Test Guideline Paragraph (¶)	Comments
General	The data generated in the OECD validation program demonstrated the ability of the test method to reproducibly detect (anti)androgenic activity and 5∝- reductase inhibition for a limited number of substances (three agonists, five antagonists, one 5∝- reductase inhibitor, and two negatives) in a combination of 17 laboratories that were instructed to use specific doses for each test substance. However, the ability of laboratories to test substances (coded for five substances only in Phase III of the validation study), to select appropriate doses, and to obtain reproducible and accurate results using the complete test method protocol (all substances were tested by oral gavage except one that was tested subcutaneously) has not been demonstrated. Especially troubling is the lack of substances that are weak agonists in the validation study. Stating that the majority of androgenic substances to which humans might be exposed are antiandrogens does not obviate the need to demonstrate the ability of the assay to reliable detect compounds with weak androgenic activity. Thus, because OECD test guidelines should be based on adequately validated test methods (OECD GD 34), this validation database is insufficient for this purpose. There are a number of inconsistencies in the terminology used to define substances tested (i.e., substances are referred to as test substance, test material, test compound, and test chemical). We recommend the use of a single term, such as “test substance.”
Title	The title should indicate that this is a screening assay as was done for the Test Guideline (TG) for the Uterotrophic assay (i.e., The Hershberger Bioassay in Rats: A Short-Term Screening Test for (Anti)Androgenic Properties).

Test Guideline Paragraph (¶)	Comments
¶ 1	This paragraph states that "extensive" intra- and inter-laboratory studies were conducted. However, considering the number of chemicals tested, the word "extensive" is inappropriate or needs to be defined. The paragraph also states that the Hershberger Bioassay was first standardized and validated by an expert committee in 1962 (TG reference no. , "R.I. Dorfman. Standard Methods Adopted by Official Organizations. New York, Academic Press (1962)). This reference is incomplete. Our search indicates that it most likely refers to a section in a series of monographs published in 1962 and edited by R.I. Dorfman entitled "Methods in Hormone Research, Vol. II, Part IV: Standard Methods Adopted by Official Organizations." If this is the correct reference, it is little more than a brief review of a version of the Hershberger protocol (evaluation of 3 androgen sensitive organs for 21 day castrate rats dosed for 7-10 days after 1 day of recovery) that was used to test 3 androgen agonists and 2 antagonists. Therefore, it would be more accurate to use this reference to state that the assay was first "standardized" by an expert committee rather than to also state that the assay was first "validated" by an expert committee. p,p' DDT should read p,p' DDE.
¶ 4	This paragraph states that the validation study demonstrated the sensitivity of the assay for antagonists and agonists, as well as a "low rate of false positives with two negative compounds." These two conclusions are without scientific merit. First, no weak agonists were included in the validation study. Thus, the sensitivity and the reliability of the assay were not adequately evaluated. Second, claiming a low false positive rate based on two compounds lacks scientific credibility.
¶ 6	The demonstration of intralaboratory reliability and repeatability is insufficient considering that, besides reference standards, only three substances (one strong agonist and two weak antagonists) were independently tested only twice within a limited number of laboratories.
¶ 8	The TG states that the growth response of the individual androgen-dependent tissues is not entirely of androgenic origin and that weight changes in target tissues should initiate actions for further clarification. However, such methods are not specifically stated in the TG. Further clarification of methods to evaluate whether a positive androgen response constitutes a relevant androgen response is essential if the Hershberger assay is to be used as a screen for (anti)androgenic activity. Controlling for specificity of response is of particular importance given potentially confounding issues that are briefly addressed in the TG.

Test Guideline Paragraph (¶)	Comments
¶ 12	A citation that supports the use of 6 animals per dose group, based on power calculations, should be added. This is especially true given the lack of weak androgens in the validation study.
¶ 13	The term "castration is humane" should be deleted from the last sentence. It should be simply stated that castration is an effective means of reducing the number of animals required to screen for these endocrine activities.
¶ 14	The term "graduated" needs to be explained or replaced. In this context, it just seems to mean that substances should be tested at a minimum of two different dose levels. The TG states that "test substance is administered daily by oral gavage for a period of ten days." This is not consistent with the statement in ¶ 43 which states "test compound is administered by oral gavage or subcutaneous injection."
¶ 21	It is stated that in the validation study, no effects were observed that could be attributed to diet. While accurate, considering that no weak agonists were tested, how can it be concluded that the detection of weak agonists would be unaffected by diet. Diets containing high levels of phytoestrogens may affect the endpoints in the Hershberger Bioassay (see comments on ¶ 24). Therefore, the TG should recommend the use of a certified low phytoestrogen content diet for the bioassay.
¶ 22	The TG should also require testing of bedding for phytoestrogen content prior to testing.
¶ 24	This paragraph states that high levels of phytoestrogens in laboratory diets have not been shown to affect the endpoints in the Hershberger Bioassay. Studies by Stoheker (Food and Chemical Toxicology, 2003, vol. 41, pp. 1175-1183) indicate that rodent diets containing phytoestrogens may indeed alter the results of the Hershberger assay and recommend the use of standardized open-formula diets devoid of phytoestrogens when performing the Hershberger Bioassay.

Test Guideline Paragraph (¶)	Comments
¶ 27	Lab competence in performing the assay should be required a priori to data generation with the assay and would serve to address some concerns about potential assay/TG performance issues and possible confounders (e.g. related to bedding concerns, phytoestrogen content of the diet, potential species/strain differences, dose-setting, control, route of administration, etc.). A plan for how this would be done should be developed. The TG recommends verification of the responsiveness of the test system on a periodic basis or prior to the study by testing testosterone proprionate and examining whether a statistically significant increase in target tissues is achieved. This only verifies the response of a potent androgen agonist. The TG should also verify the response of a potent antagonist as well as a weak agonist and antagonist.
¶ 28	The rationale for the choice of rat strain should be clearly stated.
¶ 29	The minimum number of animals per group that are needed to support an adequate statistical analysis (e.g. 6) needs to be defined since deaths are possible (see ¶ 48)
¶ 35	The TG recommends that a minimum of two test groups and a control group should be used. This is not consistent with ¶ 45 and 46 (Specific procedures for androgen agonists/antagonists) where a positive and vehicle control is specified for the agonist protocol and a positive, negative and vehicle control is specified for the antagonist protocol. Considering the high variability of certain target tissues when testing weak agonists and antagonists, the TG should also recommend the use of a substance that induces a relatively weak but statistically significant response (or a dose level of a strong agonist/antagonist) as a concurrent control for both the agonist and antagonist protocols.
¶ 36	Delete the reference to NOEL. This is a screening assay and the objective is to detect substances that are positive for (anti)androgenic activity and not to establish a NOEL, which can't be reliably estimated with only two dose levels. The TG states that a range finding study "may" be done if there are no suitable data available. The TG should state that a range finding study "should" be done if there are no suitable data available.

Test Guideline Paragraph (¶)	Comments
¶ 37	This paragraph states that the maximum limit dose should be 1000 mg/kg/dy. However, other OECD short-term *in vivo* test guidelines (specifically those for genetic toxicology) mandate 2000 mg/kg/dy for studies of less than 14 days. It is not clear why the dose level for these studies are different, especially based on the likelihood that infants and children may be especially sensitive to endocrine disruption. Furthermore, as different regulatory agencies have different limit dose requirements, the statement "The limit test applies except when human exposure data indicates the need for a higher dose level to be used" should be revised to state "The limit DOSE applies except when there is a specific regulatory mandate that a higher dose level be tested, or when human exposure data indicates the need for a higher dose level to be used."
¶ 38	This section states that range finding results can be used to select an acceptable maximum and lower doses and recommend the number of dose groups. As only two dose groups are stated to be needed, should this paragraph not say "select the acceptable maximum and minimum dose groups"?
¶ 41	A vehicle for administration of a test substance is not specified. The TG recommends the use of several different oils, which have different densities and different caloric and fat contents. This is an issue of concern, since the vehicle may affect total metabolizable energy intake and alter measured endpoints. This should be addressed by running historical controls prior to initiating testing (i.e., vehicle to be used should be tested against controls without vehicle).
¶ 43	The TG indicates that test substance is administered by oral gavage or subcutaneous injection and that animal welfare and the physical/chemical properties of the test substance should be considered when choosing the route of administration. Validation study reports for Phases I-III indicated that all substances tested, other than the testosterone proprionate reference standard, were administered by oral gavage. Therefore, the subcutaneous route of administration can not be considered as validated.
¶ 45 & 46	See comment on ¶ 35.
¶ 49	Differences in food consumption can lead to differences in the magnitude of response to test substance. Therefore, unless justified, the optional weighing of feeders to measure food consumption should be deleted and guidance should be provided for determining individual food consumption for animals that are communally housed.
¶ 53	This paragraph states that weighing liver, kidneys, and adrenals are optional but provides no rationale as to why it might be useful to weigh these organs as opposed to any other,

Test Guideline Paragraph (¶)	Comments
¶ 61	Since only an increase in weight is relevant when testing for agonist activity, the statistical test should be one-tailed. It is stated that a statistically significant increase in tissue weight is consider positive. However, ¶ 14 states that a "dose responsive" statistically significant increase in target organ weights compared to the appropriate vehicle control group indicates a positive response. These differences need to be reconciled.
¶ 62	Since only a decrease in weight is relevant when testing for antagonist activity, the statistical test should be one-tailed. It is stated that a statistically significant increase in tissue weight is consider positive. However, ¶ 15 states that a "dose responsive" statistically significant decrease in target organ weights compared to the appropriate control group indicates a positive response. These differences need to be reconciled. The criteria for classifying a compound as positive, taking into account the number of responsive tissues, needs to be included,
¶ 64	Data collected and reported should also include analysis of phytoestrogen content in feeds and bedding.
¶ 66	Historical database should also include phytoestrogen content in feeds and bedding.