Organisation for Economic Co-operation and Development

ENV/CBC/MONO(2022)43

Unclassified

English - Or. English

15 December 2022

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

Case Study on the use of Integrated Approaches to Testing and Assessment for potential Systemic Toxicity and Estrogen Receptor Activation of a Group of Bisphenols and Select Alternatives

Series on Testing and Assessment
No. 373

JT03509961

OECD Environment, Health and Safety Publications

SERIES ON TESTING AND ASSESSMENT

NO. 373

Case Study on the use of Integrated Approaches to Testing and Assessment for potential Systemic Toxicity and Estrogen Receptor Activation of a Group of Bisphenols and Select Alternatives

IOMC

INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS

A cooperative agreement among FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD

Environment Directorate

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT

Paris 2022

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 38 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in twelve different series: **Testing and Assessment**; **Good Laboratory Practice and Compliance Monitoring**; **Pesticides**; **Biocides**; **Risk Management**; **Harmonisation of Regulatory Oversight in Biotechnology**; **Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; Safety of Manufactured Nanomaterials;** and **Adverse Outcome Pathways.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (www.oecd.org/chemicalsafety/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

- **Also published in the Series on Testing and Assessment: link**

For this and many other Environment, Health and Safety publications, consult the OECD's World Wide Web site (www.oecd.org/chemicalsafety/)

or contact:

OECD Environment Directorate,
Environment, Health and Safety Division
2 rue André-Pascal
75775 Paris Cedex 16
France

E-mail: ehscont@oecd.org

Foreword

OECD member countries have been making efforts to expand the use of alternative methods in assessing chemicals. The OECD has been developing guidance documents and tools for the use of alternative methods such as (Q)SAR, chemical categories and Adverse Outcome Pathways (AOPs) as a part of Integrated Approaches for Testing and Assessment (IATA). There is a need for the investigation of the practical applicability of these methods/tools for different aspects of regulatory decision-making, and to build upon case studies and assessment experience across jurisdictions.

The objective of the IATA Case Studies Project is to increase experience with the use of IATA by developing case studies, which constitute examples of predictions that are fit for regulatory use. The aim is to create common understanding of using novel methodologies and the generation of considerations/guidance stemming from these case studies.

This case study was developed by Health Canada (Anthony Reardon, Matthew Gagne, Reza Farmahin, Shamika Wickramasuriya, Sean Collins, Marc Beal, Andrew Williams, Karen Leingartner, Andrea Rowan-Carroll, Matthew Meier, Andy Nong, Ella Atlas, Carole Yauk, Tara Barton-Maclaren) for illustrating practical use of IATA and submitted to the 2021 review cycle of the IATA Case Studies Project. This case study was reviewed by the project team.

The case study is illustrative examples, and their publication as OECD monographs does not translate into direct acceptance of the methodologies for regulatory purposes across OECD countries. In addition, the cases study should not be interpreted as official regulatory decisions made by the authoring member countries.

This document is published under the responsibility of the Chemicals and Biotechnology Committee of the OECD.

Table of contents

FIGURES

TABLES

Table 4. Lowest available *in vivo* Points of Departure values found for case study chemicals from the EPA Comptox Dashboard, JECDB, ECHA REACH Dossiers, eChemPortal, EFSA, NTP website, and EKDB database.

Abbreviations and Acronyms

AC_{50}	Half-Maximal Activity Concentration
AD	Applicability Domain
AED	Administered Equivalent Dose
ADME	Absorption, Distribution, Metabolism, Excretion
AOP	Adverse Outcome Pathway
AUC	Area Under the Curve
BER	Bioactivity Exposure Ratio
BMC	Benchmark Concentration
BMCL	Benchmark Concentration Lower bound
BMCU	Benchmark Concentration Upper bound
BMD	Benchmark Dose
CEPA	Canadian Environmental Protection Act
CERAPP	Collaborative Estrogen Receptor Activity Prediction Project
CMP	Chemicals Management Plan
Css	Steady-State Plasma Concentration
DART	Developmental and Reproductive Toxicology
DEG	Differentially Expressed Gene
EDC	Endocrine Disrupting Chemical
ER	Estrogen Receptor
GO	Gene Ontology
HTTK	High-Throughput Toxicokinetics
HTTr	High-Throughput Transcriptomics
IPA	Ingenuity Pathway Analysis
IVIVE	*in vitro* to *in vivo* extrapolation
KEGG	Kyoto Encyclopedia of Genes and Genomes
LO(A)EL	Lowest Observed (Adverse) Effect Level
LOEC	Lowest Observed Effect Concentration
NAM	New Approach Methodology
NO(A)EL	No Observed (Adverse) Effect Level
NTP	National Toxicology Program
OECD	Organisation for Economic Co-operation and Development
POD	Point of Departure
QSAR	Quantitative (or Qualitative) Structure Activity Relationship
RF	Random Forest
Tcpl	ToxCast Data Analysis Pipeline
TK	Toxicokinetics
UR	Upstream Regulators

Executive summary

Concerns over exposure to bisphenol A (BPA) and its effects on human health, particularly during early life stages, has led to the implementation of regulatory measures to reduce the risks for both health and the environment. This action has resulted in a rise in the use of BPA alternatives in recent years leading to the potential for regrettable substitution in consumer products considered “BPA-free”. This Integrated Approach to Testing and Assessment (IATA) was developed to acquire information for hazard characterization of BPA and its alternatives to facilitate chemical prioritization and inform risk assessment. BPA and a subset of potential alternatives were examined using new approach methodologies (NAMs) demonstrating the integration of *in silico* and *in vitro* methods. When paired with exposure estimates, NAMs can be used to support the evaluation of the potential risks that these substances may pose to the general population.

Here, 25 BPA alternatives were selected for the current work based on sharing structural characteristics with BPA, or if considered to be functional alternatives from their presence in shared-use applications. NAM-based points of departure (PODs) obtained using *in vitro* transcriptomics were derived from two approaches, i) general systemic toxicity (hazard-independent) PODs that do not predict the potential of specific adverse effects; and ii) estrogen receptor (ER) pathway-specific PODs from expression of genes related to the ER pathway.

For general systemic toxicity, there was agreement for most substances across each of the approaches used to derive NAM-based PODs based on transcriptomic data as estimates generally fell within one order of magnitude. Specific approaches used to derive PODs such as the 25th relative ranked gene benchmark concentration (BMC) were in agreement with the endpoints derived from available ToxCast data while other approaches based on distributions (e.g., 5th and 10th percentile) were less reliable and observed to be different by orders of magnitude. ER pathway specific analysis identified a selection of these chemicals as ER agonists, albeit with varying potency.

Within this case study, bioactivity PODs were defined with the intent of deriving protective estimates for general systemic effects, and estimates that are predictive of ER pathway specific effects from exposure to BPA and alternatives. Further to this, a weight of evidence approach was considered combining both *in silico* predictions and *in vitro* data from the ER pathway to create qualitative hazard flags for those bisphenols and corresponding alternatives that interact with the ER. The findings suggest that the substances used in this case study exhibit endocrine activity, wherein all bisphenols exhibited the potential for ER signal disruption but non-bisphenol alternatives were not observed to interact with the ER. The derived NAM-based PODs were found to be protective estimates of both general systemic *in vivo* effects and ER-specific effects using *in vitro* endpoints. These endpoints were generally more conservative when compared to *in vivo* PODs based on traditional animal data supporting their consideration in future prioritization and risk assessment frameworks.

1 Introduction

Bisphenol A (BPA) is an industrial chemical used to make polycarbonate, a hard, clear plastic, and is also used in epoxy resins that act as a protective lining on the inside of food packaging. BPA can be found in pitchers, tableware, water bottles, food storage containers, and older types of polycarbonate baby bottles. Furthermore, BPA is often found in thermal paper, including receipts, movie or theatre tickets, bus transfers, and boarding passes (Pivnenko et al., 2015). These sources provide multiple pathways for exposure and intake of BPA (e.g., oral, dermal and inhalation) from both dietary and non-dietary sources (Healy et al., 2015). BPA is known to be an endocrine disrupting chemical (EDC); the activation of the estrogen receptor is of particular concern, (Vandenberg et al., 2009), and is reported to be associated with several adverse effects such as certain cancers, reproductive and developmental problems, delays in puberty, reduced fertility, type II diabetes, metabolic alterations, and neurological disorders (Beausoleil et al., 2018; Ma et al., 2019; Pelch et al., 2019). In a recent review of endocrine disrupting chemicals, mechanistic data revealed BPA binding to ERα and ERβ resulting in transcriptional activation in estrogenic responsive (i.e., MCF-7) cells (La Merrill et al., 2020). In 2008, the Government of Canada assessed BPA under the Canadian Environmental Protection Act, 1999 (CEPA) and concluded that BPA met the criteria under sections 64 (a) and (c), and is considered as a substance that may be entering the environment in a quantity or concentration or under conditions that constitute or may constitute a danger in Canada to human life or health (Environment and Climate Change Canada and Health Canada, 2008). The final risk assessment report determined, based on animal studies, that there was a trend toward increased susceptibility to the effects of BPA when exposed during developmental stages (i.e., during pregnancy and up to a year after birth). Furthermore, the final assessment report considered infants, the most highly exposed subpopulation, as a potential vulnerable population. Subsequently, BPA was placed on Canada's list of Toxic Substances in Schedule 1 of CEPA (Environment and Climate Change Canada and Health Canada, 2008; Government of Canada, 2010). Currently, the Canada Consumer Product Safety Act protects new-borns and infants by prohibiting the manufacture, import, sale or advertising of polycarbonate baby bottles that contain BPA while other risk management actions have been implemented under different authorities in Canada[1].

International concerns over the potential adverse health effects from exposure to BPA has led to many developed countries imposing regulations (Kadasala, Narayanan and Liu, 2016). Within Europe, concern for BPA found in baby bottles led to a restriction for its use in plastic infant feeding bottles in 2011 (O.J. 2011). The regulatory actions regarding BPA, as well as consumer and market shifts toward "BPA-free" products have led to a rise in the use of BPA alternatives. As industry shifts away from using BPA, there is an increased use and abundance of alternatives resulting in a lack of clarity regarding safety when including products with the label "BPA-free". As such, there is potential for regrettable substitution (i.e., substitution of one substance with another that poses an equivalent or greater risk to environmental or human health), which is an important consideration in the context of the bisphenols and has been reviewed in-depth elsewhere (US NTP, 2017).

[1] https://www.canada.ca/en/health-canada/services/chemical-substances/challenge/batch-2/bisphenol-a/risk-management-action-milestones.html

There is a global motivation among regulatory agencies to promote the use of alternative test methods and strategies toward the common goal of reducing the use of animal models. New approach methodologies (NAMs) have been broadly described as any technology, method, approach, or combination thereof that provides chemical hazard and risk assessment information without the use of intact animals (US EPA, 2018)[2]. Among broader applications, NAMs provide a means of rapidly acquiring information on data-poor chemicals, as well as identify specific chemicals or groups of chemicals as priority candidates for chemical risk assessment using combination of *in silico* and *in vitro* methods.

With the goal of reducing the potential for regrettable substitution, there is a need to consider the broader class of bisphenols used in Canada. Health Canada has identified 25 substances on Canadian chemical inventories (i.e., Domestic Substance List (DSL) and Non-Domestic Substance List (NDSL)) that share structural characteristics with BPA or are used as functional alternatives for common applications. This IATA document was drafted by Health Canada to characterize the potential hazards posed and compare relative potencies within this group of 25 substances to demonstrate the effectiveness of scoping exercises to establish priority chemicals prior to progression to a more comprehensive risk assessment under CEPA, where appropriate.

[2] https://www.canada.ca/en/health-canada/services/chemical-substances/chemical-substances-glossary.html#n

2 Purpose

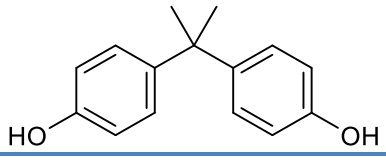
2.1. Purpose of use

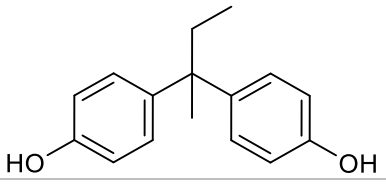
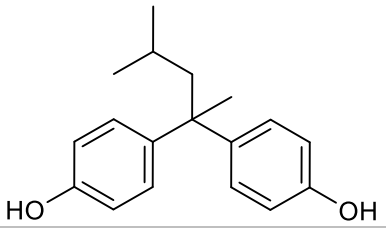
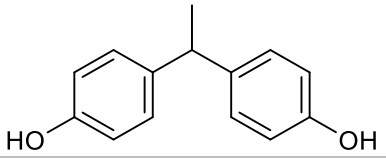
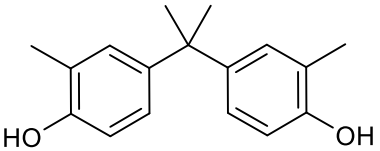
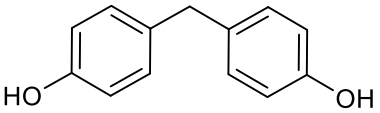
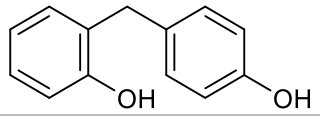
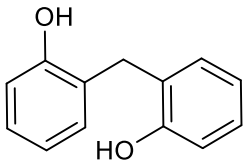
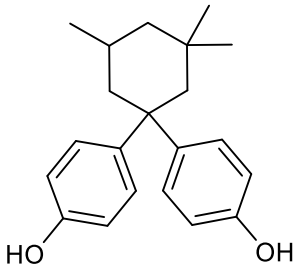
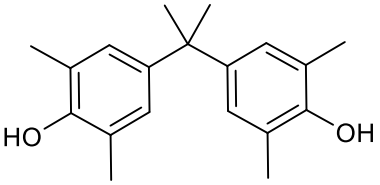
This IATA examines 25 potential BPA alternatives using NAM-based points of departure (PODs) for comparison to those derived from traditional *in vivo* data (where available). The primary purpose of this case study was to demonstrate that NAM-based PODs are protective for this group of substances and provide a weight of evidence from assessment of estrogenic activity of bisphenol compounds via high-throughput transcriptomics and predictions from computational models. Both NAM-based PODs and comparative *in vivo* data are based on oral route of exposure and, hence, the findings have primary application in oral exposure scenarios. Considerations for the application of the current approach to address other exposure routes should be the subject of future work. This work produces information for chemical screening and determines differences in relative potency that will be used to further inform hazard characterization and prioritization efforts in identifying substances for further work in Canada. These are precursor analyses to inform scoping and future risk assessment activities under CEPA. The current work is not intended to reflect the considerations for a comprehensive hazard identification or risk assessment.

2.2. Target chemicals

Target chemicals that were selected from Canadian chemical inventories were included if, i) they are structurally similar to BPA (n = 18) (i.e., contained a bisphenol substructure) (Table 1) or, ii) have evidence of potential use as functional alternatives to BPA (n = 7) (Table 2). Bisphenol-like structural filters for the chemicals considered in this case study are as follows: 1) the presence of two phenyl rings connected by not more than one atom; 2) the two phenyl rings may not be multi-ring (e.g., naphthalene, anthraquinone, etc.) and substituents on phenyl ring can be anything other than a ring structure; and 3) have no more than one free hydroxyl group on each phenyl ring. Functional alternatives were chosen based on responses to a Canadian industry survey about BPA replacements or being noted as a replacement for BPA in various international reports (US EPA, 2012; ANSES, 2013; Danish EPA, 2014; US NTP, 2017). Further details are given in the Government of Canada technical consultation document on Bisphenol A Structural Analogues and Functional Alternatives (ECCC and HC, Environment and Climate Change Canada and Health Canada, 2020).

Table 1. Target chemicals with a bisphenol substructure.

CAS Number	IUPAC Name **(Common Name)**	Chemical Structure
80-05-7	4-[2-(4-hydroxyphenyl)propan-2-yl]phenol **(Bisphenol A; BPA)**	

77-40-7	4-[2-(4-hydroxyphenyl)butan-2-yl]phenol **(Bisphenol B; BPB)**	
6807-17-6	4-[2-(4-hydroxyphenyl)-4-methylpentan-2-yl]phenol **(HPMPP)**	
2081-08-5	4-[1-(4-hydroxyphenyl)ethyl]phenol **(Bisphenol E; BPE)**	
79-97-0	4-[2-(4-hydroxy-3-methylphenyl)propan-2-yl]-2-methylphenol **(Bisphenol C; BPC)**	
620-92-8	4-[(4-hydroxyphenyl)methyl]phenol **(4,4'-Bisphenol F; 4,4'-BPF)**	
2467-03-0	2-[(4-hydroxyphenyl)methyl]phenol **(2,4'-Bisphenol F; 2,4'-BPF)**	
2467-02-9	2-[(2-hydroxyphenyl)methyl]phenol **(2,2'-Bisphenol F; 2,2'-BPF)**	
129188-99-4	4-[1-(4-hydroxyphenyl)-3,3,5-trimethylcyclohexyl]phenol **(Bisphenol TMC; BPTMC)**	
5613-46-7	4-[2-(4-hydroxy-3,5-dimethylphenyl)propan-2-yl]-2,6-dimethylphenol **(Tetramethyl Bisphenol A; TMBPA)**	

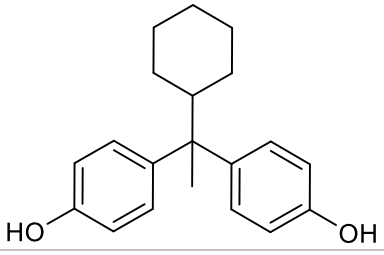
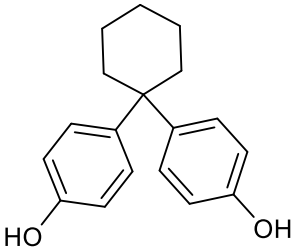
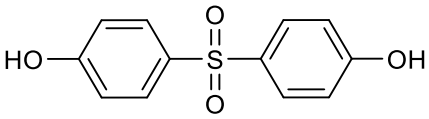
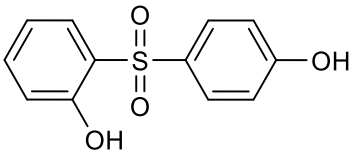
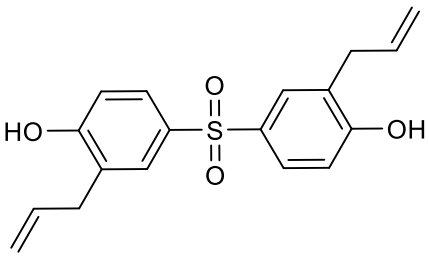
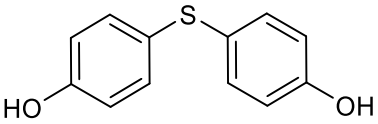
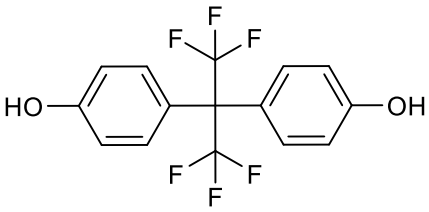
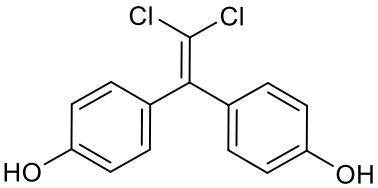
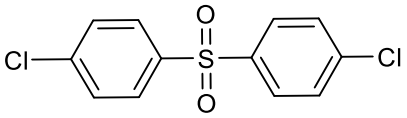
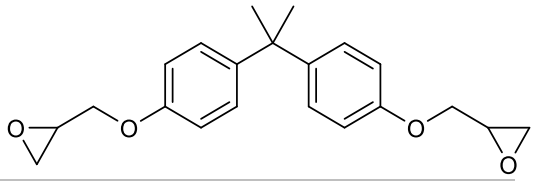
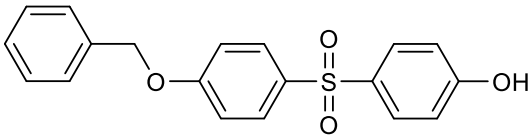
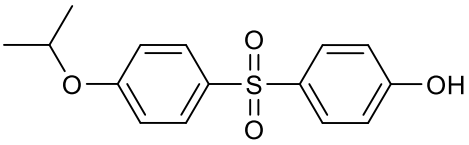
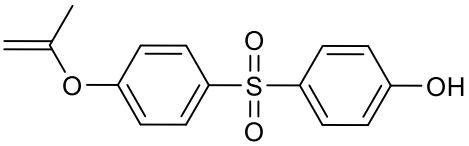
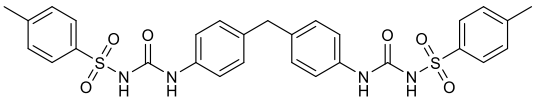
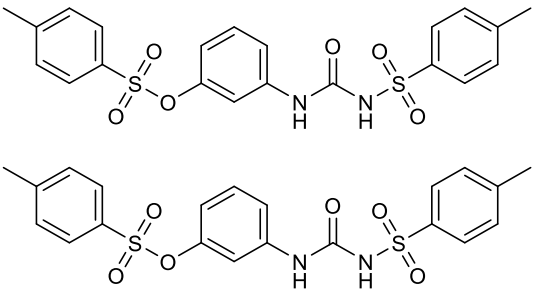
1571-75-1	4-[1-(4-hydroxyphenyl)-1-phenylethyl]phenol **(Bisphenol AP; BPAP)**	
843-55-0	4-[1-(4-hydroxyphenyl)cyclohexyl]phenol **(Bisphenol Z; BPZ)**	
80-09-1	4-(4-hydroxyphenyl)sulfonylphenol **(Bisphenol S; BPS)**	
5397-34-2	2-(4-hydroxyphenyl)sulfonylphenol **(2,4'-Bisphenol S; 2,4'-BPS)**	
41461-66-7	4-(4-hydroxy-3-prop-2-enylphenyl)sulfonyl-2-prop-2-enylphenol **(TGSA)**	
2664-63-3	4-(4-hydroxyphenyl)sulfanylphenol **(4,4'-Thiodiphenol; 4,4'-TDP)**	
1478-61-1	4-[1,1,1,3,3,3-hexafluoro-2-(4-hydroxyphenyl)propan-2-yl]phenol **(Bisphenol AF; BPAF)**	
14868-03-2	4-[2,2-dichloro-1-(4-hydroxyphenyl)ethenyl]phenol **(Bisphenol C 2; BPC2)**	

Table 2. Target chemicals that have evidence of potential use as a functional alternative to BPA.

CAS Number	IUPAC Name **(Common Name)**	Chemical Structure
80-07-9	4,4'-Dichlorodiphenyl sulfone **(DCDPS)**, or Bis(4-chlorophenyl) sulfone **(BCPS)**	
1675-54-3	2-[[4-[2-[4-(oxiran-2-ylmethoxy)phenyl]propan-2-yl]phenoxy]methyl]oxirane **(BADGE)**	
63134-33-8	4-(4-phenylmethoxyphenyl)sulfonylphenol **(BPS-MPE)**	
95235-30-6	4-(4-propan-2-yloxyphenyl)sulfonylphenol **(D-8)**	
97042-18-7	4-(4-prop-2-enoxyphenyl)sulfonylphenol **(BPS-MAE)**	
151882-81-4	1-(4-methylphenyl)sulfonyl-3-[4-[[4-[(4-methylphenyl)sulfonylcarbamoylamino]phenyl] methyl]phenyl]urea **(BTUM)**	
232938-43-1	[3-[(4-methylphenyl)sulfonylcarbamoylamino]phenyl] 4-methylbenzenesulfonate **(Pergafast 201)**	

2.3. Endpoint(s)

This IATA is intended to provide a series of NAM-based points of departure for both general systemic toxicity (POD_{NAM_SYS}) and ER pathway specific activity (POD_{NAM_ER}) for 25 substances, including BPA and its alternatives. In addition, the IATA provides qualitative hazard flags that are relevant to hazard

identification and risk assessment of structurally related bisphenols and their respective functional alternatives.

POD_{NAM_SYS} addresses the assessment of general systemic toxicity and is “hazard independent”, in that it does not predict the potential for specific adverse effects. This POD is a conservative estimate of doses where biological activity begins to occur and is considered protective against adverse effects that are typically observed at higher doses. The POD_{NAM_SYS} uses available ToxCast™ bioactivity data as well as *in vitro* high-throughput transcriptomics (HTTr) data from experiments conducted at Health Canada. As a NAM, HTTr uses gene expression profiles to rapidly evaluate numerous chemicals using *in vitro* data that, through recent decreases in costs to generate transcriptome profiles has become a feasible method to facilitate risk assessment of chemicals (Harrill *et al.*, 2019).

POD_{NAM_ER} addresses the potential for estrogen receptor (ER) activation in this group of substances. POD_{NAM_ER} is considered pathway specific as it is derived from expression of genes related to the ER pathway and ER biomarker signature from experiments conducted at Health Canada. Likewise, POD_{NAM_ER} is considered protective of the potential adverse effects that are observed from exposure to higher doses *in vivo*.

3 IATA Objectives

The central objectives for this IATA-based approach are as follows:

a. Derive a minimal concentration corresponding to a bioactivity threshold observed in targeted *in vitro* assays (ToxCast) and HTTr that can be defined for both general systemic toxicity and for ER activation.

b. Couple these minimal bioactivity threshold concentrations with *in vitro* to *in vivo* extrapolation (IVIVE) to estimate surrogate NAM-based points of departure for both general systemic toxicity (POD_{NAM_SYS}) and potential ER mediated effects (POD_{NAM_ER}).

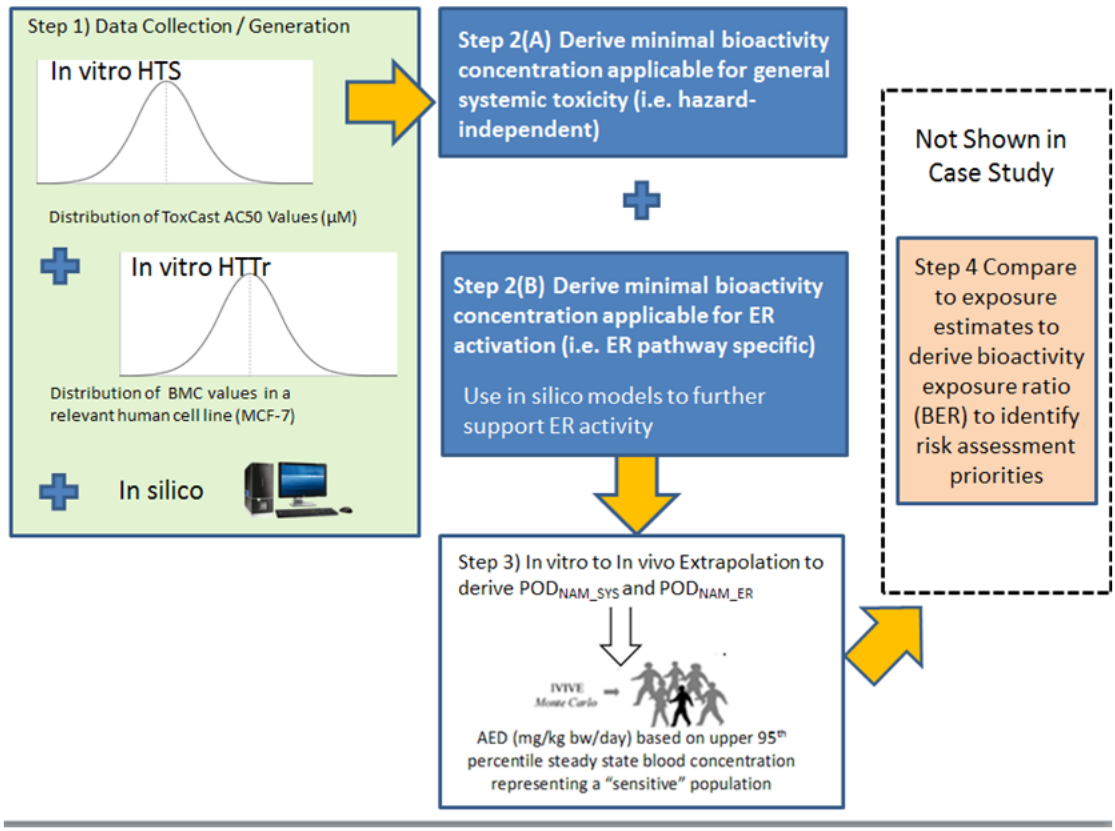
c. Demonstrate that POD_{NAM_SYS} and POD_{NAM_ER} can serve as a lower bound (i.e. protective) estimate of effect levels that may be observed *in vivo* which would be suitable for prioritization and for integration into assessment efforts particularly where *in vivo* data are lacking.

4 Approaches

4.1. IATA Workflow

This IATA case study follows the workflow outlined in Figure 1 and the general steps are briefly described here. More detail on data inputs and models used in the IATA is provided in Section 5.

Figure 1. IATA workflow for prioritizing bisphenols and related compounds using *in vitro* bioactivity for general systemic toxicity and ER activity.



The first step of the workflow is to collect and generate the required data to develop the IATA. For this case study, this step included collecting results from the ToxCast™ assay for 25 compounds of interest, where available. Additionally, HTTr work was conducted at Health Canada for 17 of these substances. HTTr testing for all 25 substances was not possible at this time due to resource and time constraints. However, efforts were made to include substances not available within the ToxCast database.

The second step of the IATA involves deriving the minimal bioactivity concentration. Two tracks of interest were considered in order to compare the general systemic toxicity to capture non-specific hazard-independent effects to pathway-specific activation of the estrogen receptor, a known target of BPA and its analogues. For general systemic toxicity, minimal bioactivity concentrations were selected from both ToxCast and *in vitro* HTTr experiments. The ToxCast concentration was selected as the 5th percentile from the distribution of half-maximal activity concentrations (AC_{50} values) from active assays (i.e., after filtering out unreliable assays). This lower-bound estimate of the ToxCast database is used to derive an administered equivalent dose (AED) using IVIVE modeling that is then compared to the population exposure level to obtain a bioactivity exposure ratio (BER); the details of this approach and methodology have been described elsewhere (Paul Friedman *et al.*, 2020). For the HTTr analysis, several techniques for determining a minimum bioactivity concentration were explored (see Section 5, subsection *ii*). For the purposes of the IATA approach, the most fitting HTTr-based minimal bioactivity concentration was determined in consideration with endpoints derived from ToxCast (see Section 6). When determining the final minimal bioactivity concentration for general systemic toxicity, the minimum from either the ToxCast approach or the HTTr approach was carried forward for IVIVE. For assessing the ER activity of the compounds, a subset of the ToxCast assays as well as the ToxCast ER area under the curve (AUC) model were used (where available). Additionally, the HTTr results were used to interrogate an ER transcriptomic biomarker (Ryan *et al.*, 2016). ER activation was also assessed using the generated differentially expressed genes (DEGs) through the identification of predicted upstream regulators using Ingenuity Pathway Analysis (IPA) platform (ER IPA target gene analysis). The minimal bioactivity concentration corresponding to ER activation was assessed by examining both the lowest observed effect concentration (LOEC) from the ER biomarker signature and through the calculation of the benchmark concentration (BMC) at the 5th percentile for target genes of ER in IPA. For the IATA, a minimum ER bioactivity concentration was selected based on the minimum of either of these two HTTr approaches and was carried forward for IVIVE. Finally, to further assess the potential for ER activation, *in silico* models were also used to generate predictions and to serve as qualitative hazard flags to support the conclusion of ER activity from both ToxCast and the HTTr approaches. All scripts and code that were used to perform differential gene expression analysis are available from the following project repository (https://github.com/R-ODAF/R-ODAF_Health_Canada).

The third step in the IATA approach converted the minimal bioactivity concentrations determined in the second step for both general systemic effects and ER activation to AEDs through the application of an HTTK model (Pearce *et al.*, 2017). These AED values serve as surrogate points of departure (POD_{NAM}) that are considered conservative lower bound estimates for effects that may be observed at higher doses *in vivo*.

Within the last step in the IATA workflow, the $PODs_{NAM}$ can be compared to exposure estimates for the population to prioritize chemicals for further risk assessment activities. Although it should be noted that exposure estimates for the substances in this case study are not provided, in practice the POD_{NAM} derived within this IATA would be compared against human exposure estimates for the general population in Canada to derive BERs. The magnitude of a BER determines the priority of a specific substance for further assessment work (i.e., chemicals with lower BERs are identified to have greater risk potential compared to those with higher BERs). It was anticipated that both exposure estimation and assessing the adequacy of the BER would be country and regulatory context specific. More information on Health Canada's exposure assessment[3] and assessing BER adequacy for prioritization[4] under the Chemicals Management Plan (CMP) are available online.

[3] https://www.canada.ca/en/health-canada/services/chemical-substances/fact-sheets/assessing-exposure-canadians-environment-substances-products.html

[4] https://www.canada.ca/en/environment-climate-change/services/evaluating-existing-substances/science-approach-document-bioactivity-exposure-ratio-application-priority-setting-risk-assessment.html

5 Data Collection / Generation (IATA STEP 1)

5.1. High-throughput *In Vitro* Screening (ToxCast)

5.1.1. Minimal Bioactivity Concentration Applicable for General Systemic Toxicity (Hazard Independent)

The methods to select a minimal bioactivity threshold from the distribution of assays within the ToxCast database followed the process outlined elsewhere (Filer *et al.*, 2017; Paul Friedman *et al.*, 2020). Briefly, *in vitro* bioactivity was obtained from the publicly available MySQL ToxCast database (invitrodb_v3.1) (US EPA, 2019) extracted using the ToxCast Data Analysis Pipeline (tcpl) (version 2.0) package (Filer *et al.*, 2017) in R (version 3.5.3) (R Core Team, 2013). The ToxCast database and the methods used for curve fitting, determining activity (hit-call) in assay endpoints, and quantifying the respective uncertainty for these methods are described in detail elsewhere (Filer *et al.*, 2017; Watt and Judson, 2018). The generic filtering criteria that were used for assays with an active hit-call (i.e., assays deemed active for given chemical) have also been described elsewhere (Paul Friedman *et al.*, 2020). The aim was to eliminate less reproducible or less reliable activity calls and respective AC_{50} values for quantitative use when deriving the POD_{NAM}. Specifically, curves were required to have less than 3 caution flags from tcpl, an AC_{50} value greater than the lowest concentration screened, and a hit percent of ≥ 50 %. For more details on caution flags from tcpl and for how the hit percent is calculated, readers are referred to Filer et al., 2017 and Watt and Judson, 2018, respectively.

The 5th percentile from the distribution of AC_{50} values from the previous step was selected to represent the *in vitro* bioactivity threshold applicable for general systemic toxicity. The 5th percentile of AC_{50} values was selected over the minimum AC_{50} value to limit the influence of potential extreme AC_{50} values that may have resulted from the limitations inherent to the generalized curve fitting process applied in tcpl. The 5th percentile balances the desire to select a conservative bioactivity threshold informed by the whole distribution of AC_{50} values without relying on potentially non-representative outlier values at the extremes (i.e., the minimum). Cytotoxicity was not considered when filtering out AC_{50} values from active assays. It is possible that some of the bioactivity observed in the distribution of active assays is confounded by cytotoxicity and the “burst” phenomenon, where large numbers of assays begin to show activity near cytotoxic concentrations. However, the 5th percentile is used to define the bioactivity threshold which is likely to be below the threshold for cytotoxicity for most chemicals.

5.1.2. Hazard Flags to Support ER Activation

ToxCast contains various targeted assays to elucidate the potential of chemicals to interact with the ER. Moreover, the ToxCast-based ER Bioactivity model integrates the concentration-response curves across 16 ToxCast ER assays into a computational model. The model, described in detail elsewhere

(Judson *et al.*, 2015; Organisation of Economic Co-Operation and Development, 2019), discriminates bioactivity from assay-specific interference and responses related to cytotoxicity. An AUC score that is relative to the activity seen for the endogenous hormone, 17β-estradiol, the estrogen receptor ligand, is provided by the computational model. A set of reference compounds, for which guideline-type studies were available for the *in vivo* uterotrophic assay, were previously used to evaluate the model, and the results demonstrated that it was highly predictive of ER activity (Browne *et al.*, 2015), supporting its application within this IATA. For eight of the 25 compounds in this case study, testing was available for all 16 assays in the model (US EPA, 2015) and agonist AUC scores ≥ 0.1 were considered positive while a model score of 0.1 > AUC > 0.001 were considered inconclusive as described in (OECD, 2018). Twelve compounds had a portion of the 16 ER assays but not the full suite required for application of the full ER AUC model. In a previous OECD IATA Case Studies project, a workflow was proposed on how to apply the AUC model without the full suite of assays available (OECD, 2019). The model could be used with as little as four assays and still maintain a high level of balanced accuracy. The minimal four assay set consists of (as outlined in OECD 2019 Annex 3) NVS_NR_hER, OT_ER_ERaERb_1440, ATG_ERa_TRANS_up, and ACEA_T47D_80hr_Postive. No additional case study chemicals had results in all four assays beyond the eight substances that had a full 16 assay set. For our purposes, the additional substances with partial ToxCast/Tox21 testing were considered to be ER “active” in ToxCast where the majority of available ER agonist assays had a positive hit call. If the majority of available ER assays were not considered active, the substance was considered “inactive”. This majority rule approach has not been validated against reference chemicals and could be influenced by assay specific sensitivities, interference and/or cytotoxicity. The ToxCast ER AUC model results and majority rule approach are both used as qualitative flags for activity in this case study.

5.2. High-throughput Transcriptomics (HTTr)

5.2.1. Gene expression analysis

HTTr data was used in this case study to support potency ranking of the various bisphenol compounds and to provide a weight of evidence in the assessment of estrogenic activity of each chemical (i.e., antagonist or agonist). Michigan Cancer Foundation-7 (MCF-7) breast cancer cells (ATCC, Manassas, VA) cultured in a medium free of estrogenic compounds were exposed to each case study chemical at ten different concentrations (0.0005, 0.001, 0.01, 0.1, 0.5, 1, 5, 10, 50, 100 µM), and three controls - solvent control (0.1% DMSO). Experimental controls 17β-Estradiol (E2; 0.0001, 0.001, 0.01, 0.1, 1, 10 nM) and non-estrogenic compound Dexamethasone (Dex; 0.0001, 0.001, 0.01, 0.1, 1 µM) were also included (data not shown). Cells were exposed in 96-well plates for 48 hours over multiple experimental replicates (n = 4). Cells were washed with phosphate-buffered saline and lysed for Templated Oligo-Sequencing (TempO-Seq®).

Gene expression was measured using the TempO-Seq Human Whole Transcriptome v2.0 Kit following the manufacturer’s instructions (BioSpyder Technologies Inc. Carlsbad, CA). TempO-Seq next-generation sequencing (NGS) libraries were sequenced using a NextSeq 500 High-Throughput Sequencing System (Illumina San Diego, California) using 50 cycles from a 75-cycle high-throughput flow cell. A median read depth of 2 million reads/sample was achieved. Following pre-processing of sequencing data to achieve high-quality data, NGS reads were aligned to the TempO-Seq Human WT probe set (22,537 probes over 19,687 genes) provided by BioSpyder using the TempO-SeqR analysis pipeline in R (version 3.1). After additional quality control measures, the data were normalized and DEGs identified following the recommendations set out by the Omics Data Analysis Frameworks for Regulatory application (R-ODAF) guidelines (Verheijen *et al.*, 2020) and the data were processed using a readily available bioinformatic pipeline (Verheijen *et al.*, 2022). The DEG cut-off was a +/- 1.5 fold change (on a linear scale) with a false discovery rate (FDR) adjusted p-value of 0.05.

5.2.2. Cell Stress, Concentration Filtering, and Cytotoxicity

DEGs were used to analyse the effect of chemicals at different concentrations to determine a POD resulting in some concentrations being omitted (see Appendix A for additional information). The DEGs for BPA (CAS-RN: 80-05-7) at 0.001 and 0.01 µM were dropped due to technical issues with the NGS libraries. The nuclear factor erythroid-2 related factor 2 Nrf2 and MTF-1 transcriptomic biomarkers have previously been published as effective tools for detecting cellular stress signalling pathways (Jackson *et al.*, 2020; Rooney *et al.*, 2020); the data supporting cell stressor biomarkers HSF1, NFkB, and HIF1a have manuscripts that are still in preparation.

Sample concentrations were subject to exclusion based on properties of solubility and cytotoxicity. Concentrations were removed if found to be insoluble in experimental media (e.g., DMSO) or cell culture media. At the end of the 48-hour exposure, cell viability was measured using the CellTiter-Blue Cell Viability Assay performed according to guidelines from the manufacturer (Promega Corp, Madison, WI, USA). Cytotoxicity was then further investigated by examining cell viability after depletion of estrogenic compounds by maintaining them for 3 days in 5% charcoal stripped FBS in phenol red free media. A full description of solubility and cytotoxicity experiments and excluded concentrations is present in Appendix B. Cytotoxicity Assay and Solubility.

5.2.3. Benchmark Concentration Models

BMC modelling was conducted prior to examining different approaches for determining the minimal bioactivity concentrations from HTTr. Briefly, the raw datasets were normalized using log2 DESeq2 normalized counts (counts were shifted by one log2(counts + 1) to avoid issues with 0's). The normalized datasets were imported to BMDExpress v2.3. The probes were pre-filtered using the Williams Trend test and a fold-change filter (≥ 1.5-fold). The filtered data were fit to curve models: Power (power term restricted to > 1), Linear, Polynomial 2°, Exponential 2°, Exponential 3°, Exponential 4°, and Exponential 5°. The criteria to determine the best fit include: (1) a nested Chi-square test cut-off of 0.05 for linear model selection; (2) the least complex model based on the Akaike Information Criterion (AIC) for the Linear and Power models; and (3) a goodness-of-fit test p value > 0.1. The BMC is the benchmark concentration that corresponds to a benchmark response (BMR) of 1 standard deviation. Post-filtering of BMC's included removing BMC's a goodness-of-fit test p value less than 0.1, BMC divided by the BMC lower bound greater than 20, BMC upper bound divided by the BMC lower bound greater than 40 and BMC's that were greater than the highest dose. Probes that met all BMC filtering criteria were converted to their corresponding Entrez Identifiers.

5.2.4. Minimal Bioactivity Concentration Applicable for General Systemic Toxicity

Several approaches were explored for deriving candidate minimal bioactivity concentrations from the HTTr data. These approaches considered, i) the median BMC of the lowest most-sensitive gene sets established by the US National Toxicology Program (NTP); ii) the lowest of either the 10th percentile or first mode BMC; iii) the 5th percentile BMC; and iv) the 25th ranked gene BMC derived from gene accumulation plots. Previous studies have indicated that NAM-based points of departure using high-throughput screening (e.g., ToxCast) are generally more conservative when compared to traditional points of departure using *in vivo* models (Paul Friedman *et al.*, 2020). The advantages of considering several approaches are two-fold: (a) it not only allows for determination of the reproducibility of transcription-based PODs between several chemicals within the same class (e.g., class of bisphenols) and (b) it allows for identification of the most conservative PODs when compared to previously

established methods (e.g., *in vitro* PODs of the 5th percentile AC50 values using ToxCast data) (Harrill *et al.*, 2021).

Approach 1 (NTP approach)

We analysed the HTTr data sets using parameters in accordance with the US NTP recommended approach (US NTP, 2018). This approach was established to achieve greater consistency in genomic dose-response modelling and to facilitate the use of genomics data in risk assessment (US NTP, 2018). The NTP, EPA, and Health Canada developed BMDExpress 2.0 to allow users to apply this approach for analysing dose-response data produced in HTTr experiments (Phillips et al., 2019).

The genes and their associated BMC/Benchmark Concentration Lower-bound (BMCL)/Benchmark Concentration Upper-bound (BMCU) values were parsed into the Gene Ontology (GO) Biological Process (Harris et al., 2008; Fabregat et al., 2016; Kanehisa et al., 2017). Gene sets that contain at least three gene BMCs (genes that pass all criteria in the analysis) and are at least 5 % populated (based on total annotated gene number) were selected. As per NTP's recommendations, the lowest (i.e., most sensitive) gene sets from this analysis, based on median BMC, were selected as the PODs (US NTP, 2018).

Approach 2: Lowest of 10th Percentile or First Mode

Another approach involves the derivation of minimal bioactive concentration using the lowest value of either the 10th percentile or first mode of the BMC distribution; these methods were adapted from Page-Lariviere (Pagé-Larivière, Crump and O'Brien, 2019). Briefly, the filtered BMC results, as described above, were also used in this analysis. The BMC derivation methods involve the 10th percentile estimate from the BMC distribution and the value of the first mode from the BMC distribution. The BMC distribution was estimated using the Sheather and Jones bandwidth selection method using a 5% minimum probability density. The first mode was determined using empirical estimates for the first and second derivative of the BMC distribution using forward, centre and backward differencing.

Approach 3: 5th Percentile BMC

The lower bound 5th percentile of all genes with BMCs was used to derive a transcriptomic bioactivity concentration to be similar to the approach used to derive the lower bound estimate bioactivity concentration from ToxCast data previously described. The 5th percentile BMC was derived similar to the 10th percentile methods. However, the BMC selected for the 5th percentile is established by the gene BMC that represents the 5th percentile that is based on the rank ordering of those genes with BMCs, rather than an estimated interpolation of the 5th percentile from the empirical BMC distribution.

Approach 4: Gene Accumulation Plots

Gene accumulation plots of individual genes with BMCs were generated following protocols from previous studies (Ramaiahgari *et al.*, 2019; Reardon *et al.*, 2021) to support potency ranking of bisphenol substances. Briefly, within these plots genes are rank-ordered sequentially from those genes with the lowest BMCs to genes with the highest BMCs, with incremental increases in gene accumulation number on the y-axis and BMC value on the x-axis. A gene accumulation curve is produced that

indicates the transcriptional activity over a given exposure concentration range, with the most sensitive genes having the lowest value BMCs. For example, bisphenols eliciting a steep curve with an increased number of accumulated genes with BMCs have increased transcriptional activity. For our purposes, the 25th rank-ordered gene from the accumulation plots was considered a conservative estimate to a minimal bioactive concentration, representing a value where transcriptional activity begins to increase exponentially. The 100th ranked gene BMC was also considered, and details are listed in Appendix C: Gene Accumulation Plot Rankings and Cut-offs. A comparison of the relative chemical potency across the 25th and 100th ranked gene BMCs indicated a consistent ranking of relative potency among these chemicals, regardless of the endpoint used demonstrating the robustness of using gene accumulation plots for potency ranking (Appendix C: Gene Accumulation Plot Rankings and Cut-offs).

5.2.5. Minimal Bioactivity Concentration Applicable for ER Activation (Pathway Specific)

Approach 1: Application of ERα transcriptomic biomarker

To classify the ER agonism or antagonism of the bisphenols using HTTr data, the expression profiles of each bisphenol were compared against an ERα biomarker signature following a previously established protocol (Ryan *et al.*, 2016). Briefly, the correlation between the bisphenol expression pattern and ERα biomarker signature was assessed using the Running Fisher algorithm. The p-value and direction of the correlation was used in the classification. P-values were converted to −log(p-value)s, where a positive value indicates ERα agonism and a negative value signifies ERα antagonism. A threshold of 3.5 was applied to capture agonists/antagonists, which is marginally more conservative compared than the published threshold of 4 (Ryan *et al.*, 2016). The lowest-observed-effect-concentration (LOEC) for each chemical where the −log(p-value) was above the threshold was reported. It should be noted that the experimental design for this study was intended to detect agonism and not antagonism since estrogen was not present. Therefore, true antagonism of the ER would not be determined with high confidence.

Approach 2: ER IPA target gene analysis

Gene expression data were analysed using IPA (QIAGEN Redwood City, www.qiagen.com/ingenuity) to identify inhibited/activated upstream regulators and their target genes. IPA Core Analysis of differentially expressed genes was carried out using the indirect and direct relationship settings, focusing on human data sources. From the three confidence levels provided by the IPA, we used “Experimentally observed” and “Highly predicted” data. Statistical significance of the overlap ($p \leq 0.01$) between the dataset and known targets of a particular upstream regulator in IPA were calculated using Fisher’s Exact tests. The z-score was calculated based on the expected relationship for directions between upstream regulators and target genes and those observed in the dataset. A z-score >2 (activated) or <-2 (inhibited) was considered statistically significant.

5.3. High throughput toxicokinetic data (HTTK) and *in vitro* to *in vivo* extrapolation (IVIVE)

The methods used to derive AEDs are based on the model described in (Wetmore *et al.*, 2012) and the methodology following the current protocol is described by Beal *et al.*, 2021. Specifically, plasma protein binding and hepatic clearance data from *in vitro* assays made available in the HTTK R library or from

in silico predictions made by ADMET Predictor v10 are used to parameterise a generic pharmacokinetic model. The model is used to estimate a steady-state plasma concentration (C_{ss}) at a given administered dose under the assumption of 100% bioavailability. A distribution of C_{ss} values is modelled using a Monte Carlo simulation that accounts for physiological variability of a healthy human population. Assuming that the relation between steady state plasma concentration and administered dose are linear, reverse dosimetry can be used to estimate the oral dose required to reach a steady state plasma concentration equal to the BMC or AC_{50}. All IVIVE modelling was performed using the R HTTK library version 2.0.3.

The *in silico* application ADMET Predictor, developed by Simulations-Plus[5] is designed to predict absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties of a chemical. Using 2D molecular structure ADMET Predictor v10 was able to predict plasma protein binding and hepatic clearance of the substances which are useful in the calculation of a $C_{ss.}$ For our purposes, the predictor was used to parameterize HTTK and IVIVE modelling when *in vitro* toxicokinetics data were not available.

5.4. Other *in vitro* Supporting Data

In this work, *in vitro* ER activity data were taken from the Collaborative Estrogen Receptor Activity Prediction Project (CERAPP) that contains information on 7523 unique substances including information on receptor binding, agonism and antagonism (Mansouri *et al.*, 2016). The data for CERAPP came from several sources including Tox21, the U.S. EPA Estrogenic Activity Database, the Japanese Ministry of Economy, Trade, and Industry database, and the ChEMBL database. Curation was applied to the database to remove any ambiguous data, cytotoxicity information, and all *in vivo* data. In this case study, the *in vitro* data were used to give qualitative support for other results as well as being used to train an in-house developed random forest model for ER activity predictions.

5.5. (Quantitative) Structure Activity Relationship [(Q)SAR]

The following (Q)SAR models were selected for this case study based on their availability (i.e., freely accessible), reliability (i.e., developed using relatively large databases), and having been subject to peer-review. However, it should be noted that the random forest (RF)-model is in preparation for publication.

5.5.1. Random Forest Models

Random Forest (RF) models were developed and trained using in-house written code. The RF models were trained using the CERAPP validation dataset and were trained to predict binding, agonism, and antagonism. The models used PubChem fingerprints, OpenBabel fingerprints (FP2, 3, and 4), and OpenBabel calculated physico-chemical properties as the descriptors. Binary and class models were developed for each activity: estrogen binding, agonism, and antagonism. For all six models, a total of 101 trees were trained up to a total possible depth of 10 nodes. After training, each model was optimized using an in-house developed genetic algorithm to alter the depth of each tree to maximize the predictive capabilities of the RF. Each RF model had an applicability domain developed to ensure confidence in the predictions. After training and optimization, all substances in this case study had all activities predicted by the RF models. More information on the RF models can be found in the QMRF (see Appendix D: RF Model QMRF).

[5] www.simulations-plus.com/Products.aspx?grpID=1&cID=11&pID=13

5.5.2. CERAPP Consensus Models

The intention of the CERAPP work was to develop a consensus model which incorporates multiple (Q)SAR models into a single prediction (Mansouri *et al.*, 2016). The models used in the CERAPP consensus model were developed from CompTox data by research groups around the world. The models were sent back to the US EPA and then weighted based on their performances. Predictions were for a binding, agonism, and antagonism both for binary and multi-class predictions. The results for the CERAPP models are available for most of the studied substances on the US EPA Comptox website.[6]

5.5.3. OECD QSAR Toolbox

This is an *in silico* application that facilitates grouping of chemicals within categories based on similarity of molecular structure, mechanism (or mode) of action, physico-chemical properties, and metabolism to[7] was used to carry out expert assessment of substances belonging to the bisphenols subgroup. Furthermore, it helped to gain some insight on potential ER binding as well as developmental and reproductive toxicity (DART) based on presence of alerts and DART decision-tree, respectively.

5.6. *In vivo* Data

In vivo toxicity data were collected, where available, for the case study chemicals. Specifically, POD values (i.e., No Observed (Adverse) Effect Level (NO(A)EL) and Lowest Observed (Adverse) Effect Level (LO(A)EL)) from oral repeat dose, developmental and reproductive toxicity studies were taken from EPA's ToxVal database that is available through the CompTox Dashboard[8], as well as from the publicly available ECHA REACH dossiers. For the purposes of this case study, all available effect levels were collected for each substance but the specific *in vivo* adverse effects were not explicitly examined from the studies. Rather, the POD values are used to compare against the POD_{NAM} values when determining whether the POD_{NAM} may be considered a protective metric for establishing BERs and useful for prioritization when *in vivo* data are not available. Moreover, *in vivo* uterotrophic data (TG 440) were collected, where available, from the NICEATM Uterotrophic Database[9]. Five chemicals out of the 25 compounds studied in this IATA case study were found in NICEATM dataset. These 5 chemicals (bisphenol A, bisphenol B, bisphenol F, bisphenol S, and BPAF) were all found to significantly increase the mean uterine weight of test groups that indicated positive responses in this bioassay. Uterotrophic assay results, where available, were used to provide a comparison to the ER activity seen in the *in silico* models, transcriptomics assays, and ToxCast results used in this case study.

[6] https://comptox.epa.gov/dashboard

[7] https://www.oecd.org/chemicalsafety/risk-assessment/oecd-qsar-toolbox.htm

[8] https://comptox.epa.gov/dashboard

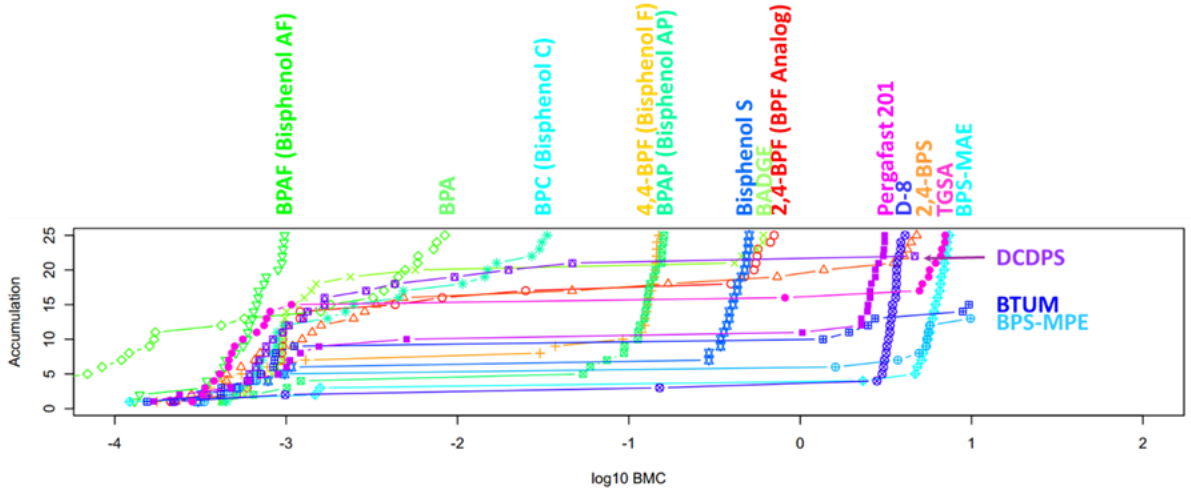
[9] https://ntp.niehs.nih.gov/whatwestudy/niceatm/test-method-evaluations/endocrine-disruptors/ref-data/edhts.html

6 Determining Minimal Bioactivity Concentrations (IATA STEP 2)

6.1. Minimal Bioactivity Threshold Applicable for General Systemic Toxicity

As described in Section 5, several approaches were examined to develop candidate minimal bioactivity concentrations applicable for general systemic toxicity (e.g., calculated from 5th percentile gene BMC, the lowest value from either the 10th percentile or 1st mode gene BMC described previously (Pagé-Larivière, Crump and O'Brien, 2019) or the NTP's lowest gene set approach (US NTP, 2018) from GO analysis. Gene accumulation plots were also used to rank order potencies of the chemicals. A cut-off at the 25th rank-ordered gene BMC from the accumulation plot is presented in Figure 2 that is representative of early changes from baseline biology. The 25th ranked gene BMC allowed for comparison of the most-potent (furthest left on the x-axis) to the least-potent (furthest right on the x-axis) substances based on the value of the BMC (µM) of the 25th ranked gene (y-axis) for a given chemical (Figure 2). Within this plot, 25th ranked gene BMC could not be derived for BPS-MPE, BTUM, and DCDPS; this was attributed to a lack of transcriptional activity elicited from exposure to these substances and an indication that this approach differentiates substances of lower potency and with decreased hazard potential.

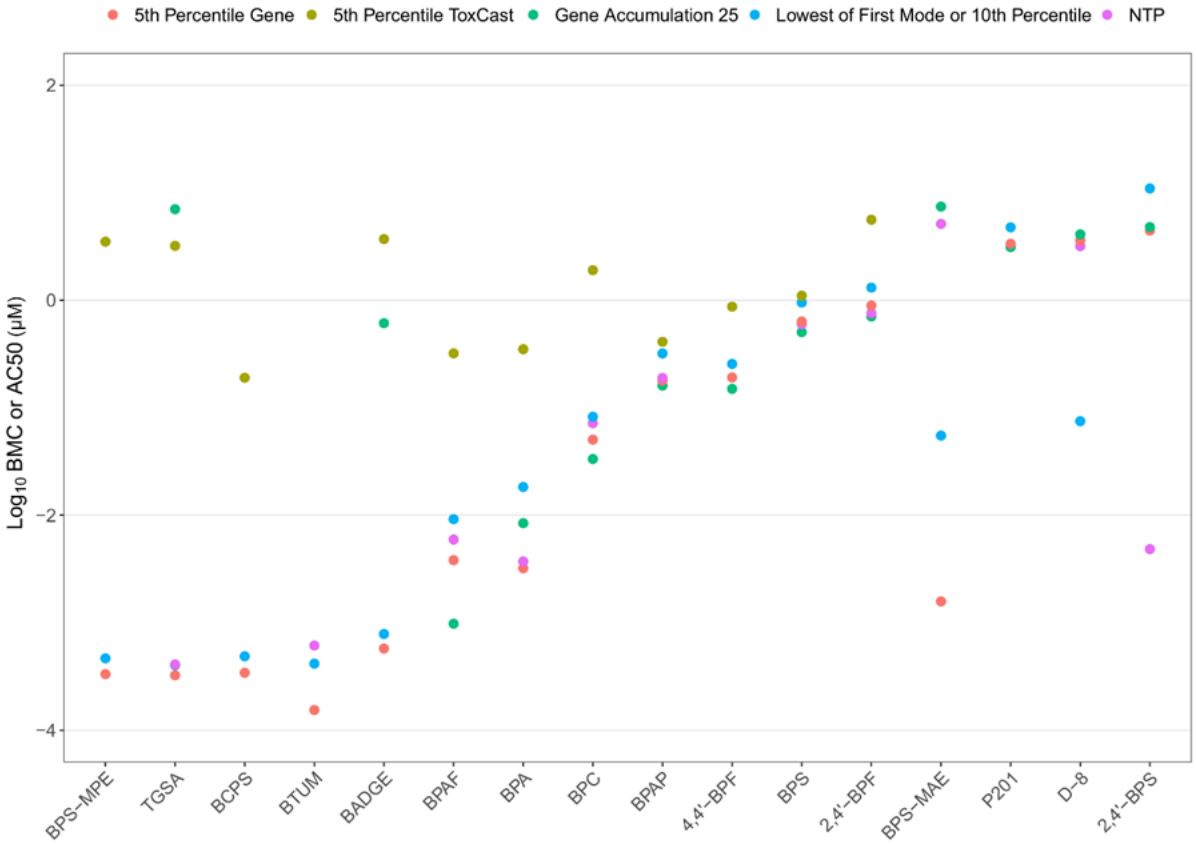
Figure 2. Gene accumulation plot of MCF-7 cells exposed to bisphenols and alternatives for 48 hrs.



Note: The log10 BMC value and each gene is represented a symbol and colour up to the 25th ranked gene BMC shown along the top of the plot.

The candidate minimal bioactivity concentrations were plotted using the log transformed BMCs for the 5th percentile gene, the lowest of either the first mode or the 10th percentile, lowest median gene set BMC (NTP approach), and the gene accumulation 25th gene BMC and depicted in Figure 3 (see Appendix C: Gene Accumulation Plot Rankings and Cut-offs for numerical values for BMCs derived from HTTr). Also included is the 5th percentile AC_{50} value from the distribution of active ToxCast assays.

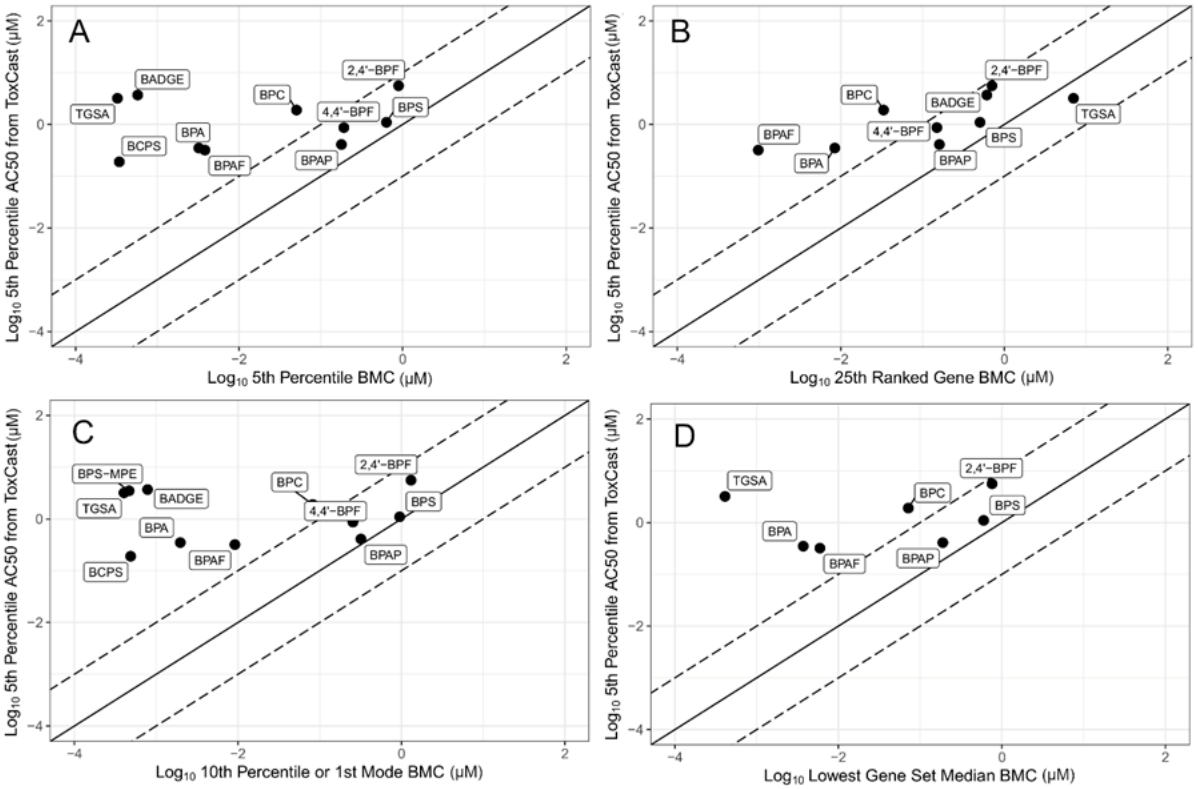
Figure 3. Comparisons of *in vitro* minimal bioactivity concentration for general systemic toxicity determined using multiple approaches.



In general, there was agreement across the HTTr approaches to estimate a minimal bioactivity concentration applicable for general systemic toxicity, with the values generally falling approximately within an order of magnitude of each other. Noted exceptions are the 5th percentile BMC and 10th percentile/1st mode for BPS-MPE, BCPS, BADGE, and to a lesser extent BPS-MAE, whose BMCs were orders of magnitude lower than other derived metrics. These substances, along with BTUM, and TGSA had fewer total genes with BMCs and could be inducing bias that manifests as a very sensitive BMC that is not an accurate representation of the minimal bioactivity concentration for these substances (Appendix E: Transcriptomic BMC values). It was also observed that most transcriptomic approaches were at least an order of magnitude lower than derived ToxCast metrics for BPAF, BPA, and BPC and the NTP approach was lower/more sensitive for 2,4'-BPS. Although there was general agreement across the HTTr approaches for a proportion of chemicals, there were discrepancies among those most potent/sensitive substances, particularly when using percentiles to derive values (i.e., the 5th or the 10th) from the distribution of genes with BMCs.

The 25th ranked gene BMC was observed to have the closest agreement with Toxcast values as 6 of the 9 portrayed substances, followed by the lowest gene set (3 of 7), the 5th percentile (4 of 10), and the lowest of the 10th percentile or 1st mode (4 of 11) approaches (Figure 4). For most scenarios, the transcriptomics approaches were more sensitive than the ToxCast approach, furthering the potential bias that is introduced for approaches deriving metrics using percentiles when excluding values at top concentrations as a result of insolubility or cytotoxicity. The 25th ranked gene BMC was carried forward for determining the minimal bioactivity concentration from the transcriptomics analysis. There are three notable exceptions, the ToxCast approach does not appear to be as sensitive as the transcriptomics approaches for BPAF, BPA, and BPC.

Figure 4. Comparison of the minimal bioactivity concentrations derived from the ToxCast and transcriptomics approaches using; A) 5th percentile gene BMC, B) the 25th ranked gene BMC, C) the lowest of the 10th percentile or 1st mode BMC, and D) the lowest and most sensitive gene set BMC (NTP approach).



Note: The solid line represents a theoretical perfect agreement between approaches whereas the dashed lines represent a ± log unit deviation.

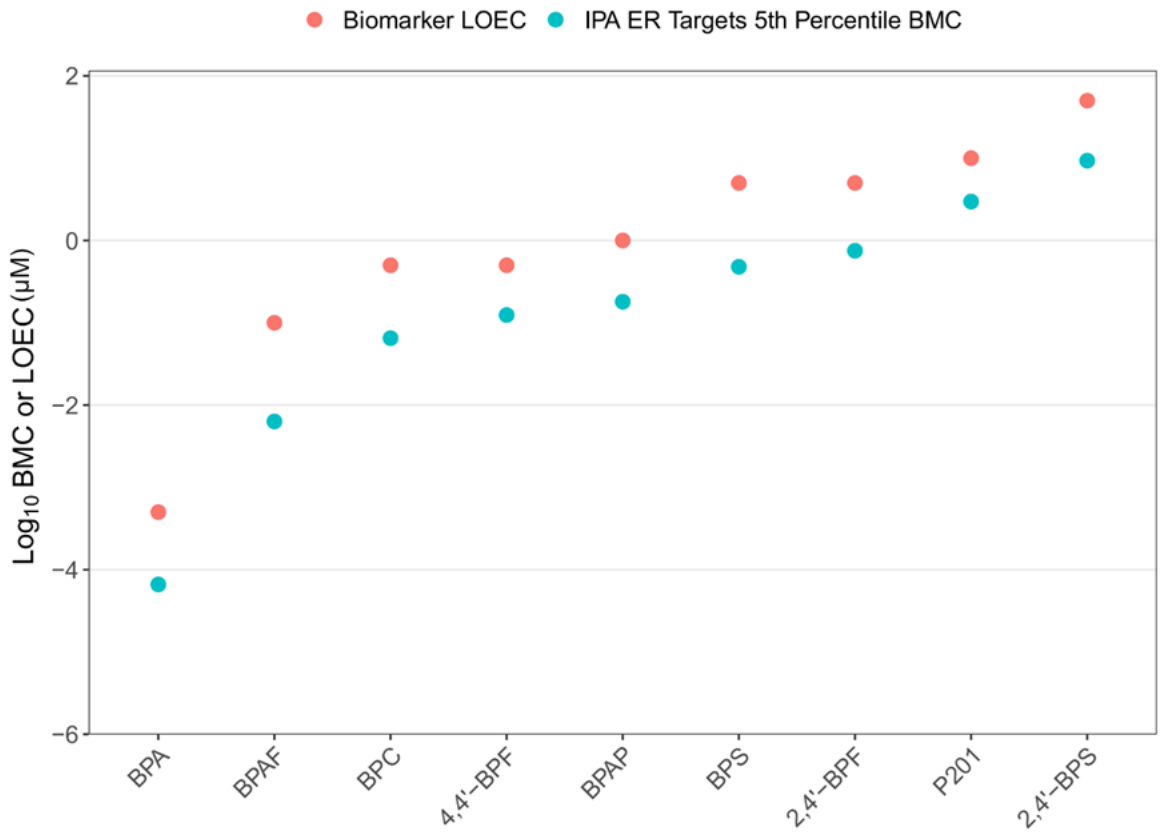
When selecting the minimal bioactivity concentration for IVIVE, the lower value of either the 25th ranked gene BMC approach examined, or the 5th percentile AC50 value from ToxCast was carried forward for derivation of POD_{NAM_SYS}.

6.2. Minimal Bioactivity Concentration Applicable for ER Activation (Pathway Specific)

The 46 gene biomarker was used for potency ranking based on ERα agonism. LOECs were used for this analysis because of the complex nature of the biomarker, which includes numerous genes. Although the biomarker LOECs were used, it should be noted that the 5th percentile BMC from the IPA ER target analysis were similar. This further increases the confidence of the ability to detect ERα activation using the 46 gene biomarker.

The total number of DEGs for each concentration of each chemical is shown in Appendix A. Cells shaded in grey represent the concentrations that were eliminated from our analysis based on a 'bioactivity burst' described above (i.e., high numbers of DEGs) or because of cytotoxicity or chemical insolubility (Appendix B. Cytotoxicity Assay and Solubility). The potential for each of the chemicals to interact with ERα was predicted using an ER-specific 46 gene biomarker as well as through IPA ER target gene analysis. Previous work on the ER transcriptomic biomarker demonstrated 94% and 93% balanced accuracy for ERα agonism and antagonism in the MCF-7 cell line (Ryan et al., 2016). Within this analysis, IPA ER upstream regulators with overlap p-value < 0.01 (Fisher's exact test) and z-scores > 2 (activated) are considered statistically significant and are shown in yellow (↑) (Appendix A). A z-score of -2 or less indicates inhibition of ER upstream regulators that was not observed at any concentrations of chemicals. In addition, cells are shaded in red to indicate the lowest concentration that significantly activated ERα using ER-specific 46 gene biomarker (also see Appendix F: ERα 46 Gene Transcriptomic Biomarker LOECs (µM) for log p-values of the transcriptomic biomarker). Using the transcriptomic biomarker approach, BPA had the lowest LOEC for ERα activation, at 0.0005 µM, followed by BPAF with a LOEC of 0.1 µM (Appendix G: IPA upstream regulator analysis to determine enrichment of genes regulated by ERα in the MCF-7 cells treated with test chemicals for 48hrs; Figure 5). Most chemicals in this study demonstrated potential for ERα agonism through the biomarker analysis at various concentrations (LOECs of 0.5 µM: BPC, 4,4'-BPF; 1 µM: BPAP; 5 µM: 2,4'-BPF, BPS; 10 µM: Pergafast201; 50 µM: 2,4'-BPS). Similar LOECs within these groups suggests similar potencies before considering IVIVE. BADGE, BCPS, D-8, TGSA, BPS-MPE, BPS-MAE and BTUM were negative in the biomarker assay, suggesting neither agonism nor antagonism.

Figure 5. Comparisons of the minimal bioactivity concentration applicable for ER activation derived from the 46-gene biomarker and IPA ER upstream regulator approaches.



Controls were included to increase confidence in the approaches and assays used in the context of the current investigation. Dexamethasone (Dex) was included as a negative control in this experiment and the analysis suggested it is not an ERα agonist as it was negative within the biomarker assay and the IPA ER upstream regulator analysis also confirmed this. Where Dex has been tested, it was not observed to be an agonist either in ToxCast/Tox21. Estradiol was also included as a positive control in this experiment. Both IPA ER upstream regulator and 46 gene biomarker analyses predicted ER activity for all different concentrations of estradiol.

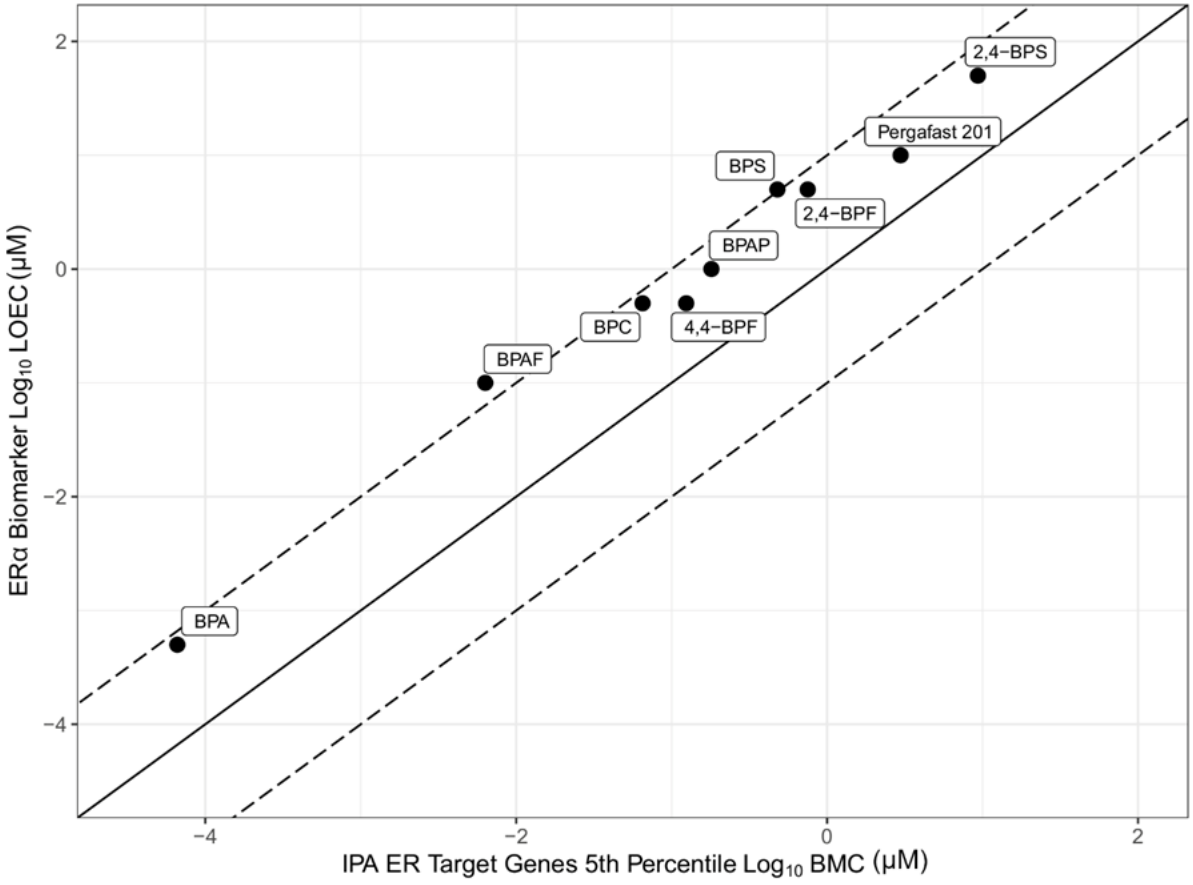
Chemicals were considered ER agonists when both the biomarker and the IPA upstream regulator analysis suggested activation. For quantitative analysis, Dex and BADGE were not considered due to the lack of IPA confirmation. Additionally, BCPS, BTUM, TGSA, BPS-MPE, BPS-MAE and D-8 were not considered further for quantitative analysis as they were not identified as inducing ER-mediated gene expression changes by the 46 gene biomarker.

Overall, out of 16 chemicals, 9 potential ER modulators were predicted by both the transcriptomic biomarker and IPA ER upstream regulator analysis (Appendix A). There was good agreement between the two different approaches of the 46 gene biomarker assay and the IPA ER target gene analysis as activation was indicated for 2,4-BPF (5 µM), 2,4-BPS (50 µM), 4,4-BPF (0.5 µM), BPS (5 µM), BPA (0.0005 µM), BPAF (0.1 µM), BPAP (5 µM), BPC (0.5 µM), and Pergfast 201 (10 µM). Exceptions for these chemicals included more sensitive activation at lower concentration in the biomarker assay for

BPAP (1 µM), and activation was only detected for D-8 using the IPA analysis of upstream regulators that was negative using the biomarker assay (Appendix A).

In addition to the LOEC derived from the 46-gene biomarker, we calculated the lowest gene BMC and the BMC at the 5th and 10th percentiles for DEGs of the IPA ER upstream regulator analysis for consideration in potency ranking and for identifying BERs (Appendix G: IPA upstream regulator analysis to determine enrichment of genes regulated by ERα in the MCF-7 cells treated with test chemicals for 48hrs). The threshold derived from the 5th percentile IPA target genes suggest BPA is the most potent ER agonist among the chemicals tested in this study. The other ER activators (ranked from the strongest to weakest) are as follows: BPAF, BPC, 4,4-BPF, BPAP, BPS, 2,4-BPF, Pergafast 201, D8, 2,4-BPS. In general, the results of these two approaches to estimate pathway specific thresholds are in good agreement with a strong correlation observable between log 5th percentile for target genes of ER upstream regulator target genes and log LOEC values for the transcriptomic biomarker (R^2 = 0.95, P= 1.6x10^{-6}; n=10). The 5th percentile of ER target genes approach appears to be more sensitive than the LOEC approach (Figure 6). A correlation plot shows the majority of the data falling within the dashed lines that represent a ± log unit deviation (Figure 6).

Figure 6. Comparison of log-transformed 5th percentile BMC for target genes of the DEGs in the ER upstream regulator gene set and the transcriptomic biomarker LOEC values.



Note: The 1:1 line is indicated by a solid black line, and the ± log unit deviation is indicated by dashed lines.

6.3. Other *In Vitro* Data

Work was done under the CERAPP project to search the literature to find *in vitro* data about ER activity (Mansouri *et al.*, 2016). Mathematical models were developed to combine multiple data points into a single concentration-response value (AC_{50}) after which, threshold values were used to determine a quantified value, such as AC_{50} values larger than 800 µM being labelled as inactive and those with AC_{50} less than 0.09 µM being labelled strong. This process was done for each receptor activity (binding, agonism, and antagonism), and the results are shown in Table 3. Due to the availability of data, some substances may have no ER information, or limited ER information.

6.4. QSAR

The two primary QSAR models used in this work for ER activity are the CERAPP consensus model developed by the US EPA and the Health Canada in-house developed RF model. Table 3 shows the results of applying these models along with the CERAPP literature results to facilitate comparisons. There are two different calls which can be made, a binary (active or inactive) and strength-based calls (ranking from Inactive to Strong). Both binary and strength-based are based on defined thresholds in the original CERAPP work. It should be noted that it is possible for substances to have binary predictions (noted by cell shading) that contradict the strength predictions (noted by the symbols). For example, BPE binary antagonist activity was predicted as inactive; however, based on the strength prediction it is a weak antagonist (Table 3, RF Antagonist column). This is due to the models being trained independently and can lead to variance in the results. It is expected that there would be some disagreement between the models, however in general, both models gave similar predictions. It was found that the antagonist predictions showed the most discrepancies as noted by a few inactive predictions from the CERAPP consensus model. In contrast, the RF model predicted half of the chemicals to be inactive, although most of the inactive substances were also predicted by the strength-based model to be weakly active antagonists. For both the CERAPP models and the in-house RF model there appears to be a trend where the substances are predicted to be both agonists and antagonists. This may be due to the agonists and antagonists having relatively similar structures, however in the training datasets there are a smaller amount of antagonist substances leading to a higher degree of difficulty to predict their activity. This means it may be possible that the antagonistic effects seen from the models may be false positives, as noted by their precisions 30.7 and 42.9% for RF and CERAPP agonism respectively and 13.0 and 7.4% for RF and CERAPP antagonism respectively.

Overall, the results strongly suggest that the substances used in this case study do have endocrine disrupting properties, in particular the ability to interact with ERs. The trend in the QSAR predictions show that all the bisphenol substances are ER binders and agonists, while roughly half of them appear to have antagonistic potential as well.

ToxCast, QSAR and *In vivo* Supporting Flags

Table 3. ER activity from CERAPP literature and consensus models, Health Canada's RF model, the OECD QSAR Toolbox, and the ERα transcriptomic biomarker.

Chemical	CERAPP Literature Binding	CERAPP Consensus Binding	RF Binding	OECD QSAR Toolbox	CERAPP Literature Agonist	CERAPP Consensus Agonist	RF Agonist	ToxCast ER Agonist AUC Score	ERa Biomarker Agonism LOEC (uM)	IPA ER Target Gene 5th Percentile BMC (uM)	CERAPP Literature Antagonist	CERAPP Consensus Antagonist	RF Antagonist	ToxCast ER Antagonist AUC Score	ERa Biomarker Antagonism LOEC (uM)	# of Active Assays ToxCast/Sum of ToxCast assays tested	NICEATM Uterotrophic Database (lowest LEL)	Weight of Evidence
BPA	++	++	+	++++	++	++	++	0.45	0.0005	0.000066		++++	++	0.0		17/18	Active LEL 2.0 mg/kg/day 1.2 fold fold increase s.c. over 3 days rat; age at first dose PND 19	Active
BPB	+++	++	+++	++++	+++	++	++	0.491				++++	++	0.00196		17/18	Active LEL 20.0 mg/kg/day 1.5 fold fold increase s.c. over 3 days rat; age at first dose PND 19	Active
HPMPP	++++	++	+++	++++	++++	++	++++					++++	++			8/9		Active
BPE	++	++	++	++++	++	++	++					++++	++			5/6		Active
BPC	++	+	+	++++	++	+	++		0.5	0.065		+++	++			5/7		Active
4,4'-	++	+	++	++++	++	++	++		0.5	0.12		++	++			5/7		Active

BPF																	
2,4'-BPF	++	+++	++	++++	++	++	++		5	0.75		++++	++			5/9		Active
2,2'-BPF		+	++	++++		++	++	0.0303				++		0.0		6/18		Inactive
BPTMC			+++	++++			++++						++					Active
TMBPA	++	+++	+		++	+++	++				++	+++	++			4/4		Active
BPAP			++	++++			++		1	0.18			++			2/2		Active
BPZ	++	+	+++	++++	++	+	++					+++	++			2/4		Active
BPS	++	+	+	++++	++	+	++	0.263	5	0.48		+++		0.0		16/18	Active LEL 20 mg/kg/day Reported as log lowest effective dose (1.9) s.c. over 3 days rat;Crj:CD(SD)IGS age at first dose PND 20 PubMed ID 17904329	Active
2,4'-BPS		+	+	++++		+	++		50	9.3		+++						Active
TGSA	+	+++	+	++++		++	++	0.0231			+	++	+	0.0449		7/18		Active
4,4'-TDP	++	+++	++	++++	++	+++	++					+++				3/4		Active
BPAF	+++	++	+++	++++	+++	++	+++	0.552	0.1	0.0063		++++	++	2.6E-5		18/18	Active LEL 4.0 mg/kg/day 1.6 fold fold increase s.c. over 3 days	Active
BPC2	++	+	++	++++		+	++					+++	++			2/4		Active
BCPS	+		+				+	0						0		1/18		Inactive
BADGE			+				++	0					++	0		2/18		Inactive
BPS-MPE			+	++++			++						++			0/2		Active
D-8			+	++++			++			3.2			++					Inactive

BPS-MAE			+	++++			++						++				Active
BTUM			+										++				Inactive
P201			+						10	3			++				Active

Notes:
Background colour indicates binary activity; White (No Data), Grey (Inactive), and other colour (blue for binding, red for agonist, green for antagonist). Symbols indicate strength; Very Weak (+), Weak (++), Moderate (+++), and Strong (++++). Red background for the biomarker indicates the significant LOEC, and indicates an active AUC score from ToxCast (AUC ≥ 0.1).
The fraction of active assays of the total assays tested in Toxcast that were used to derive the AUC score is also presented. The chemicals that had 5th percentile BMCs available from IPA ER target gene analysis are also presented in red. A white background without a symbol indicates that no data were available.

An important aspect of any QSAR model is their applicability domains (AD). An AD is the chemical space that surrounds the data used to train the model; this region of space is viewed as having a higher confidence in the predictions being correct. The CompTox dashboard only shows predictions for substances within its AD, therefore all those predictions shown above are within the CERAPP AD. For the RF model, an AD was developed using an in-house developed program using a k-nearest neighbour approach (Sahigara *et al.*, 2013). When the AD is applied, only one substance (BTUM) is outside the AD for all RF models; additionally, Pergafast 201 (P201) was found to be outside the AD for binary antagonism predictions (prediction in Table 3 'RF Antagonist' column). All other predictions presented are within the ADs of their models.

The ER Binding mechanistic profiler from the OECD QSAR Toolbox was also assessed. This profiler is not a predictive model (there is no AD); rather, the profiling is based on molecular weight and the presence of specific functional groups. The profiler outputs add to the weight of evidence for ER binding.

6.5. Weight of Evidence on ER Activity

The strength of the binary calls (indicated with a '+' symbol) varies across models; however, there is good agreement for the binary calls on ER binding across the models (Table 3). The RF model has the most conservative binary calls for ER binding and has the most coverage of the chemical space of interest. For agonism, the RF model has the best coverage for the non-bisphenols group and agrees well with the IPA ER upstream regulator and the transcriptomic biomarker binary calls, except for P201. When all 6 qualitative flags are available for ER agonism, there is agreement on the binary calls, apart from TGSA. Moreover, there is total binary call agreement for every ToxCast AUC score with active agonism, the CERAPP literature, CERAPP consensus model, and the RF model. IPA confirmation of the biomarker only consisted of agonists, and therefore, no antagonists are listed for IPA.

Minimal information is available for antagonism, with only 5 chemicals indicating CERAPP literature results. Not only is there a lack of data for the antagonistic qualitative flags, there are less agreements within the antagonism flags, notably with the limited CERAPP literature results suggesting inactivity and *in silico* models indicating activity. It should be noted that the current experimental models used for transcriptomics were not intended to detect ER antagonism that would have required co-exposure with estrogen and also did not include the control ERα antagonist ICI 182,780 required to detect antagonistic gene signatures. Thus, within the context of our investigation, ER antagonism was not considered. The ER Antagonist AUC scores indicate that bisphenols are typically inactive; however, *in silico* predictions and the CERAPP literature results indicate some potential for activity. Regardless of the limited information, TMBPA and TGSA show some consensus of antagonistic activity across the CERAPP literature results, CERAPP consensus models, and the RF model. Between the CERAPP consensus model and the RF model, the CERAPP model is overall more conservative in both binary and strength classifications, and identifies more chemicals as ER antagonists. Interestingly, many of these same chemicals are flagged as binary agonists as well with a weaker strength classification for the CERAPP agonism consensus model.

None of the AUC scores for antagonism were over 0.1 to indicate activity. All 8 of the inactive chemicals indicated by the AUC score had all 18 possible ER-specific assays tested, which increases confidence in the inactive call.

7 *In vitro* to *In vivo* Extrapolation to Derive POD_{NAM_SYS} and POD_{NAM_ER} (IATA STEP 3)

Minimal bioactivity concentrations intended to be protective of general systemic effects as well as ER pathway specific effects have been derived in this case study. Moreover, qualitative hazard flags based on a collection of *in silico* predictions and *in vitro* data for the ER pathway have also been collected and combined in a weight of evidence approach to determine the potential for the bisphenol compounds and the alternatives examined in this case study to interact with ER.

As a final step in the IATA, the bioactivity concentrations were converted to their respective AED values through the HTTK modelling and IVIVE. These AED values represent the derived POD_{NAM_SYS} and POD_{NAM_ER} and have been compared against collected PODs from animal studies for repeat dose as well as reproductive and developmental studies as previously described (Lowest POD_{Animal}, Table 4). In all cases, where such a comparison is possible based on available data, POD_{NAM_SYS} and POD_{NAM_ER} are lower than POD_{Animal} indicating that POD_{NAM} are protective (Figure 7). When considered alongside exposure estimates, NAM-based PODs are useful endpoints for prioritization of chemicals where animal data are deficient.

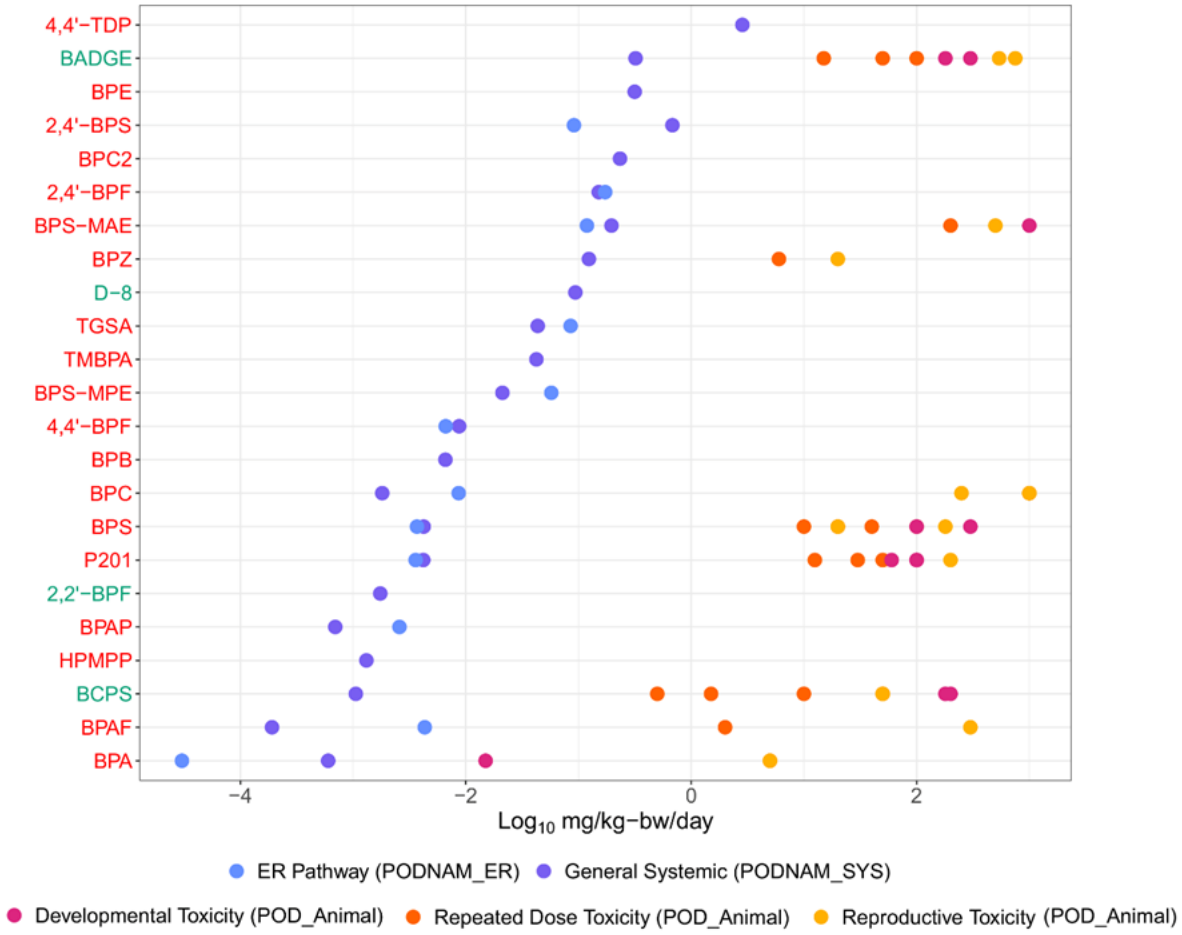
Table 4. Lowest available *in vivo* Points of Departure values found for case study chemicals from the EPA Comptox Dashboard, JECDB, ECHA REACH Dossiers, eChemPortal, EFSA, NTP website, and EKDB database.

CAS-RN (Abbreviation)	Type of *In Vivo* Study	Study Details	Reference
80-05-7 (BPA)	Developmental Neurotoxicity	LOAEL = 0.015 mg/kg-day Dosing method: drinking water rat: Wistar ToxRefDB: EPA ORD 2006	Fujimoto,T, K Kubo & S Aou. Prenatal exposure to bisphenol A impairs sexual differentiation of exploratory behavior and increases depression-like behavior in rats. Brain Research, 1068 (2006): 49-55.
1675-54-3 (BADGE)	Repeated Dose Toxicity	NOAEL = 15.0 mg/kg-day Dosing method: gavage Rat EFSA: EFSA OpenFoodTox 2004	Opinion of the Scientific Panel on Food Additives, Flavourings, Processing Aids and Materials in Contact with Food (AFC) on a request from the Commission related to 2,2-bis(4-hydroxyphenyl)propane bis(2,3-epoxypropyl)ether (Bisphenol A diglycidyl

			ether, BADGE)
80-07-9 (BCPS)	Repeated Dose Toxicity	NEL = 0.5 mg/kg-day Dosing method: feed Rat: Fischer 344 ToxRefDB: EPA ORD 2001	Chhabra, R. S., R. A. Herbert, et al. (2001). 'Toxicology and carcinogenesis studies of p,p '-dichlorodiphenyl sulfone in rats and mice.' Toxicological Sciences 60(1): 28-37. (Chronic rat)
1478-61-1 (BPAF)	Repeated Dose Toxicity	NOEL = 2.0 mg/kg-day Dosing method: gavage Rat: Sprague-Dawley HAWC: Public HAWC projects 2012	Feng Y, Yin J, Jiao Z, Shi J, Li M, Shao B. 2012. Bisphenol AF may cause testosterone reduction by directly affecting testis function in adult male rats. Toxicol Lett 211(2):201-9.
80-09-1 (BPS)	Repeated Dose Toxicity	LOEL = 10.0 mg/kg-day Dosing method: gavage Rat: Crj:CD(SD)IGS OECD Test Guideline 421 JECDB 2001	80-09-1 SIDS INITIAL ASSESSMENT PROFILE CoCAM 4, 16-18 April 2013 https://hpvchemicals.oecd.org/UI/handler.axd?id=f39d746b-5bda-4653-8afa-aff7abf87cd1
843-55-0 (BPZ)	Repeated Dose Toxicity	NOEL = 6.0 mg/kg-day Dosing method: gavage Rat: Crj:CD(SD) HAWC: Public HAWC projects 2004	Yamasaki, K, Noda, S, Imatanaka, N, et al. 2004. Comparative study of the uterotrophic potency of 14 chemicals in a uterotrophic assay and their receptor-binding affinity. Toxicol Lett 146:111-120.
232938-43-1 (P201)	Repeated Dose Toxicity	LOEL = 12.5 mg/kg-day Dosing method: gavage Rat: Wistar HAWC: Public HAWC projects 2002	ECHA. 2002. N-(p-toluenesulfonyl)-N'-(3-(p-toluenesulfonyloxy)phenyl)urea: Exp Key Repeated dose toxicity: oral.001.
97042-18-7 (BPS_MAE)	Repeated Dose Toxicity	NOAEL = 200 mg/kg-day Dosing method: gavage Rat: Crj:CD (SD) IGS OECD Guideline 407 ECHA REACH 2003	ECHA REACH dossier
79-97-0 (BPC)	Reproductive Toxicity	NOAEL = 250 mg/kg-day Dosing method: gavage Rat: Crl:CD (SD) OECD Guideline 421 ECHA REACH 2019	ECHA REACH dossier

Note: Studies were limited to the oral route from Repeated Dose, Reproductive or Developmental Toxicity studies.

Figure 7. Application of the IATA approach for bisphenols and alternatives.



Notes:
NAM based PODs (for both general systemic effects and ER pathway specific) are compared against collected animal-based PODs for repeat dose as well as developmental and reproductive studies, where available.
Acronyms for the chemicals in red are considered to be active with respect to disruption of ER signalling based on a weight of evidence approach combining *in silico*, *in vitro* and *in vivo* mechanistic assays (where available), whereas those acronyms in green are considered inactive.

The combined POD_{NAMs} from both general systemic effects and the ER pathway are the lowest for BPA providing support that BPA appears to have the greatest potential concern out of the bisphenols in the current investigation (Figure 7). All bisphenols (with the exception of 2,2'-BPF) appear to exhibit some potential for disruption of ER signalling but at higher doses than BPA. For example, the POD_{NAM_ER} for BPAP, BPS, BPAF, 4,4'-BPF and BPC are at least 2 orders of magnitude higher than the POD_{NAM_ER} established for BPA. BPA, with the exception of BPAF, also has the lowest POD_{NAM_SYS} across the bisphenols studied and BPA ranks as the most potent of the substances included within the current work. Although this suggests that there is the potential for lower hazard and relative risk for those bisphenols as compared to BPA, additional data are needed to further refine and evaluate the risk potential of these chemicals. Within the context of the current work, comparative *in vivo* estimates were obtained from a search of multiple sources containing guideline studies from ToxVal database as well as the REACH dossier. It is noted that a more comprehensive review of the current state of research regarding the effects from exposure to BPA in animal models is available in a recent report from the National Toxicology Program (NTP), (i.e., Consortium Linking Academic and Regulatory Insights on

Bisphenol A Toxicity (CLARITY-BPA Core Study)). Inclusion of exposure estimates (i.e., including additional refinements for further precision) would allow for calculation of a range of AEDs that may influence the rank of each bisphenol when compared to BPA.

Finally, as noted, several of the potential functional alternatives such as BCPS, D-8, and BADGE do not appear to interact with ER. BCPS and BADGE have multiple available PODAnimal metrics that were observed to be higher than their respective PODNAM_SYS values, suggesting that PODNAM_SYS is protective when considering general systemic effects. The PODNAM_SYS values are higher than the respective value for BPA (i.e., indicative of being less potent at inducing broad, non-specific bioactivity). This also may translate to the potential for lower risk of these substances, depending on their estimated potential for exposure.

8 Uncertainty

Factor	Uncertainty (low, medium, high)	Impact of uncertainty	Comment(s)
Deriving Minimal Bioactivity Concentration for General Systemic Effects	Medium	**Toxicodynamic Uncertainty** - Incomplete biological coverage from ToxCast and limited HTTr analysis in a single cell line can result in uncertainty. Certain toxicity mechanisms may not be adequately covered with the two *in vitro* systems. - Application of immortalized monocultures and culture conditions may fail to replicate cellular interactions *in vivo* thus missing potential mechanisms of toxicity. Moreover, monocultures may not capture inter-individual (human) variability which may underestimate the minimal bioactivity concentration. **Toxicokinetic Uncertainty** - *In vitro* systems lack metabolic competencies seen *in vivo*. Thus if tested substances become metabolically activated to a more toxic metabolite *in vivo*, this would not be captured in the *in vitro* test. However, for specific bisphenols (e.g., BPA and BPS) the main metabolites (glucuronide and sulfate conjugates) are phase II metabolites for detoxification. Therefore, in such cases the parent chemicals may be of higher toxicological concern than most of the metabolites. - Oral exposure was considered as the primary route of exposure in the current study and may be considered a limited representation of the contribution to internal concentration. The contributing influence from other exposure routes with different kinetics (e.g., dermal and inhalation) are a consideration for future work.	ToxCast assays may represent an incomplete biological space. However, many different cell lines were used for Toxcast assays, which adds to the variety of cell responses. HTTr analysis helps to reduce uncertainty since the complete transcriptome is analysed. The *in vitro* cells used in this study are human, whereas many of the *in vivo* studies are done in rats. There is inter-species variation in toxicodynamics and TK that can impact the POD comparisons. MCF-7 cells can also have genetic differences within subcultures that originate from the same batch (Kleensang *et al.*, 2016). However, the effects of the chemicals on these cells are assumed to far outweigh the effects of the genetic differences among and across batches. Cell culture conditions may alter the biological activity of cells. Furthermore, monocultures fail to replicate cellular interactions *in vivo*. *In vivo* biotransformation of compounds into metabolites are not considered in the *in vitro* ToxCast assays. If the chemical is detoxified *in vivo*, false positives can occur, and false negatives can occur when metabolites are more bioactive than parent. For the general systemic toxicity, the lowest estimate of either the ToxCast AC50 or the median transcriptomic BMC derived from four approaches was carried forward for the POD derivation. This multifaceted and conservative approach helps to reduce uncertainty for the purposes of prioritization. Cytotoxicity was not considered when filtering out AC_{50} values from active assays. As it is known that more assays within ToxCast assays are active beyond a cytotoxic concentration, this lack of cytotoxicity filtering can affect the distribution of AC50 values and subsequently the derivation of the 5th percentile value. Cytotoxicity was accounted for, see Appendix B. Cytotoxicity Assay and Solubility for details. The removal of higher potentially cytotoxic concentrations helps to better fit of the data through BMC modelling. Therefore, the values derived from a stronger model helps to reduce uncertainty and increases confidence in the BMCs used for the POD derivation.
Deriving Minimal Bioactivity Concentration	Low-Medium	**Toxicodynamic Uncertainty** - Application of immortalized monocultures and culture conditions may fail to replicate	The MCF-7 cell-line was selected for ER activation since it expresses estrogen receptors, proliferates in response to estrogens and estrogenic compounds, and retains several characteristics of differentiated mammary

for ER activation + Other Qualitative flags for ER activation		cellular interactions *in vivo* thus missing potential mechanisms of toxicity. Moreover, monocultures may not capture inter-individual (human) variability which may underestimate the minimal bioactivity concentration. **Toxicokinetic Uncertainty** - *In vitro* systems lack metabolic competencies seen *in vivo*. Thus if tested substances become metabolically activated to a more toxic metabolite *in vivo*, this would not be captured in the *in vitro* test. **Quality of *in silico* predictions**	epithelial cells. The MCF-7 cell line is from a breast carcinoma and may express altered biological activity relative to normal cells. In addition, the cell line has a high expression of the ER and is dependent on estrogen for proliferation. The experimental design was intended to detect agonism and not antagonism as the *in vitro* system did not test in the presence of estrogen which is typically necessary for detecting ER antagonists. Two HTTr approaches were used to determine ER activation. There was good agreement from the ERa biomarker and IPA ER upstream regulator analysis. Firstly, the biomarker has an established balanced accuracy of 94% and 93% for ER agonism and antagonism, respectively. Secondly, the IPA confirmation helps to increase certainty, since IPA identifies known ER upstream regulators. The minimal bioactivity concentration for ER activation was derived from choosing the lowest of the 5th percentile BMC of genes identified by IPA as regulated by ER or the LOEC identified by the 46 gene biomarker. This approach is considered conservative and helps to derive a protective POD. For the selected ER-specific assays within ToxCast there is less uncertainty. These assays have been previously validated and have been outlined in a previous IATA case study AOP. Moreover, the AUC score was used where available; this has also been validated against *in vivo* data. Augmenting, the HTTr quantitative results with ER activity from ToxCast, lowers the uncertainty for ER activation determination. The CERAPP model is a consensus model, which helps to remove uncertainty in the predictions, as errors in a one model may be overcome by the other models. The Random Forest model developed in house at Health Canada has a high accuracy (surpassing the CERAPP consensus model) making the predictions high quality. The in-house RF does have a wider applicability domain allowing predictions for more substances with a greater degree of confidence.
In vitro to *in vivo* Extrapolation to Derive POD_{NAM_SYS} and POD_{NAM_ER} IVIVE	Medium	**Toxicokinetic Uncertainty** - HTTK model makes use of *in vitro* (or *in silico*) inputs for hepatic clearance and fraction unbound. *In vitro* derived clearance parameters may not adequately address key clearance parameters observed *in vivo*. - Inter-individual (human) variability is considered in the IVIVE model as Css is determined using a Monte Carlo simulation to model a population and varies liver volume, cell density, blood flow, body weight, glomerular filtration, and intrinsic clearance. The C_{ss} at the 95th percentile of 1,000 individuals is returned by the calculation and as such already approximates a "sensitive" population that may have lower metabolic and renal clearance rates.	IVIVE models use Css, which are adapted from pharmaceutical applications, where regular dosing is expected. However, in the case of these chemicals the exposure is not well-defined (I.e., oral or dermal, acute or chronic) and it has been shown that at a given dose rate, the rate of absorption influences the kinetics for a given chemical (Belkebir et al., 2011, US EPA 1998). There are some components of chemical disposition that are not considered using this generic IVIVE model. For example, enterohepatic reabsorption is not considered in the IVIVE models and could be a factor impacting the reverse dosimetry of the bisphenol substances. Furthermore, metabolic clearance of BPA is rapid following oral exposure compared to other exposure routes and there will be limitations to using generic models to identify the Css for specific chemicals (e.g., brief and instantaneous oral exposures to BPA and related bisphenols may not be captured and may require further consideration). In addition, the generic model does not account for conjugation/deconjugation and glucuronidation kinetics. However, previous comparisons with *in vivo* data have shown these missing parameters have not impacted a large portion of chemicals, and the PODs derived for bisphenols using IVIVE appear to approximate traditional PODs (Health Canada, 2021).

			It has been demonstrated that AEDs based on ToxCast data are conservative relative to PODs derived from traditional endpoints in most cases (Paul Friedman et al., 2020). However, due to uncertainties, AEDs are intended to serve as an early screening tool in the identification of compounds of greater potential for concern. Metabolism is not adequately covered in this *in vitro* system; however, we acknowledge that this is a known uncertainty and a an explanation of metabolites and their corresponding implications from exposure has been described elsewhere: (https://www.canada.ca/en/environment-climate-change/services/evaluating-existing-substances/science-approach-document-bioactivity-exposure-ratio-application-priority-setting-risk-assessment.html).
Limited HTTr data for all case study chemicals	Low	- Chemicals that did not have HTTr data were supplemented with ToxCast data, which follow similar trends in potency. Specifically, the reduced potency for the non-bisphenol group compared to BPA exemplifies the trend of reduced potency with increased structural dissimilarity to BPA. As mentioned previously, HTTr data were generated for 16/25 chemicals due to time and resource constraints. The missing results are not anticipated to affect the overall conclusions on potency trends.	This IATA focused on methodology for hazard prioritization. However, further information gathering and/or data generation to fill additional data gaps may be necessary prior to risk assessment.
Overall uncertainty of the IATA	Medium	--	Given the intended purpose of the IATA is prioritization the level of uncertainty is acceptable. Beyond prioritization, the mechanistic analysis presented in this case study could also be used to support hazard characterization for risk assessment.

9 Case Study Conclusion

The investigation demonstrated that several approaches may be considered for establishing robust and reproducible endpoints for the purpose of establishing NAM-based PODs for bisphenols and related chemical substances. Both NAM-based PODs and comparative *in vivo* data are based on oral route of exposure whose findings have primary application in oral exposure scenarios. Of the different approaches used as estimates of minimal bioactivity concentration the 25th ranked gene BMC using high-throughput transcriptomic data agreed with endpoints derived from readily available ToxCast data and the lowest estimate from either of these approaches was carried forward in subsequent steps to establish a NAM-based POD for general systemic toxicity. The majority of observed substances also exhibited pathway-specific potential as ER modulators that were predicted using transcriptomic biomarkers or IPA upstream regulator analysis. Results of QSAR models also suggested that these substances have endocrine disrupting properties, and all bisphenols are ER binders and agonists. Overall, both general systemic toxicity and ER-specific analyses resulted in the lowest estimates for BPA, identifying it as the most potent and of greatest potential concern of the substances included in the investigation. Converting these bioactivity concentrations to respective AEDs representing the POD_{NAM_SYS} and POD_{NAM_ER} allowed comparison to PODs from comparative studies using animal models. In all cases, comparisons using NAM-based PODs were lower (i.e., more conservative) than those derived from *in vivo* models reaffirming the protective nature of the approach when considered in applications to facilitate human health risk assessment. These case-studies provide specific data in hazard characterization that can contribute to the weight of evidence assessment used in regulatory decisions for chemicals risk assessment.

10 References

ANSES and Agence nationale de sécurité sanitaire l'alimentation l'environnement et du travail (2013) 'Substitution du bisphénol A'. Available at: www.Anses.fr (Accessed: 19 March 2020).

Beal, M. A. *et al.* (2021) 'Implementing in vitro bioactivity data to modernize priority setting of chemical inventories', *ALTEX*, 39(1), pp. 123–139. doi: 10.14573/altex.2106171.

Beausoleil, C. *et al.* (2018) 'Regulatory identification of BPA as an endocrine disruptor: Context and methodology', *Molecular and Cellular Endocrinology*, 475, pp. 4–9. doi: 10.1016/j.mce.2018.02.001.

Belkebir, E., *et al.* (2011). 'Haber's rule duration adjustments should not be used systematically for risk assessment in public health decision-making', Toxicol Lett, 204: 148-55. doi:10.1016/j.toxlet.2011.04.026

Browne, P. *et al.* (2015) 'Screening Chemicals for Estrogen Receptor Bioactivity Using a Computational Model', *Environmental Science & Technology*, 49(14), pp. 8804–8814. doi: 10.1021/acs.est.5b02641.

Danish Environmental Protection Agency (2014) *Alternative technologies and substances to bisphenol A (BPA) in thermal paper receipts*. Available at: http://www2.mst.dk/Udgiv/publications/2014/03/978-87-93178-20-5.pdf.

Environment and Climate Change Canada and Health Canada (2008) *Screening Assessment for The Challenge for Phenol, 4,4' -(1-methylethylidene)bis- (Bisphenol A) - CAS No. 80-05-7*, *Ottawa (ON): Government of Canada.* Available at: http://www.ec.gc.ca/ese-ees/default.asp?lang=En&n=3C756383-1 (Accessed: 11 February 2020).

Environment and Climate Change Canada and Health Canada (2020) *Technical consultation: proposed subgrouping of bisphenol A (BPA) structural analogues and functional alternatives*. Available at: https://www.canada.ca/en/environment-climate-change/services/evaluating-existing-substances/technical-consultation-proposed-subgrouping-bpa-structural-analogues-functional-alternatives.html (Accessed: 10 June 2021).

Fabregat, A. *et al.* (2016) 'The Reactome pathway Knowledgebase', *Nucleic Acids Research*, 44(D1), pp. D481–D487. doi: 10.1093/nar/gkv1351.

Filer, D. L. *et al.* (2017) 'Tcpl: The ToxCast pipeline for high-throughput screening data', *Bioinformatics*, 33(4), pp. 618–620. doi: 10.1093/bioinformatics/btw680.

Government of Canada (2010) 'Order Adding a Toxic Substance to Schedule 1 to the Canadian Environmental Protection Act, 1999', *Canada Gazette*, 144(21).

Harrill, J. *et al.* (2019) 'Considerations for strategic use of high-throughput transcriptomics chemical screening data in regulatory decisions', *Current Opinion in Toxicology*, 15, pp. 64–75. doi: 10.1016/j.cotox.2019.05.004.

Harrill, J. A. *et al.* (2021) 'High-Throughput Transcriptomics Platform for Screening Environmental Chemicals', *Toxicological Sciences*, 181(1), pp. 68–89. doi: 10.1093/toxsci/kfab009.

Harris, M. A. *et al.* (2008) 'The Gene Ontology project in 2008', *Nucleic Acids Research*, 36(SUPPL. 1), pp. D440–D444. doi: 10.1093/nar/gkm883.

Healy, B. F. *et al.* (2015) 'Bisphenol A exposure pathways in early childhood: Reviewing the need for improved risk assessment models', *Journal of Exposure Science & Environmental Epidemiology*, 25(6), pp. 544–556. doi: 10.1038/jes.2015.49.

Jackson, A. C. *et al.* (2020) 'Identification of novel activators of the metal responsive transcription factor

(MTF-1) using a gene expression biomarker in a microarray compendium', *Metallomics*, 12(9), pp. 1400–1415. doi: 10.1039/D0MT00071J.

Judson, R. S. *et al.* (2015) 'Integrated Model of Chemical Perturbations of a Biological Pathway Using 18 *In Vitro* High-Throughput Screening Assays for the Estrogen Receptor', *Toxicological Sciences*, 148(1), pp. 137–154. doi: 10.1093/toxsci/kfv168.

Kadasala, N. R., Narayanan, B. and Liu, Y. (2016) 'International Trade Regulations on BPA: Global Health and Economic Implications', *Asian Development Policy Review*, 4(4), pp. 134–142. Available at: https://ideas.repec.org/a/asi/adprev/2016p134-142.html (Accessed: 10 June 2021).

Kanehisa, M. *et al.* (2017) 'KEGG: New perspectives on genomes, pathways, diseases and drugs', *Nucleic Acids Research*, 45(D1), pp. D353–D361. doi: 10.1093/nar/gkw1092.

Kleensang, A. *et al.* (2016) 'Genetic variability in a frozen batch of MCF-7 cells invisible in routine authentication affecting cell function', *Scientific Reports*, 6(1), pp. 1–11. doi: 10.1038/srep28994.

Ma, Y. *et al.* (2019) 'The adverse health effects of bisphenol A and related toxicity mechanisms', *Environmental Research*, 176(July), p. 108575. doi: 10.1016/j.envres.2019.108575.

Mansouri, K. *et al.* (2016) 'CERAPP: Collaborative Estrogen Receptor Activity Prediction Project', *Environmental Health Perspectives*, 124(7), pp. 1023–1033. doi: 10.1289/ehp.1510267.

La Merrill, M. A. *et al.* (2020) 'Consensus on the key characteristics of endocrine-disrupting chemicals as a basis for hazard identification', *Nature Reviews Endocrinology*, 16(1), pp. 45–57. doi: 10.1038/s41574-019-0273-8.

Official Journal of the European Union (2011) 'Commision Directive', L26(11). Available at: http://ec.europa.eu/food/fs/sc/scf/out128_en.pdf (Accessed: 10 June 2021).

Organisation of Economic Co-Operation and Development (2018) *Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption*. Paris. Available at: https://doi.org/10.1787/9789264304741-en.

Organisation of Economic Co-Operation and Development (2019) *Case Study on the Use of an Integrated Approach to Testing and Assessment for Estrogen Receptor Active Chemicals*.

Pagé-Larivière, F., Crump, D. and O'Brien, J. M. (2019) 'Transcriptomic points-of-departure from short-term exposure studies are protective of chronic effects for fish exposed to estrogenic chemicals', *Toxicology and Applied Pharmacology*, 378, p. 114634. doi: 10.1016/j.taap.2019.114634.

Paul Friedman, K. *et al.* (2020) 'Utility of In Vitro Bioactivity as a Lower Bound Estimate of In Vivo Adverse Effect Levels and in Risk-Based Prioritization', *Toxicological Sciences*, 173(1), pp. 202–225. doi: 10.1093/toxsci/kfz201.

Pearce, R. *et al.* (2017) 'Httk: R package for high-throughput toxicokinetics', *ncbi.nlm.nih.gov*.

Pelch, K. *et al.* (2019) 'A scoping review of the health and toxicological activity of bisphenol A (BPA) structural analogues and functional alternatives', *Toxicology*, 424(June), p. 152235. doi: 10.1016/j.tox.2019.06.006.

Phillips, J. R. *et al.* (2019) 'BMDExpress 2: enhanced transcriptomic dose-response analysis workflow', *Bioinformatics*. Edited by A. Valencia, 35(10), pp. 1780–1782. doi: 10.1093/bioinformatics/bty878.

Pivnenko, K. *et al.* (2015) 'Bisphenol A and its structural analogues in household waste paper', *Waste Management*, 44, pp. 39–47. doi: 10.1016/j.wasman.2015.07.017.

R Core Team (2013) 'R: A language and environment for statistical computing. R Foundation for Statistical Computing'. Vienna, Austria. Available at: http://www.r-project.org/ (Accessed: 23 April 2021).

Ramaiahgari, S. C. *et al.* (2019) 'The Power of Resolution: Contextualized Understanding of Biological Responses to Liver Injury Chemicals Using High-throughput Transcriptomics and Benchmark Concentration Modeling', *Toxicological Sciences*, 169(2), pp. 553–566. doi: 10.1093/toxsci/kfz065.

Reardon, A. J. F. *et al.* (2021) 'Potency Ranking of Per- and Polyfluoroalkyl Substances Using High-

Throughput Transcriptomic Analysis of Human Liver Spheroids', *Toxicological Sciences*, 184(1), pp. 154–169. doi: 10.1093/TOXSCI/KFAB102.

Rooney, J. P. *et al.* (2020) 'Mining a human transcriptome database for chemical modulators of NRF2', *PLoS ONE*, 15(9 September). doi: 10.1371/journal.pone.0239367.

Ryan, N. *et al.* (2016) 'Moving Toward Integrating Gene Expression Profiling Into High-Throughput Testing: A Gene Expression Biomarker Accurately Predicts Estrogen Receptor α Modulation in a Microarray Compendium', *Toxicological Sciences*, 151(1), pp. 88–103. doi: 10.1093/toxsci/kfw026.

Sahigara, F. *et al.* (2013) 'Defining a novel k-nearest neighbours approach to assess the applicability domain of a QSAR model for reliable predictions', *Journal of Cheminformatics*, 5(1), p. 27. doi: 10.1186/1758-2946-5-27.

United States Environmental Protection Agency (1998) Summary of the U.S. EPA workshop on the relationship between exposure duration and toxicity. EPA/600/R-99/081.

United States Environmental Protection Agency (2012) *Bisphenol A Altenatives in Thermal Paper*.

United States National Toxicology Program (2017) *NTP Research Report on Biological Activity of Bisphenol A (BPA) Structural Analogues and Functional Alternatives*, *National Toxicology Program*.

United States National Toxicology Program (2018) *NTP Research Report on National Toxicology Program Approach to Genomic Dose-Response Modeling*.

US Environmental Protection Agency (2015) 'Endocrine Disruptor Screening Program (EDSP) Estrogen Receptor Bioactivity'. Available at: https://www.epa.gov/endocrine-disruption/endocrine-disruptor-screening-program-edsp-estrogen-receptor-bioactivity (Accessed: 1 April 2021).

US Environmental Protection Agency (2018) *Strategic Plan to Promote the Development and Implementation of Alternative Test Methods Within the TSCA Program*.

US Environmental Protection Agency (2019) 'ToxCast Database (invitroDB)'. Available at: https://epa.figshare.com/articles/dataset/ToxCast_Database_invitroDB_/6062623 (Accessed: 23 April 2021).

Vandenberg, L. N. *et al.* (2009) 'Bisphenol-a and the great divide: A review of controversies in the field of endocrine disruption', *Endocrine Reviews*, pp. 75–95. doi: 10.1210/er.2008-0021.

Verheijen, M. *et al.* (2020) 'Towards the development of an omics data analysis framework', *Regulatory Toxicology and Pharmacology*, 112, p. 104621. doi: 10.1016/j.yrtph.2020.104621.

Verheijen, M. C. *et al.* (2022) 'R-ODAF: Omics data analysis framework for regulatory application', *Regulatory Toxicology and Pharmacology*, 131, p. 105143. doi: 10.1016/j.yrtph.2022.105143.

Watt, E. D. and Judson, R. S. (2018) 'Uncertainty quantification in ToxCast high throughput screening', *PLOS ONE*. Edited by J. H. Freedman, 13(7), p. e0196963. doi: 10.1371/journal.pone.0196963.

Wetmore, B. A. *et al.* (2012) 'Integration of Dosimetry, Exposure, and High-Throughput Screening Data in Chemical Toxicity Assessment', *Toxicological Sciences*, 125(1), pp. 157–174. doi: 10.1093/toxsci/kfr254.

Appendix A. The total number of DEGs, predicted activity of estrogen receptor (ER) assessed by examining ER biomarker signature, and IPA upstream regulators IPA analyses.

Chemicals	0.0000001 µM			0.000001 µM			0.00001 µM			0.0001 µM			0.0005 µM			0.001 µM			0.01 µM			0.1 µM			0.5 µM			1 µM			5 µM			10 µM			50 µM			100 µM		
	Biomarker	IPA ER UR	DEGs #	Biomarker	IPA ER UR	DEGs #	Biomarker	IPA ER UR	DEGs #	Biomarker	IPA ER UR	DEGs #	Biomarker	IPA ER UR	DEGs #	Biomarker	IPA ER UR	DEGs #	Biomarker	IPA ER UR	DEGs #	Biomarker	IPA ER UR	DEGs #	Biomarker	IPA ER UR	DEGs #	Biomarker	IPA ER UR	DEGs #	Biomarker	IPA ER UR	DEGs #	Biomarker	IPA ER UR	DEGs #	Biomarker	IPA ER UR	DEGs #	Biomarker	IPA ER UR	DEGs #
2.4-BPF			n.t.			n.t.			n.t.			n.t.			35			43			11			44			11			3	↑	↑	206	↑	↑	413	↑	↑	992			
2,4-BPS			n.t.			n.t.			n.t.			n.t.			122			0			49			177			22			10			29			32	↑	↑	269	↑	↑	426
4,4-BPF			n.t.			n.t.			n.t.			n.t.			3			1			19			0	↑	↑	111	↑	↑	174	↑	↑	547	↑	↑	936	↑	↑	954			
BADGE			n.t.			n.t.			n.t.			n.t.			1			0			9			17			20			15			2									
DCDPS			n.t.			n.t.			n.t.			n.t.			0			0			0			6			0			0			0									
BPS			n.t.			n.t.			n.t.			n.t.			0			15			12			1			22			18	↑	↑	363	↑	↑	499	↑	↑	648	↑	↑	1035
BPA			n.t.			n.t.			n.t.			n.t.	↑		156									0	↑	↑	1537	↑	↑	145	↑	↑	2613	↑	↑	3905						
BPAF			n.t.			n.t.			n.t.			n.t.			0			0			0	↑	↑	293	↑	↑	319	↑	↑	409	↑	↑	518	↑	↑	563						
BPAP			n.t.			n.t.			n.t.			n.t.			0			0			0			0			16	↑		22	↑	↑	385									
BPC			n.t.			n.t.			n.t.			n.t.			0			0			0			0	↑	↑	148	↑	↑	278	↑	↑	508	↑	↑	495						
BPS-MAE			n.t.			n.t.			n.t.			n.t.			394			3			11			0			122			0			1			5						
BPS-MPE			n.t.			n.t.			n.t.			n.t.			0			0			0			0			0			0			0									
BTUM			n.t.			n.t.			n.t.			n.t.			0			0			0			2			1			0			1			0						
D-8			n.t.			n.t.			n.t.			n.t.			0			0			5			3			0			16			1	↑		135						
DEX			n.t.			n.t.			n.t.			0			n.t.			0			2			14			n.t.			24			n.t.			n.t.			n.t.			n.t.
Estradiol	↑	↑	542			2	↑	↑	312	↑	↑	701			n.t	↑	↑	659	↑	↑	638			n.t.			n.t.			n.t.			n.t.			n.t.			n.t.			n.t.
Pergafast201			n.t.			n.t.			n.t.			n.t.			40			25			0			3			28			16			1	↑	↑	123	↑	↑	368	↑	↑	604
TGSA			n.t.			n.t.			n.t.			n.t.			0			0			1			0			0			0			0			1						

n.t.: not tested

Gray shading indicates concentrations removed from the analysis due to observed cell toxicity, chemical insolubility, or technical issues (details described in Appendix B. Cytotoxicity Assay and Solubility.

Red shading indicates the lowest concentration that significantly activated ERα

The table shows the numbers of differentially expressed genes that are significantly up-or down-regulated compared to their expression in control groups (FDR ≤ 0.05; absolute fold change ≥ 1.5). ER biomarker signature (46 genes) activation (log p-value> 3.5) shown in yellow arrows (↑). IPA ER upstream regulators activation (z-scores >2) are shown in yellow arrows (↑).

Appendix B. Cytotoxicity Assay and Solubility

Cytotoxicity

To determine cytotoxic concentrations, MCF-7 cells were seeded and treated as per the protocol described above, with 4 replicates per plate, for a total of 8 plates per experiment (repeated 3 times). Cell viability was measured using the CellTiter-Blue Cell Viability Assay (Promega Corp, Madison, WI, USA). At the end of the 48 hour exposure, CellTiter-Blue reagent was added to each well, and the plates were incubated at 37°C, 5% CO2 for 2 hours. The fluorescence was read using excitation wavelength of 560nm and emission of 590nm using a SpectraMax M2 (Molecular Devices LLC, San Jose, CA USA). Fluorescence readings were expressed as a % ratio to that of the control (DMSO) group. Cytotoxicity was identified when the viable cells were <50% of the control. Based on the results of this assay alone, BTPMC was identified as cytotoxic at the highest exposure concentrations (50 and 100 µM) and removed from subsequent analyses. Several chemicals had their highest exposure removed based on previously discussed criteria, (i.e., with more than 2000 DEGs, and paired with activation of two or more of the cell stressor biomarkers) that may be the result of bursts of non-specific activity that occur prior to measurable cytotoxicity

To further investigate the cytotoxicity of the bisphenols, MCF-7 cells were depleted of estrogenic compounds by maintaining them for 3 days in 5% charcoal stripped FBS in phenol red free media. The cells were then seeded in 24 well plates at 10000 cells/well in phenol red free media with 5% charcoal stripped serum. After 24 hours the cells were treated with bisphenols at the indicated concentrations for seven days, with one media change, containing bisphenols, after 3 days. The cells were then trypsinized and counted using a cell counter. Concentrations at which less than 50% of the cells were viable as compared to solvent controls were deemed cytotoxic. Several of the bisphenols were found to be cytotoxic, mostly at high concentrations: BPA at 50, 100µM, BADGE at 10µM, 2,4-BPF at 100µM, BPTMC at 10µM, BPS-MAE at 50 and 100µM, BPAP at 10, 50 and 100µM, BPTMC at 10µM and 4,4-BPF at 100µM. These concentrations were excluded from further analysis.

All chemicals were soluble in DMSO but several chemicals were not soluble in cell culture media. The following table includes a list of chemicals and their insoluble concentrations in culture media that were removed from subsequent analyses.

Appendix B summary table of substances that have been excluded

Chemical	List of excluded concentrations (µM)
2,4-BPF	100
2,4-BPS	-
4,4-BPF	100
BADGE	10, 50, 100
BCPS	10, 50, 100
BPA	0.001, 0.01, 50, 100
BPAF	50, 100
BPAP	10, 50, 100
BPC	50, 100
BPTMC	10, 50, 100
BPS-MAE	50, 100
BPS-MPE	50, 100
BPS	-
BTUM	50, 100
D-8	50, 100
DCDPS	-
TGSA	50, 100

Black = Technical issues, **Blue** = Cytotoxicity, **Green** = Insolubility

Appendix C: Gene Accumulation Plot Rankings and Cut-offs

Results of deriving transcriptomic points of departure using the 25th and 100th ranked genes from gene accumulation plots as a result of exposure to each chemical. The relative potency rank for each gene as well as the average relative rank is present on the left-hand side and each chemical is listed in the table according to its relative potency ranking.

	Relative Potency Rank by Gene			25th Rank Gene and Response (µM)				100th Rank Gene and Response (µM)			
	25th	100th	Average	Gene	BMC	BMCL	BMCU	Gene	BMC	BMCL	BMCU
BPAF	1.00	1.00	1.00	FGF12	9.79E-04	5.33E-04	2.96E-03	KIF20B	1.32E-02	3.82E-03	1.01E-01
BPA	2.00	2.00	2.00	OTUD1	8.44E-03	2.41E-03	8.61E-03	GREB1	1.08E-01	7.36E-02	1.65E-01
BPC	3.00	3.00	3.00	PRC1	3.33E-02	8.93E-03	1.02E-01	CENPK	2.74E-01	1.67E-01	5.17E-01
4,4-BPF	4.00	4.00	4.00	ELOVL2	1.50E-01	1.08E-01	2.23E-01	LONRF2	3.49E-01	3.69E-02	9.18E-01
BPAP	5.00	5.00	5.00	EPB41L4A	1.61E-01	5.40E-02	6.74E-01	HIST1H3B	4.61E-01	3.06E-01	9.13E-01
BPS	6.00	6.00	6.00	TMEM120B	5.04E-01	2.50E-01	1.15E+00	ATAD2	9.34E-01	4.19E-01	2.30E+00
BADGE	7.00	.	7.00	MR1	6.11E-01	3.71E-01	1.74E+00	.	.	.	.
2,4-BPF	8.00	7.00	7.50	LIG1	7.05E-01	2.58E-01	6.25E+00	LMNB1	1.51E+00	8.77E-01	2.62E+00
D-8	10.00	8.00	9.00	AURKB	4.10E+00	3.10E+00	6.15E+00	MAOB	5.87E+00	3.90E+00	1.18E+01
P201	9.00	9.00	9.00	AURKB	3.12E+00	1.72E+00	6.21E+00	CIT	6.09E+00	2.90E+00	2.04E+01
2,4-BPS	11.00	10.00	10.50	MSANTD3	4.79E+00	2.47E+00	1.08E+01	SALL4	2.60E+01	2.03E+01	3.58E+01
TGSA	12.00	.	12.00	HIST1H3F	7.01E+00	4.36E+00	1.78E+01	.	.	.	.
BPS-MAE	13.00	.	13.00	VTCN1	7.42E+00	4.86E+00	9.69E+00	.	.	.	.
BPS-MPE	.	.	.		.	.	.	.	.	.	.
BCPS	.	.	.		.	.	.	.	.	.	.
BTUM	.	.	.		.	.	.	.	.	.	.

. indicates that a value was unable to be derived

Appendix D: RF Model QMRF

1.QSAR identifier

1.1. QSAR identifier (title):
Estrogenic Random Forests
1.2. Other related models:
1.3. Software coding the model:
Coded in python

2.General information

2.1. Date of QMRF:
March 2021
2.2. QMRF author(s) and contact details:
Sean Collins Health Canada, 269 Laurier St., Ottawa, Ontario, Canada, K1A 0K9
Seanp.collins@canada.ca
2.3. Date of QMRF update(s):
2.4. QMRF update(s):
2.5. Model developer(s) and contact details:
Sean P. Collins, Scientific Evaluator, seanp.collins@canada.ca
2.6. Date of model development and/or publication:
Developed December 2019
2.7. Reference(s) to main scientific papers and/or software package:
In press
2.8. Availability of information about the model:
This model is not proprietary
Set up involves the installation of python and the following packages:

1) numpy
2) sklearn
3) pandas
4) opebabel
5) rdkit

2.9. Availability of another QMRF for exactly the same model:

3.Defining the endpoint - OECD Principle 1

3.1. Species:
Human
3.2. Endpoint:
Estrogen receptor binding, agonism, and antagonism
3.3. Comment on endpoint:
Endpoint values are related to AC50 values for estrogen receptor activity as described in training set materials.
3.4. Endpoint units:
A dimensional, Inactive or Active for binary. Inactive, Very Weak, Weak, Moderate, or Strong for multi-class based models.
3.5. Dependent variable:

AC50 for estrogen receptor models
3.6. Experimental protocol:
Training and test set data were taken from the Collaborate Estrogen Receptor Activity Prediction Project (CERAPP) specifically the evaluation set (Mansouri *et al.*, 2016). Evaluation datasets were randomly split into testing and training data.
3.7. Endpoint data quality and variability:
Data were originally collected, compiled, and curated by the US EPA by taken results from literature sources. Further curation work was done by the authors of the model to remove potential duplicate substances. There was noted discrepancies for estrogen receptor binding activity for substances which contained a low number of sources.
4.Defining the algorithm - OECD Principle 2
4.1. Type of model:
Classification
4.2. Explicit algorithm:
Random forest classification model with majority voting
4.3. Descriptors in the model:
Combination of PubChem fingerprints, MACCS fingerprints, OpenBabel fingerprint (FP2, FP3, FP4), and available OpenBabel calculated physical chemistry properties (logP, molar mass, melting point, etc.)
4.4. Descriptor selection:
From original set of all descriptors, pruning was done to remove descriptors which contained no variance among the training data or high correlation (Pearson correlation greater than 0.98). This process was done for each model.
4.5. Algorithm and descriptor generation:
1D and 2D descriptors
4.6. Software name and version for descriptor generation:
Openbabel V3.0.0
RDKit 2020.03.5
4.7. Chemicals/Descriptors ratio:
From 2799 / 1264 = 2.21 to 4904 / 1317 = 3.72
5.Defining the applicability domain - OECD Principle 3
5.1. Description of the applicability domain of the model:
The applicability domain uses a density kNN approach to determine the number of substances which are within a threshold of substances from the training data. Distances between substances are calculated using the principal components from the unpruned descriptors.
5.2. Method used to assess the applicability domain:
The applicability domain was optimized to balance a high number of substances and the predictive capability of the models. By altering the threshold of substances that a chemical needs to be near, the applicability domain can be tuned.
5.3. Software name and version for applicability domain assessment:
Software made in-house to accompany random forest models.
5.4. Limits of applicability:
Substances which are dissimilar to the training dataset will be outside the applicability domain. The training set is from literature data on substances which were evaluated for the estrogen activity. This leads to a bias where if a category or class of substances (e.g. PFAS) are not typically evaluated for estrogen or androgen activity, then that leaves a narrow applicability domain over the universal chemical space.
 # 6.Internal validation - OECD Principle 4
6.1. Availability of the training set:
CERAPP data is available from the US EPA, which was the data used to train these models, specifically the “Evaluation Sets” were used for this work.
6.2. Available information for the training set:

CAS-RN: Yes
Chemical Name: Yes
SMILES: Yes
Formula: Yes
InChI: Yes
MOL file: No
6.3. Data for each descriptor variable for the training set:
All
6.4. Data for the dependent variable for the training set:
All
6.5. Other information about the training set:
Contains continuous data for pertinent AC50 and AUC.
6.6. Pre-processing of data before modelling:
Continuous data are binned into classes
6.7. Statistics for goodness-of-fit:
Binary Binding Balanced Accuracy and with AD: 88.0% and 88.8%
Sensitivity without and with AD: 88.0% and 88.2%

Selectivity without and with AD: 88.8% and 89.3%

Binary Agonist Balance Accuracy and with AD: 85.7% and 86.4%
Sensitivity without and with AD: 85.7% and 86.6%

Selectivity without and with AD: 85.7% and 86.3%

Binary Antagonist Balance Accuracy and with AD: 85.0% and 85.4%
Sensitivity without and with AD: 90.5% and 89.6%

Selectivity without and with AD: 79.4% and 81.1%

Class Binding Balanced Accuracy and with AD: 70.7% and 72.2%
Sensitivity and Specificity are ill-defined for multi class models
Class Agonist Balance Accuracy and with AD: 79.1% and 80.0%
Sensitivity and Specificity are ill-defined for multi class models

Class Antagonist Balance Accuracy and with AD: 83.2% and 84.0%

Sensitivity and Specificity are ill-defined for multi class models

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:
6.9. Robustness - Statistics obtained by leave-many-out cross-validation:
6.10. Robustness - Statistics obtained by Y-scrambling:
6.11. Robustness - Statistics obtained by bootstrap:
6.12. Robustness - Statistics obtained by other methods:
7.External validation - OECD Principle 4
7.1. Availability of the external validation set:
CERAPP data is available from the US EPA, which was the data used to train these models. Specifically the “Evaluation Sets” were used for this work.
7.2. Available information for the external validation set:
CAS-RN: Yes
Chemical Name: Yes
SMILES: Yes

Formula: Yes
InChI: Yes
MOL file: No
7.3. Data for each descriptor variable for the external validation set:
Yes
7.4. Data for the dependent variable for the external validation set:
Yes
7.5. Other information about the external validation set:
Contains continuous data for pertinent AC50 and AUC.
7.6. Experimental design of test set:
Test set is a random 25% of the “Evaluation Sets” from CERAPP, which is similar data to the training data.
7.7. Predictivity - Statistics obtained by external validation:
Binary Binding Balanced Accuracy and with AD: 87.7% and 88.1%
Sensitivity without and with AD: 87.9% and 87.6%
Selectivity without and with AD: 87.6% and 88.6%
Binary Agonist Balance Accuracy and with AD: 85.9% and 86.5%
Sensitivity without and with AD: 85.6% and 87.0%
Selectivity without and with AD: 86.2% and 86.0%
Binary Antagonist Balance Accuracy and with AD: 82.0% and 82.0%
Sensitivity without and with AD: 85.5% and 83.7%
Selectivity without and with AD: 78.5% and 80.3%
Class Binding Balanced Accuracy without and and with AD: 58.7% and 61.1%
Sensitivity and Specificity are ill-defined for multi class models
Class Agonist Balance Accuracy and with AD: 66.4% and 67.6%
Sensitivity and Specificity are ill-defined for multi class models

Class Antagonist Balance Accuracy and with AD: 73.1% and 69.5%

Sensitivity and Specificity are ill-defined for multi class models

7.8. Predictivity - Assessment of the external validation set:
7.9. Comments on the external validation of the model:
8.Providing a mechanistic interpretation - OECD Principle 5
8.1. Mechanistic basis of the model:
Compounds with similar substances are expected to have similar activity.
8.2.A priori or a posteriori mechanistic interpretation:
A posteriori
8.3. Other information about the mechanistic interpretation:
9.Miscellaneous information
9.1. Comments:
9.2. Bibliography:
9.3. Supporting information:
10.Summary (JRC QSAR Model Database)
10.1.QMRF number:
To be entered by JRC
10.2. Publication date:
To be entered by JRC
10.3. Keywords:
To be entered by JRC
10.4. Comments:
To be entered by JRC

Appendix E: Transcriptomic BMC values

Chemical	Total Genes with BMCs (#)	Benchmark Concentration (µM)			
		5th Percentile Gene	Median Lowest gene set (NTP)	25th ranked Gene	1st Mode or 10th Percentile Gene
BPS-MPE	13	3.33E-04	.	.	4.67E-04
TGSA	39	3.25E-04	4.10E-04	7.01E+00	4.04E-04
BCPS	22	3.42E-04	.	.	4.88E-04
BTUM	15	1.54E-04	6.14E-04	.	4.17E-04
BPA	290	3.21E-03	8.44E-03	1.83E-02	3.72E-03
BADGE	34	5.75E-04	.	6.11E-01	7.86E-04
BPAF	609	3.82E-03	5.94E-03	9.79E-04	9.18E-03
BPC	620	5.04E-02	7.17E-02	3.33E-02	8.23E-02
BPAP	573	1.78E-01	1.89E-01	1.61E-01	3.20E-01
4,4'-BPF	861	1.91E-01	.	1.50E-01	2.55E-01
BPS	1040	6.34E-01	5.97E-01	5.04E-01	9.51E-01
2,4'-BPF	788	8.91E-01	7.57E-01	7.05E-01	1.31E+00
BPS-MAE	44	1.58E-03	5.12E+00	7.42E+00	5.50E-02
P201	661	3.35E+00	3.12E+00	3.12E+00	4.76E+00
D-8	322	3.62E+00	3.18E+00	4.10E+00	7.49E-02
2,4'-BPS	469	4.44E+00	4.84E-03	4.79E+00	1.09E+01

. Indicates that a value for the approach could not be derived

Appendix F: ERα 46 Gene Transcriptomic Biomarker LOECs (µM)

Chemical (µM)	0.0001	0.0005	0.001	0.01	0.1	0.5	1	5	10	50	100
2,4'-BPF		-0.64608	0	-1.67572	2.065502	1.444906	0.879097	6.886057	14.0655	18.45593	
2,4'-BPS		-0.67531	0	-0.72423	0	0.497983	-0.38091	-0.44346	1.841638	9.638272	12.06048
4,4'-BPF		0.691222	0.882066	0	1.055517	6.244125	8.886057	18.95861	15.23657	13.52288	
BADGE		0	0.871601	1.10513	0	0.634699	1.326979	0			
BCPS					-0.8617	0	0				
BPA		6.537602	15.88606	11.26761	0	12.55284	15.55284	17.19382	19.63827		
BPAF		0	0	0	15.76955	14.69897	17.16115	18.20066	15.82391		
BPAP				0	0	0.534171	3.69897	12.49485			
BPC			0	0	0.692932	8.481486	13.31876	20.24413	16.58503		
BPS			1.465974	0.803824	0.844664	0.932185	1.607303	17.39794	18.69897	17.63827	16.14874
BPS-MAE		-0.6169	0	0	0	-2	0	0	1.349692		
BPS-MPE				0	0		0	0	0		
BPTMC		1.651695	1.463442	4.853872	9.657577	13.76955	18.74473	12.35655			
BTUM		2.431798	0	0	0	0	-0.80549	1.091515	0		
D-8			0	1.638272	0	0	0	0	1.349692		
Dex				0	1.002177		1.935542				
Estradiol	9.853872		1.324222	13.63827	19.37675		21.79588		21.95861		
P201			0.746904	-0.22388	0.533726	0.651111	0	1.872895	3.522879	13.88606	12.95861
TGSA				0	0	0	0	0	1.356547		

Cells shaded in red represent significant (-Log(p-value) ≥3.5) concentration for ER agonism.

Appendix G: IPA upstream regulator analysis to determine enrichment of genes regulated by ERα in the MCF-7 cells treated with test chemicals for 48hrs

Chemicals	IPA Upstream Regulator Analysis: Estrogen receptor Target Genes									ERα Biomarker
						Gene with the Lowest BMC				
	# Target Genes	# Target Genes with BMC	5th percentile (µM)	10th percentile (µM)	Median BMC (µM)	Gene	BMC (µM)	BMCL (µM)	BMCU (µM)	LOEC (µM)
Estradiol	89	12	2.3E-06	2.6E-06	4.1E-06	CCN5	2.0E-06	3.6E-07	4.7E-06	1.00E-07
BPA	144	61	6.6E-05	8.7E-03	3.33	FOSL1	9.4E-09	1.9E-10	9.6E-09	5.00E-04
BPAF	93	61	6.3E-03	9.6E-03	0.03	CCNA1	9.0E-04	2.7E-04	7.6E-03	0.1
BPC	83	66	0.065	0.076	0.19	KCNK6	0.0048	0.00092	0.026	0.5
4,4-BPF	90	63	0.12	0.14	0.45	BRCA1	0.037	0.011	0.10	0.5
BPAP	55	47	0.18	0.25	0.66	ANXA1	0.054	0.006	0.18	1
BPS	99	69	0.48	0.54	1.7	KIF23	0.29	0.20	0.46	5
2,4-BPF	84	62	0.75	0.82	3.4	FAS	0.00	0.00	0.02	5
Pergafast 201	70	49	3.0	3.2	14	MME	2.5	1.5	4.5	10
D-8	11	9	3.2	3.3	4.9	KIF23	3.1	2.4	4.4	NA
2,4-BPS	48	30	9.3	10	40	LMO3	6.4	2.2	23	50

The table shows the numbers of target DEGs that were enriched in ERα upstream regulator IPA gene sets for each chemical. Only chemicals with absolute z-scores >2 (statistically significant) were included. The DEGs that fit BMC models were then identified for each chemical. The 5th, 10th, and median BMCs of target genes were calculated. A target gene with the lowest BMC was identified and BMC, BMCL, and BMCU are shown. LOEC values obtained from ERα 46 gene transcriptomic gene biomarkers were also calculated.