Regulatory Toxicology and Pharmacology 61 (2011) 185–191

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Hypothesis-driven weight of evidence framework for evaluating data within the US EPA's Endocrine Disruptor Screening Program

Christopher J. Borgert [a,b,*], Ellen M. Mihaich [c], Lisa S. Ortego [d], Karin S. Bentley [e], Catherine M. Holmes [f], Steven L. Levine [g], Richard A. Becker [h]

[a] Applied Pharmacology & Toxicology, Inc., Gainesville, FL, USA
[b] C.E.H.T, University of Florida, Dept. Physiological Sciences, Gainesville, FL, USA
[c] Environmental and Regulatory Resources (ER^2), Durham, NC, USA
[d] Bayer CropScience, Research Triangle Park, NC, USA
[e] DuPont Crop Protection, Newark, DE, USA
[f] BASF Corporation, Research Triangle Park, NC, USA
[g] Monsanto Company, St. Louis, MO, USA
[h] American Chemistry Council, Washington, DC, USA

ARTICLE INFO

Article history:
Received 23 March 2011
Available online 23 July 2011

Keywords:
Weight of evidence
Endocrine Screening Program
Endocrine disruption
Regulatory framework

ABSTRACT

"Weight of Evidence" (WoE) approaches are often used to critically examine, prioritize, and integrate results from different types of studies to reach general conclusions. For assessing hormonally active agents, WoE evaluations are necessary to assess screening assays that identify potential interactions with components of the endocrine system, long-term reproductive and developmental toxicity tests that define adverse effects, mode of action studies aimed at identifying toxicological pathways underlying adverse effects, and toxicity, exposure and pharmacokinetic data to characterize potential risks. We describe a hypothesis-driven WoE approach for hormonally active agents and illustrate the approach by constructing hypotheses for testing the premise that a substance interacts as an agonist or antagonist with components of estrogen, androgen, or thyroid pathways or with components of the aromatase or steroidogenic enzyme systems for evaluating data within the US EPA's Endocrine Disruptor Screening Program. Published recommendations are used to evaluate data validity for testing each hypothesis and quantitative weightings are proposed to reflect two data parameters. Relevance weightings should be derived for each endpoint to reflect the degree to which it probes each specific hypothesis. Response weightings should be derived based on assay results from the test substance compared to the range of responses produced in the assay by the appropriate prototype hormone and positive and negative controls. Overall WoE scores should be derived based on response and relevance weightings and a WoE narrative developed to clearly describe the final determinations.

1. Introduction

On November 4, 2010, the US EPA released its draft "Weight-of-Evidence Guidance Document: Evaluating Results of EDSP Tier 1 Screening to Identify Candidate Chemicals for Tier 2 Testing" (US EPA, 2010). The Agency stated in its guidance that it would use WoE to determine whether a chemical has the potential to interact with the estrogen, androgen, or thyroid hormone components of the endocrine system. EPA stated that the intent of the document was "...to provide a transparent scientific approach for broadly evaluating Tier 1 screening data to determine if additional Tier 2 testing is necessary." EPA asserted its draft Guidance document provided a clear statement of how EPA intended to evaluate Tier 1 data so that the Agency's methodology would be transparent to all stakeholders.

The draft EPA WoE Guidance offers only some general considerations and principles related to making WoE determinations within the Endocrine Disruptor Screening and Testing Program (EDSP), and this may be viewed by some as providing a desired degree of flexibility for accommodating expert judgments within the effluvium of regulatory analyses and decision-making under uncertainty. However, the draft Guidance falls well short in describing how a WoE approach for the EDSP will be structured, how data will be evaluated for use in WoE, how the endpoints

* Corresponding author at: Applied Pharmacology & Toxicology, Inc., 2250 NW 24th Avenue, Gainesville, FL 32605, USA. Fax: +1 352 335 8242.
E-mail addresses: cjborgert@apt-pharmatox.com (C.J. Borgert), emihaich@nc.rr.com (E.M. Mihaich), lisa.ortego@bayer.com (L.S. Ortego), karin.s.bentley-1@usa.dupont.com (K.S. Bentley), catherine.holmes@basf.com (C.M. Holmes), steven.l.levine@monsanto.com (S.L. Levine), Rick_Becker@americanchemistry.com (R.A. Becker).

0273-2300 © 2011 Elsevier Inc. Open access under CC BY-NC-ND license.
doi:10.1016/j.yrtph.2011.07.007

measured in the Agency's Tier 1 endocrine screening battery (ESB) will be weighted, or even how a weighing mechanism should be developed. A direct, transparent and objective methodology is still needed that will provide for consistency and credibility of WoE determinations made on the basis of EDSP data. A transparent and objective WoE methodology is especially necessary for the EDSP given the EPA's (and industry's) lack of experience conducting the ESB, the broad scope of the program, the significant impact inaccurate assessment could have on society and the regulated industry, and the excessive numbers of laboratory animals and costs required for Tier 2 testing.

The EDSP consists of two distinct tiers. Tier 1 is intended to determine whether a substance may interact with the endocrine system. Tier 1 consists only of screening assays, which are not sufficient alone to determine whether substances may have adverse health effects or to determine mode of action. Negative Tier 1 results would be adequate to determine that a substance is unlikely to have an effect on the estrogen, androgen or thyroid hormone systems or aromatase and steroidogenic enzymes. Positive Tier 1 results would indicate that the substance should be prioritized for Tier 2 testing. Tier 2, which consists of more apical assays, is intended to determine whether a substance may cause adverse effects, including those potentially mediated by the endocrine system, and evaluate the dose response associated with such effects. Tier 2 testing is more definitive than Tier 1 screening and negative Tier 2 results should supersede positive Tier 1 results (US EPA, 1998).

It is clear that screening assays provide qualitatively different information than definitive Tier 2 tests, and the results from these dissimilar assays should be used in a manner that is consistent with the scientific basis and purpose of each. The framework for conducting WoE evaluations for hormonally active agents proposed here is meant to operate within EPA's two-tiered EDSP and is intended to assist analysts in making the appropriate distinctions. Given the structure of EPA's EDSP, five separate WoE evaluations will be needed to assess EDSP data and to make the following determinations:

[a] determining from the Tier 1 ESB and other scientifically relevant information (OSRI) whether a substance exhibits the potential for interaction with androgen, estrogen, or thyroid pathways or aromatase and steroidogenic enzymes in vivo;
[b] determining from the Tier 1 ESB, OSRI and other information whether the substance should be further evaluated for endocrine activity in Tier 2 toxicity tests;
[c] determining from the results of Tier 2 toxicity tests whether a substance exhibits adverse effects potentially mediated by androgen, estrogen, or thyroid pathways;
[d] determining from Tier 1 ESB, OSRI, Tier 2 toxicity tests, and as necessary, additional mode-of-action experiments, whether the adverse effects observed in Tier 2 toxicity tests are a consequence of endocrine activity, and;
[e] determining whether endocrine-mediated adverse effects on humans or wildlife are possible at environmentally relevant exposure levels.

The framework for conducting WoE evaluations described here is applicable to all five of these separate determinations. This publication describes the elements of the framework, including its relationship to other published WoE approaches for endocrine active substances, the overarching scientific principles that govern data evaluation within the framework, and the two primary weighting types used to evaluate data for each WoE determination. This publication does not, however, describe the operational and technical details necessary to carry out the five individual WoE determinations. Subsequent publications will provide those. Instead, this paper focuses on the principles and processes for weighting data and illustrates how this is to be done for Tier 1 ESB data, i.e., for WoE determination [a] above.

Before delving further into the background literature and scientific principles governing the proposed framework, it is imperative to define terminology clearly so that the WoE framework can be considered in its proper context. Weed (2005) has noted that the term "weight of evidence" is used frequently in the scientific literature without being defined. According to Weed, the term is used in three categorically distinct ways: (1) metaphorical, (2) methodological, and (3) theoretical. As used in the framework proposed here, the term "weight of evidence" is both theoretical in that it labels the overall process, as well as methodological in that it describes specific methods and qualitative principles governing the use of the proposed process. In subsequent publications, various quantitative procedures will be described that might be used to weight data from the various types of studies relevant for evaluating potential endocrine activity and endocrine-mediated toxicity. Importantly, the framework proposed here incorporates step-by-step documentation and transparency of the decision process, which have been identified as elements that enhance scientific credibility (Borgert, 2007a,b; Schreider et al., 2010).

The proposed WoE approach can be summarized according to the following seven steps, the justification and background (Section 2), scientific principles (Section 3), operational details (Sections 4 and 5), and implications (Section 6) of which are explained further in this paper and in the tabular summaries available as Supplementary material:

1. define specific hypotheses to be evaluated;
2. systematically search, review and select data relevant to each hypothesis;
3. evaluate the primary validity and reliability of each study selected, and for WoE evaluations involving causality (e.g., [c] and [d] above), determine whether the data are derived from counterfactually designed studies;
4. develop quantitative or rank ordered relevance weightings (W_{REL}) for each type of assay or endpoint with respect to its sensitivity and specificity for testing the hypothesis;
5. develop quantitative response weightings (W_{RES}) based on results for the test substance compared to positive and negative controls in each assay or endpoint;
6. combine relevance (W_{REL}) and response (W_{RES}) weightings according to a pre-defined algorithm to produce an overall WoE score;
7. develop an overall WoE determination as to whether each hypothesis is supported or rejected, and how strongly, based on the overall WoE scores.

2. Background and justification

Several organizations have developed frameworks and discussed principles important for conducting WoE evaluations (Balls et al., 2006; Bars et al., 2011; Boobis et al., 2006, 2008; Damstra et al., 2002; ECETOC, 2009; Gray et al., 2001; Menzie et al., 1996) and independent investigators have published WoE frameworks and evaluations of endocrine active substances (e.g., Calabrese et al., 1997; Goodman et al., 2006, 2009; Martin et al., 2007; Rhomberg, 1998, 2008; Rhomberg and Goodman, 2008). It is beyond our scope to summarize each of these frameworks and publications, but a general overview is provided in the overview of weight of evidence frameworks in Supplementary material, which is helpful for understanding overarching issues related to developing WoE frameworks and is essential for understanding our proposed framework in the context of this previous work.