WEIGHT OF EVIDENCE GUIDANCE: EVALUATING RESULTS OF EDSP TIER 1 SCREENING TO IDENTIFY CANDIDATE CHEMICALS FOR TIER 2 TESTING

DRAFT FOR PUBLIC COMMENT

1. PURPOSE AND SCOPE OF DOCUMENT

The purpose of this document is to set forth some of the general principles, criteria, and considerations EPA generally believes to be relevant using a weight-of-evidence WoE) approach to evaluate data submitted as part of EPA's two-tiered paradigm for screening and testing chemicals for endocrine activity under the Endocrine Disruptor Screening Program (EDSP). This paper was developed by EPA to provide guidance to EPA staff and managers who will be reviewing data submitted in response Orders for Tier 1 screening that began October 29, 2009 under the Endocrine Disruptor Screening Program (EDSP). Additionally, outside parties submitting data may be interested to know how the results from Tier 1 screening are being evaluated. This paper provides general guidance and is not binding on either EPA or any outside parties. The use of language such as "will," "is," "may," "can," or "should" in this paper does not connote any requirement for either EPA or any outside parties. As such, EPA may depart from the guidance where circumstances warrant and without prior notice.

A WoE evaluation is a process where potentially relevant studies are judged in a professional manner for quality. Thereafter, a summary statement is developed indicating the potential effects of the compound, the mode of action (MOA), and other relevant information. It is not a process that simply involves tallying the number of positive and negative results within and among studies. Critical assessment of an entire body of available data is taken into account for consistency, coherence, and biological plausibility (*e.g.,* see USEPA, 2002 & 2005). Principles articulated in this document are equally applicable to a WoE evaluation of data from individual assays with multiple endpoints, as well as across the suite of assays in the Tier 1 screening battery. In addition, these principles would be generally relevant to the review of other scientifically relevant information (OSRI) submitted in response to test orders that request OSRI to be considered in lieu of designated screening assays in the Tier 1 battery. Most of the principles presented in this document are not unique to chemicals with potential endocrine activity but are commonly used for WoE evaluations conducted by EPA (USEPA 1991; 1992; 1996; 2002; 2005). The criteria discussed in this document are based on EPA's experience in developing and applying risk assessment guidelines involving cancer, reproductive and developmental toxicity, and ecological toxicity. Important considerations include the use of expert judgment formed through the scientific process, current understanding of endocrine mechanisms of toxicity, and knowledge of other fields of toxicology (*e.g.*, developmental, reproductive, neurological and immunological toxicology, and toxicokinetics). This document provides a transparent scientific approach for broadly evaluating Tier 1 screening data to determine if additional Tier 2 testing is necessary.

This document also is expected to comply with the provision in the Office of Management and Budget Terms of Clearance for the Information Collection Request for the first list of chemicals to be screened under the EDSP and direction in the House Appropriations Committee for the Interior and Environment FY 2010 report (HR 2996, http://thomas.loc.gov/cgi-bin/cpquery/R?cp111:FLD010:@1(hr180)) that directed EPA to:

"develop and publish criteria for evaluating the results of Tier 1 screening and determining whether a chemical should undergo Tier 2 analysis within one year of enactment. The process should allow for public input."

2. HISTORICAL BACKGROUND

This section provides an abbreviated overview of EPA's EDSP. A more detailed history of the program can be found at its website (http://www.epa.gov/endo) and in other documents or websites referenced herein.

In 1996, Congress amended section 408 of the Federal Food, Drug, and Cosmetic Act (FFDCA) to require EPA to:

"develop a screening program, using appropriate validated test systems and other scientifically relevant information, to determine whether certain substances may have an effect in humans that is similar to an effect produced by a naturally occurring estrogen, or other such endocrine effect as the Administrator may designate" [21 U.S.C. 346a(p)]. (http://www.epa.gov/pesticides/regulating/laws/fqpa/)

Fundamental to the EDSP is a two-tiered approach involving a battery of Tier 1 screening assays and individual Tier 2 tests designed to indentify and characterize chemicals with the potential to interact with the estrogen, androgen, and/or thyroid (E, A and/or T) hormonal systems. Tier 1 consists of a battery of complementary *in vitro* and short term *in vivo* assays designed to maximize sensitivity for detecting interactions with E, A, and/or T. Interactions with E, A, and/or T are evaluated using a range of key endpoints involving the MOA [*e.g.*, receptor binding and gene transcription, steroidogenesis, hypothalamic-pituitary-gonadal (HPG) and –thyroid (HPT) axes] across gender and various taxa (*e.g.*, rodents, amphibians, and fish) as indicated in Table 1. The diversity in endocrine endpoints within and among the Tier 1 screening assays is expected to provide corroborating information and support a WoE evaluation to yield a decision as to whether or not the chemical under evaluation requires additional testing in Tier 2.

Tier 2 testing consists of a group of individual *in vivo* tests designed to include males and females with an intact hypothalamic-pituitary-gonadal axis, multiple pathways of exposure and life-stages, and various taxa to further identify and characterize chemical-induced interactions with E, A, and/or T for risk assessment. Although the endocrine system is included, Tier 2 tests are designed to quantify dose-response relationships in a larger context of toxicity and potential adversity that may involve other biological

systems (*e.g.*, neurological, immunological, hepatic, renal, and cardiovascular) to be used for risk assessment. While the Tier 2 mammalian two-generation reproductive toxicity test is considered valid, other Tier 2 tests are at various stages in the validation process (Table 2).

Table 1: EDSP Tier 1 battery of screening assays and complementary modes of action*.

Screening Assays	*Modes of Action							
	Receptor Binding				Steroidogenesis		HPG Axis	HPT Axis
	E	Anti-E	A	Anti-A	E	A		
In vitro								
ER Binding	■	■						
ER α Transcriptional Activation	■							
AR Binding			■	■				
Steroidogenesis H295R					■	■		
Aromatase Recombinant					■			
In vivo								
Uterotrophic	■							
Hershberger			■	■				
Pubertal Male			■	■		■	■	■
Pubertal Female	■	■			■		■	■
Fish Short-term Reproduction (male & female)	■	■	■	■	■	■	■	
Amphibian Metamorphosis								■

*A mode of action is defined as a sequence of key events and processes, starting with interaction of an agent with a cell, proceeding through operational and anatomical changes, and resulting in an adverse outcome (USEPA, 2005). These assays encompass certain key events within a mode of action (*e.g.*, receptor binding) as well as certain pathways (*e.g.*, steroidogenesis) through which a chemical can interact with the E, A, or T hormonal systems.

Table 2. EDSP Tier 2 tests.

Mammalian two-generation reproductive toxicity test
*Avian two-generation toxicity test - Japanese quail two-generation toxicity test
*Amphibian growth and development test - Larval Amphibian Growth and Development Assay
*Fish multigeneration test - Medaka Multigeneration Test
*Invertebrate two-generation test - Mysid two-generation test

*Proposed Tier 2 test currently at various stages of the validation process.

Most of the proposed Tier 1 screening assays completed the validation process in 2008 and another in 2009. Subsequent to review [Federal Register Notice of January 24, 2008 (73 FR 4216)] by the Federal Insecticide Fungicide and Rodenticide Act Scientific Advisory Panel (FIFRA SAP) and, based on a final SAP report (SAP, 2008), EPA accepted the current EDSP Tier 1 battery (Table 1). Availability of test guidelines for each of the 11 screening assays in Tier 1 was published in a Federal Register Notice October 21, 2009 (74 FR 54416).

3. TESTING METHODS AND ENDPOINTS

3.1. EDSP Tier 1 screening assays

The basis for the endpoints in each of the *in vitro* and *in vivo* Tier 1 screening assays has been described (http://www.epa.gov/scipoly/oscpendo/pubs/assayvalidation/tier1battery.htm) in Integrated Summary Reports associated with the validation process (see "peer review of individual assays" at website), summarized in a Technical Review Document for SAP review of the battery (see "external review" at website) and listed in EPA test guidelines for each of the assays (see "Test guidelines for Series 890" at website). The endpoints associated with each assay were determined through the validation process to be sensitive and specific to detect interaction with the E, A, or T hormonal systems.

For chemicals having estrogen- or androgen-like activity, the *in vitro* receptor binding assays provide information on the potential binding characteristics of a compound. Comparatively, the *in vitro* estrogen receptor transcriptional activation or steroidogenic assays are cell-based and provide mechanistic as well as some functional information on induction/inhibition of gene transcription or degree of steroid hormone production, respectively.

In vivo assays integrate effects in a whole organism and provide apical as well as mechanistic information from one or multiple endpoints within an assay. Agonistic or antagonistic E-, A-, or T-dependent changes can be detected in association with reproductive development (*e.g.*, vaginal opening and preputial separation), organ weights (*e.g.*, ovaries, uterus, testes, prostate, and thyroid), and corresponding

histology of target organs. Mechanistic information such as thyroid hormones measured in the pubertal and amphibian assays and vitellogenin in the fish assay can be correlated with apical information within the same assays. The use of gonadectomized rats in the uterotrophic and Hershberger assays provides mechanistic information on specific estrogen- (*i.e.*, uterus) or androgen- (*i.e.*, prostate, seminal vesicles, levator ani-bulbocavernosus muscle, Cowper's glands, and glans penis) dependent target organs. As with the *in vitro* assays, the results of the uterotrophic or Hershberger assays are limited to E and A receptor function. On the other hand, results involving intact animals (*i.e.,* female and male pubertal, amphibian, and fish assays) generally provide more systems information, since more sites of action are involved along the HPG and HPT axes that target some of the same E-, A-, or T-dependent endpoints.

3.2. Other Scientifically Relevant Information (OSRI)

EPA's approach to the submission and use of OSRI as part of the EDSP has previously been described by EPA and is available in the Federal Docket Management System (EPA-HQ-OPPT-2007-1080-0032). In general, OSRI consists of data from assays that satisfy the same function as EDSP Tier 1 assays or may include data that are indicative of a potential consequence or adverse effect resulting from a chemical-induced change in the E, A, and/or T hormonal systems. Hence, interference of endocrine function may come from effects measured in standard toxicity test guidelines or other comparable toxicity studies. Typically, OSRI-types of studies are not designed to provide definitive information on the modes or mechanisms of toxicity, but are generally focused on measured adverse effects *(e.g.*, ability to become pregnant, duration of gestation, signs of difficult or prolonged parturition, sex ratio, or feminization or masculinization of offspring, number of pups, stillbirths, gross pathology, and histopathology of the vagina, uterus, ovaries, testis, epididymis, seminal vesicles, prostate, and thyroid) representing permanent changes with organizational or functional consequences. These studies may also encompass a range of life-stages (*e.g.*, two-generation reproductive study), treatment durations and doses, and provide information generated by relevant routes of exposure.

4. WEIGHT-OF-EVIDENCE APPROACH

In evaluating whether additional testing is warranted in Tier 2, EPA anticipates that the following key questions would typically be considered as part of EPA's WoE approach:

- Do existing data provide adequate evidence to conclude whether there is a potential for the chemical to interfere with the normal function of the E, A, and/or T hormonal systems?
- If the data indicate a potential to interact with those specific endocrine systems, which hormone system is impacted (E, A, and/or T)?

Determination of whether the evidence suggests the substance is or is not a candidate for Tier 2 testing is based on evaluation of all relevant data, including any Tier 1 results. This WoE analysis is conducted on a case-by-case basis by assessing all of the individual lines of evidence (Section 4.1) and performing an integrated analysis of the data (Section 4.2).

4.1. Analysis of individual studies

In any WoE analysis, a full evaluation of each relevant study is conducted and documented. In general, the evaluation of individual studies includes characterization of the following:

- Nature of the effect(s) seen in the study(ies) (*e.g.*, were the effects seen in *in vitro* and/or *in vivo* assays; were the effects persistent or transient changes; were the effects molecular/biochemical changes or adverse outcomes);
- specificity and sensitivity of the effect(s);
- dose- and time- dependent changes, if available;
- potency of responses and magnitude or severity of changes; and
- consistency and relationship of the different effects seen within a study.

Both statistical and biological significance of the observed effects are relevant in evaluating study results. In general, the results of relevant studies are assumed to be indicative of interactions with the endocrine system unless data are available that demonstrate otherwise (*e.g.*, evidence that the effect is not the consequence of an interaction with the endocrine system but a consequence of excessive toxicity to a non-endocrine system). To aid in determining the level of confidence in a study, the strengths of the study as well as any attendant limitations and uncertainties shall be considered.

4.2. Integrated analysis of data

Weight-of-evidence assessments must be conducted on a case-by-case basis. EPA's WoE analyses will generally include consideration of the information as follows:

- Quality of data and the extent to which effects can be replicated within a laboratory and across different laboratories;
- strengths and limitations of *in vitro* and *in vivo* results;
- number and type of effects induced and potency, magnitude, and severity of effects;
- consistency, pattern, range, and interrelationships of effects observed across studies, species, strains, and sexes;
- conditions under which effects occur (*e.g.*, dose, route, duration); and
- understanding of MOA and biological plausibility of responses.

These considerations are part of evaluating the evidence as a whole and determining whether or not a chemical has the potential to interact with the endocrine

system via E, A, and/or T hormonal pathways. In examining the balance of positive and negative results, the relative sensitivity and specificity of the measured endpoints would also be considered. Tier 1 *in vitro* screens can provide some insight into MOA. In general, however, Tier 1 *in vivo* results would carry greater weight than *in vitro* results because *in vitro* assays inherently lack physiological conditions associated with whole tissues or organs and, therefore, have nil or very limited ability to represent metabolic processes and pathways leading to endocrine disruption. The relationship between endpoints and their impact on normal endocrine function would also typically be significant factors in this determination. Concordant effects found in multiple interrelated endpoints generally imply a compromise in endocrine function, in contrast to isolated or discordant effects. Totality of the evidence is evaluated to determine whether such effects can potentially occur across taxa. Generally, consistent positive or negative effects across studies and taxa increase confidence in the determination of whether or not a chemical has the potential to interact with the endocrine system. Additionally, if marginal or weak relationships exist with regard to dose, severity, magnitude, and/or incidence, consideration of other available information may also be appropriate in determining whether further testing in Tier 2 is warranted. This could include consideration of other critical effects (non-endocrine), dose response, what is understood about the underlying basis (*i.e.*, toxicity MOA) of these critical effects (*i.e.,* non-endocrine or an endocrine MOA not covered by Tier 1) and their human relevance, and potential for exposure. Other supportive evidence may also be used in the WoE evaluation including pertinent data on related chemicals, metabolisim or toxicokinetics, and results of computational models.

4.3. Summary and conclusions of WoE approach

A summary of the WoE analysis is expected to transparently state and explain conclusions. It should explain the selection of certain studies or effects as the key basis for conclusions. In general, this characterization should be limited to the most significant and relevant data, conclusions, and uncertainties.

Summary statements for a WoE analysis should generally address key elements as follows:

- Each E, A, and T pathway, including species, gender, and life stages;
- uncertainties and the extent these uncertainties impact the conclusions;
- relative weight placed on studies and effects (*i.e.*, points at which choices are made of critical effects or studies and why); and
- inconsistent or conflicting results.

If Tier 2 testing is indicated, *i.e.,* effects are seen that are mediated through the E, A, and/or T hormonal systems, to the extent permitted by the available data, the summary should address any potential impact of the results of Tier 2 tests to risk assessment for that chemical and provide rationale for any conclusions. This may include, to the extent supported by the available data, conclusions regarding the species in which additional

testing is warranted and the likelihood that the effect may occur at a lower dose than effects seen in existing studies.

REFERENCES

EDSTAC (1998) Endocrine Disruptor Screening and Testing Advisory Committee, Final Report, Volume I-II. http://www.epa.gov/scipoly/oscpendo/pubs/edspoverview/finalrpt.htm.

.

SAP (2008) A set of scientific issues being considered by the Environmental Protection Agency regarding: The Endocrine Disruptor Screening Program (EDSP) Proposed Tier 1 Screening Battery. SAP Minutes No. 2008-03. http://www.epa.gov/scipoly/sap/meetings/2008/march/minutes2008-03-25.pdf.

USEPA (U.S. Environmental Protection Agency) (1991) Guidelines for developmental toxicity risk assessment. Federal Register 56(234):63798-63826. Available from: http://cfpub.epa.gov/ncea/raf/recordisplay.cfm?deid=23162.

USEPA (U.S. Environmental Protection Agency) (1996) Guidelines for reproductive toxicity risk assessment. Federal Register 61(212):56274-56322. Available from: http://cfpub.epa.gov/ncea/raf/recordisplay.cfm?deid=2838.

USEPA (U.S. Environmental Protection Agency) (1998) Guidelines for neurotoxicity risk assessment. Federal Register 63(93):26926-26954. Available from: http://cfpub.epa.gov/ncea/raf/recordisplay.cfm?deid=12479.

USEPA (U.S. Environmental Protection Agency) (2002) A Review of Review of the Reference Dose and Reference Concentration Processes. Risk Assessment Forum, Washington EPA/630/P-02/002F December 2002 Final Report http://www.epa.gov/raf/publications/pdfs/rfd-final.pdf

USEPA (U.S. Environmental Protection Agency) (2005) Guidelines for carcinogen risk assessment Risk Assessment Forum, Washington, DC. EPA/630/P-03/001F. Available from: http://www.epa.gov/raf/publications/pdfs/CANCER_GUIDELINES_FINAL_3-25-05.PDFp://www.epa.gov/