

Hypothesis-Based Weight of Evidence Framework Should Be Favored for Identification of Endocrine Disruptors

Ellen M. Mihaich, PhD, DABT

Christopher J. Borgert, PhD

- *Endocrine Policy Forum (EPF)*

Interpreting Endocrine Screening (and testing) Data

1. U.S. EPA noted that a WoE would be used to determine that a chemical does or does not have the potential to interact with the endocrine system (i.e., specifically the estrogen, androgen, or thyroid hormone systems).
 - EPA Endocrine Disruptor Screening Program; Policies and Procedures for Initial Screening, Federal Register, April 15, 2009 (Volume 74, Number 71), pages 17560 - 17579.
2. OECD: “All existing relevant data should be maximally used (e.g. structural; physico-chemical information; in vivo and in vitro guideline and non-guideline testing; QSAR models; computational and other non-testing assays; toxicokinetic, pharmacokinetic and toxicodynamic information; category and read-across assessment methodologies) in a WOE approach before entering any other level of the CF.”
 - OECD (2018). *Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption*, OECD Series on Testing and Assessment. OECD Publishing, Paris.
<https://doi.org/10.1787/9789264304741-en>.

Scope of this presentation

Hypothesis-Driven WoE Approach for US EPA's Endocrine Screening Program

(Borgert, Mihaich, et al., 2011. *Reg Tox Pharm* 61:185-191)

Relevance Weighting of Tier 1 Endocrine Screening Endpoints

(Borgert et al., 2014. *Birth Defects Res B Dev Reprod Toxicol*, 101(1), 90-113.)

- Developed using Endocrine Modes of Action (MoAs) as model
- The methodology also broadly applicable for:
 - Non-endocrine MoAs
 - Human health and ecological effects
- Used for published evaluations of several industrial chemicals
- Cited in 119 peer-reviewed publications, per Google Scholar
- Cited by the OECD

Hypothesis-Driven WoE Methods Ensure Transparency

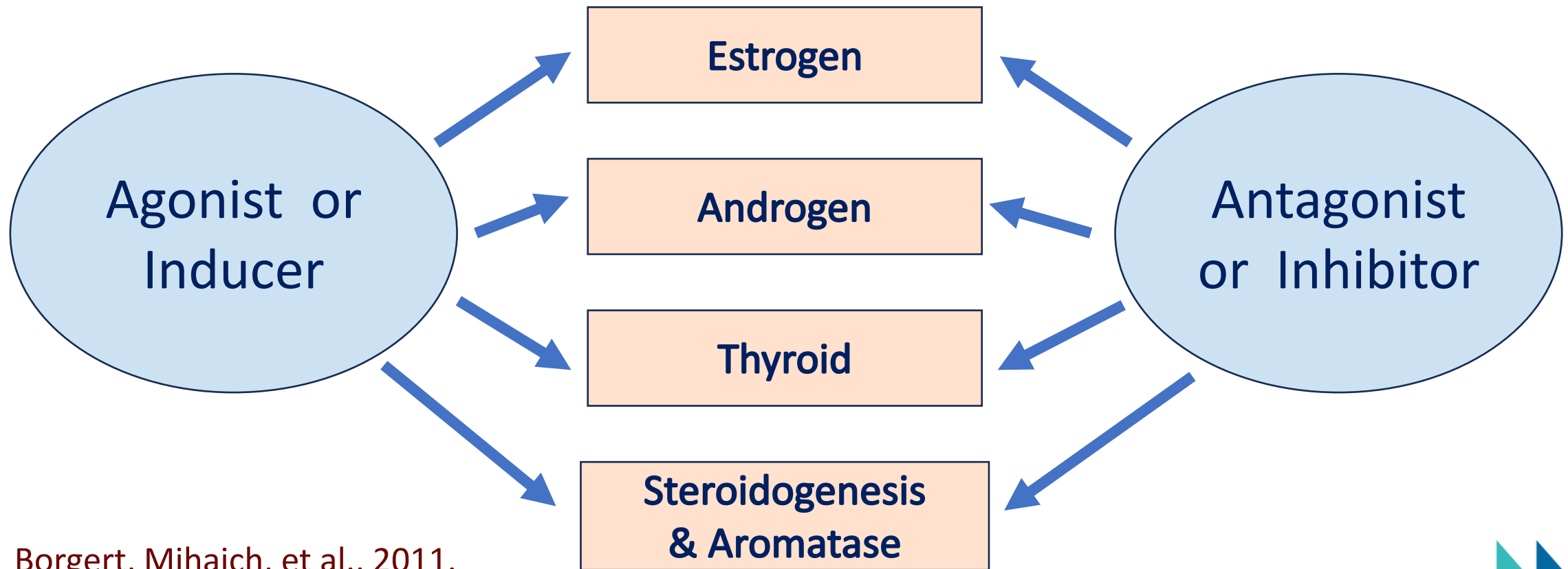
- I. Transparency of the questions asked (hypotheses)
- II. Transparency of the method used to answer:
 - a) How “evidence” will be defined;
 - b) How the evidence will be found and identified;
 - c) How the quality of the evidence will be evaluated;
 - d) How the evidence will be “weighted” [weight = “importance”];
 - e) How the evidence will be used to test the hypotheses;
 - f) How overall conclusions will be reached.

Hypothesis-Driven WoE Approach for ESB

(Borgert, Mihaich, et al., 2011. Reg Tox Pharm 61:185-191)

I. Transparency of the questions asked (hypotheses)

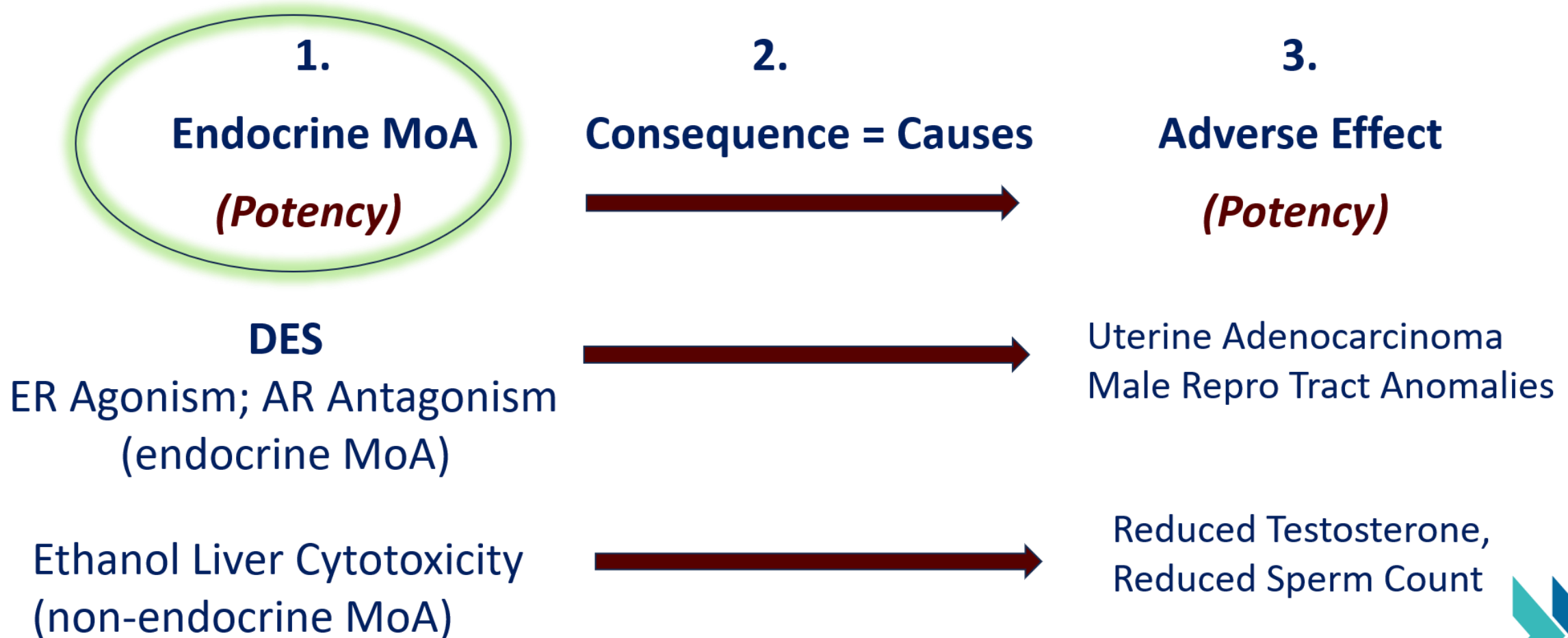
Tier 1: Endocrine Disruptor Screening Program (U.S. EPA)



Borgert, Mihaich, et al., 2011.
Reg Tox Pharm 61:185-191

Definition: “Endocrine Disruptor”

A chemical that produces an adverse effect(s) as a consequence of an endocrine mode of action (MoA).



II. a) How “evidence” will be defined:

- Endpoints from Guideline-like toxicology studies relevant to the hypotheses.
 - OECD Guideline Studies
 - U.S. EPA Guideline Studies

II. b) How to search for and identify the evidence:

- Standardized literature search terms
- Medline, Toxline, Web of Science, Google Scholar, etc.
- Inclusion Criteria
 - Expert Review of abstracts and full-text articles

II. b) How to search for the evidence:

Literature Search Terms for Estrogen Agonist Hypothesis

(estrogen OR androgen OR (uterine AND cytosol) OR (prostate AND cytosol)) AND receptor AND binding AND assay*

estrogen AND receptor AND alpha AND transcription* AND activation AND assay*

aromatase AND assay

steroidogenesis AND assay

uterotrophic AND assay

Hershberger AND assay

(Male OR female) AND pubertal AND assay

male AND intact AND assay

fish AND reproduction AND assay

amphibian AND metamorphosis AND assay

Reproductive (associated keywords with truncation for variation: sperm; spermatozoa;

spermatogen*; infertility; fertility; ovary, ovaries; ovarian; semen; testic*; testis; testes; gonad*)

Developmental (associated keywords with truncation for variation: abnormalities/drug-induced; fetal; foetal; fetus; embryo* ; prenatal; teratogen*)

Borgert CJ (2023) Hypothesis-driven weight of evidence evaluation indicates styrene lacks endocrine disruption potential. *Crit Rev Toxicol*, 53(2), 53-68.

II. b) How to identify the evidence:

Borgert CJ, 2023. *Crit Rev Toxicol*, 53(2), 53-68.

Inclusion Criteria

- Publication in English or English translation available;
- Measurement of in vivo or in vitro endpoints relevant to one or more of the hypotheses evaluated by the WoE methodology used herein;
- Endpoints measured in a study type or experimental model (e.g. regulatory guideline; experimental investigative) amenable to the WoE methodology used herein;
- Appropriate control groups against which endpoint responses to the test chemical could be compared;
- Exposure levels and endpoint responses quantified according to objective or standardized methods (i.e. not reported simply as “positive” or “negative”);
- A statistical method was applied to provide perspective on numerical differences.

II. c) How the quality of the evidence will be evaluated:

1. Evaluate Each Study according to Klimisch Categories (ToxRTool Scores)
 - Category 1: Reliable without restrictions (robust, well-documented studies, e.g., GLP-compliant).
 - Category 2: Reliable with restrictions (generally reliable but with minor limitations).
 - Category 3: Not reliable (significant methodological flaws or incomplete data).
2. Expert Review of each study: design, dosing regimens, statistical analysis, etc.

ToxRTool was developed by the European Commission to assess the reliability of toxicological data according to Klimisch categories 1, 2, or 3, enhancing transparency and consistency in evaluations.

[https://joint-research-centre.ec.europa.eu/scientific-tools-and-databases-0/toxrtool-toxicological-data-reliability-assessment-tool_en]

- Excel Spreadsheet containing Sheets for:
 - Rater identity and qualifications
 - Explanations of the study details
 - In vivo rating: maximum score of 21
 - In vitro rating: maximum score of 18

Study Number	Citation	Study Details	ToxRTool Score	Deficiency & Comment
[3]	Mellert et al., 2007	EB was administered to groups of male and female Wistar rats by gavage for 4 (n = 5/ dose/sex) and 13 weeks (n = 10/dose/sex) (OECD 408) at doses of 0 (vehicle control), 75, 250, and 750 mg/kg bodyweight/day, administered am/pm as half doses.	21 (21)	In the published manuscript, the study results are not transparent and complete for all endpoints investigated. The publication lists histological examination of the epididymides, mammary gland (female only), ovaries, prostate, seminal vesicle, testes, thyroid glands and uterus of the exposed and control animals, but the findings are not described, nor is there a general statement regarding these evaluations. The publication thus warrants a score of 20, but since the original study report contains this information and was made available, the overall score for [3] is 21 for purposes of this WoE evaluation.
12	WoE Evaluation of Styrene	Borgert CJ. 2023.	<i>Crit Rev Toxicol</i> , 53(2), 53-68.	

II. c) How the quality of the evidence will be evaluated:

- Klimisch Scores using ToxRTool

ToxRTool scores were not used to:

- eliminate studies from consideration;
- discount studies for relevance or strength (magnitude) of endpoint responses.

ToxRTool scores were used:

- to interpret the overall quality of the data,
- to interpret and resolve conflicting results between studies.

ToxRTool scores did not differ significantly for studies that reported an effect of styrene versus those that reported no effect of styrene.

- Most studies showed no response for most endpoints;
- Endpoints generally responded only at high doses; other MoAs likely involved.

II. d) How the evidence will be “weighted”:

[NOTE: “weight” means “value or importance”]

- Weight the *Relevance* of the endpoint for each endocrine MoA

Rank 1: Specific and sensitive *for the hypothesis (e.g., estrogen agonist)*

Rank 2: Less specific, but sensitive

Rank 3: Relevant, but corroborative only

- Weight the Strength of the *Response*

- Does the endpoint respond maximally to the chemical?
- What dose of the test chemical is necessary to cause a response, and is it a kinetically relevant dose?

Relevance Weighting of Tier 1 Endocrine Screening Endpoints
(Borgert et al., 2014. *Birth Defects Res B Dev Reprod Toxicol*, 101(1), 90-113.)

II. e) How the evidence will be used to test the hypotheses:

- **Endpoint responses to the test chemical are compared to the endpoint responses for known effectors (positive controls; e.g., hormones and pharmaceuticals).**
- Endpoint responses to the test chemical are compared across the various studies that measured the endpoint.
 - Patterns of response consistent with the known effector support the hypothesis, and inconsistency weakens the hypothesis.
 - A pattern of responses consistent with known effectors (e.g., hormones) supports the hypothesis; inconsistency weakens the hypothesis.
 - High relative potency of the test chemical (compared to the positive control) supports the hypothesis; low relative potency weakens it.
 - Effective doses below the kinetic maximum support the hypothesis; above the kinetic maximum weaken the hypothesis.

Table1. Estrogen Agonist Hypothesis; Guideline Toxicity Studies

Rank	Assay	Endpoint(s)	Expected Response	Respond to Styrene	No Response to Styrene
1	FSTRA	Vitellogenin	↑♂		
	Uterotrophic	Uterus Weight (Blotted or Wet)	↑		[21]
	Uterotrophic	Conversion to Estrus	↑		
	ERTA	Reporter Gene Activation	↑		[22]
	FSTRA	Behavior (sexual, mating)	Δ♂		
		Gonad histopathology	Δ♂		
		Tubercle score	↓♂		
	Female Pubertal	Age and body weight at vaginal opening	↓		
		Age at first estrus	↓		
		Ovary histopathology	Δ		
		Ovary weight	↓		
	Male Pubertal	Testis histopathology (atrophy)	Δ		
		Testis weight	↓		

Borgert CJ (2023) Hypothesis-driven weight of evidence evaluation indicates styrene lacks endocrine disruption potential. *Crit Rev Toxicol*, 53(2), 53-68.

The pattern of responses is critical to understanding MoA.

2

Repeat Dose Toxicity	Epididymis histopathology	Δ		[1c,m][15c,r] [18c,r][19s,m,r]
	Epididymis weight	$\downarrow$		[20n,s,r]
	Mammary histopathology	Δ		[1c,m] [15c,r]
	Ovary histopathology	Δ		[1c,m][2c,r] [15c,r] [18c,r][19s,m,r]
	Ovary weight	$\downarrow$		[1c,m][2c,r] [15c,r][19s,m,r]
	Prostate weight	$\downarrow$		
	Seminal vesicle weight	$\downarrow$		
	Testis histopathology (atrophy)	$\uparrow$	[3,n,r?] [20s,r $\uparrow$]	[1c,m][2c,r][11s,m] [15c,r][18c,r] [19s,m,r]
	Testis weight	$\downarrow$	[3,n,r $\uparrow$] [9s,r $\downarrow$]	[1c,m][2c,r] [10s,r] [11s,m] [15c,r] [18c,r][19s,m,r] [20n,s,r]
	Uterus histopathology	Δ		[1c,m][2c,r] [15c,r] [18c,r][19s,m,r]



Table1. Estrogen Agonist Hypothesis; Guideline Toxicity Studies

Rank	Assay	Endpoint(s)	Expected Response	Respond to Styrene	No Response to Styrene
2	Repeat Dose Toxicity	Uterus weight	↑		[2c,r][19s,m]
		Vagina histopathology	Δ		[19s,m,r]
	Developmental Toxicity	Corpora lutea	↓		
		Post-implantation loss	↑	[16m↑,h↑]	[12r,rb]
		Pre-implantation loss	↑		
	Reproductive Toxicity	Anogenital distance	Δ♂, ♀		
		Corpora lutea	↓		[4F ₁]
		Epididymis histopathology (atrophy)	Δ		
		Epididymis weight	↓		
		Estrous cyclicity	Δ		[4F ₀ ,F ₁]
		Fertility	↓♂, ♀		[2F ₀ ,F ₁ ,F ₂] [4F ₀ ,F ₁]
		Gestational length	↓♀		[4F ₀ ,F ₁]
		Implantations	↓		[4F ₀ ,F ₁]
		Litter size	↓		[2F ₀ ,F ₁ ,F ₂] [4F ₀ ,F ₁]
		Mammary histopathology	Δ♀, ♂		
		Mating index	↓♂, ♀		[4F ₀ ,F ₁]
		Ovarian follicle count in offspring	Δ		[4F ₁]
		Ovary histopathology	Δ		[2F ₁ ,F ₂] [4 F ₁]
		Ovary weight in offspring	↓		[2F ₁ ,F ₂]

Reproductive
Toxicity

Ovarian follicle count in offspring	Δ		[4F ₁]
Ovary histopathology	Δ		[2F ₁ ,F ₂] [4 F ₁]
Ovary weight in offspring	$\downarrow$		[2F ₁ ,F ₂]
Prostate histopathology (atrophy)	Δ		
Prostate Weight	$\downarrow$		
Seminal vesicle weight	$\downarrow$		
Sperm count	$\downarrow$		[4F ₀ ,F ₁]
Testis histopathology (atrophy)	Δ		[2F ₁ ,F ₂]
Testis weight (absolute)	$\downarrow$		[2F ₁ ,F ₂] [4F ₂]
Time to mating	$\uparrow \sigma, \text{♀}$		[4F ₀ ,F ₁]
Time to preputial separation	$\uparrow \sigma$		
Time to vaginal patency	$\downarrow$		[5F ₂]
Uterus histopathology	Δ		[2F ₁ ,F ₂] [4F ₁]

Table1. Estrogen Agonist Hypothesis; Guideline Toxicity Studies					
Rank	Assay	Endpoint(s)	Expected Response	Respond to Styrene	No Response to Styrene
2	Reproductive Toxicity	Uterus weight in offspring	↑		[4F ₂]
		Vagina histopathology	Δ		
3	ERBA	Displacement of Estradiol	↑		[21] [22]
	FSTRA	Behavior	Δ		
		Estradiol level	↓♀		
		Fecundity	↓		
		Fertilization success	↓♀		
		Follicular Atresia	↑		
		Gonad- somatic Index	↓♂, ↑♀		
		Testosterone level	↓♂		
	Female Pubertal	Estrous cyclicity	↑		
		Growth	↑		
	Male Pubertal	Epididymis histopathology	Δ		
		Growth	Δ♂		
		Ventral prostate weight	Δ♂		
	Steroidogenesis	Estradiol level	↑		
	Repeat Dose Toxicity	Gross pathology	Δ♂, ♀	[18c,r]	[1c,m] [15c,r]
♂= males; ♀=females; ↑-increase relative to controls; ↓-decrease relative to controls; ?-altered but not as expected. Numbers correspond to numbered citations in Reference List; r =rat; m =mouse; rb=rabbit; s = subchronic; c =chronic; n = non-guideline; F ₀ =F ₀ generation; F ₁ = F ₁ generation; F ₂ = F ₂ generation; b=bolus; i=infusion; h=Chinese hamster; H=Human					

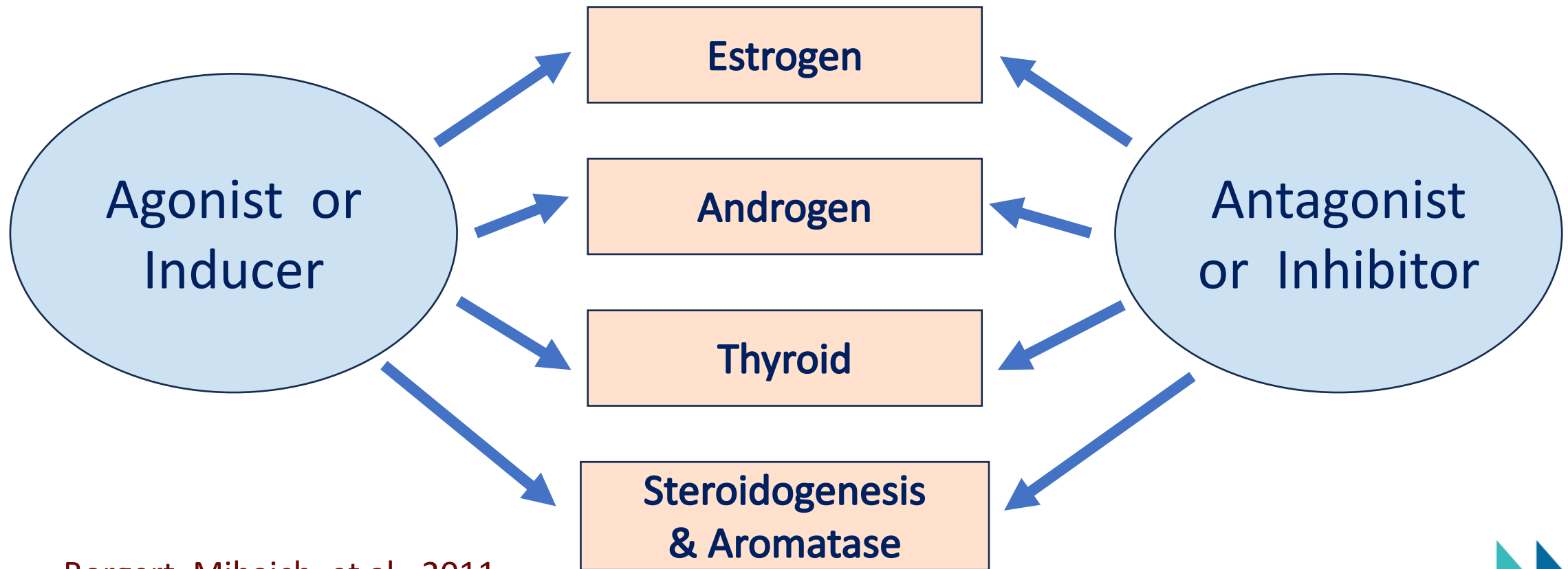
II. f) How overall conclusions will be reached:

Consider the totality of the evidence:

- Data gaps and sufficiency for interpretation,
- Pattern of effects,
- Strength of the responses,
- Doses at which responses occur,
- Alternative, more plausible MoAs for the endpoint responses?

. . . Construct similar tables for each hypothesis

Tier 1: Endocrine Disruptor Screening Program (U.S. EPA)



Borgert, Mihaich, et al., 2011.
Reg Tox Pharm 61:185-191

WoE Results for EATS Modes of Action: Percentage of Endpoint Responses to EB

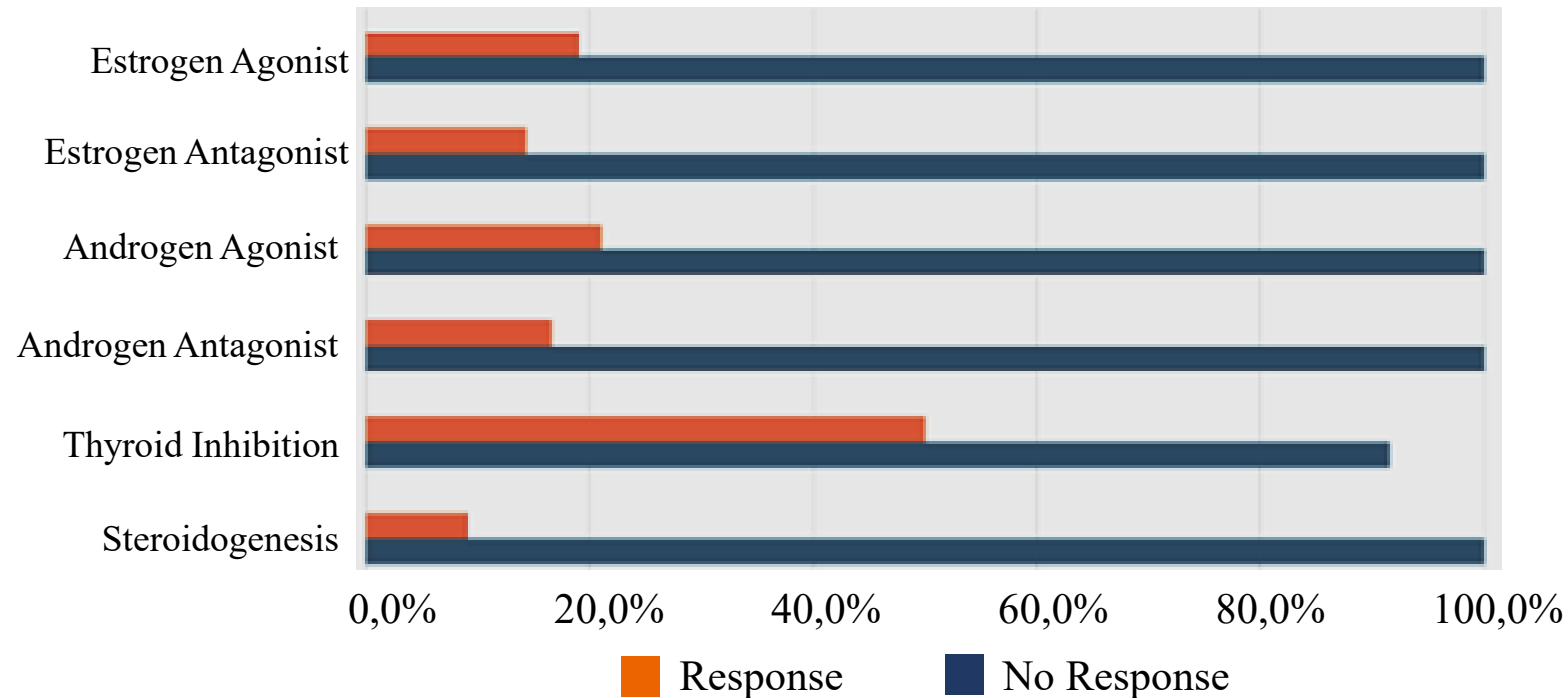


Figure 1: Graphical abstract

Original article:

**HYPOTHESIS-DRIVEN WEIGHT OF EVIDENCE EVALUATION
INDICATES ETHYLBENZENE LACKS ENDOCRINE DISRUPTION
POTENTIAL BY EATS PATHWAYS**

Christopher J. Borgert^{id}, PhD

Applied Pharmacology and Toxicology Inc, Gainesville FL, 32605 and
University of Florida College of Veterinary Medicine, Dept. Physiological Sciences,
Gainesville FL, 32610. Tel.: +1 352-219-8551

<https://dx.doi.org/10.17179/excli2024-7822>

This is an Open Access article distributed under the terms of the Creative Commons Attribution License
(<http://creativecommons.org/licenses/by/4.0/>).

CRITICAL REVIEWS IN TOXICOLOGY, 2017
<http://dx.doi.org/10.1080/10408444.2016.1269722>

REVIEW ARTICLE

**Hypothesis-driven weight-of-evidence analysis of endocrine disruption potential:
a case study with triclosan**

Ellen Mihaich^a, Marie Capdevielle^b, Daniella Urbach-Ross^b and Brian Slezak^b

^aEnvironmental and Regulatory Resources, LLC, Durham, NC, USA; ^bColgate-Palmolive Company, Piscataway, NJ, USA

CRITICAL REVIEWS IN TOXICOLOGY, 2017
<http://dx.doi.org/10.1080/10408444.2016.1272094>



REVIEW ARTICLE

OPEN ACCESS

**Weight-of-the-evidence evaluation of 2,4-D potential for interactions with the
estrogen, androgen and thyroid pathways and steroidogenesis**

B. H. Neal^a, J. Bus^a, M. S. Marty^b, K. Coady^b, A. Williams^a, J. Staveley^a and J. C. Lamb^a

^aExponent[®], Alexandria, VA, USA; ^bToxicology & Environmental Research and Consulting, The Dow Chemical Company, Midland, MI, USA



OPEN ACCESS





Kinetically-derived maximal dose (KMD) indicates lack of human carcinogenicity of ethylbenzene

Lyle D. Burgoon¹ · Christopher J. Borgert²  · Claudio Fuentes³ · James E. Klaunig⁴

Received: 5 May 2023 / Accepted: 7 November 2023 / Published online: 7 December 2023
© The Author(s) 2023

- Endocrine endpoints affected (Borgert, 2025) were likely secondary to chemical overload and liver/kidney toxicity (Burgoon et al., 2024)
- All endocrine-related ToxCast / Tox21 data for ethylbenzene = QC Omit

Problematic Issues:

Check-Box Approaches Lack Rigor and Transparency!

- “Key Characteristics” approaches are Unreliable and Unrepeatable
 - Borgert CJ (2023) Issue analysis: key characteristics approach for identifying endocrine disruptors. *Archives of Toxicology*, 97(10), 2819-2822.
 - Becker RA, Dreier DA, Manibusan MK, Cox LAT, Simon TW, Bus JS (2017) How well can carcinogenicity be predicted by high throughput “characteristics of carcinogens” mechanistic data. *Regul Toxicol Pharmacol*, 90, 185-196.

High Throughput Screening Data (HTS)

- Data Quality of ToxCast / Tox21
 - ICE (Integrated Chemical Environment) website provides data quality information

In Summary

Hypothesis-Based WoE Framework

- Uses explicit relevance rankings of endpoints derived *a priori*
 - hypothesis-based versus hypothesis-generating
- Transparent, objective and consistent interpretation of results for each hypothesis
- Systematic method to resolve inconsistencies in the data and focus on the most definitive information
- Can be applied to other modes of action and both human health and ecological data as well
- Can be updated and refined with new information

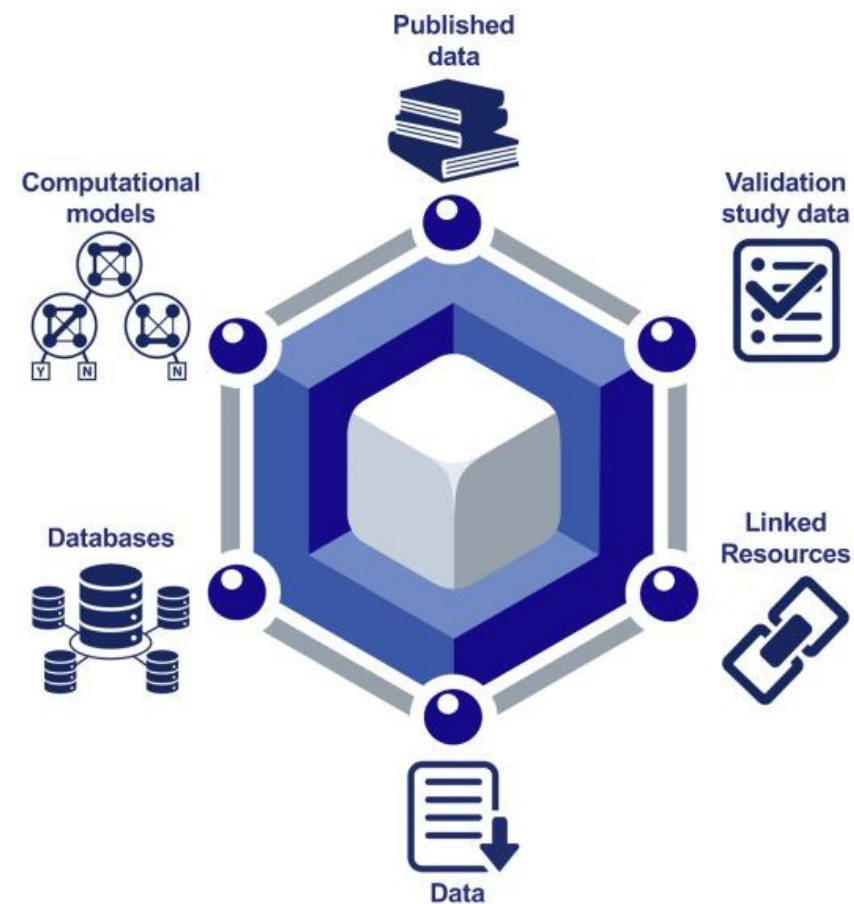
Many Thanks for Your Attention! Questions ?



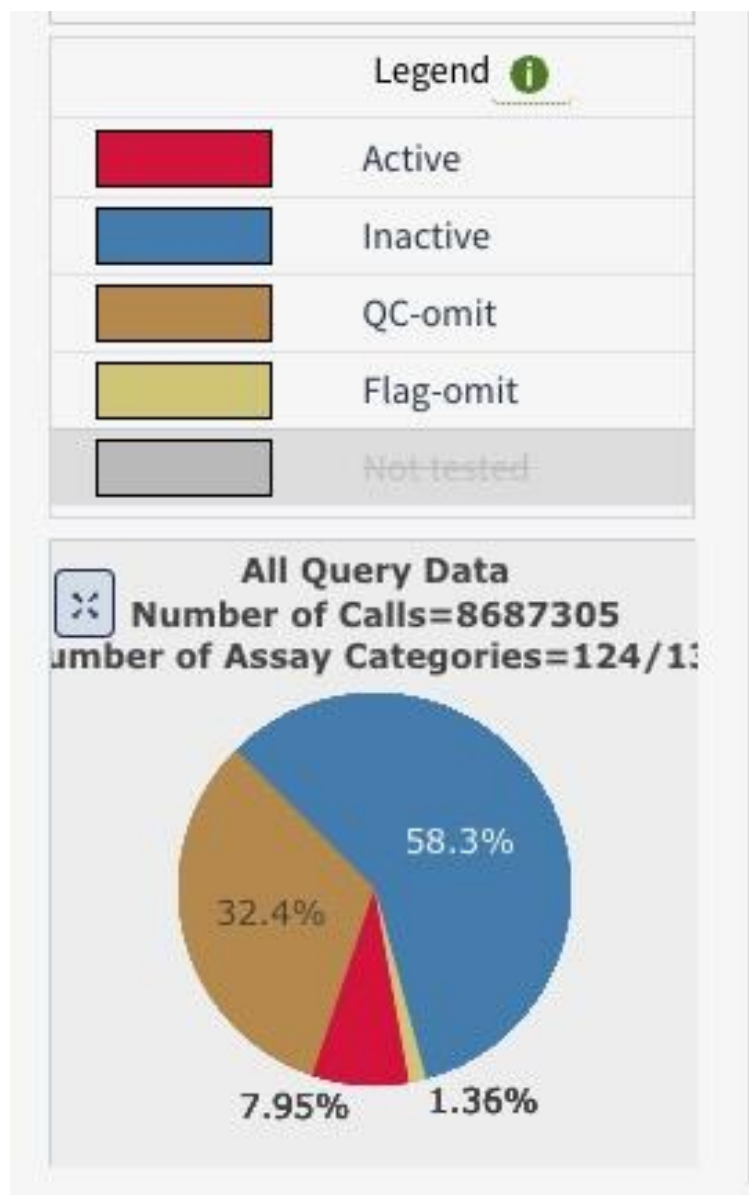
ICE (Integrated Chemical Environment)

Data quality information for ToxCast / Tox 21 data.

- U.S. National Toxicology Program/NICEATM
- Data and tools for evaluation of results from new testing approaches
- Curated in vivo and in vitro toxicity and phys/chem data, along with curated high-throughput screening data (cHTS)
 - Data quality focused
 - Minimizes use of low-confidence results



ICE Result: Proportion of Data Omitted for Poor Quality Control All In Vitro Results



Ac, Bc Cc	Low Concentration (5-30% of expected conc)
D	Purity <50%
Fc	Very Low Concentration (<5% of expected conc) (Biological Activity Unreliable)
F	Incorrect MW (Biological Activity Unreliable)
Fns	No Sample Detected (Biological Activity Unreliable)

D4 is NOT Estrogenic or an Endocrine Disruptor

Archives of Toxicology
<https://doi.org/10.1007/s00204-024-03896-y>

MOLECULAR TOXICOLOGY

Octamethylcyclotetrasiloxane (D4) lacks endocrine disruptive potential via estrogen pathways

Christopher J. Borgert¹ · Lyle D. Burgoon²

- Key estrogenic effects on developing males are *absent!* – *wrong pattern!!*
- Endocrine-sensitive endpoints affected above the KMD.
- QSAR shows excessive steric hindrance.
- Alternative MoAs can explain effects.
- Endocrine ToxCast / Tox21 data = QC Omit

Archives of Toxicology (2024) 98:1795–1807
<https://doi.org/10.1007/s00204-024-03723-4>

MOLECULAR TOXICOLOGY

The physiological and biochemical basis of potency thresholds modeled using human estrogen receptor alpha: implications for identifying endocrine disruptors

Christopher J. Borgert¹ · Lyle D. Burgoon² · John C. Matthews³

Received: 5 November 2023 / Accepted: 29 February 2024 / Published online: 5 May 2024
© The Author(s) 2024

Archives of Toxicology
<https://doi.org/10.1007/s00204-024-03914-z>

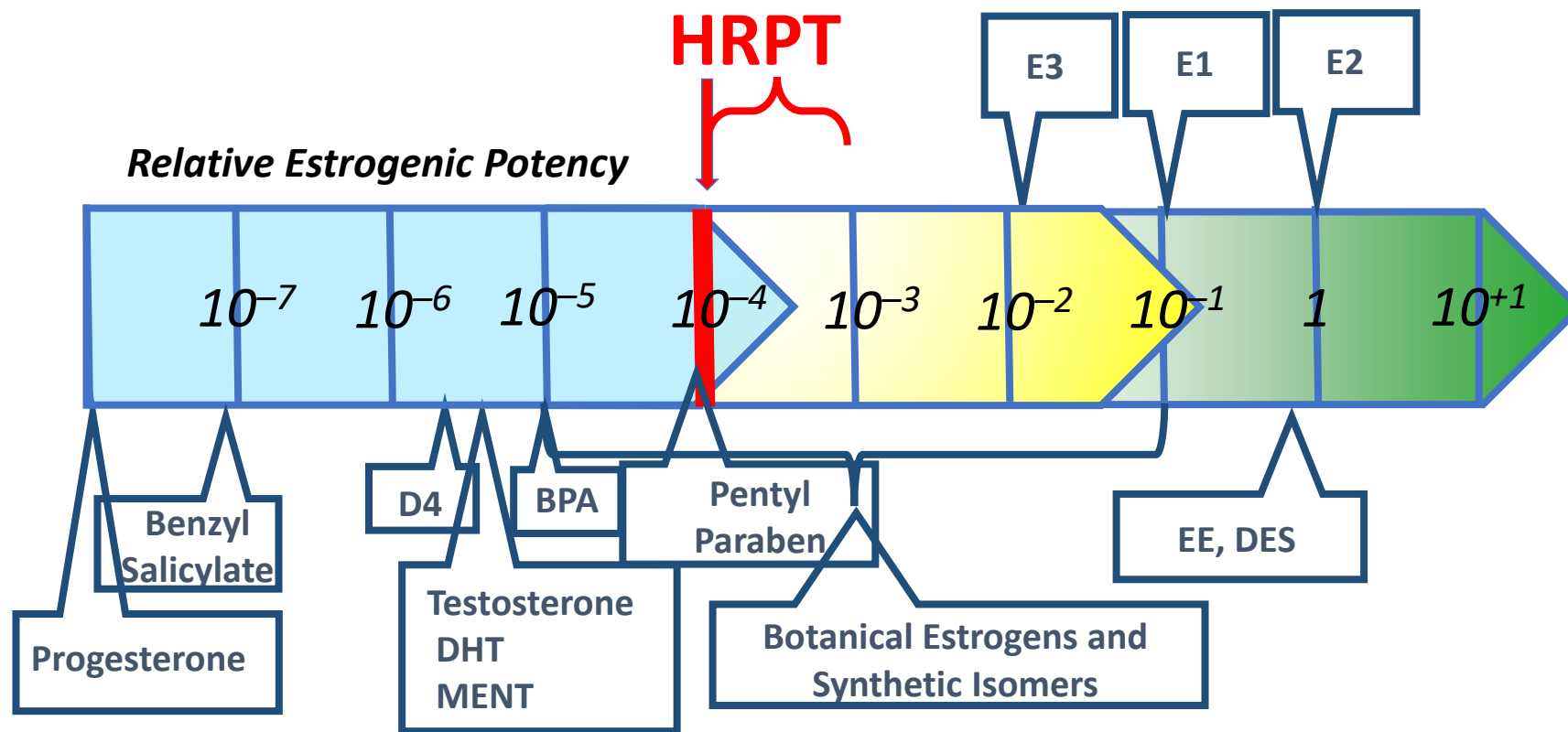
TOXICOKINETICS AND METABOLISM

Kinetically-derived maximal dose (KMD) confirms lack of human relevance for high-dose effects of octamethylcyclotetrasiloxane (D4)

Christopher J. Borgert¹ · Lyle D. Burgoon² · Claudio Fuentes³

Figure 1. Revised Human-Relevant Potency-Threshold (HRPT) for the ER α -Agonist MoA

Figure 1 was adapted from Borgert et al. (2018) showing the conservatism in the threshold region beyond the originally proposed HRPT to as high as 10^{-3} relative to E2, in accordance with receptor occupancy calculations shown in Table 2. Also added to the original figure by Borgert et al. (2018) are potency estimates for three chemicals incorrectly alleged to be exogenous estrogens. Potency data for benzyl salicylate and BPA are from Natsch et al. (2012) and data for parabens are from Fayyaz et al. (2021).



Borgert, Burgoon, Matthews, 2024. *Arch Toxicol*, 98(6), 1795-1807.