Endocrine Disruptor Screening Program Tier 1 Assays: Considerations for Use in Human Health and Ecological Risk Assessments

U.S. Environmental Protection Agency

Office of Chemical Safety and Pollution t‱ŜØŜyűĺ2y˝óh/{ttú
Office of Pesticide Programs (OPP)
Washington, DC 20460

June 2013

Table of Contents

1. Introduction

Section 408(p) of the Federal Food Drug and Cosmetic Act (FFDCA) directs the Environmental Protection Agency (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 such other endocrine effect as the Administrator may designate." In 1998, after considering public comments, external consultations and peer review, the EPA established the Endocrine Disruptor Screening Program (EDSP) as a two-tiered approach to implement the statutory testing requirements of FFDCA section 408(p) (21 U.S.C. 346a). Coordinated by the EPA, a battery of Tier 1 screening assays were developed, standardized, and validated to identify the potential of a chemical substance to interact with the estrogen, androgen or thyroid (E, A, or T) hormonal systems. These EDSP Tier 1 screening assays were externally peer reviewed by the Federal Insecticide, Fungicide, and Rodenticide Act Scientific Advisory Panel (FIFRA SAP) in March 2008. Based on the SAP recommendation, which found the proposed battery adequate to begin screening chemicals to detect the potential for interaction with the E, A, or T hormonal systems, the EPA has finalized the Tier 1 guidelines in the OCSPP Harmonized Test Guidelines Series 890 which can be found at: http://www.epa.gov/ocspp/pubs/frs/publications/Test_Guidelines/series890.htm. In accordance with FFDCA 408(p), orders were issued between October 2009 and February 2010 requiring Tier 1 screening of 67 chemicals. The status of the orders and progress of screening for each chemical can be found at: http://epa.gov/endo/pubs/toresources/index.htm.

The EDSP Tier 1 assay battery consists of eleven assays evaluating a wide range of endpoints including receptor binding, steroidogenesis, amphibian metamorphosis, fish reproduction, thyroid function, and effects on pubertal development. Typically, EDSP Tier 1 assays have been submitted to the agency in response to a test order under the EDSP. On occasion, Tier 1 assays have been submitted *in lieu* of other guideline studies required under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) as directed in the Code of Federal Regulations (40 CFR Part 158) or as a hybrid Part 158 guideline study/Tier 1 assay.

EDSP Tier 1 assays measure a range of effects at different levels of biological organization from *in vitro* molecular effects to apical endpoints *in vivo*. Tier 1 assays were designed to allow the EPA to evaluate whether a substance has the potential to interact with a hormone system, rather than to directly evaluate whether the substance will cause adverse effects in humans or ecological systems. However, the *in vivo* studies that are part of EDSP Series 890 Tier 1 Assays ("Tier 1 data") include apical endpoint data that can be relevant to evaluating a chemical's hazard, and thus, may be appropriately included as part of the weight of evidence considered in the EPA's risk assessments. This document discusses those parameters evaluated in the Tier 1 *in vivo* assays that can be indicative of plausible adverse outcomes, and that, in some instances, may be scientifically appropriate to use in risk assessments.

In evaluating the risks that pesticide exposure may pose to human health or the environment, the EPA uses a weight of evidence (WoE) approach that considers all relevant and scientifically sound data indicative of potential hazard. This WoE analysis is distinct from the analysis EPA conducts in the context of the EDSP Tier 1 evaluation, which is limited to an evaluation of a chemical's potential to interact with the estrogen, androgen, and/or thyroid (E, A, or T) hormone pathways, in order to ascertain whether additional EDSP Tier 2 data are needed. The WoE analysis described in this document only relates to the broader hazard and dose response assessment of all relevant adverse outcomes (regardless

of the mode of action) for purposes of assessing the risks that the use of a pesticide chemical may present.

Consequently, the EPA believes that apical data from the EDSP Tier 1 *in vivo* assays should generally be included as part of the weight of evidence considered in a risk assessment, provided the study is of good quality, conducted in intact animals, and yields reliable information. However, the decision to use the *in vivo* Tier 1 data in a WoE for risk assessments, as well as how it will be used—*e.g.*, whether quantitatively (*e.g.*, dose response assessment for derivation of points-of-departure) or qualitatively (*e.g.*, hazard characterization relevant to issues and questions within the risk assessment of a chemical used descriptively in WoE analysis and risk characterization)—will be made on case-by-case, after considering the specific data in question. Several factors may impact the manner in which the EPA considers the Tier 1 *in vivo* data, and the weight accorded to that data, including the overall quality of the data, physiological relevance, dose selection and dose-spacing, sensitivity of the endpoint(s) relative to other available data, relevance of the concentrations/doses, routes, and durations used in the EDSP Tier 1 assay to environmental exposures, or if Tier 1 data demonstrate effects on apical endpoints at doses/concentrations lower or in other species than those previously tested. Many of these considerations are applicable to both human health effects and environmental effects while others (*e.g.*, different taxa considerations or FQPA uncertainty factor determinations regarding level of concern for the susceptibility/sensitivity of infants and children to pesticide exposures) are more specific to either environmental effects or human health risk assessments. Some of these considerations and issues are discussed below.

2. Human Health Risk Assessments

The agency believes that it is scientifically appropriate to consider information on apical endpoints (*e.g.*, reproductive function and development, histopathology, organ weight changes, etc.) that are typically used in human health risk assessment when they are obtained from OCSPP Series 890 Tier 1 *in vivo* assays with intact animals, if a study is scientifically sound and has been conducted under appropriate conditions. These data should be interpreted and integrated with all relevant existing data and any uncertainty introduced by the use of EDSP Tier 1 *in vivo* data should be transparently characterized in the risk assessment.

2.1.1 Physiological relevance

Among the eleven assays that comprise the Tier 1 Battery, several are *in vitro* studies or *in vivo* studies using surgically-modified animals to increase the sensitivity of the assays for screening purposes. Although these assays have been validated for the purpose of determining specific MOAs and for screening purposes, their use in quantitative risk assessment present unique challenges in deriving points-of-departure and translating or extrapolating their findings to the dose response relationships of potential adverse outcomes to humans. In the case of *in vitro* studies, changes in the endpoint being assessed may not necessarily lead to an adverse outcome in an intact organism. For example, the concentration necessary to result in receptor binding and activation in an *in vitro* system may not be achieved through environmental exposure. Furthermore, biochemical perturbations at the molecular/cellular level may impact normal physiological responses in a way that may not result in adverse outcomes because of adaptive or homeostatic capacity or pharmacokinetics of intact biological systems and their limits, relative to concentration and duration of exposure. Thus, one highly relevant consideration would be to ascertain not only the level of change in the *in vitro* assay that may lead to adversity but also how the concentration causing the effect in the *in vitro* system translates to a

dose/concentration that elicits an adverse effect (akin to a lowest observed adverse effect level) in an intact organism. Considering that *in vitro* assays bypass both toxicokinetic and toxicodynamic contributions to the toxicological profile of a chemical, the EPA believes that these assays alone would not be scientifically appropriate for deriving a point-of-departure (PoD) or an FQPA factor, although they may have some utility in hazard characterization as supporting information in combination with *in vivo* apical toxicity data.

Even though assays relying on surgically-modified animals (Hershberger and Uterotrophic Assays) retain the toxicokinetic features of the organism, they also pose a challenge for interpretation. In these assays, animals have been surgically-modified to provide mechanistic information on androgen agonists, androgen antagonists and 5α-reductase inhibitors in the case of the Hershberger Assay or estrogenicity in the case of the Uterotrophic Assay. These assays have been designed to be exquisitely sensitive to perturbations of these pathways. Although test order recipients have the option of conducting the Uterotrophic Assay using immature, non-surgically modified animals, in the EDSP battery the ovariectomized (OVX) animal is preferred due to its increased specificity over the immature model. Since the majority of the Uterotrophic Assays conducted to date have used surgically-modified animals, the EPA lacks sufficient experience interpreting Uterotrophic assays conducted with intact animals to be able to provide further discussion in this document. The EPA will continue to evaluate such assays on a case-by-case basis.

Surgically manipulated systems do not represent a physiologically intact organism with its potential for compensatory capabilities. Thus, these assays explore the potential of a pesticide to interact with a system that does not accurately represent the "normal" toxicodynamic status of the animal. Likewise, the Fish Short-term Reproduction Assay (FSTRA) and Amphibian Metamorphosis Assay (AMA) utilize non-mammalian animals that are not directly applicable in a human health risk assessment. Consequently, these assays are not appropriate for endpoint selection, or derivation of points-of-departure (PoD) or to calculate an alternative FQPA factor. Nor would information from these methods be used as the sole basis of a hazard assessment, but could inform such assessments, when integrated appropriately with existing hazard and exposure information.

In summary, while the Tier 1 data are well-suited for screening the potential of a chemical to interact with the E, A, or T hormone systems, the previously described limitations of the *in vitro* tests, as well as the Hershberger, Uterotrophic, FSTRA, and AMA assays, preclude their use for derivation of points of departure used in a human health quantitative risk assessment. Consequently, when assessing the risks that chemical exposure may pose to human health, use of EDSP Tier 1 data will rely primarily on the Male and Female Rat Pubertal Assays which are *in vivo* assays using non-surgically modified mammalian animals and thus represent relevant physiological conditions.

2.1.2 Dose-Response

The ability to define the dose-response relationship of a critical effect is an important component of risk assessment, both when evaluating human health or environmental effects. Two of the features critical for a dose-response assessment are the number of doses used in the study (*i.e.*, dose selection) and the dose-spacing. The fact that the EDSP Tier 1 assays often require only 2 dose levels plus a concurrent control may limit their utility in dose response analysis, when compared to studies conducted using ≥ 3 dose groups plus a concurrent control. The number of doses in a study allows the investigators and reviewers to describe the shape and steepness of the dose-response curve (*i.e.*, slope) and the dose range responsible for any identified effects. Dose spacing is also a critical element in describing the dose

response relationship. For instance, if the dose levels are too widely spread, the PoD selected for risk assessments may be artificially low (*i.e.*, resulting in an overestimation of risk) or may not reflect the true nature of the dose-response curve. The "true" NOAEL may be close to either the lowest observed adverse effect level (LOAEL) or the study NOAEL thereby introducing additional uncertainty in the risk assessment. Either a too narrow or a too wide dose-spread may fail to fully describe the nature of the dose-response curve. The steepness of the dose-response curve may help inform decisions regarding uncertainty factors; particularly if a no observed adverse effect level (NOAEL) is not identified.

In addition, the EDSP Tier 1 assays also may employ high doses only, which can limit the utility of the information. This is because, at times, a chemical may produce toxicity that reflects a saturation of the mechanisms involved in absorption, metabolism, excretion, or homeostasis, and thus, may not accurately represent likely physiological conditions. Nonetheless, EDSP Tier 1 *in vivo* studies may be informative when considered in conjunction with the totality of the database (including Part 158 test guideline studies and OSRI) and may supplement the information used to derive PoDs and/or uncertainty factors.

2.1.3 Toxicological/Biological Relevance of Findings

In the context of risk assessment, toxicity studies are evaluated to ascertain **adverse** effects that may be the result of pesticide exposure. As described in the NRC report, *Toxicity Testing in the Twenty-first Century: A Vision and a Strategy*: "The consequences of a biologic perturbation depend on its magnitude, which is related to the dose, the timing and duration of the perturbation, and the susceptibility of the host." As such, agency reviewers are faced with the challenge of evaluating biochemical effects within a pathway and determining what magnitude and duration of perturbation may lead to an adverse health/environmental outcome. Even interpretation of traditional apical endpoints that may be modest or adaptive (*e.g.*, reductions in body weight or organ weights) or that may not be associated with any functional impairment or structural damage is important to ensure consistency in how they are incorporated in risk assessments.

Interpretations of adversity should be a WoE evaluation that considers, for example, MOA (if known), statistical and toxicological significance, the magnitude of the change, corroborating data, and toxicokinetics/toxicodynamics. Thus, the determination of adversity is a key element of the WoE analysis routinely conducted in risk assessment. There are no pre-determined response levels that would *de facto* lead to a determination of adversity. As described below, however, there are several parameter level changes that agency reviewers consider as part of the WoE analysis.

2.2 Pubertal Development and Thyroid Function in Intact Juvenile/Peripubertal Male Rats Assay (OCSPP 890.1500)

For the Pubertal Development and Thyroid Function in Intact Juvenile/Peripubertal Male Rats Assay (OCSPP 890.1500)(*i.e.*, "male Pubertal Assay"), there is a decreasing order of confidence for the following endpoints as potentially "adverse:" androgen-dependent organ weights>preputial separation (PPS)>serum testosterone. Where the data demonstrate multiple effects on androgen-dependent endpoints, such as a change in the weight of androgen dependent tissues, the timing of PPS, along with a significant change in serum testosterone, the EPA would generally have a high level of confidence that the effects should be considered "adverse." Similarly, the EPA would typically be concerned by data demonstrating a decrease/increase in androgen-dependent organ weights plus a change in PPS, even in the absence of changes in serum testosterone. In contrast, the EPA may have less confidence in concluding that an increase/decrease in serum testosterone, accompanied by an approximately 2 day

advance/delay in PPS without a change in androgen-dependent organ weights should be considered to be adverse effects. Similarly, certain effects would typically cause minimal concern, such as a delay or early onset of puberty that is ≤ 2 days in the absence of corroborating findings (*e.g.*, changes in serum testosterone, or androgen-dependent organ weights).

When evaluating these data, one important consideration is whether the effects seen are dose-dependent as data that demonstrate a dose-dependent change generally increase the confidence that the changes are treatment related and not spurious or within normal variation. It is also important to consider any other systemic effects that may be occurring, particularly body weight decrements (≥ 10%). For instance, a decrease in testosterone occurring in conjunction with a decrease in androgen-dependent absolute organ weight and a delay in puberty onset may not necessarily reflect a hormone mediated effect, but may actually be due to a decrease in body weight.

2.3 Pubertal Development and Thyroid Function in Intact Juvenile/Peripubertal Female Rats Assay (OCSPP 890.1450)

As was the case for the Pubertal Male Assay, there is a decreasing order of confidence in the endpoints to consider as potentially "adverse" in the Female Pubertal Assay: vaginal opening (VO)>cyclicity>organ weight. Early VO of ≥ 2 days has the potential to raise concern. However, the dose response in the study can affect the interpretation of the VO data. For example, data demonstrating a two-day delay at the high dose only would not typically give rise to as great a level of concern as data demonstrating a dose-dependent delay. An early VO may indicate estrogenic properties of the compound, provided there was corroborating information from other lines of evidence. If the test compound has estrogenic properties, it can affect the vaginal tissue directly to induce an early opening accompanied by a prolonged-persistent cornification of the vaginal epithelium and a lack of ovarian cycles present from VO until term (pseudoprecocious puberty *i.e.*, not a true puberty but an indicator of an estrogenic effect). An early onset of vaginal opening is likely to be reflected in altered reproductive function in later life as a result of the test chemical's effect on the brain. Ovulation would in turn likely be impaired along with an increased likelihood of early reproductive senescence. These effects should be considered in conjunction with the data from other studies (*e.g.*, Multigeneration Reproduction Toxicity Study [OCSPP 870.3800]) since there is a probability that "estrogenic" effects will be noted in other standard reproductive tests in the female. Other information available to determine whether a treatment-induced change (advance or delay) in puberty for two days is potentially adverse includes cyclicity. Information on "first estrus" and ovarian cyclicity data (number of cycles, etc.) collected prior to sacrifice can provide additional insight. Thus, delayed first estrus and a disruption of the ovarian cycle should typically elevate concern that the compound is disruptive to the reproductive system. In contrast to the above, if puberty is delayed for 1.5 - 2 days (and, if this is statistically significant, but not accompanied by any change in body weight) and there is **no** difference in estrous onset or cyclicity onset, the concern that the chemical may have an adverse effect would typically be lessened.[1]

Endocrine responsive organ weights are typically less reliable in the female pubertal (as opposed to the male pubertal assay) because they fluctuate with the estrous cycle and the stage of the estrous cycle is not controlled for in the female pubertal assay (*i.e.*, all females are killed on post-natal day [PND] 41 & 42). But these same considerations indicate that, if a change in uterine or ovarian weight is detected it

[1] The onset of estrus cyclicity or first estrus should be interpreted in light of any change in VO. Because first estrus is linked to VO, there will also typically be a delay in the first Estrus (indicative of a completion of the first ovarian cycle). This may simply reflect a developmental delay that has minimal consequences. Such a conclusion would be further supported if the female has a cycling pattern similar to controls.

would typically be considered adverse as this would reflect a robust effect of the test chemical.[2] Caution should be used when evaluating compounds exhibiting estrogenic properties as they may also cause anorexia. Thus, a decrease in body weight may be important and should not be used *de facto* to discount findings.

2.4 Thyroid Hormone Assessments

Historically, EPA's Office of Pesticide Programs (OPP) has used biochemical changes for endpoint/PoD selection when such changes can be linked to apical endpoints indicative of potential adverse health outcomes in humans. Biochemical changes that have been used in risk assessment include acetylcholinesterase inhibition, increases in liver enzymes indicative of hepatocellular damage (*e.g.*, alanine aminotransferase, aspartate aminotransferase, or alkaline phosphatase), and serum thyroid hormone changes. Thyroid hormone assessments are an integral part of both the Male and Female Pubertal Assays. For example, the developing nervous system is dependent on adequate amounts of thyroid hormones, and neurological impairments can potentially occur when the deficiency is present during brain development. Thus, an approximate 20% increase in TSH or decrease in thyroxine (T_4) is generally considered to be toxicologically relevant. As with all other endpoints, changes in thyroid hormones (manifested as changes in TSH and/or T_4) as well as histopathology and organ weight changes should be considered in context with the totality of the data in determining whether an effect is adverse. Some of the factors considered in this context include dose selection, variability of the assay, variation of the endpoint *per se*, and consistency with effects seen throughout the database.

2.5 FQPA Factor Considerations

The EDSP Tier 1 data may also be relevant to the agency's conclusions regarding the magnitude of the FQPA safety factor to be applied to protect infants and children.

To ensure the safety of infants and children with respect to pesticide chemical residues, FFDCA section 408(b)(2)(C) provides that "an additional tenfold margin of safety for the pesticide chemical residue and other sources of exposure shall be applied for infants and children to take into account potential pre- and postnatal toxicity and completeness of the data with respect to exposure and toxicity to infants and children. Notwithstanding such requirement for an additional margin of safety, the Administrator may use a different margin of safety for the pesticide chemical residue only if, on the basis of reliable data, such margin will be safe for infants and children." [21 USC 346a(b)(2)(C).] In making this judgement, the statute requires the EPA to consider "available information concerning the special susceptibility of infants and children to pesticide chemical residues…" [21 USC 346a(b)(2)(C)(i)(II)]. More directly, section 408(b)(2)(D)(viii) requires the agency to consider "such information as the Administrator may require on whether the pesticide chemical may have…endocrine effects," in determining the safety of a tolerance…" [21 USC 346a(b)(2)(D)(viii)]. Thus, to the extent EDSP Tier 1 data provide information relevant to these determinations, the EPA must consider it.

In the case of EDSP Tier 1 assays, it is important to consider that they are intended to identify substances that have the potential to interact with the estrogen, androgen, or thyroid hormone pathways. The fact that a substance may interact with a hormone pathway does not mean that any potential exposure may cause adverse effects in humans. As a result, the impact of EDSP Tier 1 data on FQPA

[2] Weighing organs in the morning (as instructed in the guideline) is less sensitive and may obscure findings. Uterine weight undergoes the greatest change beginning around noontime on the day of vaginal proestrus and the guidelines ask that the animals be killed two hours after the last dose (done at 9 AM).

Safety factor determinations will be decided on a case-by-case basis using a WoE approach that takes into consideration all relevant information in the database, including data submitted to fulfill Part 158 data requirements, as well as OSRI, and Tier 1 data for the chemical under evaluation.

3. Ecological risk assessment

The agency generally believes that it is scientifically appropriate to use information on apical endpoints (*e.g.*, survival, growth, and reproduction) that are typically used in ecological risk assessment when they are obtained from OCSPP Series 890 Tier 1 *in vivo* assays with intact animals, provided the study is scientifically sound and has been conducted under appropriate conditions. Once the reliability of a study has been established, the decision of whether to use the apical endpoint information only qualitatively or perhaps quantitatively would be made on a case-by-case basis and may be influenced by numerous factors, including whether the Tier 1 data demonstrate effects on apical endpoints at lower concentrations or in other species than were previously tested, whether the concentrations tested are informative in the context of estimated environmental concentrations (EECs), and the degree of uncertainty in risk conclusions that may result from using the information quantitatively. Further, risk assessment does not rely on a single test result, and thus apical endpoint data from Tier 1 *in vivo* assays would be integrated appropriately with existing hazard and exposure information.

OPP previously released guidance on determining and documenting whether information from open literature sources may be appropriate for qualitative or even quantitative use in some cases for ecological risk assessment.[3] Similar to the Tier 1 EDSP assays, open literature studies are often conducted without the specific intent of establishing endpoints for risk assessment; however, they may still contain valid and useful information when used in the context of all of the available data. Therefore, the quality criteria summarized in the open literature review guidance, in addition to assay-specific considerations (*e.g.*, validity and performance criteria) described in the Series 890 Test Guidelines and Standard Evaluation Procedures (SEPs), may be useful to the reviewer in determining whether a given study provides valid and useful information for risk assessment. Some common considerations and uncertainties are discussed below. Any uncertainty introduced by the use of Tier 1 *in vivo* data would be transparently characterized in the risk assessment.

3.1 Selection of maximum test concentration

For ecological risk assessment of conventional pesticides, long-term or chronic ecotoxicity data are of greatest utility when the doses tested in the study include test concentrations greater than or equal to a relevant EEC of an active ingredient. If no effects are observed in a study, but the substance has not been tested at concentrations up to and including an EEC, there is uncertainty about whether effects may occur at environmentally relevant concentrations. This may result in conservative risk conclusions because the risk assessor may presume chronic risk in the absence of adequate data.

None of the existing, publicly available test guidelines for long-term or chronic exposure of aquatic vertebrates discuss the importance of considering EECs when selecting the maximum test concentration (see **Table 1**). The Tier 1 EDSP guidelines for fish[4] and frog,[5] which are intended as screens for

[3] USEPA. 2011. *Procedures for Screening, Reviewing, and Using Published Open Literature Toxicity Data in Ecological Risk Assessments*. May 2011.
[4] USEPA. 2009. *Endocrine Disruptor Screening Program Test Guidelines OPPTS 890.1350: Fish Short-Term Reproduction Assay*. EPA 740-C-09-007. October 2009. http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2009-0576-0007.

potential endocrine interaction but may provide useful information about effects on traditionally assessed endpoints such as growth and reproduction, each recommend that the maximum test concentration be based on one of the following: the solubility limit, 100 ppm, or the highest test concentration that results in less than 10% mortality. The latter may be based upon range finding information or existing data for similar species. Based on these recommendations, the maximum concentration selected may be either higher or lower than the maximum concentration that would be selected in either of the FIFRA part 158 data requirements: a fish early life stage (ELS) or fish full life cycle (FFLC) test, which is intended specifically to inform risk assessment (see **Table 1**). However, the maximum test concentration is likely to be lower in the Tier 1 assays than in the part 158 studies because (1) mortality greater than 10% may cause a treatment to be discarded in the Tier 1 assays, and (2) for the Amphibian Metamorphosis Assay, the African clawed frog (*Xenopus laevis*) tadpoles may be more sensitive to the test compound than the fish species and life stages typically used in acute (*e.g.*, 96-hour) toxicity and range finding studies.

Therefore, the use of apical endpoints from Tier 1 assays with fish and frog in ecological risk assessment may result in uncertainty with respect to risk conclusions in the following cases:

1) The maximum test concentration in the Tier 1 assay is below the EEC for a comparable duration of exposure and no effects on apical endpoints are observed.
2) The maximum test concentration in the Tier 1 assay is below the range of concentrations previously tested in fish ELS or FFLC studies for a given active ingredient, and no effects on apical endpoints are observed in the Tier 1 assay.
3) The maximum test concentration in the Tier 1 assay is below the solubility limit and no effects on apical endpoints are observed.

3.2 Spacing of Test Concentrations

It is generally recommended that the Tier 1 EDSP assays with frog and fish be conducted with at least three test concentrations, each separated by approximately 3x to 10x (see Table 1). Depending on the slope of the concentration-response curve (if estimated) for a given endpoint, and the nature of the concentration response (*e.g.*, monotonic or non-monotonic), the use of data from assays with widely spaced test concentrations may result in an overly conservative (*i.e.*, protective) endpoint for risk assessment. Alternatively, data generated by assays with more narrowly spaced concentrations may fail to fully characterize the nature of any concentration-response. However, in screening level risk assessment for conventional pesticides, concentration-response data (slopes, etc.) for quantitative use are typically only obtained from acute toxicity tests and from plant toxicity tests; concentration-response curves for pesticides are rarely generated using data from long-term or chronic animal toxicity tests and are not currently used in risk estimation.[6]

[5] USEPA. 2009. *Endocrine Disruptor Screening Program Test Guidelines OPPTS 890.1100: Amphibian Metamorphosis (Frog)*. EPA 740-C-09-002. October 2009. http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2009-0576-0002.

[6] Regression-based endpoints may be used by other agency offices in the assessment of industrial chemicals.

Table 1: Recommendations for Test Concentrations in EDSP and non-EDSP Assays with Aquatic Vertebrates.

Series 890 Test Guideline	Minimum Number of Test Concentrations	Maximum Test Concentration	Recommended Spacing
890.1100 Amphibian Metamorphosis Assay (AMA)	3	the lowest of 1) solubility limit 2) 100 ppm 3) the highest test concentration that results in less than 10% mortality	3x to 10x
890.1350 Fish Short-Term Reproduction Assay (FSTRA)	3	the lowest of 1) solubility limit 2) 100 ppm 3) the highest test concentration that results in less than 10% mortality	3x to 10x[1]
Series 850 Test Guideline			
850.1400 Fish Early Life Stage Test (ELS)	5[2]	$\leq$ 96-hour LC_{50} or $\leq$ 10 ppm	$\leq$ 3.2x
850.1500 Fish Full Life Cycle Test (FFLC)	5[2,3]	based on range finding or existing data; $\geq$ LOAEC[3]	based on range finding, should bracket endpoint of interest (NOAEC or EC_x)[3]

1. USEPA. 2011. Corrections and Clarifications to Technical Aspects of the Test Guidelines for the Tier 1 Endocrine Disruptor Screening Program Assays. March 2011. Available online at http://www.epa.gov/endo/pubs/toresources/clarificationdoc.pdf.
2. Unless being conducted as a limit test.
3. Based on internally proposed revisions to draft test guideline. The existing public draft test guideline (1996) does not give explicit recommendations for study design.

3.3 Relevance of endpoints

Effects on survival, growth, and reproduction of wildlife species are routinely evaluated and endpoint values (*i.e.*, NOAEC, LOAEC) are determined for these parameters which are then used in risk estimation (see **Table 2**). Additional information on sublethal effects (*e.g.*, clinical signs, including biochemical alterations) other than the traditional measures of growth (*e.g.*, length, weight, weight gain) and reproduction are incorporated into risk characterization but are not typically utilized quantitatively (*i.e.*, in the calculation of risk quotient values). Data for non-mammalian taxa are primarily obtained from Series 850 guideline studies, but may also be drawn from a variety of other sources, including non-EPA guideline studies (*e.g.*¸OECD Test Guidelines) and open literature. Mammalian toxicology data for use in ecological risk assessment are primarily obtained from OCSPP Series 870 Health Effects Test Guideline studies, which are reviewed by the Health Effects Division. The degree to which information from any given source is considered scientifically valid and appropriate for use in ecological risk assessment of pesticides is generally documented in a Data Evaluation Record (DER) or an Open Literature Review Summary (OLRS).

The Tier 1 EDSP assays with fish and frog (Series 890) are *in vivo* studies that combine observations of traditionally assessed apical endpoints (see **Table 2**) with other developmental biomarkers, histopathology, and (for the fish) biochemical observations from either the same individual or a subset of individuals under the same exposure conditions (*i.e.*, from the same replicate). Similarly, the Tier 1 EDSP assays with female and male pubertal rats (OCSPP 890.1450 and 890.1500, respectively) provide information on mammalian growth and mortality in addition to developmental, histopathological, and biochemical endpoints. Thus, the Tier 1 assays provide information above and beyond the type of

information that would typically be obtained by conducting a Series 850 guideline study. Although not the singular purpose of the Tier 1 assays, such information can support a determination of whether there are effects on apical endpoints within the range of concentrations tested.

The agency has stated that the potential for endocrine interaction and the need for Tier 2 testing will be determined as part of a weight of evidence evaluation that considers the available battery of Tier 1 assays and any Other Scientifically Relevant Information (OSRI) available to the agency. However, the determination of whether an apical endpoint, such as survival, body weight, length, or reproduction, is affected as a result of chemical exposure in a given assay is more straightforward and is independent of any conclusions regarding potential endocrine interaction. Therefore, data regarding these types of endpoints could be confirmed and utilized in hazard characterization and risk assessment prior to and irrespective of the conclusions of the weight of evidence evaluation for the full Tier 1 battery. As with any other data source, the degree to which information on apical endpoints would be appropriate and useful in risk assessment would rely upon a case-by-case evaluation of whether the study was scientifically sound, was conducted under appropriate conditions, and provides relevant information.

Table 2: Endpoints Common to EDSP and non-EDSP Assays with Aquatic Vertebrates.

		Examples of Similar Endpoints and Clinical Signs											
Series 890 Test Guideline	**Recommended Species**	**Survival***	**Fecundity***	**Fertility***	**Egg viability***	**Time to hatch**	**Time to swim up/ metamorphosis**	**Length***	**Weight***	**Behavior**	**Morphology (e.g., deformities)**	**Secondary sex characteristics**	**Gonadal morphology**
890.1100 Amphibian Metamorphosis Assay (AMA)	African clawed frog (*Xenopus laevis*)	X						X	X	X	X		
890.1350 Fish Short-Term Reproduction Assay (FSTRA)	Fathead minnow (*Pimephales promelas*)	X	X	X				X	X	X	X	X	X
Series 850 Test Guideline													
850.1400 Fish Early Life Stage Test (ELS)	Rainbow trout (*Oncorhynchus mykiss*) Fathead minnow (*Pimephales promelas*) Sheepshead minnow (*Cyprinodon variegatus*) Silverside spp. (*Menidia spp.)*	X			X	X	X[1]	X	X	X	X		
850.1500 Fish Full Life Cycle Test (FFLC)	Fathead minnow (*Pimephales promelas*) Sheepshead minnow (*Cyprinodon variegatus*)[2]	X	X	X	X	X	X[1]	X	X	X	X	X	X

* Endpoints marked with an asterisk are commonly used to calculate risk quotient (RQ) values. Other endpoints may be used in risk characterization.

1. For example, rainbow trout (*Oncorhynchus mykiss*) and medaka (*Oryzias latipes*).
2. Based on internally proposed revisions to draft test guideline. The existing public draft test guideline (1996) does not give explicit recommendations for study design.

4. Conclusions

In conclusion, the decision to use Tier 1 data in risk assessments and the manner in which it may be used will be made on a case-by-case basis, influenced by a number of factors. The EPA does not intend that data from any given Tier 1 assay will be solely relied upon in making an assessment, but will be interpreted and integrated with all pertinent existing data (*e.g.*, Part 158 data). In keeping with the NRC report vision, the goal is to focus resources on the evaluation of the more biologically plausible and sensitive adverse effects of exposures of concern rather than on full characterization of every possible adverse effect irrespective of relevance for risk-assessment and risk-management needs. Given that the Tier 1 *in vivo* studies generally represent apical endpoints that are routinely evaluated in Part 158 studies (Table 3) and used in agency risk assessments, existing guidance should be followed for interpreting and integrating these data into risk assessments. This paper is not intended to be prescriptive or to provide a checklist for the evaluation of Tier 1 data in the context of risk assessment. The purpose is to delineate some of the basic concepts/principles generally used by agency scientists in evaluating all toxicological data that are also applicable to Tier 1 data. The first and foremost consideration would be the overall quality of the data, followed by a WoE analysis of whether the effect(s) observed demonstrate(s) a plausible adverse human health or environmental outcome. In the context of ecological risk assessment, this determination may be made regardless of the ability to discern a specific mechanism of action if the data demonstrate an effect on an apical endpoint such as survival, growth, or reproduction.

Table 3: Mammalian Toxicity Cross Walk with Part 158 Required Studies and the *In Vivo* Non-surgically Modified EDSP Protocols

Parameter	Male Pubertal	Female Pubertal	Repro	Dev Tox	DNT	Subchronic	Chronic	Carcinogenicity
Test Species/Sex								
Mouse								X
Rat	X	X	X	X	X	X	X	X
Rabbit				X				
Dog						X	X	
Age at Start of Study								
PND 22		X						
PND 23	X							
Prior to PND 25								
PND 42								
6 weeks								
8-9 weeks			X	X	X	X	X	X
Route of Administration								
Oral								
Gavage	X	X	X	X	X			
Dietary			X		X	X	X	X
Drinking water			O	O		O	O	
Dermal						X		
Subcutaneous								
Inhalation						X		
Duration of Exposure								
Acute								
Subchronic	X	X		X	X	X		
Chronic							X	X
Generations			X					
Endpoints								
Mortality			X	X	X	X	X	X
Moribundity			X	X	X	X	X	X
Body Weight	X	X	X	X	X	X	X	X
Food Consumption			X	X	X	X	X	X
Clinical Signs of Toxicity			X	X	X	X	X	X
FOB			O			X	O	
Survival of Offspring			X		X			
Preputial separation (age)	X		X		X			

Parameter	Male Pubertal	Female Pubertal	Repro	Dev Tox	DNT	Subchronic	Chronic	Carcinogenicity
Sperm morphology			X					
Sperm motility			X					
Vaginal opening (age)		X	X		X			
Vaginal cytology (estrous cyclicity)		X	X					
# of corpora lutea				X				
# of implantation sites			X	X				
# of viable pups at day of cesarean				X				
# of viable pups at day of birth			X		X			
# of non-viable pups			X	X	X			
Sex ratio of offspring			X	X				
Anogenital distance			X					
Male Reproductive Performance			X					
Female Reproductive Performance			X					
External malformation/anomalies			X	X	X			
Visceral malformation/anomalies				X				
Skeletal malformation/anomalies				X				
Morphometrics					X			
Hormones	X	X	O			O	O	
Clinical pathology						X	X	X
Opthalmological examination						X	X	X
Gross necropsy			X	X		X	X	
Organ Weight								
Liver	X	X	X			X	X	X
Kidneys	X	X	X			X	X	X
Adrenals	X	X	X			X	X	X
Testes	X		X			X	X	X
Individ Accessory Male Sex Organs	X		X			X	X	X
Ovaries		X	X			X	X	X
Uterus		X	X			X	X	X
Uterus - Gravid				X		X	X	X
Thyroid (with parathyroid)						X	X	X
Lungs						X	X	X
Brain			X		X	X	X	X
Peripheral Nerve						X	X	X
Spleen			X			X	X	X
Thymus			X			X	X	X
Lymph Nodes						X	X	X

Parameter	Male Pubertal	Female Pubertal	Repro	Dev Tox	DNT	Subchronic	Chronic	Carcinogenicity
Heart						X	X	X
Pituitary	X	X	X			X	X	X
Target Organs			X			X	X	X
Histopathology								
Liver			X			X	X	X
Kidneys			X			X	X	X
Adrenals			X			X	X	X
Testes	X		X			X	X	X
Individ Accessory Male Sex Organs	X		X			X	X	X
Vagina			X					
Ovaries		X	X			X	X	X
Uterus		X	X			X	X	X
Thyroid (with parathyroid)	X	X				X	X	X
Lungs						X	X	X
GI Tract						X	X	X
Brain			X	X		X	X	X
Pituitary			X			X	X	X
Target Organs			X			X	X	X
Urinalysis						X	X	X

X = Required
0 = Optional

Blue area = EDSP assays
Unshaded area = Part 158 toxicity studies