Birth Defects Research (Part B) 101:114–124 (2014)

Research Article

Interpreting Estrogen Screening Assays in the Context of Potency and Human Exposure Relative to Natural Exposures to Phytoestrogens

Richard A. Becker,[1]* Sean M. Hays,[2] Christopher R. Kirman,[3] Lesa L. Aylward,[4] and Kimberly Wise[1]

[1]American Chemistry Council, Washington, District of Columbia
[2]Summit Toxicology, LLP, Lyons, Colorado
[3]Summit Toxicology, LLP, Orange Village, Ohio
[4]Summit Toxicology, LLP, Falls Church, Virginia

While the Environmental Protection Agency and the Organization for Economic Cooperation and Development have developed validated *in vitro* and *in vivo* screening assays to measure interaction of substances with estrogen, androgen and thyroid pathway components, to date, methods to contextualize such results in terms of potencies and actual human exposures are lacking. To place endocrine screening results in the context of potency and human exposure, we propose a method that entails (1) calculating a benchmark dose for a response measured in an endocrine screen; (2) estimating the human urinary concentration (biomonitoring equivalent, BE) expected to correspond to this dose (BE_{BMD}); (3) deriving the exposure:activity ratio (EAR) by comparing actual urinary values from human biomonitoring studies (e.g., National Health and Nutrition Examination Survey (NHANES)) to the BE_{BMD}. Using OECD uterotrophic assay validation studies and NHANES results, we calculated EARs for genistein ($EAR_{GEN} = 6.6 \times 10^{-4}$) and bisphenol A ($EAR_{BPA} = 8.8 \times 10^{-7}$). The EAR_{GEN} is more than 700-fold greater than the EAR_{BPA}. Not only can these methods be applied to additional endocrine assays and compounds, they can contribute to weight of evidence decisions regarding the need for additional endocrine screening and testing—substances with low EARs may not warrant additional testing. *Birth Defects Res (Part B)* 101:114–124, 2014.

Key words: ***endocrine disruption; risk assessment; biomonitoring***

INTRODUCTION

Alterations in hormonal homeostasis from possible interactions between chemical exposures and the estrogen receptor have been the focus of extensive toxicologic research and regulatory investigation. For example, the U.S. Environmental Protection Agency (EPA) has been engaged in developing the Endocrine Disruptor Screening Program (EDSP) for more than a decade. Under the EDSP, EPA has developed a Tier 1 screening battery (EDSP-ESB) consisting of 11 different assays, including both *in vitro* and *in vivo* methods, for screening chemicals to identify substances with the potential to interact with components of the estrogen (E), androgen (A), and thyroid (T) systems (EPA, 2013a). Through a global collaborative effort, complementary test guidelines for endocrine screening assays have also been developed by the Organization for Economic Cooperation and Development (OECD, 2013a,2013b). While the endpoints in the EDSP-ESB and the corresponding OECD test guidelines are not considered to be adverse effects, these assays have been developed to be highly sensitive to detect interactions of chemicals with components of the EAT pathways. Compared with *in vitro* assays, the *in vivo* EDSP-ESB assays have the advantage that they include *in vivo* physiologic systems, integrated biologic responses, and intact *in vivo* metabolic systems, including intestinal and first-pass metabolism (Borgert et al., 2003; Charles et al., 2007). Among the *in vivo* EDSP-ESB assays, the rat uterotrophic assay has proven in the joint OECD-EPA validation studies to be a reliable and reproducible test, and this assay allows for a comparison across selected compounds of relative estrogenic potency on an administered dose basis (Kanno et al., 2001, 2003). However, as noted above, this assay has been optimized as a screening-level method, and does not provide information of the potential health effects, if any, of exposure to chemicals at low, environmentally relevant levels. EPA's EDSP was launched in earnest in 2009, with EPA requiring EDSP-ESB screening

*Correspondence to: Richard A. Becker, Regulatory and Technical Affairs, American Chemistry Council, 700 2nd Street, NE, Washington, DC 20002. E-mail: Rick_Becker@americanchemistry.com

Received 08 October 2013; Accepted 08 November 2013

Published online in Wiley Online Library (wileyonlinelibrary.com/journal/bdrb) DOI: 10.1002/bdrb.21085

of 50 pesticide active ingredients and 2 pesticide inerts (see http://www.epa.gov/endo/). Screening of this first set of substances is complete, and results are undergoing review and analysis by EPA. EPA is now poised to require EDSP screening of a second set of more than 100 substances, consisting of both pesticides and commodity chemicals (EPA, 2013b). The chemicals selected for these first two rounds of EDSP screening were identified by EPA based on consideration of potential for exposure, not potential for endocrine activity or toxicity.

In EPA's EDSP, substances that result in positive responses in the EDSP-ESB, as determined by an overall weight of the evidence evaluation, are candidates for more definitive testing in multigenerational tests, in one or more taxa (mammalian, avian, amphibian, invertebrate, etc.), to determine potential adverse effects and dose–response for use in risk assessment. As with any screening assay, or battery of assays that will be used for prioritization (in this context prioritization refers to further testing), it is important to understand what "triggers" a response worthy of further testing. To date, little attention in the EDSP has focused on the use of exposure information to inform decision-making with respect to selection of substances for Tier 2 testing. Methods that can place response levels in high throughput screening assays in an exposure context (Aylward and Hays, 2011; Wetmore et al., 2012; Wambaugh et al., 2013) have been developed, and we draw from these in this article to propose methods whereby quantitative measures of human exposure can be used to assist in EDSP decision making. This involves determining the "relevant" response(s) level in a screening assay and assessing how to interpret the margin between the dose leading to this response level and actual human exposures with respect to a decision of whether further testing is, or is not, warranted. In the process of developing this approach, we have identified other issues that merit further consideration.

While concern in the regulatory arena has focused primarily on assessing the potential for synthetic chemicals to interact with components of the EAT endocrine systems, a wide variety of compounds naturally occurring in the diet, such as polyphenols, also can interact with the estrogen receptor. More than 500 polyphenol compounds in the diet, including flavonoids, phenolic acids, stilbenes, and lignans, have been characterized (Pérez-Jiménez et al., 2010). Epidemiologic studies have linked polyphenols in the diet with health protective effects, including reduced rates of cardiovascular disease (Basli et al., 2012; Khurana et al., 2013; Kishimoto et al., 2013). Many of these compounds also have demonstrable estrogenic activity in various assay systems, both *in vitro* and *in vivo* (Kanno et al., 2001, 2003). In addition, compounds with potent interactions with the estrogen receptor (notably estrone, 17-β-estradiol, and estriol) are of course produced endogenously, and endogenous production varies widely over life stages and within and between individuals. Furthermore, synthetic estrogens used in contraceptives and endogenous estrogens are incompletely removed by publicly owned treatment works, and even though these substances have been widely detected in numerous studies of water bodies in the United States and Europe (Caldwell et al., 2010) estimates demonstrate a wide margin of exposure (MOE) for children compared to exposures to naturally occurring estrogens in milk (Caldwell et al., 2010).

Interactions between chemicals (endogenous, natural dietary, synthetic chemicals, and pharmaceutics) and the estrogen pathway are extremely complex. The biologic system in humans includes multiple estrogen receptors, numerous feedback loops, homeostatic regulation, interactions with other steroid hormone receptors, and varies with life stage and other physiologic factors. Variations in hormone levels occur normally, and perturbations outside of the range of normal homeostasis by exogenous chemicals (whether natural dietary or synthetic) are likely to require the exceedance of a threshold dose and will be dose dependent (Borgert et al., 2013).

Direct interpretation of phytoestrogen intakes in humans and human effects is challenging. Overall, while it appears that dietary phytoestrogen exposures benefit human health (Tham et al., 1998), in the context of epidemiology and risk assessment, testing for correlations between polyphenol intake and disease risks has been complicated by the errors associated with estimating polyphenol intake rates among individuals and potential bias and inaccuracies associated with dietary recall (Pérez-Jiménez et al., 2010). Biomonitoring for polyphenol exposures via measurement of parent compounds or metabolites in urine has been advocated as a more reliable means of assessing intakes (Pérez-Jiménez et al., 2010). However, although estimates of the relative estrogenic potency for many of these compounds are available on both an *in vitro* and *in vivo* basis (see Borgert et al., 2003), there is no direct ability to interpret the measured levels of phytoestrogens and other potentially estrogenic compounds detected in urine in terms of the doses required to induce a relevant *in vivo* estrogenic response.

APPROACH FOR CONTEXTUALIZING ENDOCRINE SCREENING RESULTS IN TERMS OF POTENCY AND HUMAN EXPOSURE

Although numerical values of urinary concentrations of different chemicals can be compared on a mass per volume basis, such comparisons are not biologically meaningful because substances differ in absorption, distribution, metabolism, excretion, and in potency to interact with receptors; a receptor response is dependent upon dose of the active moiety at the receptor and potency of the molecule. We propose a method for contextualizing endocrine screening assay results in terms of potency and human exposure that entails:

(1) Calculating a biologically relevant dose for a specific response measured in an endocrine screening assay (e.g., a benchmark dose [BMD]);
(2) Estimating the human urinary concentration (biomonitoring equivalent [BE]) expected to correspond to this dose (BE_{BMD}, defined as "activity");
(3) Deriving the exposure:activity ratio (EAR) by comparing measured urinary concentrations from human biomonitoring studies, such as NHANES to the BE_{BMD}; and
(4) To assist with interpreting the significance of endocrine screening assay results, comparing the EAR

for a given substance to an EAR for a reference chemical.

For substances that are rapidly metabolized and excreted into the urine, to translate a biologically relevant dose of an EDSP-ESB assay into a concentration in urine corresponding to that exposure, we draw upon the framework of BEs (Hays et al., 2007, 2008). BEs are estimates of the concentration of a chemical or its metabolite in a biological medium (usually blood or urine) that is consistent with steady-state exposure at an exposure level of interest, for example, a reference dose, tolerable daily intake, a cancer risk-specific dose (RSD), or, in this case a BMD for estrogenic responses. That is, BE values are the biomarker concentration in urine or blood *equivalent* to a specified external dose level. While conventionally BE values have been derived to correspond to a reference dose or tolerable daily intake, BE values have also been derived corresponding to a point of departure dose level (reviewed in Hays et al., 2008). In this analysis, to illustrate the application of these approaches, BE values corresponding to benchmark doses (BE_{BMD} values, also defined as "Activity" in the EAR) for the estrogen receptor mediated proliferation of rat uterine tissue are calculated and presented. These BE_{BMD} values are then converted to the corresponding concentrations in urine, which can be compared to actual urinary levels measured in NHANES (or other human biomonitoring studies) to understand how human exposures compare to the actual doses that produce uterotrophic responses in the screening assay.

These calculations require two major types of data for each compound in the analysis:

(1) Estimates of the urinary excretion fraction in humans for each compound. That is, at steady state in humans, what proportion of administered dose is excreted in urine as the analyte of interest?
(2) Estimates of benchmark doses across all of the compounds of interest for an estrogen-sensitive response in an *in vivo* experimental system that relies upon an oral exposure route.

Given these two pieces of information, a steady-state urinary concentration consistent with the benchmark dose (BE_{BMD}) can be estimated for each compound of interest. While urinary excretion fraction is used in this example, physiologically based pharmacokinetic (PBPK) models, when available, can also be used.

What is a response and an appropriate degree of response in an endocrine screening assay that is relevant and useful for this approach? By design, the EDSP-ESB assays were selected and optimized to be highly sensitive, to err on the side of producing false positives (EPA, 1998), and the endpoints measured are not adverse effects per se, but instead indicative of the potential of a chemical to interact with one or more components of E, A, or T pathways. Therefore, a positive response in an endocrine screening assay does not indicate the potential for an actual positive response *in vivo* in apical endpoints in EDSP Tier 2 test systems or in humans or wildlife. This raises the question, what is an appropriate response level of benchmark response (BMR) in a screening assays? A default convention for a continuous response metric is to use 1 SD from control as a BMR (EPA, 2012a), and we have used this 1 SD response level in this article.

To illustrate application of this approach for contextualizing endocrine screening results in terms of potency and human exposure, we calculated BMDs for two dietary phytoestrogen compounds, genistein (GEN) and daidzein (DZN), and for one synthetic chemical, bisphenol A (BPA), using dose–response data from the rat uterotrophic screening assay. We then used available urinary excretion fraction data for these compounds to calculate the corresponding urinary steady-state concentration—the $BE_{Ut\text{-}BMD}$ that we define here as Activity—for each compound. This allowed us to compare the measured urinary concentrations in NHANES for the 2009 to 2010 sampling cycle to the $BE_{Ut\text{-}BMD}$ values for each compound to calculate an EAR. In addition, we propose a metric to help interpret the significance of endocrine screening results, potency, and exposure for a specific compound by comparing the EAR_{CMPD} to the EAR of a reference substance. To illustrate this, we use GEN as the reference agent and derived the relative endocrine exposure activity quotient ($REEAQ = EAR_{CMPND} \div EAR_{GEN}$). For example, to assist in weight of evidence decision making and priority setting for more advanced (Tier 2) testing in EPA's EDSP, an REEAQ of 1 could be interpreted to indicate an effective exposure and activity level equivalent to that predicted for typical GEN exposure levels in a normal diet in the U.S. population; an REEAQ larger than 1 indicates greater effective exposure and activity levels than that associated with typical GEN exposures (the larger the value, the higher the priority); as REEAQ decreases below 1, the value indicates decreasing relative exposure and activity levels than that experienced from typical GEN exposures. As dose–response data for additional compounds are generated in the rat uterotrophic assay, or other relevant endocrine assays, and urinary excretion fraction data or sufficient pharmacokinetic information are available or can be estimated, EAR and REEAQ values can be developed for these compounds too.

METHODS

Briefly, the approach here is as follows:

(1) A relatively robust and validated response of interest in an oral, *in vivo* assay in mammals from the commonly used EDSP-ESB was identified (uterine weights in the rat uterotrophic assay);
(2) A consistent response level (1 SD from control mean response) was identified as the BMR level of interest, and benchmark doses (BE_{BMD}), also defined as "Activities" were calculated for each compound of interest;
(3) Urinary excretion fraction (the fraction of ingested compound excreted as the measured biomarker) data were identified for each compound of interest from published literature;
(4) Urinary steady-state concentrations of the biomarker for each compound of interest corresponding to its benchmark dose (BE_{BMD}) were calculated;
(5) EAR_{CMPND} were calculated for each compound (CMPND) by comparing population urinary concentrations from the U.S. NHANES dataset to the respective Activity (BE_{BMD}) values, to enable interpreting

results of endocrine screening in both the context of potency and human exposure.

Dose–Response Data

Review of the available data on both pharmacokinetics and screening for estrogenic activity of phytoestrogen compounds highlights several issues that will impact efforts to develop urinary benchmarks for interpreting urinary concentrations in terms of dose levels that lead to estrogen receptor-mediated proliferation of rat uterine tissue. One issue is that the relative activity of these compounds *in vivo* may be quite different from that observed *in vitro* (Borgert et al., 2003, Charles et al., 2007). These differences in activity likely result from several factors, including *in vivo* metabolism (both Phase I and Phase II) that results in differences across compounds in proportion of free, active compound available to produce estrogenic effects *in vivo*, compared to the situation observed in assays conducted *in vitro*, which often do not include a full metabolic capability. In particular, these *in vitro* assays do not capture intestinal conjugation that occurs following dietary exposure and which decreases the free compound available for *in vivo* activity. As a result, only data from assays employing an oral route of exposure in intact animals are considered here.

Based on this criterion, the orally administered rat uterotrophic assay was identified as likely to provide the most robust available dataset. However, even though this assay figures prominently in current endocrine disruptor screening efforts, robust dose–response data in the uterotrophic assay for the NHANES phytoestrogen analytes were available only for GEN (see, e.g., Kanno et al., 2003; Carbonel et al., 2011), equol (reviewed in [Schwen et al., 2012]), and a reference compound, ethinyl estradiol (Kanno et al., 2001). More limited data are also available for daidzein (DZN) (Diel et al., 2000; Rachon et al., 2007).

There are two versions of the uterotrophic assay, one involving immature animals and the other using mature, ovariectomized animals. In both versions, dosing of chemical lasts at least 3 days, and absolute and relative increases in uterine weight are used as response metrics. While it has been suggested that the assay using immature rats is preferred to the mature/ovariectomized rats for animal welfare reasons (OECD, 2007) the dose–response achieved from the two versions appears compatible (Carbonel et al., 2011), and EPA currently prefers the ovariectomized rat assay for the EDSP (EPA, 2009).

Robust oral dose–response data in the rat uterotrophic assay are available for a range of synthetic compounds that have been assessed for potential estrogenic activity, including BPA, nonylphenol, octylphenol, methoxychlor, several phthalate esters, parabens, and o,p′-DDT and ethinyl estradiol (see, e.g., the OECD validation studies described by Kanno et al., 2003). For the purposes of this analysis, a BE_{BMD} value for BPA is included for comparison to the values calculated for GEN and DZN.

The rat uterotrophic assay dataset for GEN (Kanno et al., 2003) offers robust datasets for BMD modeling. In these studies, multiple laboratories ran identical studies under three protocols: intact juvenile animals following oral exposure or subcutaneous injection (protocols A and B, respectively), or in ovariectomized adult animals using subcutaneous injection (protocol C) to assess the reproducibility of the assays and dose–response results. Because the exposure route of interest for human exposure is via oral exposure, the data from protocol A utilizing intact juvenile animals and dosing via the oral route from Kanno et al. (2001 and 2003) were selected for the estimation of oral BMD values. The wet uterine weights were tabulated for each dose and laboratory. The response data were expressed in terms of percent control for each study. For each dose group, pooled arithmetic mean and SD response values were calculated across studies.

Pooled mean response levels as percent of controls for each tested dose across all studies included in Kanno et al. (2001, 2003) for the analytes of interest were calculated using equation (1):

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \ldots n_i\bar{x}_i}{n_1 + n_2 + \ldots n_i} \quad (1)$$

Similarly, pooled SDs were calculated using equation (2):

$$S^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots (n_i - 1)s_i^2}{(n_1 + n_2 + \ldots n_i) - i} \quad (2)$$

Benchmark dose values for GEN and BPA were calculated using the pooled means and SDs.

The available dose–response data for DZN were less robust. Two studies were available that used the ovariectomized adult rat via oral exposure; however, no datasets from the immature rat model via oral exposure were identified (Diel et al., 2000; Rachon et al., 2007). Rachon et al. (2007) used a subchronic dietary administration protocol that was quite different from the OECD standard protocol used by Kanno et al. (2001, 2003). Diel et al. (2000) used a 3-day oral gavage administration protocol that was more consistent with the OECD protocol. As a result, dose–response data from this study were selected for benchmark dose modeling for DZN.

U.S. EPA Benchmark Dose Modeling Software (version 2.4) was used to model each dataset. A 1 SD change was used as the BMR rate. The BMD and lower-bound confidence limit on the BMD (BMDL) were calculated using a range of dose–response models. The best-fitting model was chosen in each case as determined by visual inspection, and statistical evaluation criteria including Akaike information criteria, and chi-square goodness-of-fit p values. Because the purpose of the analysis is to compare relative EARs across compounds, the change of 1 SD from control means provides a consistent benchmark that does not rely upon an assumption that the estrogen receptor agonist exhibit full activity or the same maximal activity. The BMD, rather than BMDL, was used as the basis for comparison across compounds since the BMD provides the best estimate of the dose associated with the BMR for each compound, while the BMDL reflects the statistical uncertainty associated with the datasets.

Urinary Excretion of Biomarkers and Activity (BE_{BMD}) Calculation

Both DZN and GEN are biomonitored in humans by measuring parent compound in urine. Deriving BEs for compounds excreted in urine requires estimating the steady-state excretion of compound(s) in urine using a mass-balance approach (Aylward and Hays, 2008). The BE is calculated by estimating the steady-state urinary concentration of the compound of interest associated with ongoing exposure at the target exposure level. In this effort, the BMD (rather than the BMDL) was selected as the dose level of interest for each compound. This reflects the interest in this effort in examining the relative potency of each compound compared to one another. Given this focus, the best estimate of the BMD, rather than a lower uncertainty bound (which is impacted by study design factors), was of greatest interest. The calculation of the BE_{BMD} relies upon an estimate of the fraction of an oral dose excreted in urine (F_{UE}—fraction of urinary excretion) as the measured analyte(s) (parent compound and/or metabolite(s)) and assumptions regarding daily urine volume or creatinine excretion rate.

$$BE = \frac{BMD \times F_{UE}}{V_{24}\, or\, Cr_{24}} \quad (3)$$

For this evaluation, data on the urinary excretion fraction for each compound in humans are required. For both GEN and DZN, numerous studies have been identified that provide estimates of F_{UE} in humans (fraction of parent compound excreted as GEN and DZN). A recent review by Pérez-Jiménez et al. (2010) provides an extensive overview of the available datasets. The F_{UE} estimates tabulated by Pérez-Jiménez are only for 24-hr collections of urine. However, the phytoestrogens of interest have short half-lives (e.g., DZN, 4.7 hr, GEN, 5.7 hr) and 24-hr collections are likely sufficient to capture full urinary excretion fractions.

In vivo, the active moiety in binding to the estrogen receptor is the parent BPA molecule. But BPA is rapidly metabolized following oral exposure by glucuronidation and sulfation, and these conjugates are excreted in the urine. The NHANES technique for quantifying BPA in urine includes a deconjugation step to generate free BPA before quantification, and thus NHANES urinary BPA results reflect total BPA (CDC, 2009). The urinary excretion fraction of BPA (parent BPA plus deconjugated BPA metabolites) has also been determined in several controlled human dosing studies and is determined to be 1.0 (all orally administered BPA is recovered in urine within approximately 24 hr—reviewed in [Krishnan et al., 2010]). The F_{UE} estimates for GEN, DZN, and BPA are compiled in Table 1.

Table 1
Weighted Average and Range of Fraction of Urinary Excretion (F_{UE}) Estimates for DZN and GEN (Reviewed in Pérez-Jiménez et al., 2010) and BPA (Tham et al., 2005; Tinwell et al., 2008)

Compound	Average F_{UE} (%)[a]	Range (%)
DZN	37	8.3–60
GEN	12	1.7–47
BPA	100	-

[a]Pérez-Jiménez et al. (2010) reported a weighted mean recovery percentage in urine across available studies, with weighting based on number of subjects included in each study.

Table 2
Benchmark Dose (BMD) Modeling Results for Dose–Response Datasets

Compound	Best-fit model	BMD (BMDL) mg/kg-day
GEN	Exponential	10.1 (6.6)
DZN	Hill	99 (31)
BPA	Exponential	59 (33)

All BMD values correspond to a response level of a 1 SD change from the control mean.

Human Biomonitoring Data (NHANES)

The Centers for Disease Control and Prevention (CDC) has been generating biomonitoring data for a range of polyphenols and synthetic compounds as part of the NHANES program. The polyphenols biomonitored for in urine in NHANES include GEN, DZN, equol, o-desmethylangolensin, enterolactone, and enterodiol. NHANES also reports urinary concentrations of synthetic compounds including BPA and various other compounds potentially of interest. CDC provides summaries of these data as a function of age and gender, and percentiles for the distribution of measured values among each group are provided.

Calculation of EARs

The EAR, the ratio of the measured general population urinary concentration reported in NHANES to the BE_{BMD}, is calculated as follows for each analyte of interest:

$$EAR = \frac{[\text{Biomarker concentration}]}{BE_{BMD}} \quad (4)$$

Comparison of EAR values across compounds allows assessment of the relative activity of each compound compared to the urinary concentration associated with ongoing exposure at the BMD.

RESULTS

The BMD software was used to model the various datasets for each compound. Using a BMR of 1 SD, the best-fit model was chosen based on Akaike information criteria, *p*-value, and visual inspection of model fits. The resulting BMD and BMDLs are presented in Table 2 for the compounds included in this analysis.

Utilizing the F_{UE} estimates in Table 1 and the BMDs from Table 2, BE_{BMD} values were calculated for the phytoestrogens GEN and DZN as well as for BPA (Table 3). The Activity (BE_{BMD}) values are presented in Table 3. They range from approximately 50 mg/l for GEN to approximately 2300 for BPA. The Activity (BE_{BMD}) for DZN has an intermediate value of approximately 1400 mg/l. Because of the differences in both the BMD values and the differences in urinary excretion fraction, this calculation

Table 3
Biomonitoring Equivalent (BE_{BMD}) Values in Human Urine for the Uterotrophic Benchmark Doses of GEN, DZN, and BPA

Factor	Units	GEN	DZN	BPA
BMD	mg/kg-day	10.1	99	59
F_{UE}	Unitless	0.12	0.37	1
Average daily urinary volume	l/kg-day	0.026	0.026	0.026
BE_{BMD}	mg/l	47	1409	2269

Table 4
Exposure Activity Ratio (EAR) and Relative Endocrine Exposure Activity Quotient (REEAQ) Calculations

Parameter	GEN	DZN	BPA
BMD (mg/kg-day)	10.1	99	59
BE_{BMD} (mg/l)	47	1409	2269
NHANES 2009–2010 biomonitoring data			
50th percentile (mg/l)	0.031	0.058	0.002
Exposure Activity Ratio (EAR)	6.6×10^{-4}	4.1×10^{-5}	8.8×10^{-7}
$REEAQ = EAR_{CMPND}/EAR_{GEN}$	1	6.2×10^{-2}	1.3×10^{-3}
(1/REEAQ)	1	16	750

indicates that, using urinary concentrations as a metric to compare uterotrophic response potencies from oral exposures, GEN is approximately 30 times more potent than DZN and approximately 45 times more potent than BPA.

Biomonitoring data for GEN, DZN, and BPA are available from the NHANES exposure report (CDC, 2013). The 50th population percentiles from the most recent NHANES cycle were used in the calculation of population EARs for these compounds (Table 4). We did not use the 95th percentiles because substances with short half-lives in the body are expected to exhibit substantial intraindividual variability, and thus NHANES biomonitoring upper percentiles are not informative of long-term average biomarker concentrations for individuals (Aylward et al., 2012). EARs at the 50th percentile of the population urinary biomonitoring data ranged from 6.6×10^{-4} for GEN to 8.8×10^{-7} for BPA.

DISCUSSION

There are a significant number of challenges faced in interpreting the endocrine screening assays in EPA's EDSP and OECD's test guidelines program (see, e.g., Borgert et al., 2011a). To address these challenges, many activities have focused on evaluating the various endpoints in each assay (Juberg et al., 2013; EPF, 2013a) and for integrating results across the battery of assays to reach conclusions regarding the potential of an agent to have E, A, or T activity (Borgert et al., 2011b; EPA, 2011; EPF, 2013b). In EPA's two-tiered EDSP, substances determined to have E, A, or T activity in the EDSP-ESB, based on a weight of evidence determination, are slated to be considered for definitive testing for adverse effects and dose–response in Tier 2 multigenerational tests in rats, birds, frogs, fish, and mysid. These EDSP Tier 2 tests are complex, costly, and require the use of large number of experimental animals.

As the studies of Tinwell et al. (2013) illustrate, "everyday life chemicals" can yield positive results in numerous validated endocrine screening assays. Since potential risk is a function of hazard and exposure, approaches to inform Tier 2 decision making should include consideration of exposures. However, until now, little focus has been paid to approaches that would enable use of exposure information as an element within decision making in endocrine screening and testing. We present an approach for using human biomonitoring data to help place the dose associated with an activity of substances in an endocrine screening assay within the context of actual human exposure levels. Although the activity of a substance in an EDSP-ESB assay is not a measure of risk or health hazard, understanding how such a dose compares to human exposures provides a valuable additional metric that can be used in decision making.

In addition to the EAR, we propose a metric to help interpret the significance of endocrine screening results, potency, and exposure for a specific compound by comparing the EAR_{CMPD} to the EAR of a reference substance. To illustrate this, we use GEN as the reference agent and derived the REEAQ ($REEAQ = EAR_{CMPND} \div EAR_{GEN}$). The REEAQ can assist in weight of evidence decision making and priority setting for more advanced (Tier 2) testing in EPA's EDSP. For example, an REEAQ of 1 could be interpreted to indicate an effective exposure and activity level equivalent to that predicted for typical GEN exposure levels in a normal diet in the U.S. population; an REEAQ larger than 1 indicates greater effective exposure and activity levels than that associated with typical GEN exposures (the larger the value, the higher the priority); as REEAQ decreases below 1, the value indicates decreasing relative exposure and activity levels than that experienced from typical GEN exposures. Table 4 presents the REEAQ ($REEAQ = EAR_{CMPND} \div EAR_{GEN}$). The $REEAQ_{BPA/GEN}$ is 1.3×10^{-3} and the $REEAQ_{DZN/GEN}$ is 6.2×10^{-2}. Since an $REEAQ_{BPA/GEN}$ of 1 would indicate a priority equivalent to that of exposure to GEN in a normal diet in the U.S. population, the $REEAQ_{BPA/GEN}$ of 1.3×10^{-3} suggests, taking into account both potency and human exposure, a 750-fold ($1/1.3 \times 10^{-3}$) lower priority for typical exposures experienced by the U.S. population to BPA compared to the priority for exposure to the normal level of GEN consumed in a healthy diet.

Thus, the EAR and REEAQ metrics can be used within the EDSP to contribute to and support Tier 2 decisions. Based on considerations of both human exposure and potency, substances with small EARs and REEAQs would indicate low priority for further endocrine testing. The approach illustrated here can be applied to additional compounds provided that they have both consistent dose–response data from an oral rat uterotrophic assay and a measured or estimated urinary excretion fraction. BE_{BMD} values could then be estimated and used in an EAR assessment of urinary biomonitoring data for additional natural dietary or synthetic compounds. Using GEN as a reference compound for estrogenic assays, we propose that any substance with an $REEAQ_{CMPND/GEN}$ less than 1 in EDSP Tier 1 assays would be a low priority and may

not warrant further EDSP Tier 2 testing for estrogenic endpoints.

Specific Discussion of the EAR Methodology Used for GEN, DZN, and BPA

As with other compounds, urinary biomonitoring data for these compounds is especially valuable as an exposure metric because these data capture both variation across the population and exposures resulting from a variety of sources that may be very difficult to estimate or measure on an external dose basis. We did not use the population 95th percentile urinary concentrations for these compounds (GEN, DZN, and BPA) because (1) all are rapidly eliminated (generally, urinary elimination half-lives are less than 6 hr (Tham et al., 2005, Tinwell et al., 2008; Pérez-Jiménez et al., 2010), (2) within-day intraindividual variation in urinary concentrations are likely to be substantial, with 100-fold or more changes in urinary concentrations expected depending on the relationship between time of exposure and time of urinary sampling (Teeguarden et al., 2011; Aylward et al., 2012), and (3) variations within and across individuals in hydration status can substantially influence biomarker concentration. Thus, extremes in the distributions of population urinary biomonitoring data concentrations are likely not to be accurate markers for longer term average concentrations. Because the BE_{BMD} values calculated here are estimates of steady-state *average* urinary concentrations corresponding to the BMD exposure levels, assessment of central tendency estimates of the population biomonitoring data, such as medians or geometric means, are likely to be a more appropriate use of the BE_{BMD} values, and 95th percentile concentrations are likely to substantially overestimate longer term exposure levels for the general population. An analysis directed toward the comparative exposure and activity levels at peak exposure concentrations (internal as well as reflected by upper bound biomonitoring data for urinary concentrations) would require more detailed pharmacokinetic data on all of the compounds of interest to simulate and estimate peak urinary concentrations associated with a given BMD for each compound. This analysis may be of interest, but would constitute a refined assessment that might be triggered as a follow-up to the initial screening-level assessment presented here.

Use of this approach requires an understanding of toxicokinetics of the substances of interest in the test system and in humans. For example, BPA's biliary clearance differs between rats and humans. In both humans and rats, the BPA–glucuronide conjugate is readily formed during first pass metabolism in the intestinal wall and liver following oral absorption. In rats, a portion of the BPA–glucuronide is excreted in bile into the gut, where it can be deconjugated and free BPA reabsorbed, resulting in enterohepatic recycling, and, a longer half-life in rats. Humans, on the other hand, have a higher molecular weight threshold for biliary clearance and thus the BPA–glucuronide in humans is not eliminated in the bile, but instead gets cleared by glomerular filtration. Orally administered BPA is nevertheless rapidly excreted in both rats and humans, and the F_{UE} approach used in this method accounts for such species differences. But caution still needs to be taken for other compounds, for substances with longer half-lives, and for comparison where the route of administration in the endocrine screening test method differs from that of humans. For lipophilic substances with longer half-lives, human biomonitoring studies typically measure levels in blood, not urine. While this method can still be employed for such substances, it requires use of toxicokinetics, not the F_{UE}, for converting the BMD from an endocrine assay to a corresponding blood level. Using such techniques, we have published many BEs for blood concentrations of long half-life lipophilic substances, including dioxins, PCBs, DDT, and used these to interpret human biomonitoring data (as reviewed by Aylward et al., 2013).

Other sources of uncertainty inherent in this approach include the choice of F_{UE} value used in this assessment. As reviewed by Pérez-Jiménez et al. (2010), measured and estimated urinary excretion fractions for GEN and DZN vary from study to study, with a fairly wide range of values reported for both compounds (see Table 1). The range of values reported may represent imprecision due to individual study design factors. However, variation in the metabolism of phytoestrogen compounds is well recognized due to both differing sources of the compounds (different food or supplements) as well as inherent inter- and intraindividual variation in metabolism (Pérez-Jiménez et al., 2010). Thus, there is likely to be substantial variation in the "true" urinary excretion fractions for these two compounds. The weighted mean estimates, weighted by study size, calculated by Pérez-Jiménez et al. (2010) were used here in the calculation of the BE_{BMD} values. However, for a given individual and sample collection, the true F_{UE} value could vary from the mean estimate used here, potentially by a factor of three to four in either direction. This would result in correspondingly lower or higher BE_{BMD} values, and correspondingly higher or lower EAR estimates.

In addition, average values for urinary flow rates were used here in the calculation of the BE_{BMD} values; as previously noted, flow rates for children may be somewhat higher (per kg bodyweight per day), while those for older adults may be somewhat lower. These variations would result in modest changes in age-specific values for BE_{BMD} values. However, in a screening framework, such differences are relatively small.

Although EARs are not risk metrics per se, because the activities measured in the ESB assays are not considered to be adverse effect endpoints, it is still informative to consider the EAR values compared to an MOE risk metric. The MOE is a comparison of a No Observed Adverse Effect Level (NOAEL) or BMDL (derived from an adverse effect) to the actual or estimated level of human exposure; MOE = (NOAEL or BMDL) ÷ (human exposure). Typically, in a risk assessment context, setting an acceptable MOE involves consideration of toxicokinetics and toxicodynamics (the method used to extrapolate from animal to human, the existence of potentially sensitive human subpopulations, as well as the relative adversity of the benchmark dose). When using a NOAEL or BMDL, often the acceptable MOE of 100 is considered as a benchmark in a risk assessment context, based on application of a 10-fold interspecies uncertainty factor coupled with a 10-fold intraspecies uncertainty factor. When allometric scaling or PBPK modeling is used to adjust a NOAEL or BMDL from

an animal study to a human dose, the toxicokinetic portion of the interspecies extrapolation factor is accounted for, and an acceptable MOE of 30 or lower may be used (Aylward et al., 2011; EPA, 2012b). To compare an EAR value to an MOE benchmark requires inverting the EAR. By comparison to these acceptable MOE values, the inverted EARs for GEN ($1/6.6 \times 10^{-4} = 1{,}500$) DZN ($1/4.1 \times 10^{-5} = 24{,}000$) and BPA ($1/8.8 \times 10^{-7} = 1{,}130{,}000$) suggest a substantial margin of safety for these compounds, a conclusion consistent with that of an assessment based on external exposure estimations rather than biomonitoring data (Bolt et al., 2001).

Extending the EAR Methodology to the Other EDSP Screening Assays

The EAR method for contextualizing endocrine screening assay results in terms of potency and human exposure can be adapted and applied to other EDSP assays and to similar endocrine assays, such as those in ToxCast™, when data is available. Table 5 summarizes alternative approaches for deriving Exposure and Activity components of the EAR based on availability of exposure data and the EDSP screening assay type. For the mammalian EDSP assays, BMDs can be calculated for virtually any endpoint that responds to a test article, and corresponding human exposures can be derived in the same manner as illustrated here for the uterotrophic assay. It is important to note, however, that for some substances, the relevant biomonitoring medium is blood, not urine, and additional dosimetry calculations, such as PBPK modeling will be needed for such substances to relate an oral BMD to an equivalent blood BE_{BMD}. The duration of treatment differs across EDSP *in vivo* assays, and depending upon the compound, steady-state concentrations may not be achieved. This is likely to be the case for lipophilic substances with long half-lives (Hays et al., 2007), and this may require adjustments to be made, such as reducing the derived BMD by a factor that can correct for this.

For the EDSP *in vitro* assays, several approaches can be used to relate the *in vitro* activities in endocrine screening assays to human exposures. All approaches assume steady state. With the BE approach, human biomonitoring data of concentrations in blood can be directly compared to the *in vitro* BMD concentration. A second approach (Rotroff et al., 2010; Wetmore et al., 2012) would employ *in vitro* to *in vivo* extrapolation model(s) to estimate the daily human oral dose necessary to produce *in vivo* blood concentrations equivalent to the *in vitro* BMD, and then this can be compared to a measured or modeled estimate of the actual human oral exposure. For the fish and amphibian EDSP assays, BMDs would be derived as applied doses (concentrations in water) and then compared directly to measured or modeled environmental concentrations.

The approach described here can also be applied to high-throughput screening results from ToxCast endocrine screening assays. Wetmore et al. (2012) calculated the daily human oral dose for steady-state *in vivo* blood concentrations in humans equivalent to the *in vitro* concentration at 50% of maximum activity (or lowest effective concentration value) for more than 500 *in vitro* ToxCast assays. This allowed the authors to calculate a metric essentially the inverse of the EAR discussed here. However, these authors did not specifically address endocrine assays or use GEN as a reference substance for judging the significance of activity:exposure estimates. Previously, it was proposed that results of the ToxCast screening assays could be evaluated using ToxPi, an interpretation and visualization tool that assembles results of related assays into different groups, arrays these groups into a specific sections of a pie chart, and then for each compound, the area of each section reflects its activity in ToxCast assays (Reif et al., 2010). However, ToxPi has been criticized because it normalizes each analysis of each assay in relation to the compound with the highest level of activity among the set of substances evaluated and does not require use of standard reference compounds (Patlewicz et al., 2013). This relativistic approach falls well short of benchmarking or anchoring ToxCast results to meaningful potencies or exposures. Based on our results, we recommend that suitable reference standards be agreed upon for each of the EDSP-ESB screening assays, and other similarly relevant ToxCast endocrine assays, to assure consistent normalization of results and enhance the use of such metrics as EARs and REEAQs in data analysis and decision making.

In this analysis, we used BMR in the uterotrophic assay of 1 SD difference from control. But perhaps this is too sensitive for a screening assay. Wetmore et al. (2012) used benchmarks of 50% of maximum activity or lowest effective concentration value in their comparison to human exposures of the results of more than 500 ToxCast assays. While arguably it may be less important which endpoint or benchmark is used across a wide swath of different assays for a specific program, such as EPA's EDSP, agreement on a specific endpoint (or endpoints) from each assay, on the appropriate benchmark for each endpoint and on a specific set of reference compounds, would be beneficial by promoting uniform comparisons and fostering consistent decision making. In any case, when comparing compounds, it will be important to use the same endpoint, BMR, and reference agent, and if there is full transparency, to some degree, this may solve the issue. However, simply knowing the BMD for a consistent BMR across compounds is of limited utility. These BMDs should be put in an exposure context and, if feasible, anchored to an EAR for a suitable reference agent.

In addition to the EAR metric, we introduce the REEAQ as a metric to place an EAR of a synthetic chemical into the context of an EAR for a reference chemical, in this instance we use GEN ($REEAQ = EAR_{CMPND} \div EAR_{GEN}$). For BPA, the $REEAQ_{BPA/GEN}$ is 1.3×10^{-3}, indicating a much lower potency:exposure profile for BPA than for typical human exposures to the natural, dietary phytoestrogen GEN. As discussed above, ground truthing endocrine screening results using the REEAQ as a comparison of the potency and exposure of naturally occurring agents in the normal human diet can assist in weight of evidence decision making as to whether additional Tier 2 testing may, or may not, be warranted. For other assays that evaluate estrogenic activity, the $REEAQ = EAR_{CMPND} \div EAR_{GEN}$ can be used, provided results for GEN have been obtained in each assay. For antiandrogenic and antithyroid activities, similar Relative Exposure Activity Quotients could be derived if reference substances could be agreed

Table 5
Methods for Deriving the Exposure and Activity Components to Calculate Exposure:Activity Ratios (EAR) for all 11 EDSB ESB Assays

EDSP-ESB assays	Methods for the Exposure and Activity components of the EAR
Uterotrophic Hershberger Male pubertal Female pubertal	When human biomonitoring data are available: for Exposure use the actual biomonitoring data and for Activity calculate the BMD and convert to human BE_{BMD} (using the same procedures as described in detail above for the uterotrophic assay) In the absence of human biomonitoring data: for Exposure use calculated human oral exposure values (e.g., from pesticide registration evaluations or similar risk assessments) and for Activity calculate the BMD and convert to human oral BMD using allometric scaling or PBPK modeling
ER binding ERα transcriptional activation AR binding Steroidogenesis Aromatase	When human biomonitoring data are available: for Exposure use the actual human blood/serum biomonitoring data[a] and for Activity use the AC_{50} values In the absence of human biomonitoring data: for Exposure use calculated human oral exposure values (e.g., from pesticide registration evaluations or similar risk assessments) and for Activity use in vitro to in vivo extrapolation to calculate the human oral equivalent dose necessary to produce a steady-state in vivo concentration equivalent to in vitro AC_{50} (Rotroff et al, 2010; Wetmore et al., 2012)
Fish short-term reproduction Amphibian metamorphosis (frog)	When environmental monitoring data are available: for Exposure use the actual monitoring data and for Activity calculate the BMD (applied concentration) In the absence of environmental monitoring data: for Exposure use calculated exposure values (e.g., from pesticide registration evaluations or similar environmental risk assessments) and for Activity calculate the BMD (applied concentration)

[a]While it is appropriate to compare human blood/serum concentrations to test article concentrations in media used for *in vitro* screening, it is not appropriate to use human urine concentrations for such comparisons. If only human urine biomonitoring results are available, then for determining Exposure, urine concentrations first need to be converted to blood/serum concentrations or the urine biomonitoring data should not be used at all and instead calculated human oral exposure values should be used.

upon. Glucosinolates, antithyroid agents (EFSA, 2008) found primarily in Brassica vegetables (Brussels sprouts, cauliflower, broccoli, etc.), may be candidates for a thyroid reference compound. Adult human exposures to total glucosinolates have been estimated (Krul et al., 2002) to range from approximately 0.25 to 2.5 mg/kg-day for 60 kg adult (15 mg/day [United States], 50 mg/day [United Kingdom], 150 mg/day [Japan]). Of the more than 100 glucosinolates, 5-vinyloxazolidine-2-thione, whose antithyroid effects have been studied in preweanling rats (Lewerenz et al., 1992) and is commercially available, may be an appropriate reference chemical candidate. A potential antiandrogen reference compound is 3,3-diindolylmethane, formed from indole-3-carbinol during digestion of cruciferous vegetables. 3,3-Diindolylmethane has been characterized as a "naturally occurring pure androgen antagonist" (Le et al., 2003), and is also commercially available.

While the challenges and issues of evaluating simultaneous exposures to mixtures (see, e.g., Borgert et al, 2012) are beyond the scope of this analysis, it should be noted that this approach potentially allows a screening level cumulative assessment, in a manner similar to the Hazard Index method, if dose additivity is assumed:

$$CUM_{EAR} = \sum_{i=1}^{n} EAR_i \qquad (5)$$

Such CUM_{EAR} values could be calculated on an individual-by-individual basis for the three compounds included in this analysis if the NHANES datasets provided measured values for these three analytes in the same individuals. However, urinary BPA was measured in a one-third subset of the full NHANES dataset in 2009 to 2010, while the phytoestrogens were measured on a different one-third subset. So, cumulative exposures and EARs cannot be assessed across the phytoestrogens and environmental phenols using this dataset on an individual basis. None the less, a cumulative assessment is possible on the basis of the population median values. Clearly, the cumulative $CUM_{EAR(GEN+DZN+BPA)}$ ($6.6 \times 10^{-4} + 4.1 \times 10^{-5} + 8.8 \times 10^{-7} = 7.02 \times 10^{-4}$) is dominated by the value for GEN, with DZN contributing less than 6% and BPA less than 0.2% to the value of $CUM_{EAR(GEN+DZN+BPA)}$. If cumulative assessment were to be an element in the decision making process for whether or not to undertake additional EDSP testing beyond Tier 1, then, interpreting the CUM_{EAR} in the framework of the maximum cumulative ratio (Price and Han, 2011) would suggest little value in additional testing of substances with EAR values 2 or more orders of magnitude lower than the EAR_{GEN}.

This framework can also be used to address interest in human exposure to putative endocrine active substances. Relative urinary levels of different potentially endocrine active compounds can be assessed and compared via the EAR approach. When endocrine screening results are available for compounds in the NHANES survey, this approach allows the population biomonitoring data across compounds of interest, including both natural dietary polyphenols as well as synthetic compounds, to be assessed in a coherent and consistent framework.

The framework and approach presented here suggests two areas of additional data collection or refinement of existing data collection procedures. First, improvements

in understanding pharmacokinetics in humans should be a priority. While there has been considerable necessary discussion about study designs, ethics, informed consent, etc. in controlled dosing experiments in humans, improving knowledge on, and data of, urinary excretion fraction in humans would clearly enhance the scientific basis for interpreting biomonitoring studies of both natural and synthetic compounds. Second, in the U.S. NHANES program, it would be helpful if the population subsets selected for analyte measurement took into account potential cumulative exposure interests for substances that have similar receptor-mediated modes of action. In the case of environmental phenols and phytoestrogen compounds, potential interactions with the estrogen receptor are of interest in both categories of compounds, and analysis of both classes of compounds in a specific subset of the NHANES participant population would be helpful in assessment of cumulative exposures in this context.

ACKNOWLEDGMENTS AND CONFLICT OF INTEREST STATEMENT

L.L.A., S.M.H., and C.R.K. received funding to support this research from the American Chemistry Council (ACC) and its Center for Advancing Risk Assessment Science and Policy. The authors had complete control over the design, conduct, interpretation, and reporting of the analyses included in this manuscript. The contents of this manuscript are solely the responsibility of the authors and do not necessarily reflect the views or policies of ACC. L.L.A., S.M.H., and C.R.K. are independent partners in Summit Toxicology LLP, a toxicology, risk assessment, and pharmaceutical consulting firm, and have worked on risk assessment issues for a wide variety of governmental, trade association, and industry clients. R.A.B. and K.W. are employed by ACC, a trade association of chemical manufacturers.

REFERENCES

Aylward LL, Hays SM. 2008. Biomonitoring equivalents (BE) dossier for 2,4-dichlorophenoxyacetic acid (2,4-D) (CAS No. 94–75–7). Regul Toxicol Pharmacol 51(3 Suppl):S37–S48.

Aylward LL, Hays SM. 2011. Consideration of dosimetry in evaluation of ToxCast™ data. J Appl Toxicol. 31(8):741–751.

Aylward LL, Becker RA, Kirman CR, Hays SM. 2011. Assessment of margin of exposure based on biomarkers in blood: an exploratory analysis. Regul Toxicol Pharmacol 61:44–52.

Aylward LL, Kirman CR, Schoeny R, Portier CJ, Hays SM. 2013. Evaluation of biomonitoring data from the CDC National Exposure Report in a risk assessment context: perspectives across chemicals. Environ Health Perspect 121:287–294.

Aylward LL, Kirman CR, Adgate JL, McKenzie LM, Hays SM. 2012. Interpreting variability in population biomonitoring data: role of elimination kinetics. J Expo Sci Environ Epidemiol 22(4):398–408.

Basli A, Soulet S, Chaher N, Mérillon JM, Chibane M, Monti JP, Richard T. 2012. Wine polyphenols: potential agents in neuroprotection. Oxid Med Cell Longev 2012:805762, pages 1–14.

Bolt HM, Janning P, Michna H, Degen GH. 2001. Comparative assessment of endocrine modulators with oestrogenic activity: I. Definition of a hygiene-based margin of safety (HBMOS) for xeno-oestrogens against the background of European developments. Arch Toxicol 74(11):649–662.

Borgert CJ, LaKind JS, Witorsch RJ. 2003. A critical review of methods for comparing estrogenic activity of endogenous and exogenous chemicals in human milk and infant formula. Environ Health Perspect 111(8):1020–1036.

Borgert CJ, Mihaich EM, Quill TF, Marty MS, Levine SL, Becker RA. 2011a. Evaluation of EPA's Tier 1 endocrine screening battery and recommendations for improving the interpretation of screening results. Regul Toxicol Pharmacol. 59:397–411.

Borgert CJ, Mihaich EM, Ortego LS, Bentley KS, Holmes CM, Levine SL, Becker RA. 2011b. Hypothesis-driven weight of evidence framework for evaluating data within the US EPA's Endocrine Disruptor Screening Program. Regul Toxicol Pharmacol 61:185–191.

Borgert CJ, Sargent EV, Casella G, Dietrich DR, McCarty LS, Golden RJ. 2012. The human relevant potency threshold: reducing uncertainty by human calibration of cumulative risk assessments. Regul Toxicol Pharmacol 62:313–328.

Borgert CJ, Baker SP, Matthews JC. 2013. Potency matters: thresholds govern endocrine activity. Regul Toxicol Pharmacol 67:83–88.

Caldwell DJ, Mastrocco F, Nowak, E, Johnston J, Yekel H, Pfeiffer D, Hoyt M, DuPlessie BM, Anderson PD. 2010. An assessment of potential exposure and risk from estrogens in drinking water. Environ Health Perspect. 118:338–344.

Carbonel AA, Simões RS, Santos RH, Baracat MC, Simões Mde J, Baracat EC, Soares Júnior JM. 2011. Effects of high-dose isoflavones on rat uterus. Rev Assoc Med Bras 57(5):534–539.

CDC. 2009. Laboratory procedure manual, analyte: benzophenone-3, bisphenol A, 2,4-dichlorophenol, 2,5-dichlorophenol, orthophenylphenol, methyl-, ethyl-, propyl-, and butyl parabens, 4-tertoctylphenol, 2,4,5-trichlorophenol, 2,4,6- trichlorophenol, triclosan, Matrix: Urine, Method No: 6301.01. Available at: http://www.cdc.gov/nchs/data/nhanes/nhanes_05_06/eph_d_met_phenols_parabens.pdf. Accessed Nov. 25. 2013.

CDC. 2013. Fourth National Report on Human Exposure to Environmental Chemicals, updated tables, March, 2013. Available at: http://www.cdc.gov/exposurereport/pdf/FourthReport_UpdatedTables_Mar2013.pdf. Accessed Nov. 25. 2013.

Charles GD, Gennings C, Tornesi B, Kan HL, Zacharewski TR, Bhaskar Gollapudi B, Carney EW. 2007. Analysis of the interaction of phytoestrogens and synthetic chemicals: an in vitro/in vivo comparison. Toxicol Appl Pharmacol 218(3):280–288.

Diel P, Schulz T, Smolnikar K, Strunck E, Vollmer G, Michna H. 2000. Ability of xeno- and phytoestrogens to modulate expression of estrogen-sensitive genes in rat uterus: estrogenicity profiles and uterotropic activity. J Steroid Biochem Mol Biol 73(1–2):1–10.

EFSA. 2008. EFSA panel on contaminants in the food chain: glucosinolates as undesirable substances in animal feed. Scientific Panel on Contaminants in the Food Chain. EFSA J 590:1–76.

EPA. 1998. EDSTAC final report, chapter three, conceptual framework and principles. Page 3–16. Available at: http://www.epa.gov/endo/pubs/edstac/chap3v14.pdf. Accessed Nov. 25, 2013.

EPA. 2009. Endocrine Disruptor Screening Program Test Guidelines, OPPTS 890.1600: uterotrophic assay. Available at: http://www.epa.gov/ocspp/pubs/frs/publications/Test_Guidelines/series890.htm. Accessed Nov. 25, 2013.

EPA. 2011. Endocrine disruptor screening program, weight-of-evidence: evaluating results of EDSP Tier 1 screening to identify the need for Tier 2 testing. Available at: http://www.epa.gov/endo/. Accessed Nov. 25, 2013.

EPA. 2012a. Benchmark dose technical guidance. Available at: http://www.epa.gov/raf/publications/pdfs/benchmark_dose_guidance.pdf. Accessed Nov. 25, 2013.

EPA. 2012b. EPA 2012. TSCA workplan chemical risk assessment, *N*-methylpyrrolidone: paint stripping use. December 2012. Available at: http://www.epa.gov/oppt/existingchemicals/pubs/TSCA_Workplan_Chemical_Risk_Assessment_of_NMP.pdf. Accessed Nov. 25, 2013.

EPA. 2013a. Endocrine Disruptor Screening Program (EDSP). Available at: http://www.epa.gov/endo/index.htm. Accessed Nov. 25, 2013.

EPA. 2013b. Series 890—endocrine disruptor screening program test guidelines. Available at: http://www.epa.gov/ocspp/pubs/frs/publications/Test_Guidelines/series890.htm. Accessed Nov. 25, 2013.

EPA. 2013b. Overview of the second list of chemicals for Tier 1 screening. http://www.epa.gov/endo/pubs/prioritysetting/list2facts.htm. Accessed Nov. 25, 2013.

EPF. 2013a. Endocrine Policy Forum comments on "Endocrine Disruptor Screening Program (EDSP) Evaluation of Tier 1 Screening Assays and Battery Performance, May 20, 2013." Available at: http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPP-2013-0075-0013. Accessed November 25, 2013.

EPF. 2013b. Endocrine policy forum comments on weight-of-evidence: evaluating results of the EDSP Tier 1 screening. Available at: http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPP-2013-0230-0011. Accessed Nov. 25, 2013.

Hays SM, Becker RA, Leung HW, Aylward LL, Pyatt DW. 2007. Biomonitoring equivalents: a screening approach for interpreting biomonitoring results from a public health risk perspective. Regul Toxicol Pharmacol 47(1):96–109.

Hays SM, Aylward LL, LaKind JS, Bartels MJ, Barton HA, Boogaard PJ, et al. 2008. Guidelines for the derivation of biomonitoring equivalents: report from the Biomonitoring Equivalents Expert Workshop. Regul Toxicol Pharmacol 51(3 Suppl):S4–S15.

Juberg DR, Borghoff SJ, Becker RA, Casey W, Hartung T, Holsapple MP, Marty MS, Mihaich EM, Van Der Kraak G, Wade MG, Willett CE, Andersen ME, Borgert CJ, Coady KK, Dourson ML, Fowle III JR, Gray LE, Lamb JC, Ortego LS, Schug TS, Toole CM, Zorrilla LM, Kroner OL, Patterson J, Rinckel LA and Jones BR. 2013. Lessons Learned, Challenges, and Opportunities: The U.S. Endocrine Disruptor Screening Program. ALTEX. In Press.

Kanno J, Onyon L, Haseman J, Fenner-Crisp P, Ashby J, Owens W. 2001. The OECD program to validate the rat uterotrophic bioassay to screen compounds for in vivo estrogenic responses: phase 1. Environ Health Perspect 109(8):785–794.

Kanno J, Onyon L, Peddada S, Ashby J, Jacob E, Owens W. 2003. The OECD program to validate the rat uterotrophic bioassay. Phase 2: coded single-dose studies. Environ Health Perspect 111(12):1550–1558.

Khurana S, Piche M, Hollingsworth A, Venkataraman K, Tai TC. 2013. Oxidative stress and cardiovascular health: therapeutic potential of polyphenols. Can J Physiol Pharmacol 91(3):198–212.

Kishimoto Y, Tani M, Kondo K. 2013. Pleiotropic preventive effects of dietary polyphenols in cardiovascular diseases. Eur J Clin Nutr 67(5):532–535.

Krishnan K, Gagne M, Nong A, Aylward LL, Hays SM. 2010. Biomonitoring equivalents for bisphenol A (BPA). Regul Toxicol Pharmacol 58(1):18–24.

Krul C, Humblot C, Philippe C, Vermeulen M, van Nuenen M, Havenaar R, Rabot S. 2002. Metabolism of sinigrin (2-propenyl glucosinolate) by the human colonic microflora in a dynamic in vitro large-intestinal model. Carcinogenesis 23:1009–1016.

Le H, Schaldach CM, Firestone GL, Bjeldanes LF. (2003). Plant-derived 3,3′-diindolylmethane is a strong androgen antagonist in human prostate cancer cells. J Biol Chem 2278:21136–21145

Lewerenz H-J, Bleyl DWR, Schnaak W. 1992. Effects of 5-vinyloxazolidine-2-thione on lactating rats and their sucklings. Nahrung 36:388–393.

OECD. 2007. OECD Guideline for the Testing of Chemicals. Uterotrophic bioassay in rodents: a short-term screening test for oestrogenic properties. Available at: http://www.epa.gov/endo/pubs/uterotrophic_OECD_guideline.pdf. Accessed Nov. 25, 2013.

OECD. 2013a. Series on testing and assessment: testing for endocrine disrupters. Available at: http://www.oecd.org/chemicalsafety/testing/seriesontestingandassessmenttestingforendocrinedisrupters.htm. Accessed Nov. 25, 2013.

OECD. 2013b. OECD guidelines for the testing of chemicals; Section 4, health effects and Section 2, effects on biotic systems. Available at: http://www.oecd.org/env/ehs/testing/oecdguidelinesforthetestingofchemicals.htm. Accessed Nov. 25, 2013.

Patlewicz G, Simon T, Goyak K, Phillips RD, Rowlands JC, Seidel SD, Becker RA. 2013. Use and validation of HT/HC assays to support 21st century toxicity evaluations. Regul Toxicol Pharmacol. 65(2):259–68.

Pérez-Jiménez J, Hubert J, Hooper L, Cassidy A, Manach C, Williamson G, Scalbert A. 2010. Urinary metabolites as biomarkers of polyphenol intake in humans: a systematic review. Am J Clin Nutr 92(4):801–809.

Price PS, Han X. 2011. Maximum cumulative ratio (MCR) as a tool for assessing the value of performing a cumulative risk assessment. Int J Environ Res Public Health 8(6):2212–2225.

Rachon D, Vortherms T, Seidlova-Wuttke D, Wuttke W. 2007. Dietary daidzein and puerarin do not affect pituitary LH expression but exert uterotropic effects in ovariectomized rats. Maturitas 57(2):161–170.

Reif DM, Martin MT, Tan SW, Houck KA, Judson RS, Richard AM, Knudsen TB, Dix DJ, Kavlock RJ. 2010. Endocrine profiling and prioritization of environmental chemicals using ToxCast data. Environ Health Perspect 118(12):1714–20

Rotroff DM, Wetmore BA, Dix DJ, Ferguson SS, Clewell HJ, Houck KA, Lecluyse EL, Andersen ME, Judson RS, Smith CM, Sochaski MA, Kavlock RJ, Boellmann F, Martin MT, Reif DM, Wambaugh JF, Thomas RS. 2010. Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening. Toxicol Sci 117:348–358.

Schwen RJ, Nguyen L, Plomley JB, Jackson RL. 2012. Toxicokinetics and lack of uterotropic effect of orally administered S-equol. Food Chem Toxicol 50(5):1741–1748.

Teeguarden JG, Calafat AM, Ye X, Doerge DR, Churchwell MI, Gunawan R, et al. 2011. Twenty-four hour human urine and serum profiles of bisphenol a during high-dietary exposure. Toxicol Sci 123(1):48–57.

Tham DM, Gardner CD, Haskell WL. 1998. Clinical Review 97. Potential health benefits of dietary phytoestrogens: a review of the clinical, epidemiological, and mechanistic evidence. J Clin Endocrinol Metab 83(7):2223–35. Available at: http://jcem.endojournals.org/content/83/7/2223.long. Accessed Nov. 25, 2013.

Tham DM, Gardner CD, Haskell WL, Volkel W, Bittner N, Dekant W. 2005. Quantitation of bisphenol A and bisphenol A glucuronide in biological samples by high performance liquid chromatography-tandem mass spectrometry. Drug Metab Dispos 33(11):1748–1757.

Tinwell H, Colombel S, Blanck O, Bars R, Volkel W, Kiranoglu M, Fromme H. 2008. Determination of free and total bisphenol A in human urine to assess daily uptake as a basis for a valid risk assessment. Toxicol Lett 179(3):155–162.

Tinwell H, Colombel S, Blanck O, Bars R. 2013. The screening of everyday life chemicals in validated assays targeting the pituitary-gonadal axis. Regul Toxicol Pharmacol. 66:184–196.

Wambaugh JF, Setzer RW, Reif DM, Gangwal S, Mitchell-Blackwood J, Arnot JA, Joliet O, Frame A, Rabinowitz J, Knudsen TB, Judson RS, Egeghy P, Vallero D, Cohen Hubal EA. 2013. High-throughput models for exposure-based chemical prioritization in the ExpoCast project. Environ Sci Technol. 47(15):8479–8488.

Wetmore BA, Wambaugh JF, Ferguson SS, Sochaski MA, Rotroff DM, Freeman K, Clewell HJ 3rd, Dix DJ, Andersen ME, Houck KA, Allen B, Judson RS, Singh R, Kavlock RJ, Richard AM, Thomas RS. 2012. Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicol Sci 125(1):157–174.