tolerated Dose could miss important endocrine activity that would be evident at low doses. A footnote in the protocol to restate the EPA position on this issue should be considered. | The Agency is unaware of methods to identify where non-monotonicity in the dose-response curve should be expected and thus tested. It is not clear that testing at a large number of doses spaced (for example) at order-of-magnitude intervals would be sufficient to identify low-dose effects as such effects could occur entirely between the intervals tested. Also, as explained in the response to comment 12.b.2, it is more often than not difficult to estimate a representative environmental exposure level. All of these considerations conflict with the goal of making screening assays reasonably quick and inexpensive.<br><br>The Agency considered including a statement of position on this issue in the protocol but believes that such a policy statement would not be germane to the |

| | | | |
|---|---|---|---|
| | | | scientific protocol and thus declined to add such a statement. |
| 12.b.5 | HP | Only two doses will be used in the assay. Both are based on body weight with the second being half of the first. By basing the doses used for the assay on the maximum tolerated dose (MTD), both are likely to be quite high compared to what humans and wildlife could reasonable expect to be exposed to. Some of the most concerning findings with endocrine disrupting compounds have occurred at doses that are well within the realm of human exposure and far lower than would be chosen by the parameters of this assay (Alworth et al., 2002; Gioiosa et al., 2007; Goodman et al., 2006; Kato et al., 2003; Rubin et al., 2006; Rubin et al., 2001; vom Saal, 2006; vom Saal and Hughes, 2005). One of the recommendations in the Final Report of the Endocrine Disruptors Low-Dose Peer Review (2001, NEIHS) was to replicate and validate "low dose" studies. Although the argument for "low dose effects" is controversial, employment of a low dose in the assay is well justified and would address this issue. The data within the Integrated Summary Report illustrate the critical need for the employment of a lower dose. Atrazine, Bisphenol-A and methoxychlor all produced significant effects at substantially | See the response to comment 12.b.2. The inability to specify a "low dose level" that, if tested, would alleviate concerns for effects at "low dose levels" is a problem which requires substantial further research and should not delay the screening of chemicals. |

| | | | |
|---|---|---|---|
| | | lower doses than what would likely be used under the current protocol guidelines. Inclusion of a “low dose” would also increase the robustness of the assay. For example, within the Integrated Summary Report the observed effects of methoxychlor at 25 and 50 mg/kg/day are argued to confirm “the transferability of the assay and provide evidence that the assay is sensitive.” (Page 27, line 7-8). The use of a “low dose” as defined by the Endocrine Disruptors Low-Dose Peer Review (2001, NEIHS) or a similar protocol would significantly strengthen the assay and help to clarify whether or not “low dose” effects are of genuine concern. If left out, this controversy will continue to linger and doubt about the safety of the test compounds will remain even after they are subjected to testing with this assay.<br><br>It should be noted that most of the data regarding low dose effects have come from animals exposed during the gestational or neonatal period. Therefore it is unclear if low dose effects would be observed when the exposure begins just prior to puberty, as proposed in the current protocol. However, there is sufficient data to warrant the inclusion of a low dose. | |
| 12.b.6 | BD | Page 4, second paragraph: Guidance is | The setting of the low dose for 2-CNB is |

| | | | |
|---|---|---|---|
| | | given in this paragraph for setting the low dose to be used in the assay. For the interlaboratory validation study with 2-CDNB, it was not explained why the low dose selected did not follow this guidance. | acknowledged to be an inconsistency that should not have occurred. |
| 12.b.7 | DF | Designation of the Maximum Tolerable Dose level is unclear to me (p. 4 line 2), since the protocol is presumably to be applied generally to compounds with both known and unknown general toxicity profiles. The way this is worded implies that preliminary studies are done to determine the MTD prior to a full scale assay for endocrine system effect, which may not actually be the case. | In most protocols, gathering of information on which to base dose-levels is not considered part of the protocol itself. For conventional *in vivo* mammalian studies using adult animals, there is often a preliminary database allowing an educated guess of dose levels that might be appropriate to examine. The lack of such information for pubertal animals should not be considered a deficiency of the pubertal assay. |
| 12.b.8 | HP | Finally, the use of high doses may explain why, to date, no compound has produced a negative result in this assay. The highest dose to be used is defined as a statistically significant reduction in body weight with "no clinical signs of toxicity." The acceptable "signs of toxicity" are not identified or discussed in the protocol but should be, perhaps in an appendix. In general, the use of body weight to define dose is problematic for several reasons, most of which have already been addressed previously by Goldman et al, but again highlights the need for a positive control group within this assay. It is well established that estradiol administration significantly reduces body weight. Because a decrease in body weight | The Agency believes that it would be inappropriate to specify every potential sign of toxicity that could be used as a basis to claim that MTD had been exceeded since it would be difficult to produce an exhaustive and exclusive list.<br><br>Body weight as an indicator of approaching but not exceeding MTD: See the response to comment 12.b.2.<br><br>Positive controls: See the response to comments 7.1 and 7.2. |

| | | | |
|---|---|---|---|
| | | of 10% or more can result in the disinclusion of subjects or treatment groups, the employment of a positive control group would help clarify whether or not the MTD was reached or exceeded, and whether or not the laboratory was conducing the assay properly. | |
| | | | |
| | | | |
| c. Vehicle/solvent | | | |
| 12.c.1 | JB | Injection vehicle. Test Method, Section VI, para 2. Corn oil is the preferred vehicle. However, it is not stated that this must be of pharmaceutical grade. Relying on the experimenter to make judgments of clarity, sedimentation and odor, without specifying a common source or grade will likely lead to differences among laboratories. Likewise, leaving up to the lab the choice of corn oil, water or carboxymethylcellulose is a mistake. Although this reviewer is not a pharmacokinetics expert, the solvent would be likely to influence the rate of uptake into the circulatory system. In addition, more information needs to be given regarding the method of making up of the solutions. Should the solutions be warmed or not? Should they be subjected to sonication? Etc | Inclusion of precise and inflexible rules for choosing a vehicle/solvent is inappropriate since there are so many factors to consider. The Agency will change the wording in the protocol to provide guidance on some of the factors to consider, but the study director is more likely than the Agency to be in a position to make a final appropriate choice. The Agency retains the right to reject a study if highly inappropriate choices have been made.<br><br>Guidance on selection of vehicle/solvent will be changed to the following:<br><br>“The test substance is dissolved or suspended in a suitable vehicle. Consideration should be given to the following characteristics: Effects on the absorption, distribution, metabolism, or retention of the test substance; effects on the chemical properties of the test substance which may alter its toxic characteristics; and effects on the food or water consumption or the nutritional status of the animals. |

| | | | |
|---|---|---|---|
| | | | Use of vehicles with potential intrinsic toxicity should be avoided (e.g., acetone, DMSO). If corn oil is used, it must be clear and free of sediment. It should have a bland odor, free from rancid, musty, metallic, putrid or any other undesirable odor. Other solvents such as water or carboxymethylcellulose may be used where appropriate. Gentle warming may be used to assist solubilization but the solution must not be administered warm and the solution should be checked to make sure that precipitation did not occur upon cooling. Use of intermediate solvents (e.g., ethanol) to assist in solubilization is not appropriate. If the test substance is not soluble in any of the conventional solvents, it is administered as a suspension. Sonication may be used to assist in suspending particles. It is important that the dosing solution or suspension be well-mixed to keep the chemical well-distributed prior to and throughout dosing, and care must be taken to ensure that the particle size of insoluble substances does not interfere with delivery of the full dose through the gavage tube or needle tip." |
| 12.c.2 | JB | *[Regarding the transferability study]* Page 15: It is stated that the ethynyl estradiol was dissolved in ethanol prior to dilution in corn oil. It is essential that the amount of ethanol be stated. In fact, since it is not in the protocol, it should not have been part of the procedure, because it may have resulted in the presence of ethanol in one group, and not the others. | The protocol will be clarified to disallow use of intermediate solvents such as the ethanol used in this case. See the response to comment 12.c.1. |

| | | | |
|---|---|---|---|
| | | | |
| | | | |
| d. Strain of rat | | | |
| 12.d.1 | BD | The controversial issues of sensitivity differences of rat strains and the potential impact of diet were discussed and EPA has provided reasoned explanations for their decision to recommend the Sprague-Dawley rat….. These decisions will no doubt be reviewed as additional data become available. | Agree. No change in protocol. |
| 12.d.2 | DF | Rat strain choice for these assays is still very much open for debate. While it is apparent that the EPA is aware of potential strain influences in sensitivity to endocrine disrupting chemicals, the assay only proposes to use Sprague Dawley rat (Crl:CD(SD)). This choice is based largely on comparing only Wistar and Sprague-Dawley rats, and experience with the male prepubertal assay. However, there are several publications demonstrating increased sensitivity of Fisher F344 rats to estrogens and BPA relative to at least SD rats in multiple endpoints as pointed out in Appendix 12, several publications in the open literature, and as noted by the ISR. This point should not be dismissed in considering the utility of the assay based only on practical considerations. | The Agency agrees that an optimal rat strain for the endpoints in this assay may not have been identified yet. However, additional research to identify the best strain for use with the endpoints specified in this assay was judged likely to take a significant amount of time and would have delayed the initiation of testing under the Screening Program. The strain marked as "preferred" in the protocol has been shown to be sensitive across many chemicals representing several modes of action. In the Agency's judgment, proceeding with the current strain was more appropriate than delaying testing for the additional research it would take to find an optimal strain for the endpoints included in this assay. |
| 12.d.3 | DR | The animal strains used are not the most | Although criticisms have been raised that the |

| | | | |
|---|---|---|---|
| | | sensitive to estrogenic compounds, which makes it further difficult in judging the suitability of analytical methods. | preferred strain is not the most sensitive to estrogens in other assays, data have not been provided showing which strains are better for the endpoints and modes of action covered by the female pubertal assay. The research necessary to determine a more appropriate strain would have considerably delayed testing under the Screening Program. |
| | | | |
| | | | |
| e. Source of animals | | | |
| 12.e.1 | JB | Supply of animals for the experiment and confound of stress exposure: First and foremost, this reviewer has a major concern in the way that the animals are received. In Section IV, paragraph 2 of the protocol, it is stated that rats are "bred in-house or purchased from a supplier as "timed pregnant" dams with arrival at the laboratory on gestation day (GD) 7, 8, 9 or 10". The use of timed pregnant animals in a reproductive study, or for that matter in most studies, is contraindicated, because shipping is a stressor, and gestation is a time of vulnerability to stress for both the fetus and the mother. Therefore, this introduces a major confound in the protocol. Depending on how and when rats are supplied (shipped "timed-pregnant" vs. bred in lab, some animals will not be exposed to a stressor, some to a major stressor. In addition, the developmental age prenatally | Prolonged stress during the latter part of gestation does affect pubertal maturation, and for this reason the Agency specified that the dams must arrive at the laboratory no later than gestation day 10. The data obtained in the interlaboratory validation and other studies confirm that the assay performs appropriately when the stress of transport is limited to the initial period of gestation. |

| | | | |
|---|---|---|---|
| | | that the rats are exposed to the stressor will vary depending upon day of pregnancy that the rats are shipped. It is likely, but should be determined if this is a confounding factor or not, that prenatal stress influences the fetus's physiology. If there are influences, which is this reviewer's expectation, then use of "timed pregnant" females should be considered unacceptable. In-house breeding is complicated and may require additional facilities that some contract laboratories have. Therefore, it may decrease the number of laboratories equipped to do the experiments. However, use of timed pregnant animals is a serious flaw in the design. If the experiments were submitted to an endocrine journal of which I am editor, they would not be accepted. | |
| 12.e.2 | JB | A secondary problem with use of animals derived from mothers shipped while pregnant is the possibility that the stress of shipping compromises maternal care of the F1 generation. Since quality of maternal care is a prerequisite to normal development, and suboptimal maternal care can have epigenetic effects on the offsprings' subsequent response to hormones (Weaver et al., 2004), and perhaps xenoestrogens, use of timed-pregnant rats presents a major problem. There are so many interactions between the | Effects of compromised maternal care are minimized by specifying in the protocol that the pups needed for the study be taken from the middle of the weight distribution across all litters when the pups are distributed to treatment groups, and also by specifying that litters with fewer than 8 pups are to be excluded from the study. The alternative of requiring that only labs with in-house colonies may perform this assay is, in the Agency's judgment, unnecessarily restrictive in light of the data showing that the assay performs acceptably when transport of dams is limited to early gestation. |

| | | | |
|---|---|---|---|
| | | hypothalamo-pituitary-adrenal axis and the hypothalamo-pituitary-gonadal axis that the influence of stressors on the dependent variables has to be considered in the design of the protocol. | |
| 12.e.3 | DF | Within an experiment, juvenile rats should be obtained from either pregnant females from an in-house breeding or purchased from a supplier but not mixed in a study or when comparing between studies (p. 2 line 19). Inadvertent prior exposure to chemicals during gestation and the perinatal period, or maternal stressors, can influence later responses to these same or different chemicals to the offspring (e.g. Newbold et al. Reproductive Toxicology 23:290–296 (2007)). Therefore, the source of animals used in the studies ought to be standardized. | The protocol will be changed to require that dams obtained and transported from an external supplier may not be used in the same study as dams bred in-house. |
| 12.e.4 | HP | Animals can either be shipped timed pregnant or born in-house. This introduces gestational stress as a source of variability. Transport stress during pregnancy can delay parturition by 24-48 hours, reduce maternal care and expose the developing pups to stress hormones during gestation. All of this can affect the growth and overall well-being of the offspring. Maternal stress should be minimized as much as possible. As such, all females should be impregnated in-house with care to avoid any unnecessary | See the responses to comments 12.e.1, 2, and 3. |

| | | | |
|---|---|---|---|
| | | disturbance. Cage changes and all other intrusions should be minimized as much as possible. | |
| 12.e.5 | HP | Transport stress is also a concern. Pups born to dams delivered "timed pregnant" will experience more gestational stress than those born to mothers impregnated in-house. This may also affect inter-laboratory data collection and consistency. | See the response to comment 12.e.1. |
| | | | |
| | | | |
| f. Defining day of birth | | | |
| 12.f.1 | HP | The day of birth is defined as the "morning" of PND 0. This definition is too broad. Under this definition, an animal born at 12:10 pm on a Monday could be listed as PND 0 on Tuesday and cross fostered with animals born at 11:45 am on Tuesday. This means litters could contain animals born 18-23 hours apart. This is a problem in a protocol where many of the data interpretations are to be based on body weight. A specific time frame for the definition of PND 0, and some discussion is needed as to how animals born close to that window will be dealt with. | The protocol says that "PND 0 is defined as the day on which the pup is first seen, assuming that the cages are checked for new births daily, in the morning." While the situation described by the reviewer is possible, the Agency is relying on the fact that rats give birth at night rather than at all hours. It would be unreasonable to require that the laboratory record the hour of birth of each pup, particularly when parturition is most likely to occur in the middle of the night. |
| | | | |
| | | | |
| g. Distribution of animals to groups | | | |
| 12.g.1 | BD | Page 3, last sentence of the second | The protocol will be changed to specify that enough |

| | | | |
|---|---|---|---|
| | | paragraph: It is indicated that placement of litter mates in the same group should be avoided, but the last sentence in the fifth paragraph suggests that there may be situations where this could occur. It seems that either a stronger directive should be given (i.e., Do not place litter mates in the same group) or guidance on how litter mates in the same group should be reported and handled in the statistical analysis provided. | dams should be ordered to avoid the need for placing littermates in the same group. The Agency expects that including only a few extra dams will suffice. In the Agency's experience with this assay, it is rare that littermates need to be placed in the same group. A strict prohibition against this situation might inappropriately cause an entire study to be dismissed. Modification of statistical procedures to cover the rare case does not appear warranted. |
| 12.g.2 | JB | Littermate effects. Test Method, Section IV, para 4: It is stated "Avoid placing littermates in the same group." Since litter-effects can be so robust, this statement should be considerably stronger than it is. If the experiment is worth doing, then the protocol should be very clear that placing littermates in the same group is unacceptable. | See response to comment 12.g.1. |
| | | | |
| | | | |
| h. Route of administration | | | |
| 12.h.1 | BD | One thing that is unclear is why the oral route was selected as the only possible route of exposure for this assay. While there is likely to be limited information on the majority of chemicals that will be tested in the Endocrine Disruptor Screening Program (EDSP), in cases where either planned use or pharmacokinetic data are available, it would seem that, at least in some cases, | The Agency agrees that in some cases, pharmacokinetic data or uses of the chemical may suggest that routes of exposure other than oral may be appropriate to test. However, the Agency does not have a database against which to validate this assay using other routes of exposure. Validating the assays for additional routes of exposure may be an appropriate activity for the future but should not delay initiation of the Screening Program. |

| | | | |
|---|---|---|---|
| | | other routes might be preferred. If the goal of the assay is primarily to detect endocrine system activity that would then be further investigated and defined in higher tier studies, it would seem that a route that results in higher systemic exposure to the test chemical might be selected even if that route is not the major route of exposure to be expected in humans. At any rate, a statement of why the oral route is specified (e.g., because it is likely to be the primary route of human exposure, etc.) would be appropriate. | |
| | | | |
| | | | |
| i. Husbandry other than diet | | | |
| 12.i.1 | JB | Animal housing. Animals are housed in clear plastic cages. Problems with leaching of bisphenol A from some plastic cages have been documented (Howdeshell et al., 2003). According to these authors, polycarbonate and polysulfone cages leach bisphenol A, but polypropylene cages do not. Since this presents another potential confound, the use of particular plastics and methods for cleaning them should be given a great deal of thought, and the protocol should be very proscriptive in what is acceptable. | Howdeshell et al. (2003) tested leaching of bisphenol A from cages by allowing 250 ml of water (2.5 cm depth in a standard 29 x 19 x 13 cm cage) to stand undisturbed for one week. This does not appear to be a scenario of sufficient concern to warrant restriction of cage material for this screening assay. The authors also noted that the BPA to which mice might have been exposed from the polycarbonate cages and water bottles used in their study did not cause a positive response in the uterotrophic assay, further suggesting that this variable is not likely to be of significant concern in the pubertal assay. |
| 12.i.2 | DF | Specific plastic used for caging (p.2 line 2) should be described in more detail. I found | See the response to comment 12.i.1. |

| | | | |
|---|---|---|---|
| | | at least one article describing the leaching of potential endocrine disrupting chemicals from polycarbonate tubs used in rat or mouse housing, and this can be drastically increased after exposure to alkaline washing conditions and/or high temperature (Koehler et al. Lab Animal 32: 24-27, 2003; Everitt and Foster ILAR 45: 417-424 2004). | |
| 12.i.3 | DF | It is also interesting to note that even differences in caging can lead to significant differences in age of vaginal opening and estrous cycle parameters (Firlit and Schwartz, Biol. Reprod. 16:441-444, 1977) indicating the effect of non-specific stressors on these parameters although the exact mechanism ins not understood, to my knowledge. | The Agency has specified that clear plastic cages be used and has specified approximate dimensions of the cages, so these potential variables are controlled. See the response to comment 12.i.1 concerning further specification of the type of plastic. |
| 12.i.4 | JB | Test Method, Section IV, para 6: The second statement is very ambiguous, and should be more directive. In addition, the statement about littermates states that "littermates should not be in the same group." Since animals were assigned to groups in paragraph 4, this seems misplaced and therefore a possible source of confusion. | The protocol will be changed to read as follows (addition in italics): "…the preferred procedure is to kill all the animals on a single day to close the in-life portion of the study, but *if the number of planned necropsies is considered too large to allow careful measurement of endpoints on one day (e.g.,when multiple chemicals are being tested simultaneously)* kills may be conducted over two days rather than one, with half of each group killed on each day. *Day of kill is assigned to individuals at the time of distribution of the pups to groups; kills over two days must not be adopted as a matter of convenience on the day of necropsy.*" |

| | | | |
|---|---|---|---|
| | | | For distribution of littermates, see response to comment 12.g.1. |
| 12.i.5 | BD | Is the statement that the AAALAC guideline extremes are “only marginally tolerable for the pubertal assay supported by published data that could be cited? | The footnote stating that the AAALAC guideline extremes are “only marginally tolerable” will be deleted. |
| 12.i.6 | BD | With regard to guidance on the culling of litters, it might be useful to directly state that litters smaller than 8 can be used without adjustment. | On the contrary, the protocol already states that “[a]ny litters with fewer than 8 total pups/litter (i.e., including both males and females) … are excluded from the study.” |
| 12.i.7 | DF | Specific water bottles to be used for ad lib drinking are not described (see previous concerns re: housing and washing and care of plastics) (p.2 line 8). | The protocol will be changed to prohibit use of polycarbonate water supply equipment. |
| 12.i.8 | HP | Great care is taken to consider variation in body weight when interpreting the data. In contrast, relatively little information is given to as to how stress was minimized or how housing conditions differed between laboratories. This is a considerable problem. Housing of other animals in the facility, particularly non-human primates, dogs, or cats can increase rodent stress substantially. | The protocol will be changed to specify that the study must be conducted in AAALAC-approved facilities, and that care should be taken to minimize stress from all sources including noise, other species housed nearby, or other disturbances. |
| 12.i.9 | HP | Housing conditions, including diet and day of cage changes, should be matched as closely as possible between laboratories. | The Agency believes it has adequately controlled the most important variables in the assay while still allowing a certain amount of flexibility to accommodate individual labs’ needs. While absolute uniformity is ideal, the Agency is under no illusion that absolute uniformity is achievable when the assay is run in a |

| | | | |
|---|---|---|---|
| | | | large number of laboratories. The interlaboratory validation study showed that the same overall result is obtained on a chemical at the current level of specificity of assay conditions. |
| | | | |
| | | | |
| j. Diet | | | |
| 12.j.1 | JB | Diet: The ISR (page 78-79) makes the argument that, although phytoestrogens in the diet may have effects on vaginal opening, this is unlikely to be a concern, because control groups will be exposed to the phytoestrogens as well. This demonstrates a lack of understanding of physiology. Just for the sake of argument, take the case of the presence of a hypothetical antiestrogen with no estrogenic effects in the feed. The antiestrogen might be expected to block the effects of an estrogenic test compound, but be without effect in the control group. This would then lead to a false negative. One can come up with all sorts of scenarios in which a compound in the feed would influence the experimentals, but not the controls. For example, a drug that has a permissive effect on action of an estrogen. It would have no effect in controls, but a very dramatic effect in the experimentals. I must respectfully disagree with the conclusion that it is "prudent to set a limit on the concentration | The Agency would prefer to have data showing the need for a quite reduced or phytoestrogen-free diet, rather than having to invoke a "hypothetical antiestrogen" in the feed. As indicated in the ISR, the Agency agrees that there is still considerable uncertainty over the effects of phytoestrogens in feed on the endpoints in the female pubertal assay. Research in this complex area*, which perhaps should be coupled with research on rat strain responses, should not delay implementation of the Screening Program.<br><br>*(see, for example, Latendresse JR et al. 2001. Polycystic kidney disease induced in F1 Sprague-Dawley rats fed *para*-nonylphenol in a soy-free, casein-containing diet. Toxicol Sci 62:140-147.) |

| | | | |
|---|---|---|---|
| | | of phytoestrogens in feed used in the pubertal assay." The cost for requiring that phytoestrogen-free or at least quite reduced would seem to be minimal contrasted with the cost of each of these studies. | |
| 12.j.2 | JB | Although the protocol states limits of genistein allowed in the diet, Teklad 7012C was used, which is not routinely analyzed for genistein, and the experimenters did not have this analyzed. Furthermore, because phytoestrogens in the diet are a potential source of confound, the protocol could be much more directive. I see no reason that a specific diet could not be required. | This comment refers to the TherImmune 1 (transferability) study, which was completed before the Agency recognized the appropriateness of measuring and capping the level of phytoestrogens in feed. |
| 12.j.3 | BD | The detection of estrogenic activity of methoxychlor at 12.5 mg/kg in the multi-dose study (Appendix 8) is cited as an example of the sensitivity of the assay (ISR page 47, lines 14-17), but the three laboratories in the interlaboratory comparison study did not detect activity at this dose. It is of interest to note that the study that did detect activity at 12.5 mg/kg appeared to use the diet with the highest phytoestrogen level, although it is certainly not clear what factors might have contributed to the discrepant result. | The reviewer raises a good point, that the sensitivity of the female pubertal assay of methoxychlor was not consistent between these two studies. This may raise concerns that very weak compounds (weaker than methoxychlor, which was reliably detected at 50 mg/kg) will not be detected consistently, even when tested at high dose levels. The Agency believes that future improvements to the assay may increase the consistency of detection of weak compounds, but that the Screening Program should not be delayed further to re-validate an improved protocol at this time. |
| 12.j.4 | BD | The controversial issues of sensitivity differences of rat strains and the potential impact of diet were discussed and EPA has provided reasoned explanations for their | See response to comment 12.j.1. |

| | | | |
|---|---|---|---|
| | | decision to… set an approximate limit of 300 ppm genistein-equivalents of phytoestrogens. These decisions will no doubt be reviewed as additional data become available. | |
| 12.j.5 | BD | What is the basis for the statement that totally synthetic diets are not appropriate? Is this because data indicate that they are not suitable, or because there are insufficient data to support their use? Most studies (for example, reference 58 in the ISR) appear to use the older AIN-76 formulations, and there appear to be less data for the revised AIN-93 diet formulation. AIN-93 has been used in reproductive studies (*e.g.* Collins *et al*., Effects of flaxseed and defatted flaxseed meal on reproduction and development in rats, Food and Chemical Toxicology **41**: 819-834, 2003). In that published study, data presented for sexual development for the female controls on the AIN-93 diet seems to fall in line with performance criteria given for the female pubertal protocol. | The footnote that totally synthetic diets are not appropriate will be removed from the protocol. |
| 12.j.6 | BD | The recommendation of the limit on the level of phytoestrogens is based on data from the uterotrophic assay. Recent data reported by Thigpen *et al.* http://www.ehponline.org/docs/2007/10165/abstract.html indicates that the CD Sprague-Dawley rat may be insensitive to | The Agency agrees that diet and strain may need to be studied together as potential improvements to the protocol are investigated. |

| | | | |
|---|---|---|---|
| | | phytoestrogens relative to the F344 rat. Response to an estrogen challenge was not reported in that study, but it would appear that there would be concern with the level of phytoestrogens specified in the current protocol if strains other than the Sprague-Dawley are used in the future. | |
| 12.j.7 | BD | The diet that the timed pregnant females were fed by the supplier prior to shipment to the test laboratory should be reported. | The protocol will be changed to add this requirement to the protocol. |
| 12.j.8 | HP | The diet should be free of phytoestrogens. As written, diets containing up to 300 µg/kg are allowed. The presence of phytoestrogens, even in such small quantities, needlessly impairs the sensitivity of the assay and introduces inter-laboratory variability. A number of phytoestrogen-free diets are now readily available from all of the major lab diet manufacturers so there is no reason not to exclude this potentially problematic source of endocrine disrupting compounds. | The protocol allows up to 300 ug/g, not ug/kg.<br><br>The Agency is not aware of data showing that the presence of phytoestrogens impairs the sensitivity of the assay and introduces inter-laboratory variability when present at levels below the cap set in the protocol. See also the latter part of comment 12.j.3. |
| | | | |
| | | | |
| k. Water | | | |
| 12.k.1 | JB | It is stated that tap water in not acceptable, and deionized water is preferred. Since all labs have access to deionized water, why not simply require it to standardize the methods as much as possible? As with food, there is evidence that some water | The protocol will be changed to require use of deionized water. |

| | | | |
|---|---|---|---|
| | | supplies can contain compounds with estrogenic properties. It would be most prudent to state requirements. | |
| | | | |
| | | | |
| l. Vaginal smears | | | |
| 12.l.1 | JB | There are many issues in doing vaginal smears that must be considered and are not covered in the protocol (Becker et al., 2005). First and foremost, there is no discussion of how this is to be done. The technique of vaginal lavage is a bit of a craft. If the cervix is stimulated, the animal enters pseudopregnancy (aka the progestational state), an anestrous period of twice daily surges of prolactin, rescue of the corpus luteum, and elevated progesterone levels. The females do not cycle, and the vaginal smear would look like diestrous stage. Doing these by an inexperienced technician without knowledge of the problems will result in pseudopregnancy, which would confound the results of effects of xenoestrogens.<br>In addition, reading the slides also that takes some practice. These are not all-or-none of one cell type or another. Typically, sample photomicrographs are included in protocols to facilitate the task of the technician and to make the assessment of cell type more repeatable (Becker et al., 2005). There are | The protocol will be changed to add the following reference on how to perform vaginal lavage:<br><br>Cooper RL, Goldman JM, Vandenbergh JG. Monitoring of estrus cyclicity in the laboratory rodent by vaginal lavage. In: Methods in Toxicology. Vol III, Part B. Female Reproductive Toxicology. Edited by Chapin RE and Heindel J. Academic Press: Orlando. 1993. pp. 45-56. |

| | | | |
|---|---|---|---|
| | | also times during the light: dark cycle that result in greatest consistency, since for example, the proestrous stage of the cycle lasts only 12-14 hours (Becker et al., 2005). | |
| | | | |
| | | | |
| m. Method of kill | | | |
| 12.m.1 | JB | Decapitation: Test method, Section X. para 2. It is stated that the preferred method of kill is by decapitation without any form of anesthesia. More discussion is needed, since this statement conflicts with the statement in the Guide for the Care and Use of Laboratory Animals:<br>"Euthanasia is the act of killing animals by methods that induce rapid unconsciousness and death without pain or distress. Unless a deviation is justified for scientific or medical reasons, methods should be consistent with the 1993 Report of the AVMA Panel on Euthanasia (AVMA 1993 or later editions). In evaluating the appropriateness of methods, some of the criteria that should be considered are ability to induce loss of consciousness and death with no or only momentary pain, distress, or anxiety; reliability; nonreversibility; time required to induce unconsciousness; species and age limitations; compatibility with research objectives; and safety of and emotional effect on personnel. | The protocol will be changed to state that the preferred method of kill in the **female** pubertal assay is by injectable anesthetic followed immediately by decapitation in order to obtain a sufficient volume of blood for the $T_4$ and TSH measurements (approximately 1 ml). If necessary to obtain a sufficient volume of blood, the decapitation may be performed after anesthesia has been achieved but before death. Carbon dioxide is not an anesthetic. A less preferred but still acceptable method of kill is by decapitation. The 2007 AVMA Guidelines on Euthanasia ( http://www.avma.org/issues/animal_welfare/euthanasia.pdf ), which replace the 2000 Guidelines, state "[Decapitation] is conditionally acceptable if performed correctly, and it should be used in research settings when its use is required by the experimental design and approved by the Institutional Animal Care and Use Committee." The need for a large volume of blood requires decapitation at some point. |

| | | | |
|---|---|---|---|
| | | Euthanasia might be necessary at the end of a protocol or as a means to relieve pain or distress that cannot be alleviated by analgesics, sedatives, or other treatments. Protocols should include criteria for initiating euthanasia, such as degree of a physical or behavioral deficit or tumor size, that will enable a prompt decision to be made by the veterinarian and the investigator to ensure that the end point is humane and the objective of the protocol is achieved. Euthanasia should be carried out in a manner that avoids animal distress. In some cases, vocalization and release of pheromones occur during induction of unconsciousness. For that reason, other animals should not be present when euthanasia is performed.<br>The selection of specific agents and methods for euthanasia will depend on the species involved and the objectives of the protocol. Generally, inhalant or noninhalant chemical agents (such as barbiturates, nonexplosive inhalant anesthetics, and CO2) are preferable to physical methods (such as cervical dislocation, decapitation, and use of a penetrating captive bolt). However, scientific considerations might preclude the use of chemical agents for some protocols. All methods of euthanasia should be reviewed and approved by the | |

| | | | |
|---|---|---|---|
| | | IACUC.<br>It is essential that euthanasia be performed by personnel who are skilled in methods for the species in question and that it be performed in a professional and compassionate manner. Death should be confirmed by personnel who can recognize cessation of vital signs in the species being euthanatized. Euthanatizing animals is psychologically difficult for some animal-care, veterinary, and research personnel, particularly if they are involved in performing euthanasia repetitively or if they have become emotionally attached to the animals being euthanatized. When delegating euthanasia responsibilities, supervisors should be aware of this as a p0tential problem for some employees or students."<br>Therefore, while decapitation would be acceptable in this case, more discussion is needed before referring to it as "The preferred method…" In addition, as discussed in the "Guide for the Care and Use…" this is not something left to untrained personnel without regard to the animals' and the technician's welfare. | |
| 12.m.2 | BD | Page 6, section X, Necropsy: The statements here in footnote 8 and on page 7 regarding carbon dioxide asphyxiation as inhumane seem to be at odds with current AVMA Panel recommendations | See the response to comment 12.m.1. References to carbon dioxide asphyxiation will be removed from the protocol. |

| | | | |
|---|---|---|---|
| | | http://www.avma.org/issues/animal_welfare/euthanasia.pdf . Those recommendations indicate that carbon dioxide is an acceptable method of euthanasia while decapitation is conditionally acceptable with scientific justification. Decapitation can be scientifically justified here, and the comments on carbon dioxide asphyxiation seem unnecessary | |
| 12.m.3 | DF | One significant source of stress that can be avoided is to sacrifice the animals at an area of the laboratory away from the rest of the animals to be sacrificed that day. (p. 6 line 23) | The protocol will be changed to require that the animals be moved to a holding room separate from the room in which the kills and/or necropsies are performed, on the day before the kills are to be performed. The holding room should be undisturbed except for the removal of the next individual to be killed. Only the animal which is to be killed next should be transferred from the holding room to the room in which the kill is performed, and the time for transfer and kill should be as brief as possible. |
| 12.m.4 | HP | The use of decapitation without anesthesia is inappropriate and unnecessary. The Office of Laboratory and Animal Welfare (OLAW) within the National Institutes of Health (NIH) does not generally support the use of decapitation without anesthesia unless there are extenuating circumstances that require it. Use of CO2 asphyxiation will not alter any of the outcomes in the protocol and should therefore be used. | See the response to comment 12.m.1. |
| 12.m.5 | HP | Inconsistency within the data may also result from variation in animal stress. | See the response to comment 12.m.3. |

| | | | |
|---|---|---|---|
| | | Without sufficient controls to minimize stress it is likely that inter-laboratory variability will continue to be a problem particularly when large batches of animals have to be sacrificed on a single day and there is thus a lot of human activity in the vivarium. | |
| | | | |
| | | | |
| n. Necropsy, dissection, histopathology | | | |
| 12.n.1 | JB | Necropsy. Test Method, Section X. para 1. It is stated that "On the day of kills, moving the cages or otherwise stressing the animals unnecessarily should be avoided so that variations in stress-related hormone levels are minimized." Although stress-related hormones are not being assayed in this study, this statement should still be more directive. There is no need to move or clean cages on the day of euthanasia, and it should be prohibited. | The protocol will be changed to remove the reference to stress-related hormones since the female pubertal assay does not measure any such hormones. See also the response to comment 12.m.3. |
| 12.n.2 | JB | Dissection: Test Method, Section X, para 5. "The uterus is then place on a paper towel." Usually filter paper is used, since it does not stick to the wet uterus. | The protocol will be changed to require filter paper rather than paper towel. |
| 12.n.3 | JB | Test Method, Section X, para 5. "Measures to prevent drying out may be necessary if such organs cannot be weighed immediately." Protocol should state how drying out will be prevented, and should be very directive that any drying is unacceptable. In any laboratory that I know | The protocol will be changed to suggest that if organs cannot be weighed immediately after removal, they may be placed in a weigh-boat and a moist paper towel used to cover the weigh-boat but that the paper towel must not come into contact with the organs at any time. |

| | | | |
|---|---|---|---|
| | | of, these would be weighed immediately at the time of dissection. The fact that they were allowed to dry out by at least one of the contract laboratories indicates to me that these laboratories do not always fully understand standard laboratory procedure. If they are to be used, they must be told details in very specific detail. | The Agency notes for the record that the problem appeared in a study in which several chemicals were tested simultaneously, so the necropsies at times outpaced the weighings due to the large number of organs to be processed. The problem is not expected to appear in a single-chemical study. |
| 12.n.4 | BD | In the list of endpoints under organ weights on page 1, consider indicating that the thyroid is weighed after fixation. While this is a standard practice for protocols evaluating the thyroid, highlighting this in the list of endpoints would be helpful. | The protocol will be changed to indicate in the list of endpoints that the thyroid is weighed after fixation. |
| 12.n.5 | BD | Also, consider indicating in Section II of the protocol that kidney histology is optional and add clinical (serum) chemistry, blood urea nitrogen and creatinine as optional endpoints.<br>…<br>Page 8, Section XII Histology: Add "(optional)" after "kidney | The Agency has decided to require the kidney histology and blood chemistry endpoints, and the protocol will be revised to make this clear and consistent throughout the protocol. |
| 12.n.6 | BD | Page 8, third paragraph in Section XII: There is mention of reporting changes in numbers of primary and atretic follicles. Is the ovarian evaluation intended to be a qualitative evaluation of a single section, or are step sections necessary? | See the response to comment 8.18. The method involves taking five random sections from each ovary. |
| 12.n.7 | HP | It is unclear how the blood is to be collected and prepared. The assay states that trunk blood should be collected by "inversion over | The protocol will be changed to clarify that either plasma or serum may be used for the hormone determinations depending on the measurement kit |

| | | | |
|---|---|---|---|
| | | a funnel" but the type of collection tube is not given. The choice of assay to measure plasma hormone levels will influence the type of collection tube to be used. In some cases, a siliconized or EDTA lined tube is required. This should be specified in the protocol. | selected. Thus, the blood may be collected either in serum separation tubes if serum will be used, or in tubes containing EDTA if plasma-based methods have been chosen. |
| | | | |
| o. Adjustment for weight at weaning | | | |
| 12.o.1 | JB | Adjustment for weight at weaning. The basis for adjusting organ weights at termination of the study for covariance with body weight at weaning is something that this reviewer does not understand. Since the groups are matched for body weight at weaning, this should relatively be a constant. Furthermore, to make this adjustment, one would have to know that weight at weaning is highly predictive of body weight at 42 days of age in untreated rats. | The reviewer's points are well-taken, but the adjustment for body weight at weaning is still judged by the Agency to be appropriate. The groups are matched for body weight at weaning, thus removing much of the variability that might otherwise need to be accounted for, as the reviewer points out. The Agency has found that perhaps because of this, use of unadjusted weights provides results that are similar to those obtained when weights adjusted for body weight at weaning are used. However, the Agency has found that additional sensitivity is obtained when the weights are adjusted for weaning weight. |
| | | | |
| | | | |
| p. Wording | | | |
| 12.p.1 | JB | Test Method, Page 17, para 5. The term "estrus" is used as in "age at first estrus"; because this is ambiguous, and often refers to behavioral estrus, it should state "vaginal estrus." | The protocol will be changed to refer to "vaginal estrus". |