Science Approach Document

Bioactivity Exposure Ratio: Application in Priority Setting and Risk Assessment

Health Canada

March 2021

Synopsis

This Science Approach Document (SciAD) presents a quantitative risk-based approach to identify substances of a greater potential concern or substances of low concern for human health. This approach considers high-throughput *in vitro* bioactivity together with high-throughput toxicokinetic modelling to derive an *in vitro* based point of departure ($POD_{Bioactivity}$). The purpose of this SciAD is to demonstrate that $POD_{Bioactivity}$ can provide a lower bound estimate for *in vivo* based effect levels derived from oral repeat-dose, developmental, and reproductive studies considered under the Chemicals Management Plan (CMP). Thus, $POD_{Bioactivity}$ can serve as a protective surrogate in the absence of traditional hazard data. When compared to exposure estimates to establish a bioactivity exposure ratio (BER), it is envisioned that the approach outlined in this SciAD would be used for future chemical prioritization and screening level assessment activities under the *Canadian Environmental Protection Act, 1999* (CEPA; Environment Canada and Health Canada 2014).

Health Canada has examined a subset of 46 chemicals that were previously assessed under the CMP to compare $POD_{Bioactivity}$ with points of departure from oral toxicity studies conducted in animals ($POD_{Traditional}$). This was done to demonstrate confidence in using *in vitro* bioactivity as a surrogate lower bound estimate of *in vivo* adverse effect levels. This comparison was specifically conducted for oral repeat-dose, developmental, and reproductive studies. The $POD_{Bioactivity}$ was lower than the lowest $POD_{Traditional}$ cited in the risk assessment for 43 of the 46 of the chemicals examined. These findings are consistent with other published case studies using similar methodology. The analysis presented in this SciAD, along with other available case studies, provide evidence that using a $POD_{Bioactivity}$ would be equal to or be more protective than using a $POD_{Traditional}$ when used to support modern approaches for priority setting and screening level risk assessments.

Of the 46 substances with $POD_{Bioactivity}$, BERs were calculated for 41 substances with available quantitative exposure data. This approach identified 35 substances where the BER indicated that the substances have greater potential for concern. Substances were considered to have potential concern if the $POD_{Bioactivity}$ was within 1000-fold of their maximum estimated exposure value.

When applying the BER approach for priority setting, substances with a BER of less than 1000 would be considered for further action under the CMP. This could include information gathering, the generation of additional data as well as more in depth risk assessment as considered appropriate. For screening level risk assessments, in the absence of other indicators of hazard and using human exposure estimates that take into account all potential sources of exposure, a BER of greater than 1000 may be used as a line of evidence to support a decision of not toxic under section 64(*c*) of CEPA.

A consultation period on this SciAD is being provided to the public, providing an opportunity for comments and additional information in advance of this approach being

applied in modernized prioritization and risk assessment efforts. The publication of this scientific approach will assist the government in identifying data-poor substances that are of low concern or support the identification of substances for further action.

Table of Contents

Synopsis i
1. Introduction 6
2. Background 7
2.1 *In vitro* bioactivity in a tiered testing and assessment strategy 7
2.2 Existing case studies comparing *in vitro* bioactivity to animal studies 9
3. Rationale for the approach 11
4. Methods 12
4.1 Substance selection 12
4.2 Extraction of $POD_{Traditional}$ from assessments 12
4.3 Derivation of *in vitro* $POD_{Bioactivity}$ 13
4.4 Calculation of BER 17
5. Results 19
5.1 Comparing $POD_{Bioactivity}$ with $POD_{Traditional}$ 19
5.2 BERs derived from $POD_{Bioactivity}$ and exposure values 22
6. Uncertainty Factors (UFs) to Determine Adequacy of BER 24
6.1 Deriving the $POD_{Bioactivity}$ ($UF_{Bioactivity}$) 24
6.2 Cell-based limitations (UF_{Cells}) 26
6.3 Inter-individual (human) variability (UF_{Human}) 27
6.4 Summary of Hazard UFs 28
6.5 Exposure estimates 29
7. Application of the approach under the CMP 29
7.1 Substance Exclusion Considerations 31
7.2 Research Needs 33
References 34
Annex 1: Overview of additional in vitro data generation for HTTK 45
Annex 2: Tcpl and httk R package functions used for the approach 47
Annex 3: Substances with a $POD_{Traditional}$ lower than $POD_{Bioactivity}$ 49
Appendix A: Substance list and results for CMP comparative case study 56
Appendix B: Comparison of $POD_{Traditional}$ for developmental and reproductive effects to $POD_{Bioactivity}$ 58

List of Abbreviations

3compartmentss – Three-Compartment Steady-State

AC_{50} – Half-Maximal Activity Concentration

AED – Administered Equivalent Dose

APCRA – Accelerating the Pace of Chemical Risk Assessment

A*STAR – Agency for Science, Technology and Research

BER – Bioactivity Exposure Ratio

BMD – Benchmark Dose

BPA – Bisphenol A

CAS RN – Chemical Abstracts Service Registry Numbers

CEPA – Canadian Environmental Protection Act

Cl_{int} – Intrinsic Hepatic Clearance

CMP – Chemicals Management Plan

CNS – Central Nervous System

C_{ss} – Steady-State Concentration

DSL – Domestic Substances List

ECHA – European Chemicals Agency

EPA –Environmental Protection Agency

F_{gutabs} – Fraction of Gut Absorption

F_{up} – Fraction Unbound to Plasma Protein

HBCD – Hexabromocyclododecane

HIPPTox – High-Content-Imaging-Based Phenotypic Profiling

HTTK – High-Throughput Toxicokinetics

HTTK-Pop – Virtual Population Generator for HTTK

IRAP – Identification of Risk Assessment Priorities

IVIVE – *in vitro* to *in vivo* Extrapolation

LO(A)EL – Lowest Observed (Adverse) Effect Level

MOE – Margin of Exposure

NAMs – New Approach Methodologies

NO(A)EL – No Observed (Adverse) Effect Level

OECD – Organisation for Economic Co-operation and Development

POD – Point of Departure

REACH – Registration, Evaluation, Authorisation and Restriction of Chemicals

SAR – Screening Assessment Report

SciAD – Science Approach Document

tcpl – ToxCast Data Analysis Pipeline

ToxCast – Toxicity Forecaster

ToxRefDB – Toxicity Reference Database

ToxValDB – Toxicity Value Database

TTC – Threshold of Toxicological Concern

UF – Uncertainty Factor

1. Introduction

Following the categorization of substances on the Domestic Substances List (DSL), which was completed in 2006, approximately 4,300 of the 23,000 substances on the DSL were identified for assessment under three phases of the Chemicals Management Plan (CMP). Each phase of the CMP built on lessons learned from the previous phase and progressively introduced streamlined assessment approaches with demonstrated value for efficiently identifying and rapidly assessing low priorities. Application of these assessment approaches thereby allowed resources to be focused on substances and groups of substances of higher priority.

Canada's CMP has provided the opportunity to explore the integration of novel approaches and emerging data to gain efficiencies for chemical assessments. Aligned with the global efforts to modernize chemical testing and assessment, method development is now underway that will be used to inform the future of chemicals management in Canada. This includes the exploration and implementation of New Approach Methodologies (NAMs) to inform chemical prioritization and risk assessment activities. NAMs most often refer to novel, non-animal or alternative test methods, technologies, and/or innovative approaches developed to support chemical risk assessment (Harrill et al. 2019). NAMs can be employed as part of an overall testing and assessment strategy to reduce, refine, or replace vertebrate animals in toxicity testing.

In 2016, the Government of Canada convened a meeting of the CMP Science Committee to discuss the scientific considerations for integrating NAMs within the CMP, specifically, to identify priorities for risk assessment. Among the various methods reviewed, the committee members discussed using *in vitro* bioactivity coupled with human exposure estimates to derive a bioactivity exposure ratio (BER). The committee was supportive of NAMs and the BER concept for use in priority setting, as supplemental lines of evidence in risk assessment and as a high-throughput risk approximation/classification tool (CMP Science Committee 2017).

Under the Accelerating the Pace of Chemical Risk Assessment (APCRA) initiative (Kavlov et al. 2018), Health Canada collaborated with the U.S. Environmental Protection Agency (EPA) and the European Chemicals Agency (ECHA), among other international regulators, to discuss progress and build case studies on using quantitative metrics, derived from NAMs, for prioritization, screening level assessments, and more in-depth risk assessments. This included a large scale retrospective analysis that developed a workflow to derive an *in vitro* bioactivity-based point of departure ($POD_{Bioactivity}$) to compare with points of departure derived from animal studies ($POD_{Traditional}$). A goal of the analysis was to demonstrate that a bioactivity-based POD can be used as a lower bound estimate for *in vivo* based effect levels and as such would be protective when carried forward in the application of the BER approach (Paul Friedman et al. 2019). Building on the collaborative advancements, the approach developed under the APCRA

was applied to substances that have completed risk assessments under the CMP as a proof of concept for broader application moving forward.

The purpose of this Science Approach Document (SciAD) is to demonstrate that *in vitro* bioactivity can provide a lower bound estimate for *in vivo* based effect levels derived from oral repeat-dose, developmental, and reproductive studies considered under the CMP. It is envisioned that the approach outlined in this SciAD would be used for future chemical prioritization and screening level assessment activities under *Canadian Environmental Protection Act, 1999* (CEPA; Environment Canada and Health Canada 2014).

The approach described in this SciAD does not determine the genotoxic potential of a chemical which is also an important consideration in identifying substances of concern or a substance's hazard potential. A complementary approach to screen for potential genotoxic carcinogens is under development and incorporates additional higher throughput *in vitro* genotoxicity assays.

This SciAD was prepared by staff in the CEPA Risk Assessment Program at Health Canada, and has undergone external written peer review and consultation. The reviewers were Dr. Michael Waters, Dr. Joan Garey, and Jennifer Flippin. While external comments were taken into consideration, the final content and future application of the approach remain the responsibility of Health Canada. The critical information and considerations upon which the SciAD are based are given below.

2. Background

2.1 *In vitro* bioactivity in a tiered testing and assessment strategy

A major challenge when conducting risk assessments for existing chemicals under CEPA, and particularly for future prioritization and assessment activities, is the limitation of available *in vivo* toxicological data for many of the substances on the DSL. This limitation is not unique to risk assessment in Canada and new tiered approaches are being proposed internationally that make use of *in vitro* assays in human cells and other NAMs to effectively prioritize substances for further testing and assessment (Thomas et al. 2013; 2019). In such a tiered framework, an initial screen integrates a panel of high-throughput *in vitro* assays to probe early biological events perturbed by each substance. For example, assays that measure endpoints such as cytotoxicity, cellular viability, and transcriptional activity are among those used in the screening process. The assays are chosen to evaluate a broad range of biochemical and cellular targets implicated in adverse health outcomes and testing is done using numerous concentrations in order to derive a concentration-response relationship. If an assay is deemed to be active (i.e., response increases with concentration), the AC_{50} value (the concentration required to elicit 50% of maximal activity) is reported. Each AC_{50} concentration can then be converted to an administered equivalent dose (AED) through reverse dosimetry using simplified and conservative toxicokinetic models. For a given compound there may be

multiple AEDs, but only the AED corresponding to the *in vitro* bioactivity threshold (e.g., 5th percentile) is chosen to represent the $POD_{Bioactivity}$. BERs, which are conceptually synonymous with margin of exposure (MOE), are calculated as the ratio of $POD_{Bioactivity}$ to human exposure levels. BERs can be used to guide chemical prioritization; for example, if a BER is numerically greater than a determined cut-off then the substance can be considered a low or lower priority for risk assessment. In contrast, if a BER is below a specified numerical cut-off then further action, including exploration of targeted testing strategies in later tiers or in depth risk assessment, can be justified.

Currently, the largest *in vitro* bioactivity dataset is maintained under the auspices of the Toxicity Forecaster (ToxCast) program. At present, the database of ToxCast results contains information covering approximately 1,400 *in vitro* assay endpoints, with varying amounts of data for nearly 10,000 chemicals. ToxCast has continuously evolved since its inception and the lessons learned are guiding a strategic blueprint for next-generation risk assessment (Thomas et al. 2019). Advances in computational toxicology have also improved the analytical pipeline of ToxCast with the incorporation of dose curve fitting algorithms and rigorous filters that diminish the influence of experimental noise (Filer et al. 2017), providing robust AC_{50} values for modelling AEDs. Thus, ToxCast continues to grow as a powerful tool and dataset upon which the regulatory community can draw on in early chemical screening.

High-throughput toxicokinetics (HTTK) is essential for *in vitro* to *in vivo* Extrapolation (IVIVE), to translate bioactivity concentrations (e.g., AC_{50} in µM) into human relevant doses (i.e., AED in mg/kg bw/day) for further interpretation and application. Ranking of chemicals using AC_{50} measures alone has limited utility for prioritizing chemicals and provides a different rank order of chemical toxicities than that observed using AEDs, suggesting that the potential hazard of each chemical may be misrepresented without IVIVE (Rotroff et al. 2010). HTTK approaches and models used in quantifying AEDs were largely developed by the pharmaceutical industry, but have been carefully adapted for IVIVE in screening environmental and industrial chemicals (Rotroff et al. 2010; Wetmore et al. 2012; 2015b; Wetmore 2015a). In an effort to make the pharmaceutical-based toxicokinetics methods more conservative and to broaden their applicability domain, simplified assumptions have been applied in the development of HTTK models (Wambaugh et al. 2018). There are two high-throughput assay measurements used for HTTK-based determination of chemical disposition throughout the body: *in vitro* hepatic metabolic clearance and plasma protein binding. The *in vitro* toxicokinetics measurements are used to estimate a steady-state concentration in the plasma (C_{ss}) from a constant daily dose of 1 mg/kg bw/day (see section 4.3 Step 4 for details). Due to the linear assumptions employed, the AC_{50} and C_{ss} values can be used to calculate AEDs, and ultimately derive the $POD_{Bioactivity}$, which is used to derive a BER.

2.2 Existing case studies comparing *in vitro* bioactivity to animal studies

During the generation of high-throughput bioactivity data, the Toxicity Reference Database (ToxRefDB), consisting of data from several types of *in vivo* studies (i.e., chronic toxicity, multigenerational, prenatal developmental), was developed in order to provide a resource for the validation of *in vitro* results (Watford et al. 2019; Thomas et al. 2019; Knudsen et al. 2009; Martin et al. 2009a; Martin et al. 2009b). The initial chemicals encompassed by ToxRefDB were selected to overlap with the early phases of the ToxCast program (77% overlap between chemicals) (Richard et al. 2016). ToxRefDB has been expanded over the years and currently contains approximately 6,000 studies for validation of *in vitro* models. Comparisons between ToxRefDB and ToxCast results have been effective in demonstrating the applicability of high-throughput screening tools in modelling *in vivo* hazard and risk. For example, an analysis of 59 ToxCast Phase I chemicals using 600 ToxCast *in vitro* assays, *in vitro* measurements for rat hepatic clearance, and plasma protein binding measurements demonstrated that the minimum *in vitro* rat AED was equal to or lower than 94% of rat Low Effect Levels in ToxRefDB (Wetmore et al. 2013). Anchoring of *in vivo* and *in vitro* studies using the ToxRefDB and other case studies can not only support the applicability of these approaches but also improve the robustness of developed models.

Another case study focused on the flame retardant hexabromocyclododecane (HBCD) (Gannon et al. 2019a). This work examined BERs for HBCD across progressive tiers of information from broad *in vitro* bioactivity (tier 1) to *in vivo* toxicogenomics (tier 2) and finally traditional apical effects observed in an animal 28-day *in vivo* toxicity study (tier 3). Tier 1 screened 821 ToxCast endpoints and identified 93 endpoints with AC_{50} values. A toxicokinetics model for persistent chemicals was developed and applied to calculate AEDs for these endpoints (Moreau and Nong 2019). The BER was determined to be 810 based on the lowest significant assay and 95% confidence limit predicted in the Canadian population (this same exposure estimate was applied across tiers). The tier 2 study (Farmahin et al. 2019) used toxicogenomics to evaluate the livers of male and female rats exposed to 250, 1250, and 5,000 mg/kg diet/day for 28 days following the Organisation for Economic Co-operation and Development (OECD) test guideline 407 (OECD 2008). Tier 2 Benchmark Doses (BMDs) were based on differentially expressed genes or pathways following previously recommended approaches (Farmahin et al. 2017; NTP 2018). Hence, the tier 2 BER was derived from high-content transcriptomics, an additional source of information that can be used in assessing chemical bioactivity, and the BER was determined to be 96,000. For tier 3, conventional apical endpoints were examined in the tissues of animals from tier 2 and the literature was assessed (Gannon et al. 2019b). BMD analyses were performed on all observed effects and used to derive candidate PODs. The lowest apical rodent BMD was used to derive a MOE of 150,000, concordant with tier 2. In the example of HBCD, it was determined that tier 1 was more conservative than tiers 2 & 3 by two orders of magnitude, and the derived POD would be protective of potential health effects.

Although several small case studies comparing *in vitro* bioactivity-based AEDs to *in vivo*-based effect levels were conducted in the earlier phases of ToxCast (Judson et al. 2011; Paul-Friedman et al. 2016; Tilley et al. 2017; Blackwell et al. 2017; Corsi et al. 2019; Turley et al. 2019), a need amongst international regulators was identified to examine a large number of chemicals covering a diverse chemical space in order to gain confidence in using such an approach for regulatory efforts. The first APCRA case study was conducted to retrospectively evaluate how $POD_{Bioactivity}$ compares to $POD_{Traditional}$ from animal studies (Paul Friedman et al. 2019). For this work, data from the intersection between chemicals in ToxCast, the HTTK library (Pearce et al. 2017), and the Toxicity Value Database (ToxValDB) (Williams et al. 2017) were used. ToxValDB contains information from several sources, including ToxRefDB, and was used to inform the *in vivo* $POD_{Traditional}$. Moreover, additional $POD_{Traditional}$ values were provided by collaborating governments. Specifically, Health Canada provided chemical data that was collected as part of Canada's CMP, ECHA contributed data from the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) program, and the European Food Safety Authority (EFSA) provided data from their human health assessments. Additional *in vitro* data were provided based on high-content-imaging-based phenotypic profiling (HIPPTox) data from the Agency for Science, Technology and Research (A*STAR) program where available. The HIPPTox platform uses high-content and -throughput cellular imaging, combined with machine learning algorithms, to identify predictive phenotypic markers of toxicity (Lee et al. 2018).

In total, 448 chemicals had the required information to compare *in vitro* and *in vivo* PODs. ToxCast data were available for all 448 chemicals and HIPPTox data were available for 57. In determining the $POD_{Bioactivity}$, the minimum of either the 5^{th} percentile filtered ToxCast AC_{50} values and HIPPTox 10% effect concentration (EC_{10}) was used as the bioactivity concentration in µM. Where HIPPTox EC_{10} values were available, the minimum of the HIPPTox value was used because the HIPPTox value was thought to be more indicative of adversity rather than a conservative threshold for bioactivity in ToxCast. However, as ToxCast does not have complete biological coverage of all tissues, the HIPPTox model could provide information on lung, kidney, and liver that might not be indicated by ToxCast. IVIVE using HTTK modelling converted the bioactivity concentration to the steady state AED in mg/kg/day ($POD_{Bioactivity}$). The majority of the $POD_{Bioactivity}$ values were found to be conservatively protective compared to the $POD_{Traditional}$ (400/448 chemicals). The median difference between the PODs on an arithmetic scale was approximately 100-fold. Three compounds had a $POD_{Bioactivity}$ that was higher than the $POD_{Traditional}$ by greater than two orders of magnitude and these were all described as organophosphate pesticides. Moreover, 24 substances of the 48 substances that had a higher $POD_{Bioactivity}$ than $POD_{Traditional}$ contained chemical structure features indicative of carbamate or organophosphate pesticides. It was determined that this approach, with the current assays used, may not be suitable for prioritizing organophosphates and carbamates. Overall, the results show that the models can be used as conservative screening tools for the other compound classes evaluated in the case study.

One of the key achievements of this case study was the development of a generic workflow, applicable for a broad chemical space, which can be used to derive $POD_{Bioactivity}$ intended to be applied to prioritization or screening-level chemical assessments. This collaborative effort significantly informs the methods and provides the foundation for the approach presented in this SciAD.

3. Rationale for the approach

In evaluating the potential for human health effects of a substance, a risk assessment determines a level at which adverse health effects occur, applies factors to account for areas of uncertainty, and compares human exposure estimates against this level to determine risk. Health effects are considered to be adverse if they result in functional impairment or pathological lesions that may affect the lifespan of the organism, its ability to reproduce, or reduce the ability of the organism to respond to an additional challenge (US EPA 2011; Lewis et al. 2002; IPCS 2004). In contrast, the proposed approach, using *in vitro* assays to derive a $POD_{Bioactivity}$, does not determine a level at which adverse health effects would occur. Rather, it uses perturbations observed in *in vitro* assays covering a broad biological range of possible biochemical and cellular targets that may form the basis of events in an adverse outcome pathway but are not indicative on their own of an adverse health effect. It is expected that these initial biological perturbations occur at lower concentrations than the concentrations at which downstream adverse health effects manifest following longer term *in vivo* exposures (Becker et al. 2015; Honda et al. 2019). Thus, it is biologically plausible for the approach to yield lower bound estimates of effect levels observed *in vivo*.

The underlying premise of this approach is that a minimal concentration corresponding to bioactivity observed in a broad range of *in vitro* assays can be coupled with IVIVE to estimate a surrogate point of departure ($POD_{Bioactivity}$). The $POD_{Bioactivity}$ is intended to be a lower bound (i.e. protective) estimate of effect levels that could be observed *in vivo* independent of the biological events or adverse outcome pathways involved (Paul Friedman et al. 2019). *In vitro* bioactivity is derived from the ToxCast database (version 3) and while numerous biochemical and cellular assays are employed, it is acknowledged that it does not cover all biological targets or processes (see discussion of uncertainties in section 6).

In order to build confidence in using human *in vitro* data as a surrogate for effect levels observed in animal studies, it is important to understand how the derived $POD_{Bioactivity}$ compares to $POD_{Traditional}$ values. For these comparisons, PODs derived from rodent studies were used, as traditional assessments commonly rely on rodent data to characterize hazard and the potential for risk to human health. A case study is presented here that compares the $POD_{Bioactivity}$ with $POD_{Traditional}$ derived from *in vivo* studies collected for previously assessed chemicals under CEPA. The comparison is limited to only animal studies where the route of exposure was oral. Although other sources of *in vivo* data may currently be available, the data used to define $POD_{Traditional}$ in this case study are limited to studies that were available and examined by Health

Canada scientific evaluators at the time of the risk assessment. BERs are then derived to demonstrate the utility of the approach in risk-based prioritization and future assessment activities.

4. Methods

4.1 Substance selection

A total of 46 existing substances that were previously assessed under CEPA were chosen to illustrate the application and utility of the approach. The chemical name and Chemical Abstracts Service Registry Numbers (CAS RN[1]) for these substances are available in Appendix A (Table A-1). Chemical selection was predicated on the availability of the required information to apply the approach. Namely, the chemical required the availability of *in vitro* bioactivity data in the ToxCast database, HTTK assay data required for IVIVE, and previous risk assessments under CEPA where traditional PODs (or toxicity data collected as part of ongoing assessment activities under the third phase of CMP) were examined. The limiting data source for expanding the chemicals under the analysis was the availability of HTTK assays required to conduct IVIVE for the AED estimates. Only 29 chemicals were identified as meeting the above criteria using the available data in the HTTK package (version 1.8) (Pearce et al. 2017) in R (R Core Team 2013). In order to expand the number of chemicals used to illustrate the approach, an external contractor was used to generate HTTK *in vitro* data for an additional 17 chemicals to then be used in the HTTK R package (see Annex 1 for details).

4.2 Extraction of $POD_{Traditional}$ from assessments

Risk assessment reports for the 46 chemicals were examined and $POD_{Traditional}$ information was extracted from oral repeat-dose studies (of various durations) as well as from developmental and reproductive toxicity studies cited within the assessments. The study type, duration, species, strain, exposure method (gavage, diet or drinking water) and source_id (reference) were extracted from the assessments and indexed by CAS RN and chemical name. Where possible for the $POD_{Traditional}$ collected, both the no observed (adverse) effect level (NO(A)EL) and lowest observed adverse effect level (LO(A)EL) for each study were recorded. Moreover, for developmental or reproductive toxicity studies, the $POD_{Traditional}$ were separated based on findings in the offspring or parental animals. A short descriptive text passage was extracted where available in the assessment that describes the observed effects at the LO(A)EL (e.g. target organ

[1]The Chemical Abstracts Service Registry Number (CAS RN) is the property of the American Chemical Society, and any use or redistribution, except as required in supporting regulatory requirements and/or for reports to the Government of Canada when the information and the reports are required by law or administrative policy, is not permitted without the prior written permission of the American Chemical Society.

effects, histopathological findings, clinical chemistry parameters or other general findings (such as body weight changes)). The effects were broadly classified by sub-type. Findings in developmental toxicity studies were broadly labeled as 'developmental' if an effect was observed in the offspring from prenatal and/or postnatal exposure. This includes specific effects such as structural malformations, but this label also includes general effects such as reductions in body weight or gain. Effects were classified as 'reproductive' if the reproductive organs were the targets in repeat-dose studies and/or if fertility parameters from reproductive toxicity studies were affected. Other effects from toxicity studies were broadly labeled as 'systemic'. This label was applied for chemicals that affected multiple sites/parameters or single organs beyond the site of contact. $POD_{Traditional}$ values that were expressed as parts per million in the diet or drinking water or mg/kg in the diet were converted to a mg/kg bw/day using the conversion factors described in Health Canada (1994).

There were two types of $POD_{Traditional}$ values used to make comparisons to the derived $POD_{Bioactivity}$. Specifically, the lowest (minimum) $POD_{Traditional}$ across all toxicity studies examined and the $POD_{Traditional}$ used as the basis for risk characterization were extracted from assessments previously published under CEPA. The $POD_{Traditional}$ used for risk characterization is selected based on exposure considerations, such as duration and route of exposure or sub-population of interest, typically when deriving an MOE to assess risk. Lastly, for a more refined analysis, $POD_{Traditional}$ values associated with effects broadly classified as developmental or reproductive were compared to the $POD_{Bioactivity}$.

4.3 Derivation of *in vitro* $POD_{Bioactivity}$

The methods for deriving AEDs and the subsequent $POD_{Bioactivity}$ closely follow the methods outlined in Paul Friedman et al. (2019). A generic workflow was developed that follows these broad steps:

1) Extract bioactivity data from the ToxCast database for each chemical of interest.
2) Apply generic filtering criteria to remove AC_{50} values from curve fits of ToxCast data that may be less quantitatively informative.
3) Calculate the 5^{th} percentile from the distribution of AC_{50} values from active assay endpoints to represent a lower *in vitro* bioactivity threshold per chemical.
4) Use HTTK modelling to estimate an AED corresponding to the *in vitro* bioactivity threshold to represent the $POD_{Bioactivity}$.

Each step in the process is summarized below in more detail.

STEP 1: Extraction of In vitro Bioactivity from ToxCast Database

In vitro bioactivity for the substances examined under this approach were obtained from the publically available MySQL ToxCast database (invitrodb_v3) (US EPA 2015)

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). More details of how this step was performed are provided in Annex 2.

STEP 2: Applying generic assay filtering criteria

Generic filtering criteria for assays with an active hit-call (i.e., assays deemed active for given chemical) were determined and a rationale described in detail in Paul Friedman et al. (2019). The aim is to eliminate less reproducible or less reliable activity calls and respective AC_{50} values for quantitative use when deriving the $POD_{Bioactivity}$. For assays that are removed during the filtering process, their modeled AC_{50} concentration is less likely to be an informative value due to artefacts of the tcpl package automated curve-fitting process. The first portion of the filtering process removes assays that had three or more caution flags and a hit percent (i.e., the % of 1,000 bootstrap curve-fits runs that are classified as a hit) of less than 50% (both conditions needed to be met for filtering to apply). Assays with three or more caution flags have been found to be more susceptible to lower curve-fit reproducibility, which can be observed when looking at the hit-percent. Thus, the first filtering step removes the least likely reproducible fits and activity calls. The second step in the filtering process is to remove the tcpl curve fit categories 36 and 45. Fit category 36 corresponds to Hill model fits where the model top (top of the curve fit) is less than or equal to 1.2 times the threshold cut-off for a positive response and an AC_{50} value less than or equal to the lower limit of the concentration range screened. In other words, the maximal fitted response (or efficacy) is only slightly above the threshold where an assay is considered active and an AC_{50} value was estimated to be below the lowest concentration screened. Similarly, fit category 45 indicates the same criteria but for a gain-loss model fit. These fit categories are thought to be less quantitatively informative because the efficacy is borderline and the estimated AC_{50} is in a concentration range where there are no actual data to inform the slope of the curve. No cytotoxicity based filtering of the ToxCast data was performed. For the 46 CMP case study chemicals, the percent of active assay endpoints post-filtering and the total number of assays available are represented in Figure 4-1.

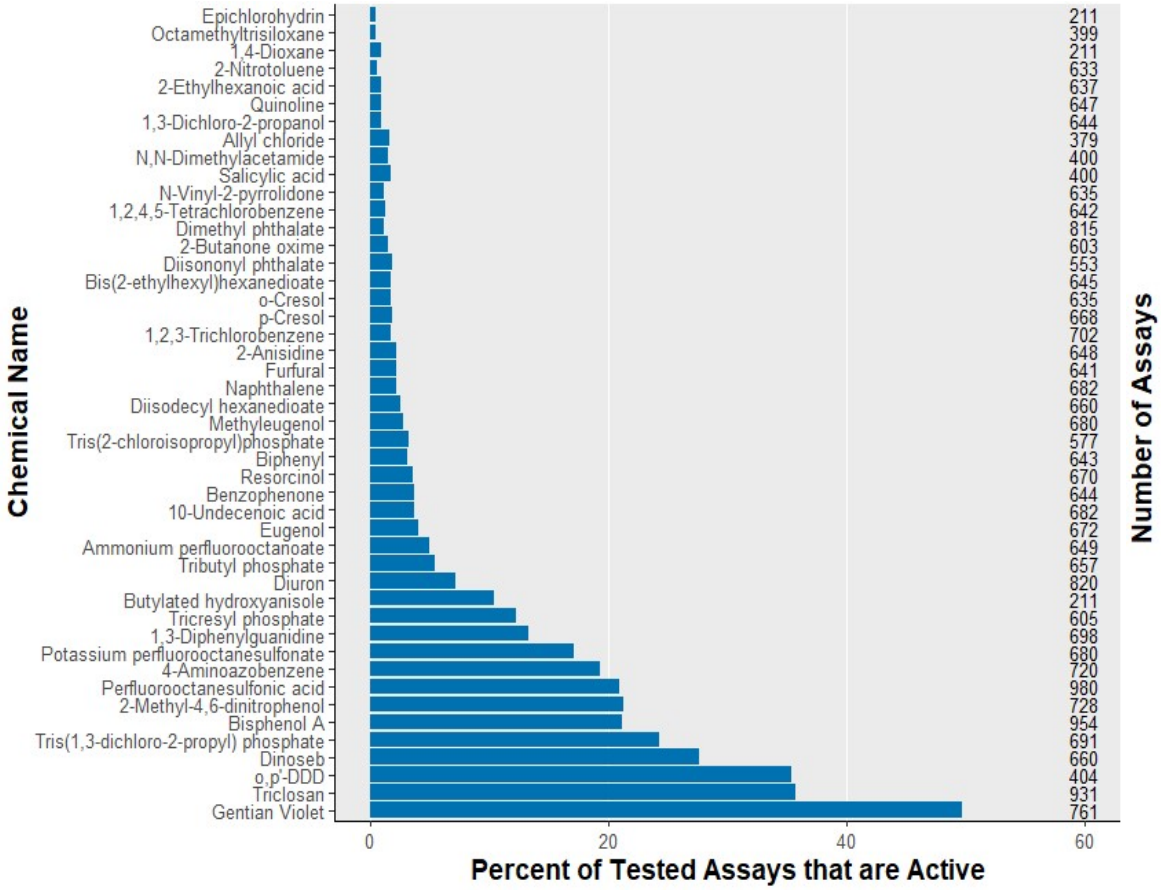

Figure 4-1: ToxCast assays available for each chemical. The percent of tested assays that are active (post-filtering) is represented by blue bars and the number of assays each chemical was tested in is shown along the far right axis of the plot.

STEP 3: Calculate the 5th percentile from the distribution of AC_{50} values from active assays to represent in vitro bioactivity threshold

The 5th percentile from the distribution of AC_{50} values from the previous step was selected to represent the *in vitro* bioactivity threshold for calculation of the AED and to represent the $POD_{Bioactivity}$. The 5th percentile of AC_{50} values was selected over the minimum AC_{50} value in order 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 the tcpl R package. Selection of the 5th percentile is an attempt to balance the desire to select a conservative bioactivity threshold informed by the whole distribution of AC_{50} values and not to solely rely on potentially non-representative outlying values at the extremes (i.e. the minimum). The estimate of the 5th percentile of the AC_{50} values was calculated using R (R Core Team 2013). By default, R uses the Type 7 algorithm for quantile estimation, which is intended for continuous samples (Hyndman and Fan

1996). As mentioned, 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. Of the 46 chemicals examined in the CMP case study, the 5th percentile is below the tcpl estimated lower bound concentration for cytotoxicity for 92% of the chemicals.

STEP 4: Calculate AED corresponding to the in vitro bioactivity threshold which represents the $POD_{Bioactivity}$

The *in vitro* bioactivity threshold is not particularly useful alone to identify substances of concern as it does not provide an indication of what human exposure would be necessary to induce the change in observed bioactivity. Thus, *in vitro* bioactivity can be extrapolated to an AED through IVIVE to provide a more useful metric termed here the $POD_{Bioactivity}$.

IVIVE modelling was done using the HTTK package (version 1.8) (Pearce et al. 2017) in R (R Core Team 2013). HTTK contains all the tissue and physiological data parameters required to perform human toxicokinetic modelling via the oral or intravenous dosing routes. Furthermore, HTTK contains physico-chemical and *in vitro* data for over 1,000 chemicals. The native HTTK R package had some data available for 29 out of the 46 CMP compounds selected for analysis. The missing data needed for IVIVE were acquired by *in vitro* pharmacokinetics assays performed by Paraza Pharma Inc. (Montréal, QC) (details Annex 1).

The three compartment steady-state model ("3compartmentss") in HTTK, modified from work in Wetmore et al. (2012; 2015b) and Wetmore (2015a), was used for IVIVE modelling of the 46 compounds via the oral route. The three compartments consist of the gut, liver, and rest of the body. This model is simple and intended to be applied for a broad range of chemicals. The specific parameters used in the "3compartmentss" model are hepatic clearance, fraction unbound in plasma protein, molecular weight, and steady state prediction.

When applied in HTTK, the 3compartmentss model predicts a C_{ss} in the plasma based on a dose of 1 mg/kg/-bw/day. The C_{ss} is calculated using the equation:

$$C_{ss} = k_{dose} \Big/ \left(f_{up} Q_{gfr} + \frac{\left(Q_{liver} + Q_{gut}\right) f_{up} Cl_{metabolism}}{\left(Q_{liver} + Q_{gut}\right) + f_{up} Cl_{metabolism} / R_{blood2plasma}} \right)$$

Where k_{dose} is the constant dose rate (mg/kg bw/day)
F_{up} is the fraction of chemical unbound to plasma protein
Q_{gfr} is the glomerular filtration rate

Q_{liver} is blood flow to liver
Q_{gut} is blood flow to gut
$Cl_{metabolism}$ is hepatic clearance in whole liver
$R_{blood2plasma}$ is the ratio of chemical blood concentration to plasma concentration

Once calculated, this C_{ss} is used to estimate the AED (mg/kg bw/day). At steady state, plasma concentration increases linearly with dose. Thus, the AED for a given *in vitro* bioactivity concentration (in this case the 5th percentile of AC_{50} values from ToxCast derived in step 3) can be extrapolated using the formula:

$$AED = bioactivity\ concentration\ (\mu M) \times \frac{1\ \frac{mg}{kg}}{\frac{day}{C_{ss}\ (\mu M)}}$$

The requisite *in vitro* parameters required for the 3compartmentss model in HTTK are intrinsic hepatic clearance (Cl_{int}), which is scaled to approximate $Cl_{metabolism}$, and F_{up}. Incorporation of permeability assay data in Caco-2 cells, which measures fraction of compound absorbed by gut (F_{gutabs}), can also improve the steady-state concentration determinations for a small fraction of chemicals (Wetmore et al. 2012). However, full absorption is assumed by HTTK when F_{gutabs} data is not available (details in Annex 1).

The model uses a Monte Carlo simulator, known as the Virtual Population Generator for HTTK (HTTK-Pop) (Ring et al. 2017), to account for inter-individual variation in the human population. The physiological metrics used by HTTK-Pop are from the National Health and Nutrition Examination Survey (NHANES) data (Johnson et al. 2014). Different demographics and subgroups can be used in the sampler by changing gender, age limit, body weight, renal function, and ethnicity parameters. Alternatively, the entire US population can be modeled by default. The Monte Carlo simulation varies several parameters including: liver volume, cell density, blood flow, body weight, Q_{gfr}, and C_{lint}. A coefficient of variation of 30% is used for each parameter by default. The C_{ss} at the 95th percentile of 1,000 individuals is returned by the function. The default HTTK-Pop parameters were used to predict a C_{ss} for each of the 46 compounds prior to IVIVE. Thus, the $POD_{Bioactivity}$ is the AED that is calculated based on the 5th percentile AC_{50} divided by the 95th percentile C_{ss}. Based on the formula above, a higher C_{ss} at the 95th percentile returns a lower dose estimate (i.e., the 95th percentile reflects the most susceptible individuals requiring a lower dose to achieve the same C_{ss}), and thus, deriving an AED based on the 95th percentile C_{ss} is a conservative approach.

4.4 Calculation of BER

A BER is determined by comparing a $POD_{Bioactivity}$ to an exposure estimate and this value may be used to identify whether a substance should be prioritized for further action; this could include data gathering, targeted testing, or additional scoping and assessment. Approaches using BER have been demonstrated in other publications, in

which a $POD_{Bioactivity}$ is compared with an exposure value as described above (Gannon et al. 2019; Paul Friedman et al. 2019).

Screening Assessment Reports (SARs) for the 46 chemicals were examined and exposure estimates for all exposure scenarios were extracted, including all exposure routes. In some cases, the exposure estimates included in the BER exercise were modified from the values presented in the exposure scenarios and/or in the MOE calculation in the SARs. For example, in some SARs, MOEs were based on comparison of air concentrations to lowest observed adverse effect concentrations. In these cases, original exposure estimates (as air concentrations) were converted to doses in mg/kg bw/day based on duration of exposure of specific events. In other cases, exposure values used in derivation of MOEs in SARs may have been averaged over a time period to better match the duration of exposure in the $POD_{Traditional}$ (e.g., intermittent use product (1/week) exposure averaged to “per day”); non-averaged “per event” exposure estimates were used in the BER calculation. As a result of these approaches, some exposure values calculated here and applied in this analysis are higher than the values used in the original SARs.

The exposure estimates were organized separately by: 1) daily intake based on environmental media (including air, water, soil, and food), 2) exposure as a result of daily or intermittent consumer product use (e.g., cosmetics, paint, do-it-yourself products, textiles), and 3) exposure estimates based on biomonitoring data. The following are examples of key factors recorded in spreadsheets indexed by CAS RN and chemical name: route of exposure, duration of exposure, subpopulation, and dermal absorption factor. In the case of products available to consumers, separate exposure estimates were generated for products considered to result in intermittent exposure (e.g., an exposure estimate based on products used infrequently such as a wall paint) and those with daily or chronic exposure (e.g., skin moisturizer). In cases where inhalation was identified as a primary route of exposure, air concentrations (mg/m^3) were converted to doses (mg/kg bw/day) based on duration of exposure and body weights for specific subpopulations, as appropriate. In the case of perfluorinated compounds, the biomonitoring data was reported as µg/mL in both the plasma and serum. Therefore, the same reverse dosimetry applied in converting AC_{50} to AED, based on plasma C_{ss}, was applied to the highest plasma concentration reported in biomonitoring data to facilitate comparisons in units of mg/kg bw/day.

For each substance, the maximum exposure value (mg/kg bw/day) for each exposure estimate type (i.e., environmental media, products available to consumers used intermittently (acute), products available to consumers used daily (chronic), and biomonitoring data) was identified. This maximum exposure was compared to the $POD_{Bioactivity}$ to derive a BER for each substance.

Quantitative exposure estimates were available for 41 of the 46 substances. For five of the 46 substances only a qualitative description of human exposure was available within the SAR. For these assessments, a qualitative approach was used to describe

exposure. A qualitative approach for exposure assessment was used for various reasons including negligible exposures or where a substance was determined to have low hazard potential. Deriving a quantitative exposure estimate for these substances would not have had a meaningful impact on the screening assessment conducted under CEPA. Therefore, using the $POD_{Bioactivity}$ and exposure estimates available from the CMP assessment reports, BERs could be derived for 41 substances. BER derivation was not possible for the five other substances previously assessed under the CMP.

5. Results

5.1 Comparing $POD_{Bioactivity}$ with $POD_{Traditional}$

Two comparisons were made between $POD_{Bioactivity}$ and $POD_{Traditional}$ (Figure 5-1). The first comparison uses the minimum $POD_{Traditional}$ extracted from the assessment reports, and the second comparison uses the $POD_{Traditional}$ carried forward for risk characterization. On a log scale, the metric for the comparison between the two types of PODs is the $log_{10}POD$ ratio, which is the difference between the $log_{10}POD_{Traditional}$ and the $log_{10}POD_{Bioactivity}$. Alternatively, the POD ratio can be calculated on the arithmetic scale by dividing the $POD_{Traditional}$ by the $POD_{Bioactivity}$.

For the first comparison using the minimum $POD_{Traditional}$, the $POD_{Bioactivity}$ was lower than the minimum $POD_{Traditional}$ (i.e. lowest NO(A)EL or LO(A)EL) examined during the risk assessment for ~93% of the 46 chemicals covered under this case study (i.e. $log_{10}POD$ ratio > 0), with the median value for the $log_{10}POD$ ratio being 2.24 (range -0.73 to 6.43). This median value translates to $POD_{Bioactivity}$ being ~100-fold lower than the $POD_{Traditional}$ on an arithmetic scale (all values in Appendix A). Three chemicals were found to have minimum $POD_{Traditional}$ values lower than the $POD_{Bioactivity}$ (p-cresol, bisphenol A, and allyl chloride). These chemicals are examined in detail in Annex 3.

The degree of conservatism offered by the $POD_{Bioactivity}$ can further be evaluated through the second comparison, which uses the $POD_{Traditional}$ values that were carried forward for risk characterization within the CEPA risk assessments. The POD used for MOE derivation at the time that the risk assessment was conducted may not always be the lowest POD across all animal studies collected and depends both on the quality and extent of the hazard dataset as well as on the information available regarding sources, uses, handling, and disposal of the substance(s). Accordingly, the $POD_{Traditional}$ was selected based on having sufficient quality and relevance to inform the risk characterization for the route and exposure scenario(s) critical to the determination of whether or not the substance meets the criteria as defined in section 64 of CEPA 1999. Moreover, during risk characterization, different PODs can be used to derive distinct MOEs for different subpopulations and across a variety of possible exposure scenarios. The PODs used for risk characterization were identified for 38 chemicals from the case study; for the remaining 8 assessments, the risk characterization was qualitative without the identification of a key $POD_{Traditional}$ for the derivation of an MOE. For these chemicals, the $POD_{Bioactivity}$ was lower than the lowest $POD_{Traditional}$ used for risk

characterization for all but one chemical (allyl chloride) with the median value for the log_{10}POD ratio being 2.32 (range of 0.47 to 7.3). On an arithmetic scale, the median value translates to the $POD_{Bioactivity}$ being 209-fold lower than the $POD_{Traditional}$ used for risk characterization.

Finally, a comparison was made looking specifically at PODs that could be associated with effects broadly classified as developmental or reproductive. There were 31 chemicals with LO(A)ELs associated with a developmental effect (i.e. the description of effects at the LO(A)EL were broadly considered to be developmental). Of the 31 chemicals, 22 also had a NO(A)EL (Appendix B; Figure B-1). For these chemicals, the $POD_{Bioactivity}$ was lower than the lowest developmental $POD_{Traditional}$ for all but one chemical (i.e., bisphenol A) with a median value for the log_{10}POD ratio being 3.025 (range -0.73 to 8.08). A similar analysis was conducted for reproductive effects. A POD considered to be related to reproductive effects was available for 21 chemicals (Appendix B; Figure B-2) and the $POD_{Bioactivity}$ was below the lowest reproductive $POD_{Traditional}$ for 20 of the substances with a median value for the log_{10}POD ratio being 3.57 (range -0.73 to 7.76). Bisphenol A had a log_{10}POD ratio less than zero for both endpoint sub-types as it had a $POD_{Traditional}$ that was considered to be related to both development and reproduction and although lower, these low dose studies were within the range of the derived $POD_{Bioactivity}$ (discussed in Annex 3).

These findings are similar to that observed for the broader analysis conducted on 448 substances under the APCRA case studies initiative. Of the 448 substances, 90% had a $POD_{Bioactivity}$ that was less than the $POD_{Traditional}$ value with a median log_{10}POD ratio of 2. The range of log_{10}POD ratios found was -2.7 to 7.5. However, once organophosphate and carbamate chemicals are excluded from the analysis, only 24 chemicals had a log_{10}POD ratio of less than zero and none were below -2 (i.e. $POD_{Bioactivity}$ was 100 fold higher than $POD_{Traditional}$ at the extreme of the analysis).

Taken together, these case studies demonstrate that a bioactivity-based POD can be used as a lower bound estimate for oral based effects levels from animal studies and as such would be a protective surrogate when carried forward in the application of the BER approach (Paul Friedman et al. 2019). Steps can be taken to account for substances where the $POD_{Bioactivity}$ may not be lower, such as exclusion of certain chemical classes all together (i.e. organophosphates or carbamates), review of quality control information and physico-chemical properties, or by consideration of certain uncertainty factors (UFs) when using the approach (see sections 6 and 7 below).

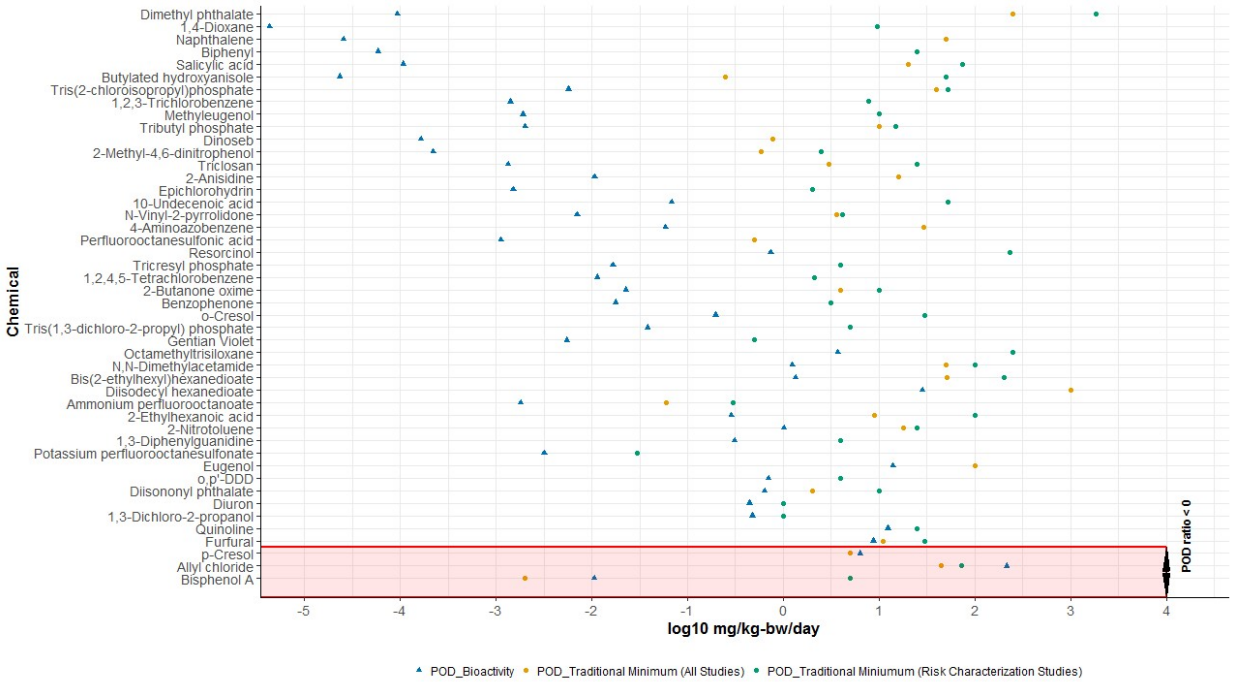

Figure 5-1. Comparison of ToxCast derived $POD_{Bioactivity}$ with $POD_{Traditional}$ from animal studies. $POD_{Bioactivity}$ represents the 5th percentile of AC_{50} values from ToxCast assays with a positive hit-call converted to an AED using reverse dosimetry and high-throughput toxicokinetic information in the HTTK package in R. $POD_{Traditional}$ represents the minimum NO(A)ELs and LO(A)ELs identified in previous risk assessments.

5.2 BERs derived from $POD_{Bioactivity}$ and exposure values

BERs for 41 substances (reflecting those substances for which quantitative exposure estimates were available of the 46 substances) were determined by comparing the $POD_{Bioactivity}$ and the maximum exposure values for each source. On a log scale, the metric for the comparison between the two values is the $log_{10}BER$, which is the difference between the $log_{10}POD_{Bioactivity}$ and the log_{10}Exposure value (i.e., $log_{10}BER = log_{10}POD_{Bioactivity} - log_{10}$Exposure and $BER = POD_{Bioactivity}$/Exposure).

For the 41 substances, the $log_{10}BER$ was found to be less than 0 for 17 substances, and less than 2 for 31 substances (Figure 5-2). In both of these cases, the BER may indicate that more assessment is required on these substances as there is a potential concern. In some cases, low $log_{10}BERs$ (e.g., < 2) were primarily driven by high exposure values (e.g., bis(2-ethylhexyl)hexanedioate (or DEHA)), whereas in other cases, a low $log_{10}BER$ was driven by a very low $POD_{Bioactivity}$ (e.g., 1,4-dioxane). This is an important result as it highlights both the importance of a risk-based approach and consistent considerations and drivers to that of traditional risk assessment. Seven of the eight substances with exposure data, that were originally concluded as a potential risk to human health under CEPA section 64(*c*), were identified with a $log_{10}BER$ of less than 2. Quinoline, also concluded to meet criteria 64(*c*) under CEPA had a very high $log_{10}BER$, based on a very low exposure value and a high $POD_{Bioactivity}$. In this case, the 64(*c*) conclusion was driven by the genotoxic potential of quinoline and was not based on a quantitative risk characterization.

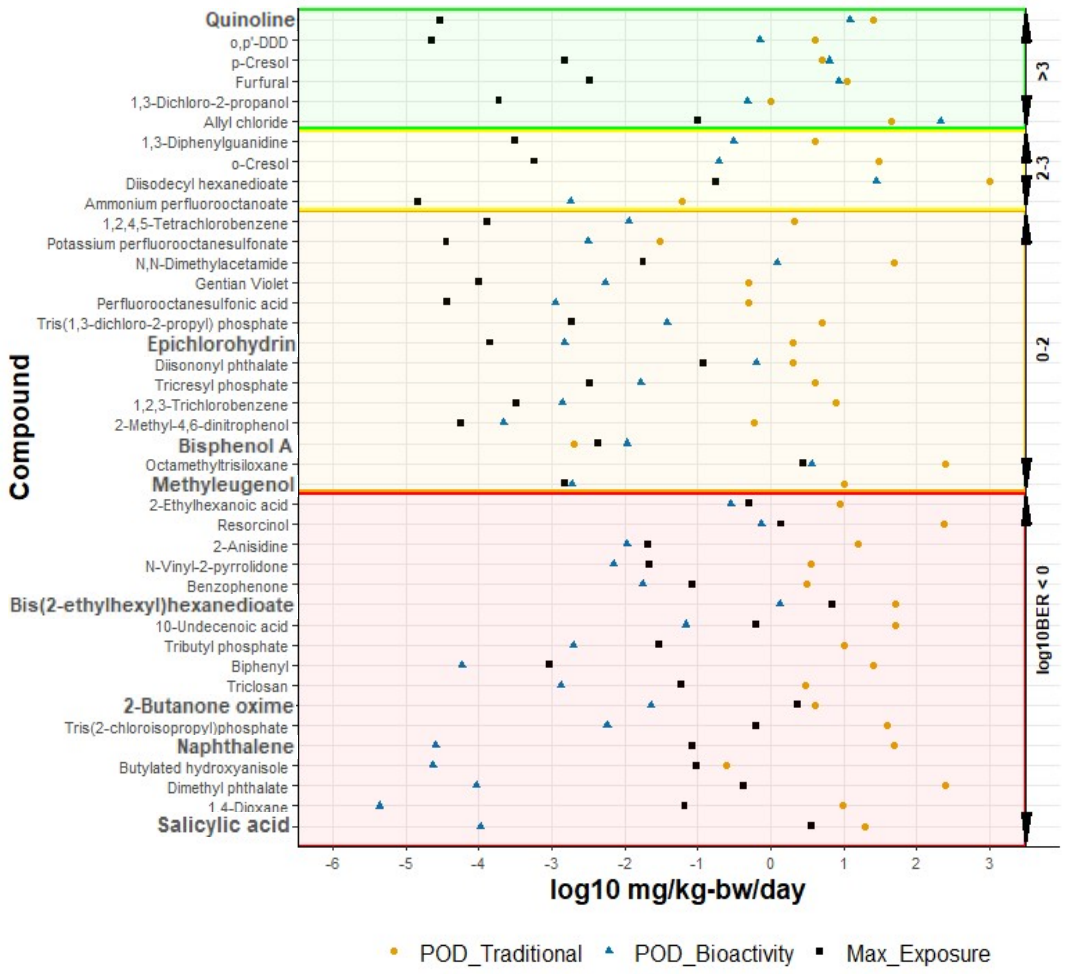

Figure 5-2. Comparison of Toxcast derived $POD_{Bioactivity}$, $POD_{Traditional}$, and maximum exposure values. Chemicals are arranged by descending $log_{10}BER$ and placed into four bins: $log_{10}BER < 0$ ($POD_{Bioactivity}$ < exposure), $log_{10}BER$ 0- 2, $log_{10}BER$ 2-3, and $log_{10}BER > 3$. Assessed chemicals with a toxic under section 64 of CEPA 1999 are bolded. 2-Nitrotoluene was the other toxic substance but does not appear in the figure due to lack of exposure data.

It should be noted that several (17/41) of the exposure scenarios for substances with low $log_{10}BERs$ are considered intermittent or acute scenarios (not daily exposures). For the purposes of this approach, BERs based on intermittent exposures were still reported; however, these substances used intermittently may not reach a steady state level in the plasma as assumed in the prediction of the $POD_{Bioactivity}$ values. Adaptations to the HTTK models and modifications to the workflow (such as using maximum plasma concentration) (Wambaugh et al. 2018) may be applied in future applications for handling cases with intermittent exposure patterns.

6. Uncertainty Factors (UFs) to Consider When Determining Adequacy of BER

The primary application context of the BER approach is as a risk-based screening tool to support prioritization and rapid risk assessment activities. As such, various decisions related to how the $POD_{Bioactivity}$ was calculated incorporated conservative considerations (e.g. selecting the 5^{th} percentile of AC_{50} values from the assays and using the C_{ss} at the 95^{th} percentile in a population for IVIVE). Nonetheless, there are several areas of uncertainty inherent in the approach that will be qualitatively described. Default UFs are proposed that can be used to help determine what an adequate target BER may be when using the $POD_{Bioactivity}$ for prioritization and screening level risk assessment purposes. Here we propose factors that can be broadly applied to the hazard component of the BER based on lessons learned in this analysis and that conducted under the APCRA. The uncertainties can be divided into three broad categories associated with deriving the $POD_{Bioactivity}$, use of cell-based assays, and inter-individual variability (i.e. human variability). We also note that defining the adequacy of a BER will also be case dependent and must also account for the uncertainty associated with the exposure predictions used to compare against the $POD_{Bioactivity.}$

6.1 Deriving the $POD_{Bioactivity}$ ($UF_{Bioactivity}$)

The current ToxCast test battery consists of nearly 1,400 assay endpoints (Richard et al. 2016), but uncertainty remains as to whether these assays comprehensively encompass the toxicological space by accurately quantifying the potencies of all possible effects (i.e. incomplete biological space). Future planned expansion of the ToxCast program through the incorporation of additional assays (e.g. toxicogenomics) will further serve to diminish the uncertainty surrounding the toxicological space. For present applications, the use of exclusion criteria addressing the chemicals outside the domain of applicability of the assay endpoints and UFs are recommended.

A source of uncertainty for predicting PODs using *in vitro* bioactivity data pertains to the confidence in the assays. Unidentified factors, such as inter-lab variability, may affect the accuracy and precision of AC_{50} predictions by individual assays. This in turn can impact the bioactivity concentration and subsequent AED calculations. For these reasons, although similar uncertainties exist with traditional data, a percentile-based approach was chosen to account for spurious measurement inaccuracies of individual assays that may underestimate the bioactivity threshold.

The biological models associated with *in vitro* assays are limited in terms of biological complexity. This adds to the uncertainty of interpreting an assay outcome. Specifically, there exist uncertainties in establishing a clear link between key events captured by the *in vitro* assays and adverse disease outcomes. However, as the Adverse Outcome Pathway database (Ankley et al. 2010; Villeneuve et al. 2014) grows, the associations between key events and adverse outcomes will be more comprehensive. This will

provide a clearer indication of which assays and perturbations are more likely to lead to adverse effects. Furthermore, the incorporation of high content data assays, including transcriptomics, have the potential to map the biological pathways perturbed by the test compounds. However, it is important to note the function of the high-throughput assays, as outlined in this approach, is to be used as a screening tool rather than a predictor of specific hazards (Thomas et al. 2013), and in this regard the assays are useful in support of chemical prioritization.

The direct comparison of *in vitro* PODs to those predicted *in vivo* here is challenging as the *in vitro* measurements were done using human cells, whereas *in vivo* studies were performed using rodents. Inter-species differences in toxicodynamics and toxicokinetics will compound the variability between the POD*s*. Thus, discrepancies between *in vitro* and *in vivo* studies may be due to an inaccurate *in vitro* prediction, uncertainty in IVIVE, and uncertainty in the *in vivo* (animal) model. Reproducing results using available rodent NAM models could help to explain any differences between *in vitro* and *in vivo* results, as the qualitative and quantitative concordance would be higher between *in vivo* results and rodent-based NAMs. However, continued use of human cell models may be more relevant for human health-based assessments and is a more appropriate use of resources.

The reverse dosimetry used to estimate the AEDs is another source of uncertainty in using bioactivity as a protective POD (Wambaugh et al. 2019). Specifically, the IVIVE method, HTTK model, and assay measurements chosen can impact the C_{ss} value used in calculating the AED. There is uncertainty around choosing a steady state model in performing IVIVE. A steady state model has a propensity to be more accurate for pharmaceuticals, which typically have a regular dosing schedule and are designed to be readily absorbed via the oral route. There is more uncertainty for diverse industrial and environmentally relevant compounds with sporadic exposures that also may not be readily absorbed. However, previous results have shown this approach to be sufficiently robust for the chemical space of interest (Wambaugh et al. 2015; 2018) and, consistent use of the steady state model across all classes can be viewed as a precautionary approach. It is acknowledged that the simplified high-throughput IVIVE strategy has limitations (Wetmore 2015a). For example, the model does not consider factors such as active renal re-absorption and enterohepatic recirculation. However, comparisons with *in vivo* results demonstrate that the missing parameters are only expected to impact a small fraction of compounds being screened.

The predictive ability of the HTTK models may be affected by unique chemical properties outside the chemical space examined in previous work. This is especially true for diverse classes of environmental and industrial compounds. The IVIVE may not be suitable for certain test substances, such as those that bioaccumulate or fail to reach steady state (Wambaugh et al. 2015) or those where the main routes of exposure are dermal or through inhalation. For other substances, the parameters used in the IVIVE model may not be the most accurate. For example, two types of hepatic clearance can be used in C_{ss} calculation: restrictive (dependent on F_{up}) and non-restrictive

(independent of F_{up}). Given that there is no way to predict which type of clearance applies to a given substance (Yoon et al. 2014), restrictive clearance was chosen as this assumption performs well in HTTK models (Honda et al. 2019) and it is on the conservative side. The same uncertainty applies to other parameters built into the models. A comparison between *in vitro* predicted C_{ss} and *in vivo* C_{ss} values in the literature demonstrated that methods and assumptions similar to the ones applied here are appropriate for >85% of suitable chemicals, in that their predicted C_{ss} are within 10× of the *in vivo* C_{ss} (Wambaugh et al. 2015; 2018). Implementation of future refinements will improve the predictive ability of rapid IVIVE and reduce the uncertainty with these models.

If the limited toxicological space in ToxCast and the approaches used for IVIVE were a significant factor driving the $POD_{Bioactivity}$ to be a less conservative estimate for $POD_{Traditional}$, one would expect a much higher proportion of chemicals in the presented case studies to reflect this. Indeed, no chemicals in the CMP specific analysis of 46 chemicals had a $POD_{Bioactivity}$ that was 10-fold higher than the $POD_{Traditional.}$ In the worst scenario, the $POD_{Bioactivity}$ was 5.4-fold higher than $POD_{Traditional}$. Therefore, a UF of less than 10 (i.e., 3) is proposed, which when applied, scales the $POD_{Bioactivity}$ to an equal or lower dose to $POD_{Traditional}$ for nearly all chemicals examined.

6.2 Cell-based limitations (UF_{Cells})

Cell-based assays provide efficient and biologically relevant assessment of chemical toxicity mechanisms (NASEM 2015). However, cell-based assays also have limitations that increase the uncertainty associated with deriving the $POD_{Bioactivity}$. Specifically, there is uncertainty surrounding the biotransformation of parent compounds to metabolites *in vivo* that are not accounted for in the *in vitro* assays. The *in vitro* assays often lack metabolic competence for xenobiotics, and this can lead to false-positives, if the chemical is detoxified *in vivo*, or false-negatives, if the metabolite is bioactive (DeGroot et al. 2018). This has proven to be a challenge for ToxCast in its ability to assess the acetylcholinesterase inhibition of organophosphates and related compounds (Aylward and Hays 2011). The uncertainty surrounding metabolism may diminish as ToxCast expansion incorporates methods to account for metabolism.

Additional uncertainties exist regarding the conditions in which the cells are cultured. For example, there is evidence that antibiotics used in cell culture to prevent contamination can alter biological activity of the cells (Ryu et al. 2017). Another consideration is that cell lines at high passage number (i.e., number of times cells have been subcultured) can have altered cell morphology, growth rates, and response to stimuli compared to cells with a lower passage number (ATCC 2010, Kwist et al. 2016). Lastly, the assays each consist of one cell line (monocultures) that are unable to replicate cellular interactions within biological systems. Emerging approaches focused on the use of co-cultures, consisting of different cell types (i.e., organotypic tissue models), or organ-on-a-chip technologies, may help to better replicate human physiological conditions.

A portion of the cell-based assays in ToxCast use immortalized human cancer cells (US EPA 2019b), such as HeLa (human cervical cancer cells), BG1 (ovarian cancer), and HepG2 (liver cancer). These cell lines contain genetic alterations that make them amenable to assessing biological activity in culture. As a consequence of these alterations, there is an added dimension of uncertainty regarding the relevance of the bioactivity in these cell lines to human physiology and human health assessments. As these cell lines are derived from a single individual tumour, they fail to capture inter-individual human variability (discussed more as its own uncertainty, referred to as UF_{Human}, in section 6.3).

Taking into consideration that the cell-based assays only form a component of the broader ToxCast program, and there are redundancies in endpoint evaluation, a UF of 3 is proposed to account for cell-based uncertainty. A UF of 3, as opposed to 10, is also proposed to acknowledge that some of the uncertainties related to using cell-based assays are already captured in $UF_{Bioactivity}$ and UF_{Human}.

6.3 Inter-individual (human) variability (UF_{Human})

Typically in risk assessment, regardless if the $POD_{Traditional}$ is based on animal or human data, a default UF of 10 is applied to account for the inter-individual (human) variability. This factor accounts for general differences within humans related to physiology and metabolism, which may result in sensitive subpopulations due to variations in age, sex, and genetic susceptibility among other considerations. This factor is generally considered to encompass both toxicodynamic and toxicokinetic differences in a population.

ToxCast mostly makes use of human cell lines (discussed in section 6.2) or targets (receptors); however, these assay systems may not be able to account for all the variability in a population, which can translate to uncertainty around toxicodynamics. Moreover, there is uncertainty regarding the inter-individual variability covered by the toxicokinetics approach. Specifically, intrinsic clearance rate declines with age and populations aged 65 and over have been shown to have lower AEDs using the population simulator (Ring et al. 2017). In contrast, younger age groups (19 or younger) have higher clearance rates and this is interpreted as lower risk in the AED determination. This is somewhat accounted for in this approach as the IVIVE method applied here uses a Monte Carlo simulation to model population effects when calculating AED 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. By default, the modelling parameters represent the US population so there is some uncertainty on how this applies to the Canadian population, but it is expected to be a reasonable approximation.

For this approach, the standard UF of 10 will be applied to account for inter-individual human variability. This is likely to be conservative as the toxicokinetic portion of this factor is already at least partially accounted for in the HTTK model used.

6.4 Summary of Hazard UFs

Case studies, such as the one presented in this SciAD and conducted under the APCRA, comparing *in vitro* and *in vivo* PODs have been effective in demonstrating that the current battery is adequately predictive of a lower bound estimate of effect levels based on *in vivo* adverse outcomes used in risk assessments in the vast majority of cases. In the APCRA case study, after excluding organophosphate and carbamates, only four chemicals were found to have $POD_{Bioactivity}$ 10-fold higher than $POD_{Traditional}$ (i.e. 99% of chemicals had a $\log_{10}POD$ ratio of greater than -1). The median $\log_{10}POD$ ratio in the APCRA case study was 2, meaning that typically the $POD_{Bioactivity}$ is already 100-fold lower than $POD_{Traditional.}$ Through the combination of UFs from the three categories of uncertainties, a total UF rounded to 100 (Table 6-1) can be considered when using the $POD_{Bioactivity}$ to derive the BER for prioritization and screening assessment purposes (see section 7). This combined factor is considered conservative and is expected to cover the potential gaps in biological space covered by the ToxCast assays along with the uncertainties associated with using cell-based assays and IVIVE methods.

Table 6-1. Proposed Uncertainty Factors related to the $POD_{Bioactivity}$ to aid in determining the adequacy of BER.

Type	Factor	Rationale
Deriving $POD_{Bioactivity}$ ($UF_{Bioactivity}$)	3	Incomplete biological space covered by assays in ToxCast. Uncertainties associated with the three compartment model to estimate C_{ss} using *in vitro* toxicokinetic parameters.
Immortalized Monocultures and Culture Conditions (UF_{Cells})	3	Considers effects of using monocultures and immortalized cell lines, as well as culture conditions, on endpoint measurements. Limitations of single cell type as a surrogate for systemic effects as well as limited metabolic competence.
Inter-individual (human) variability (UF_{Human})	10	Inter-individual variability related to toxicodynamics and toxicokinetics. Note this is likely conservative as the normal toxicokinetic portion of this factor is already at least partially accounted for in the HTTK model.
TOTAL	~100	

6.5 Exposure estimates

Exposure estimates are based on information available at the time of each screening assessment report. Updated information may be available on these substances; however, for the purposes of this analysis, it was considered appropriate to compare the information that formed the basis of the original risk assessment conclusions with the values derived using this approach.

Substances that have route-specific risk issues may not be identified using this approach (i.e. local portal of entry effects rather than systemic effects). In addition, the characterization of inhalation exposures as a dose (mg/kg bw/day) may also prevent identification of risk issues associated with peak exposure (e.g., peak air concentrations resulting in a health effect). Thus, the uncertainty of exposure estimates should be considered on a case-by-case basis.

7. Application of the approach under the CMP

Using the UFs outlined above, several recommendations can be made regarding the use of this approach to support priority setting and risk-based screening assessments. The BER approach can be applied and substances can be "binned" for consideration under the Identification of Risk Assessment Priorities (IRAP), in problem formulation, or other testing and assessment related activities particularly where there is a paucity of *in vivo* data available on which to base decisions (Figure 5-2; Table 7-1). Provided there is adequate confidence in the exposure prediction, the $\log_{10}$BERs for known toxics analyzed in this SciAD suggest a BER of less than 100 ($\log_{10}$BER < 2) would indicate that the chemical is a higher priority for further action. For these chemicals, the BER may not be adequate to account for the uncertainties inherent in the approach. For the purposes of prioritization, BER values that approach the total UF (i.e., BER between 100 and 1,000 or $\log_{10}$BER between 2 and 3) would be considered marginal and would then be more closely scrutinized for a decision regarding further action. When the ratio of bioactivity to exposure is high such that a BER is greater than or equal to 1,000 ($\log_{10}$BER >3), substances would not be considered a priority using the BER. However, prioritization and risk assessment both generally consider multiple lines of evidence related to chemical hazard and exposure as available. Thus, there may be other indicators of hazard and exposure that could result in a recommended action or decision beyond what is indicated by the BER independently, particularly if higher tier data are available (e.g. *in vivo* studies).

Table 7-1. Proposed BER thresholds for use in prioritization and assessment

BER[1]	Use	Rational
<1	Trigger for further consideration	Exposure is higher than the $POD_{Bioactivity}$ suggesting a potential

		concern. The substance would be considered a priority for further action.
1-100	Trigger for further consideration	The BER may not be adequate to account for the uncertainties inherent in the approach used to derive $POD_{Bioactivity}$ and to account for inter-individual variability. There may be a potential concern; further investigation warranted.
100-1,000	Case-by-case consideration for prioritization	The BER is approaching a threshold that may not account for the uncertainties inherent in the approach used to derive $POD_{Bioactivity}$ and the inter-individual variability. These substances should be considered on a case-by-case basis for prioritization alongside any additional supporting information (i.e., elements of exposure estimates).
>1,000	Not considered a current priority For certain screening level risk assessments, in the absence of *in vivo* data or when other indicators of potential hazard are limited, a BER of greater than 1000 may be used as a line of evidence to support a decision of not toxic under section 64(*c*) of CEPA.	The ratio of bioactivity to exposure is high and the substance is not considered to be a priority for further action unless there are other relevant hazard or exposure indicators to support a prioritization or assessment decision

[1] BER based on $POD_{Bioactivity}$ provided there is adequate confidence in the exposure prediction

Using the UFs outlined in Table 6-1 to guide the development of bins in Table 7-1 would result in a total of 35 out of 41 compounds, with quantitative exposure data, being triggered or prioritized for further consideration. The high proportion of prioritized compounds is unsurprising considering that these compounds had been previously identified as priorities for assessment.

It is anticipated that the approach of generating BERs will evolve to incorporate additional sources of NAM data. As further *in vitro* and high content assays advance within the context of the BER approach, these technologies, and the data generated, may be considered as available for the ongoing expansion of the approach. For example, high-throughput screening based on transcriptomic data may be used to

support BER derivation as part of a tiered testing scheme (Thomas et al. 2013; Mezencev and Subramaniam 2019).

As experience with the BER approach increases, data may be used to support risk assessment within the existing Health Canada framework. It is envisioned that complementary screening tools, such as *in silico* Quantitative Structure-Activity Relationships (QSARs) and the Threshold of Toxicological Concern (TTC), will serve to identify genotoxics and other chemical classes not amenable to the BER approach described here.

7.1 Substance Exclusion Considerations

The approach described in this SciAD may have less applicability to certain classes of substances and/or exposure scenarios. Consequently, a determination of whether a substance will be excluded from future application of the approach will be made on a case by case basis as the approach is applied. Known considerations for excluding a substance are described below. However, as the intention of this approach is to continue to adapt the methods as greater experience is gained with substance classes or as new IVIVE models become available, these considerations will continue to evolve.

Volatile Chemicals

Volatile, low molecular weight compounds may not be well screened in the current ToxCast battery. This limitation is well known for *in vitro* based test methods and was identified when examining allyl chloride. It is likely that these types of chemicals require different assays, conditions, or chemical management to observe bioactivity based on physico-chemical properties than what is typically applied in ToxCast.

Moreover, these types of chemicals can have high acute/peak intermittent exposures where the primary route of concern is inhalation. A more suitable generalized IVIVE model intended for these types of compounds and exposures would be preferable to convert observed relevant bioactivity into a human equivalent dose. Finally, the $POD_{Bioactivity}$ was compared against only $POD_{Traditional}$ from orally conducted toxicity studies in both the APCRA case study and the work presented here. Determining whether $POD_{Bioactivity}$ can provide a lower bound estimate for effect levels observed in animal inhalation studies would require further analysis.

Thus, volatile, low molecular weight chemicals will be considered for exclusion when moving forward with the currently described approach.

Organophosphates and Carbamates

As previously mentioned, this SciAD builds on work that was conducted as part of an international collaboration under the APCRA initiative (Paul Friedman et al. 2019). The APCRA case study specifically examined if certain chemical classes more commonly

had their respective $POD_{Bioactivity}$ higher than available $POD_{Traditional}$ from animal studies. To do this, the 448 chemicals were labelled with their structural features (i.e., chemotypes) based on the use of a publically available structural feature set known as ToxPrint developed by Altamira (Altamira, Columbus, OH USA) and Molecular Networks (Molecular Networks, Erlangen, GmbH) under contract from the U.S. Food and Drug Administration (Yang *et al.*, 2015). Statistical methods (odds-ratio (OR) with Fisher's exact test) were then used to determine if a particular structural feature was significantly associated with chemicals where the $POD_{Bioactivity}$ was greater than the $POD_{Traditional}$ (i.e., $\log_{10}$POD ratio < 0). Structural features related to organophosphate and carbamate chemistries had an OR ≥ 3 and p-value ≤ 0.05. Moreover, 24 of the 48 substances in the APCRA case study with a $\log_{10}$POD ratio < 0 were considered to be organophosphates or carbamates with 21 of these having a clear indication of being a pesticide. Finally, only 3 of the 48 substances had a $\log_{10}$POD ratio of less than -2, all of which were organophosphate insecticides (Paul Friedman et al. 2019).

Three organophosphate compounds (tributyl phosphate, tricresyl phosphate and tris(2-chloroisopropyl)phosphate) were examined in the CMP case study presented within this SciAD. For all three compounds, the $POD_{Bioactivity}$ was lower than the available $POD_{Traditional}$. As pointed out in the APCRA case study, while ToxCast does have several assays that can measure acetylcholinesterase inhibition (Padilla et al. 2012; Sipes et al. 2013), there has been previous work describing the inability of these assays to accurately reflect acetylcholinesterase inhibition potency (Aylward and Hays 2011). Furthermore, organophosphate metabolites can be more potent acetylcholinesterase inhibitors, which may not presently be captured in the ToxCast *in vitro* assays due to limitations in metabolic capacity. Lastly, there is a broad spectrum of other potential targets vulnerable to organophosphates and carbamates (Casida and Quistad, 2004) that may not be captured by ToxCast assays, resulting in inaccurate hazard characterization.

Given these concerns, future application of this approach under the CMP, will exclude organophosphate and carbamate compounds.

Case Study Chemical Space

The APCRA case study and the examination of chemicals presented within this SciAD provide an indication of where $POD_{Bioactivity}$ is a lower bound estimate for possible effect levels observed *in vivo*. Moving forward, for each chemical where this approach is applied, verification will be performed to ensure that relevant structural features (e.g. chemotypes) and physicochemical properties fall within the existing case study chemical space. An example of such a chemical space analysis can be seen in the Health Canada SciAD for the TTC-based Approach for Certain Substances (Health Canada, 2016).

7.2 Research Needs

The results presented for the selected chemicals in this SciAD demonstrate the utility of using BERs in priority setting and risk assessment. However, for broader application of the approach, some key data gaps will need to be addressed and are discussed here as future research needs. The primary research needs involve establishing the domains of applicability of the existing HTTK and ToxCast models and databases.

The HTTK models and database have a defined scope. For example, the models are limited mainly to oral absorption, but there is ongoing work to expand to other routes of exposure, such as inhalation (Linakis et al. 2020, accepted). Further research to expand the routes of exposure captured by these models will greatly increase the domain of applicability of HTTK. Another consideration is that the HTTK database itself has an enrichment of specific chemical classes, such as pesticides. It is unclear how chemicals outside this chemical space will perform using the HTTK models. Characterization of the chemical space that HTTK encompasses is identified as an immediate research need, as it will aid evaluators in identifying appropriate chemical classes to which HTTK can be applied. For the chemical space presenting with unique kinetic properties, it will be important to determine the necessary model adjustments or identify more complex and suitable models to extrapolate *in vitro* bioactivity to *in vivo* equivalent doses.

Currently, the main data element limiting wider use of the BER approach is the requirement that the chemical under investigation be present in the ToxCast database. For broader application, other suitable pieces of information will need to be utilized. Thus, research identifying the key assays that can inform bioactivity is required. One promising approach is the application of transcriptomics. Case studies have already demonstrated that transcriptomics is a powerful tool in deriving conservative BERs (Gannon et al. 2019a; Harrill et al. 2019). In the absence of ToxCast data, transcriptomics studies can be used to provide high-content information to establish the relevant AOPs activated by the chemical being evaluated. This type of data alone, or in conjunction with other bioassays, will be useful as an alternative data source to generate conservative BERs for quantitative risk assessment applications.

Research into *in silico* models offers the most potential in expanding the applicability of the BER workflow. Specifically, *in silico* predictions can be used to fill in the data gaps for chemicals that lack HTTK data, ToxCast data, or both. There are multiple *in silico* models that can predict the required parameters for running HTTK models. For example, an active area of research is the identification of appropriate simulators of metabolism and clearance. Evaluation of the performance of these models is an important next step in addressing the data gap for HTTK input. Addressing the ToxCast data gap for many compounds is a more complex endeavor. However, advancing machine learning algorithms or adaptations to read-across approaches, may be useful to create weight-of-evidence when combined with other pieces of information. Lastly, exploration of novel exposure models will help to advance the BER workflow, as exposure information is often a limiting data gap in BER derivation.

A limitation of the BER approach in its current form is the lack of endpoints that adequately capture genotoxicity. This was noted by the high BER for quinoline, a compound that was previously identified as toxic substance due to its genotoxic potential. Ongoing research at Health Canada is focusing on identifying assays and endpoints that provide insight into genotoxicity (i.e., mutagen, clastogen, and aneugen assays). This work will form the basis for a complementary approach and SciAD that quantifies genotoxic BERs.

Overall, this SciAD demonstrates the utility of *in vitro* bioactivity data in quantitative risk-based prioritization and assessment. The research needs described here will be explored and addressed in an ongoing basis. The BER approach is to be regarded as dynamic, and it will continue to evolve as new sources of information become available and as the individual research needs are addressed.

References

Al'meev KS, Karmazin VE. 1969. Pathomorphological changes in the organs of animals under the impact of allyl alcohol and allyl chloride. Faktory Vnesh Sredy Ikh Snach Sdorov'ya Naseleniya 1:31-35. [cited in OECD 1996].

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, et al. 2010. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol and Chem 730-741

ANSES. 2014. Annex XV Restriction Report: Proposal for a Restriction: 4.4'-isopropylidenediphenol (bisphenol A; BPA). France.

[ATCC] American Type Culture Collection. 2010. Passage number effects in cell lines. ATCC Tech Bulletin No. 7.

[ATSDR] Agency for Toxic Substances and Disease Registry. 2008. Toxicological Profile for Cresols. Atlanta (GA): U.S. Department of Health and Human Services, Public Health Service.

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

Becker RA, Ankley GT, Edwards SW, Kennedy SW, Linkov I, Meek B, Sachana M, Segner H, Van Der Burg B, Villeneuve DL, et al. 2015. Increasing scientific confidence in adverse outcome pathways: application of tailored Bradford-Hill considerations for evaluating weight of evidence. Regul Toxicol Pharmacol 72:514-537.

[BRRC] Bushy Run Research Centre/Hazleton Laboratories. 1988. Project Report 51-508, Developmental toxicity evaluation of o-, m-, or p-cresol administered by gavage to New Zealand White rabbits, June, 1988 (at the request of CMA) EPA/OTS0517695 [cited in OECD 2005].

[BRRC] Bushy Run Research Center/Hazleton Laboratories. 1989. Final Project Report 52-512, Two-generation reproduction study of p-cresol (CAS No. 106-44-5) administered by gavage to Sprague-Dawley (CD) rats. NTIS Report No. OTS0529224 [cited in CIR Expert Panel 2006].

Blackwell BR, Ankley GT, Corsi SR, DeCicco LA, Houck KA, Judson RS, Li S, Martin MT, Murphy E, Schroeder AL, et al. 2017. An “EAR” on environmental surveillance and monitoring: A case study on the use of exposure–activity ratios (EARs) to prioritize sites, chemicals, and bioactivities of concern in Great Lakes waters. Environ Sci Technol 51(15):8713-8724.

Brown J, Watt ED, Setzer RW, Judson R,Paul-Friedman K. 2018. Defining uncertainty in publically available high-throughput screening data from the ToxCast program. Presented at 57th Annual Meeting of the Society of Toxicology, San Antonio, Texas, March 11 - 15, 2018.

Cagen SZ, Waechter JM, Jr., Dimond SS, Breslin WJ, Butala JH, Jekat FW, Joiner RL, Shiotsuka RN, Veenstra GE, Harris LR. 1999. Normal reproductive organ development in CF-1 mice following prenatal exposure to bisphenol A. Toxicol Sci 50(1):36-44.

Carr R, Bertasi F, Betancourt A, Bowers S, Gandy BS, Ryan P, Willard S. 2003. Effect of neonatal rat bisphenol A exposure on performance in the Morris water maze. J Toxicol Environ Health A 66(21):2077-88.

Casida JE, Quistad, GB. 2004. Organophosphate toxicology: safety aspects of nonacetylcholinesterase secondary targets. Chem Res Toxicol 17(8):983-998.

Ceccarelli I, Della SD, Fiorenzani P, Farabollini F, Aloisi AM. 2007. Estrogenic chemicals at puberty change ERalpha in the hypothalamus of male and female rats. Neurotoxicol Teratol 29(1):108-115.

[CIR Expert Panel] Cosmetic Ingredient Review Expert Panel. 2006. Cosmetic Ingredient Review. Final report on the safety assessment of Sodium-p-chloro-m-cresol, p-chloro-m-cresol, chlorothymol, mixed cresols, m-cresol, o-cresol, p-cresol, isopropyl cresols, thymol, o-cymen-5-ol, and carvacrol. Int J Toxicol 25 (Suppl 1):29-127.

CMP Science Committee. 2017. Considerations for integrating new approach methodologies within the Chemicals Management Plan: November 2016 committee report. Ottawa (ON): Government of Canada.

Corsi SR, De Cicco LA, Villeneuve DL, Blackwell BR, Fay KA, Ankley GT, Baldwin AK. 2019. Prioritizing chemicals of ecological concern in Great Lakes tributaries using high-throughput screening data and adverse outcome pathways. Sci Total Environ 686:995-1009.

Darwich AS, Neuhoff S, Jamei M, Rostami-Hodjegan A. 2010. Interplay of metabolism and transport in determining oral drug absorption and gut wall metabolism: a simulation assessment using the “Advanced Dissolution, Absorption, Metabolism (ADAM)” model. Curr Drug Metab 11:716-729.

DeGroot DE, Swank A, Thomas RS, Strynar M, Lee M, Carmichael PL, Simmons SO. 2018. mRNA transfection retrofits cell-based assays with xenobiotic metabolism. J.Pharmacol Toxicol Methods 92:77-94

[ECCC, HC] Environment and Climate Change Canada, Health Canada. 2016. Screening assessment: Internationally Classified Substance Grouping: Cresol (phenol, methyl-) Substances. Government of Canada.

[ECCC, HC] Environment and Climate Change Canada, Health Canada. 2009. Screening assessment for the Challenge: 1-Propene, 3-chloro- (3-Chloropropene). Chemical Abstracts Service Registry Number: 107-05-1. Ottawa (ON): Government of Canada.

[ECCC, HC] Environment and Climate Change Canada, Health Canada. 2008. Screening assessment for the Challenge: Phenol, 4,4' -(1-methylethylidene)bis- (Bisphenol A). Chemical Abstracts Service Registry Number 80-05-7. Ottawa (ON): Government of Canada.

Ema M, Fujii S, Furukawa M, Kiguchi M, Ikka T, Harazono A. 2001. Rat two-generation reproductive toxicity study of bisphenol A. Reprod Toxicol 15(5):505-523.

Environment Canada, Health Canada. 2014. Approach for identification of chemicals and polymers as risk assessment priorities under Part 5 of the Canadian Environmental Protection Act, 1999 (CEPA 1999). Ottawa (ON): Government of Canada.

Farmahin R, Gannon AM, Gagné R, Rowan-Carroll A, Kuo B, Williams A, Curran I, Yauk CL. 2019. Hepatic transcriptional dose-response analysis of male and female Fischer rats exposed to hexabromocyclododecane. Food Chem Toxicol 133:1102623

Farmahin R, Williams A, Kuo B, Chepelev NL, Thomas RS, Barton-Maclaren TS, Curran IH, Nong A, Wade MG, Yauk CL. 2017. Recommended approaches in the application of toxicogenomics to derive points of departure for chemical risk assessment. Arch Toxicol 91:2045-2065.

Filer DL, Kothiya P, Woodrow Setzer R, Judson RS, Martin MT. 2017. Tcpl: The ToxCast pipeline for high-throughput screening data. Bioinformatics 33:618-620.

Funabashi T, Kawaguchi M, Furuta M, Fukushima A, Kimura F. 2004. Exposure to bisphenol A during gestation and lactation causes loss of sex difference in corticotropin-releasing hormone-immunoreactive neurons in the bed nucleus of the stria terminalis of rats. Psychoneuroendocrinology 29(4):475-485.

Gannon AM, Moreau M, Farmahin R, Thomas RS, Barton-Maclaren TS, Nong A, Curran I, Yauk CL. 2019a. Hexabromocyclododecane (HBCD): A case study applying tiered testing for human health risk assessment. Food Chem Tox In Press: https://doi.org/10.1016/j.fct.2019.110581.

Gannon AM, Nunnikhoven A, Liston V, Rawn DFK, Pantazopoulos P, Fine JH, Caldwell D, Bondy G, Curran I. 2019b. Rat strain response differences upon exposure to technical or alpha hexabromocyclododecane. Food Chem Tox 130:284-307.

Harrill J, Shah I, Setzer RW, Haggard D, Auerbach S, Judson R, Thomas RS. 2019. Considerations for strategic use of high-throughput transcriptomics chemical screening data in regulatory decisions. Curr Opin Toxicol 15:64-75.

He F, Jacobs JM, Scaravilli F. 1981. The pathology of allyl chloride neurotoxicity in mice. Acta Neuropathol 55(2):125-133. [cited in OECD1996].

Health Canada. 1994. Human health risk assessment for priority substances. Ottawa (ON): Minister of Supply and Services Canada. Cat. No.: En40-215/41E.

Health Canada. 2016. Threshold of Toxicological Concern (TTC)-based Approach for Certain Substances.

Hirose M, Inoue T, Asamoto M, Tagawa Y, Ito N. 1986. Comparison of the effects of 13 phenolic compounds in induction of proliferative lesions of the forestomach and increase in the labeling indices of the glandular stomach and urinary bladder epithelium of Syrian golden hamsters. Carcinogenesis 7(8):1285-1289 [cited in ATSDR 2008].

Honda GS, Pearce RG, Pham LL, Setzer RW, Wetmore BA, Sipes NS, Gilbert J, Franz B, Thomas RS, Wambaugh JF. 2019. Using the concordance of in vitro and in vivo data to evaluate extrapolation assumptions. PloS one 14(5):e0217564.

Howdeshell KL, Furr J, Lambright CR, Wilson VS, Ryan BC, Gray LE, Jr. 2007. Gestational and lactational exposure to ethinyl estradiol, but not bisphenol A, decreases androgen-dependent reproductive organ weights and epididymal sperm abundance in the male long evans hooded rat. Tox Sci 102(2):371-382.

Hyndman RJ, Fan Y. 1996. Sample Quantiles in Statistical Packages. Am Stat. 50:361-365.

[IARC] IARC Working Group on the Evaluation of the Carcinogenic Risk of Chemicals to Humans. 1985. Allyl compounds, aldehydes, epoxides and peroxides. IARC Monogr Eval Carcinog Risk Chem Hum 36:39-54.

Ishido M, Yonemoto J, Morita M. 2007. Mesencephalic neurodegeneration in the orally administered bisphenol A-caused hyperactive rats. Toxicol Lett 173(1):66-72.

[IPCS] International Programme on Chemical Safety. 2004. IPCS Risk Assessment Terminology. Geneva (CH): World Health Organization.

[JECFA] Joint FAO/WHO Expert Committee on Food Additives. 2011. Safety evaluation of certain food additives and contaminants. WHO Food Additive Series 64. Prepared by the 73rd meeting of the Joint FAO/WHO Expert Committee on Food Additives [p. 207-253: Addendum to Phenol and phenol derivatives].

Johnson CL, Dohrmann SM, Burt VL, Mohadjer LK. 2014. National health and nutrition examination survey: sample design, 2011-2014. Vital Health Stat 2(162):1-33

Judson RS. 2019. Package'tcpl' Version 2.0.2: Reference Manual.

Judson RS, Kavlock RJ, Setzer RW, Cohen Hubal EA, Martin MT, Knudsen TB, Houck KA, Thomas RS, Wetmore BA, Dix DJ. 2011. Estimating toxicity-related biological pathway altering doses for high-throughput chemical risk assessment. Chem Res Toxicol 24(4):451-462.

Kavlock R. 2016. Practitioner Insights: Bringing New Methods for Chemical Safety into the Regulatory Toolbox; It is Time to Get Serious. Bloomberg BNA Daily Environment Report (223 B-1).

Kavlock RJ, Bahadori T, Barton-Maclaren TS, Gwinn MR, Rasenberg M, Thomas RS. 2018. Accelerating the pace of chemical risk assessment. Chem Res Toxicol 31(5):287-290.

Kawai K, Murakami S, Senba E, Yamanaka T, Fujiwara Y, Arimura C, Nozaki T, Takii M, Kubo C. 2007. Changes in estrogen receptors alpha and beta expression in the brain of mice exposed prenatally to bisphenol A. Regul Toxicol Pharmacol 47(2):166-70.

Knudsen TB, Martin MT, Kavlock RJ, Judson RS, Dix DJ, Singh AV. 2009. Profiling the activity of environmental chemicals in prenatal developmental toxicity studies using the US EPA's ToxRefDB. Reprod Toxicol 28:209-219

Kubo K, Arai O, Ogata R, Omura M, Hori T, Aou S. 2001. Exposure to bisphenol A during the fetal and suckling periods disrupts sexual differentiation of the locus coeruleus and of behavior in the rat. Neurosci Lett 304(1-2):73-6.

Kwist K, Bridges WC, Burg KJ. 2016. The effect of cell passage number on osteogenic and adipogenic characteristics of D1 cells. Cytotechnology 68(4):1661-1667.

Kwon S, Stedman DB, Elswick BA, Cattley RC, Welsch F. 2000. Pubertal development and reproductive functions of Crl:CD BR Sprague-Dawley rats exposed to bisphenol A during prenatal and postnatal development. Toxicol Sci 55(2):399-406.

Lee JY, Miller JA, Basu S, Kee TZ, Loo LH. 2018. Building predictive in vitro pulmonary toxicity assays using high-throughput imaging and artificial intelligence. Arch Toxicol 92(6):2055-2075.

Lewis R, Billington R, Debryune E, Gamer A, Lang B,Carpanini F. 2002. Recognition of adverse and nonadverse effects in toxicity studies. Tox Pathol 30:66-74.

Martin MT, Judson RS, Reif DM, Kavlock RJ, Dix DJ. 2009a. Profiling chemicals based on chronic toxicity results from the U.S. EPA ToxRef Database. Environ Health Perspect 117:392-399.

Martin MT, Mendez E, Corum DG, Judson RS, Kavlock RJ, Rotroff DM, Dix DJ. 2009b. Profiling the reproductive toxicity of chemicals from multigeneration studies in the toxicity reference database. Toxicol Sci 110:181-190.

Mezencev R, Subramaniam R. 2019. The use of evidence from high-throughput screening and transcriptomic data in human health risk assessments. Toxicol Appl Pharmacol 380:114706.

Microbiological Associates (Microbiological Associates Inc., Mulligan LT). 1988. Subchronic toxicity of para-cresol in Sprague-Dawley rats, MBA Chemical No: 25, Study No. 5221.08, Final Report, Bethesda (MD) (at the request of Research Triangle Institute: Dennis Dietz), NTIS Report No. PB88-195292 [cited in OECD 2005].

Moreau M, Nong A. Evaluating hexabromocyclododecane (HBCD) toxicokinetics in humans and rodents by physiologically based pharmacokinetic modeling. Food Chem Tox In Press: https://doi.org/10.1016/j.fct.2019.110785.

Morrissey RE, George JD, Price CJ, Tyl RW, Marr MC, Kimmel CA. 1987. The developmental toxicity of bisphenol A in rats and mice. Fundam Appl Toxicol 8(4):571-582.

Nagao T, Saito Y, Usumi K, Yoshimura S, Ono H. 2002. Low-dose bisphenol A does not affect reproductive organs in estrogen-sensitive C57BL/6N mice exposed at the sexually mature, juvenile, or embryonic stage. Reprod Toxicol 16(2):123-30.

Nagel SC, vom Saal FS, Thayer KA, Dhar MG, Boechler M, Welshons WV. 1997. Relative binding affinity-serum modified access (RBA-SMA) assay predicts the relative in vivo bioactivity of the xenoestrogens bisphenol A and octylphenol. Environ Health Perspect 105(1):70-76.

[NASEM] National Academies of Sciences, Engineering, and Medicine. 2015. Application of modern toxicology approaches for predicting acute toxicity for chemical defense. National Academies Press.

National Cancer Institute. 1978. Bioassay of allyl chloride for possible carcinogenicity. Bethesda (MD): US Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health. Carcinogenesis Technical Report Series No. 73; DHEW Publication No. (NIH) 78-1323. [cited in IARC 1985].

Negishi T, Kawasaki K, Suzaki S, Maeda H, Ishii Y, Kyuwa S, Kuroda Y, Yoshikawa Y. 2004. Behavioral alterations in response to fear-provoking stimuli and tranylcypromine induced by perinatal exposure to bisphenol A and nonylphenol in male rats. Environ Health Perspect 112(11):1159-1164.

[NTP] National Toxicology Program. 2018. NTP Research Report on National Toxicology Program Approach to Genomic Dose-Response Modeling.

[NTP] National Toxicology Program. 1992. NTP report on the toxicity studies of cresols (CAS Nos. 95-48-7, 108-39-4, 106-44-5) in F344/N rats and B6C3F1 mice (feed studies). Research Triangle Park (NC): National Toxicology Program. NIH Publication No. 92-3128. NTP Tox 9. [Cited in ATSDR 2008]

[OECD] Organisation for Economic Co-operation and Development. 2008. Test Guideline 407 Repeated Dose 28-Day Oral Toxicity Study in Rodents.

[OECD] Organisation for Economic Co-operation and Development. 2005. M/P-CRESOL CATEGORY. m/p-Cresol (CAS No: 15831-10-4); m-Cresol (CAS No: 108-39-4); p-Cresol (CAS No: 106-44-5)SIDS Initial Assessment Report for SIAM 16, Paris, 27.

[OECD] Organisation for Economic Co-operation and Development. 1996. SIDS Initial Assessment Report for 4th SIAM (Tokyo, 20–22 May 1996). Chloropropene, CAS No. 107-05-1. UNEP Publications. Including accompanying SIDS (IUCLID) dataset, dated February 17, 2003.

Padilla S, Corum D, Padnos B, Hunter DL, Beam A, Houck KA, Sipes N, Kleinstreuer N, Knudsen T, Dix DJ., et al. 2012. Zebrafish developmental screening of the ToxCast Phase I chemical library. Reprod Toxicol 33(2):174-87.

Palanza PL, Howdeshell KL, Parmigiani S, vom Saal FS. 2002. Exposure to a low dose of bisphenol A during fetal life or in adulthood alters maternal behavior in mice. Environ Health Perspect 110 (Suppl 3):415-422.

Paul Friedman K, Gagne M, Loo H, Karamertzanis P, Netzeva T, Sobanski T, Franzosa J, Richard A, Lougee R, Gissi A, et al. 2019. Utility of in vitro bioactivity as a lower bound estimate of in vivo adverse effect levels and in risk-based prioritization. Tox Sci 173(1):202-225.

Paul Friedman K, Papineni S, Marty MS, Yi KD, Goetz AK, Rasoulpour RJ, Kwiatkowski P, Wolf DC, Blacker AM, Peffer RC. 2016. A predictive data-driven framework for endocrine prioritization: a triazole fungicide case study. Crit Rev Toxicol 46(9):785-833.

Pearce RG, Setzer RW, Strope CL, Wambaugh JF, Sipes NS. 2017. Httk: R package for high-throughput toxicokinetics. J Stat Softw. 79:1.

R Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (version 2.15).

Richard AM, Judson RS, Houck KA, Grulke CM, Volarath P, Thillainadarajah I, Yang C, Rathman J, Martin MT, Wambaugh JF. 2016. ToxCast chemical landscape: paving the road to 21st century toxicology. Chem Res Toxicol 29:1225-1251.

Ring CL, Pearce RG, Setzer RW, Wetmore BA, Wambaugh JF. 2017. Identifying populations sensitive to environmental chemicals by simulating toxicokinetic variability. Environ Int 106:105-118.

[RIVM] Rijksinstituut voor Volksgezondheit en Milieu. 1991. Voorstel voor de humaan-toxicologische onderbouwing van C-(toetsings) waarden. [Proposal for the foundation of human-toxicological C-(for review) values]. Vermeire TG, van Apeldoorn ME, de Fouw JC, Janssen PJCM. RIVM-report no. 725201005, February 1991. Bilthoven ((NL): National Institute for Public Health and the Environment.

Rotroff DM, Wetmore BA, Dix DJ, Ferguson SS, Clewell HJ, Houck KA, LeCluyse EL, Andersen ME, Judson RS, Smith CM. 2010. Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening. Toxicol Sci 117: 348-358.

Ryu AH, Eckalbar WL, Kreimer A, Yosef N, Ahituv N. 2017. Use antibiotics in cell culture with caution: genome-wide identification of antibiotic-induced changes in gene expression and regulation. Sci Rep 7(1):7533.

Sipes NS, Martin MT, Kothiya P, Reif DM, Judson RS, Richard AM, Houck KA, Dix DJ, Kavlock RJ, Knudsen TB. 2013. Profiling 976 ToxCast chemicals across 331 enzymatic and receptor signaling assays. Chem Res Toxicol 26(6):878-895.

Thomas RS, Bahadori T, Buckley TJ, Cowden J, Deisenroth C, Dionisio KL, Frithsen JB, Grulke CM, Gwinn MR, Harrill JA. 2019. The next generation blueprint of computational toxicology at the US Environmental Protection Agency. Toxicol Sci: 169(2):317-332.

Thomas RS, Philbert MA, Auerbach SS, Wetmore BA, Devito MJ, Cote I, Rowlands JC, Whelan MP, Hays SM, Andersen ME. 2013. Incorporating new technologies into toxicity testing and risk assessment: moving from 21st century vision to a data-driven framework. Toxicol Sci 136:4-18.

Tilley SK, Reif DM, Fry RC. 2017. Incorporating ToxCast and Tox21 datasets to rank biological activity of chemicals at Superfund sites in North Carolina. Environ Int 101:19-26.

Timms BG, Howdeshell KL, Barton L, Bradley S, Richter CA, vom Saal FS. 2005. Estrogenic chemicals in plastic and oral contraceptives disrupt development of the fetal mouse prostate and urethra. Proc Natl Acad Sci U S A 102(19):7014-9.

[TRL] Toxicity Research Laboratories. 1986. Subchronic neurotoxicity study in rats of ortho-, meta-, and para-cresol. Unpublished data submitted by Toxicity Research Laboratories to EPA [cited in ATSDR 2008].

Turley AE, Isaacs KK, Wetmore BA, Karmaus AL, Embry MR, Krishan M. 2019. Incorporating new approach methodologies in toxicity testing and exposure assessment for tiered risk assessment using the RISK21 approach: Case studies on food contact chemicals. Food CHem Toxicol 134:110819.Tyl RW, Myers CB, Marr MC, Thomas BF, Keimowitz AR, Brine DR, Veselica MM, Fail PA, Chang TY, Seely JC, et al. 2002. Three-generation reproductive toxicity study of dietary bisphenol A in CD Sprague-Dawley rats. Toxicol Sci 68(1):121-146.

Tyl RW, Myers CB, Marr MC. 2007. Two-generation reproductive toxicity evaluation of bisphenol A (BPA; CAS No. 80-05-7) administered in the feed to CD-1® Swiss mice (modified OECD 416). Research Triangle Park (NC): RTI International Center for Life Sciences and Toxicology.

Tyl RW. 1988. Developmental toxicity evaluation of o-, m-, or p-cresol administered by gavage to New Zealand white rabbits. Chemical Manufacturers Association. Submitted to the U.S. Environmental Protection Agency under TSCA Section 4. OTS0517695 [cited in ATSDR 2008].

[US EPA] US Environmental Protection Agency. 2019a. The ToxCast(TM) analysis pipeline (tcpl) an R package for processing and modeling chemical screening data (Version 2.0).

[US EPA] US Environmental Protection Agency. 2019b. ToxCast Data Generation: Overview of ToxCast Assays. Last updated on March 27, 2019.

[US EPA] US Environmental Protection Agency. 2015. ToxCast & Tox21 MySQL Database invitrodb (Version 3).

[US EPA] US Environmental Protection Agency. 2011. Integrated risk information system glossary.

Villeneuve DL, Crump D, Garcia-Reyero N, Hecker M, Hutchinson TH, LaLone CA, Landesmann B, Lettieri T, Munn S, Nepelska M. 2014. Adverse outcome pathway (AOP) development I: strategies and principles. Toxicol Sci 142:312-320.

Wambaugh JF, Hughes MF, Ring CL, MacMillan DK, Ford J, Fennell TR, Black SR, Snyder RW, Sipes NS, Wetmore BA. 2018. Evaluating in vitro-in vivo extrapolation of toxicokinetics. Toxicol Sci 163:152-169.

Wambaugh JF, Wetmore BA, Pearce R, Strope C, Goldsmith R, Sluka JP, Sedykh A, Tropsha A, Bosgra S, Shah I. 2015. Toxicokinetic triage for environmental chemicals. Toxicol Sci 147:55-67.

Wambaugh JF, Wetmore BA, Ring CL, Nicolas CI, Pearce RG, Honda GS, Dinallo R, Angus D, Gilbert J, Sierra T, et al. 2019. Assessing Toxicokinetic Uncertainty and Variability in Risk Prioritization. Toxicol Sci 172(2):235-251.

Watford S, Pham LL, Wignall J, Shin R, Martin MT, Friedman KP. 2019. ToxRefDB version 2.0: Improved utility for predictive and retrospective toxicology analyses. Reprod Toxicol 89:145-158.

Watt ED, Judson RS. 2018. Uncertainty quantification in ToxCast high throughput screening. *PLoS ONE* 13.

Wetmore BA. 2015a. Quantitative in vitro-to-in vivo extrapolation in a high-throughput environment. Toxicology 332:94-101.

Wetmore BA, Wambaugh, JF, Allen, B, Ferguson, SS, Sochaski, MA, Setzerm RW Houck, KA, Strope, CL, Cantwell, K, Judson, RS, LeCluyse, E, Clewell, HJ, Thomas, RS and Andersen, ME (2015b) Incorporating high-throughput exposure predictions with dosimetry-adjusted in vitro bioactivity to inform chemical toxicity testing, Toxicological Sciences 148 (1),121–136.

Wetmore BA, Wambaugh JF, Ferguson SS, Li L, Clewell III HJ, Judson RS, Freeman K, Bao W, Sochaski MA, Chu T. 2013. Relative impact of incorporating pharmacokinetics on predicting in vivo hazard and mode of action from high-throughput in vitro toxicity assays. Toxicol Sci 132:327-346.

Wetmore BA, Wambaugh JF, Ferguson SS, Sochaski MA, Rotroff DM, Freeman K, Clewell III HJ, Dix DJ, Andersen ME, Houck KA. 2012. Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicol Sci 125:157-174.

Williams AJ, Grulke CM, Edwards J, McEachran AD, Mansouri K, Baker NC, Patlewicz G, Shah I, Wambaugh JF, Judson RS, et al. 2017. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. J Cheminform 9(1):61.

Yang C, Tarkhov A, Marusczyk J, Bienfait B, Gasteiger J, Kleinoeder T, Magdziarz T, Sacher O, Schwab CH, Schwoebel J, et al. 2015. New Publicly Available Chemical Query Language, CSRML, To Support Chemotype Representations for Application to Data Mining and Modeling. J Chem Inf Model 55(3):510-528.

Yoon M, Efremenko A, Blaauboer BJ, Clewell HJ. 2014. Evaluation of simple in vitro to in vivo extrapolation approaches for environmental compounds. Toxicol *in vitro* 28:164-170.

Annex 1: Overview of additional in vitro data generation for HTTK

The native HTTK R package had some data available for 29 out of the 46 CMP compounds selected for analysis. The missing data needed for IVIVE were acquired by *in vitro* pharmacokinetics assays performed by Paraza Pharma Inc. (Montréal, QC).

F_{up} values unavailable in HTTK were measured by the plasma protein binding assay. Briefly, 10 µM of each test compound were added to human plasma (BioIVT, Westbury, NY, USA) and then aliquoted in triplicate onto a high throughput dialysis 96-well plate. Dialysate buffer and plasma are separated on the plates by a semi-permeable cellulose membrane that is impermeable to protein-bound chemicals. After a six hour incubation at 37°C with gentle agitation, the plasma and buffer samples were analyzed by either LC-MS/MS or GC-MS (chromatography method dependent on compound). F_{up} was determined using the formula:

$$F_{up} = \frac{Concentration\ in\ buffer}{Concentration\ in\ plasma}$$

Cl_{int} measurements for each compound were determined using the hepatocyte stability assay. Briefly, primary human hepatocytes (BioIVT) were incubated with compounds on a 96-well plate at 37°C at a final cell density of 0.5×10^6 cells/mL and a final compound concentration of 1 & 10 µM (sensitive compounds) or 10 & 30 µM (low sensitivity compounds). At the selected time-points (15, 30, 60, 120, and 240 minutes) the reactions were terminated. Samples were then analyzed by either LC-MS/MS or GC-MS (chromatography method dependent on compound) to determine the metabolic *in vitro* half-life. As standard practice, a threshold of 720 minutes (3× the incubation period) was applied as an upper limit half-life where it could not be determined experimentally for slow metabolizers. Cl_{int} values were calculated using the equation:

$$Clint = \frac{\ln 2\ /\ in\ vitro\ t_{1/2}}{0.5} \times 1{,}000$$

Where *in vitro* $t_{1/2}$ is the average half-life time, between both concentrations tested, in minutes
0.5 is the cell density in millions of cells/mL
1,000 is the conversion factor of mL to µL

To ensure that observed stability was not affected by cytotoxicity, the CellTiter-Glo® luminescent cell viability assay was performed. There were no signs of overt cytotoxicity in the hepatocytes at the chemical concentrations tested.

F_{gutabs} values could be determined using the Caco-2 permeability assay for 16 of the chemicals. Briefly, Caco2 monolayers were formed in 12-well Transwell plates. Test compounds were added to the donor chamber of the plates and cells remained incubated at 37°C. The donor chambers were the apical side of the Caco-2 monolayer for A-to-B assay or the basolateral side for B-to-A assay. Aliquots were taken from the receiver chamber at 30, 60, and 90 minute incubation time-points for LC-MS/MS analysis. The apparent permeability (P_{app}) for each compound were determined using the formula:

$$P_{app} = \frac{dQ/dt}{A \times C_i \times 60} \times 100$$

Where dQ/dt is the net rate of appearance in receiver compartment
A is the area of each Transwell (1.12 cm^2)
C_i is the initial concentration of compound in donor chamber
60 is the conversion factor for minutes to seconds

F_{gutabs} were then estimated using an empirical model (Darwich et al. 2010) with the following equations:

$$P_{eff} = 10^{(0.6532 \times \log P_{app} - 0.3036)}$$

$$F_{gutabs} = 1 - (1 + 0.54 \times P_{eff})^{-7}$$

Where P_{eff} is the effective human jejunum permeability coefficient

All of the chemicals tested had full or near full absorption ($F_{gutabs} \approx 1$). Therefore, the default assumption of full absorption in the HTTK package was applied for chemicals that were tested as well as those that could not be tested by the Caco-2 permeability assay.

Annex 2: Tcpl and httk R package functions used for the approach

Extraction of In vitro Bioactivity from ToxCast Database

The tcplLoadChem() function within the tcpl package was used to load the chemical list of 46 substances and tcplPrepOtpt() and tcplLoadData() functions were used to extract level 5, 6, and 7 data from the MySQL ToxCast database. Level 5 data includes hit-call (hitc) information for the chemicals for the assay endpoint and the AC_{50} values from the winning models used in the curve fitting process (modl_ga). Level 5 extraction was limited to assay endpoints that had an active hit-call (hitc =1) and where the chemical was tested in multiple concentration format (type =’mc’). Deriving a POD from concentration response curves was deemed more appropriate for this approach rather than including the single concentration tests, which may only be suitable for qualitative approaches. Within the ToxCast database multiple samples of a chemical may have been tested for a given assay endpoint. The function tcplSubsetChid() was used for these cases which applies a series of logic to subset the level 5 data so only a single chemical-assay endpoint pair are carried forward for further analysis (Judson 2018). Level 6 and 7 information was extracted for each given assay (on a sample id or ‘spid’ basis) which provides caution flags and uncertainty information, respectively, for the curve fits and hit-calls. Caution flags provide an indication of curves that may not have quantitatively informative AC_{50} values (see table of caution flags below).

ToxCast caution flags and descriptions

Caution Flag ID	Description
6	single point hit where activity is only at the highest concentration tested
7	single point hit with activity not at highest concentration tested
8	inactive assay with multiple medians above baseline
10	noisy curve, relative to the assay
11	hit with borderline activity
12	inactive assay with borderline activity
18	modelled AC_{50} less than lowest concentration tested
15	where gain-loss is the winning model, the gain AC_{50} is less than lowest concentration tested and the loss AC_{50} is less than the mean concentration
16	hit-call potentially confounded by overfitting (hit-call would change after small N correction in AIC values)
17	hit-calls with efficacy values less than 50%

Level 7 information attempts to quantify the level of uncertainty associated with the reproducibility of the curve fitting process and subsequent determination of hit-call, which is described in detail elsewhere (Watt and Judson 2018; Brown et al. 2018). Briefly, bootstrap methods are used to introduce normally distributed noise to the concentration-response values for each assay and then the curve fitting process is repeated using the three ToxCast models (i.e. constant, Hill, gain-loss). This resampling

process is repeated for 1,000 runs and for each run the winning model is determined along with the hit-call and modelled point-estimates (e.g. AC_{50} value) using the same methods for level 5 processing in tcpl. It is possible that the hit-call can change between runs and summary statistics are generated including the 'hit-percent' which provides an indication of the probability of the assay being classified as a hit after accounting for normally distributed noise. If the hit percent is low (i.e. below 50%) there is less confidence when classifying the assay as a hit (Watt and Judson 2018; Paul Friedman et al. 2019). Cytotoxicity in specific assays can have a confounding effect on interpreting the results of the observed bioactivity in ToxCast. Tested substances across ToxCast have been observed to exhibit non-specific activation of many targets measured in the assays as the cells approach death which has been termed the "burst" phenomenon (US EPA 2019a). The tcplCytoPt() function applies methodology for predicting chemical-specific cytotoxicity which makes use of up to 79 assay endpoints in the calculation ("burst assays"). The prediction essentially provides a concentration window where the chemical may be cytotoxic and includes a lower bound estimate to provide context for the "burst" phenomenon (US EPA 2019a). This function was used to extract the lower bound cytotoxicity prediction (cyto_pt_um) for the 46 chemicals in the case study.

Calculate administered equivalent dose (AED) corresponding to the in vitro bioactivity threshold which represents the PODBioactivity

The whole IVIVE process (C_{ss} and AED calculation) can be done in HTTK (version 1.8) using the function calc_mc_oral_equiv(). In assessing the 46 compounds, default parameters were used in calc_mc_oral_equiv() with the exception of output.units= "uM" and well.stirred.correction =T, which uses a well-stirred correction in the calculation of hepatic clearance (assumes clearance relative to amount of chemical unbound in whole blood as opposed to plasma).

Annex 3: Substances with a $POD_{Traditional}$ lower than $POD_{Bioactivity}$

p-Cresol

Bioactivity

Bioactivity for p-cresol was measured across 668 assay endpoints in ToxCast and was only active in 12 assays after the filtering criteria were applied (Table A3-1). p-Cresol was tested in 79 assays related to cytotoxicity and showed no activity; thus cytotoxicity is not expected to confound the bioactivity observed. Assay endpoints related to interactions with steroid hormones or nuclear receptors appear to be the most common bioactivity observed. The AC_{50} values across the active assays converted to their respective AEDs were found to range from 4.24 to 806.13 mg/kg bw/day. The AED for the 5^{th} percentile of AC_{50} values from the active assays which is the basis for the $POD_{Bioactivity}$ was estimated to be 6.33 mg/kg bw/day.

Table A3-1. Overview of 12 active assays endpoints for p-cresol grouped based on their intended target family

Intended Target Family	Sub Group	Number of Active Assays
cytokine	plasmogen activator	1
gpcr	rhodopsin-like receptor	2
steroid hormone	progestagens	2
steroid hormone	Estrogens	2
nuclear receptor	non-steroidal	3
dna binding	HMG box protein	1
dna binding	NF-kappa B	1

In vivo animal studies

p-Cresol was included as part of a group of cresols (ortho, para, meta, and mixed), and the final screening assessment was published in May 2016 under the CMP (ECCC, HC, 2016).

Many of the toxicity studies used in the group assessment were conducted using an isomer mixture of cresols. The details for the studies where the test material was p-cresol alone are available in the CMP assessment (ECCC, HC, 2016), and the results of these studies are summarized in Figure A3-1.

The lowest available oral $POD_{Traditional}$ for p-cresol is a NO(A)EL of 5 mg/kg bw/day based on a LO(A)EL identified in maternal rabbits at 50 mg/kg bw/day in a developmental toxicity study related to central nervous system (CNS) effects. However, the risk characterization in the assessment for cresols made use of available toxicity data across the various isomers of cresols as well as testing on mixtures. For non-

cancer effects, a MOE for all cresols was derived using a dose of 30 mg/kg bw/day which was determined to be a NO(A)EL for CNS effects and was protective for other effects observed at higher doses related to development.

Comparison of $POD_{Bioactivity}$ with $POD_{Traditional}$

The derived $POD_{Bioactivity}$ value of 6.33 mg/kg bw/day is similar to the lowest oral $POD_{Traditional}$ of 5 mg/kg bw/day, which is based on a maternal NO(A)EL from a developmental toxicity study in rabbits related to CNS effects (BRRC 1988; Tyl et al. 1988). It is important to note that the $POD_{Bioactivity}$ is lower than the LO(A)EL from the same study, and thus, the lower $POD_{Traditional}$ may be attributed to the doses used in the study. Moreover, the $POD_{Bioactivity}$ is lower than the 30mg/kg bw/day CNS effects-based NO(A)EL that was used for the MOE calculation in the assessment based on a weight of evidence across multiple cresols. The $POD_{Bioactivity}$ was also ~60 fold lower than the LO(A)EL identified from the developmental toxicity and the NO(A)EL identified for the reproductive toxicity study. Given the variability and issues with dose selection for $POD_{Traditional}$ derivation, the $POD_{Bioactivity}$ can still be considered a comparable lower bound estimate for the $POD_{Traditional}$ for this substance.

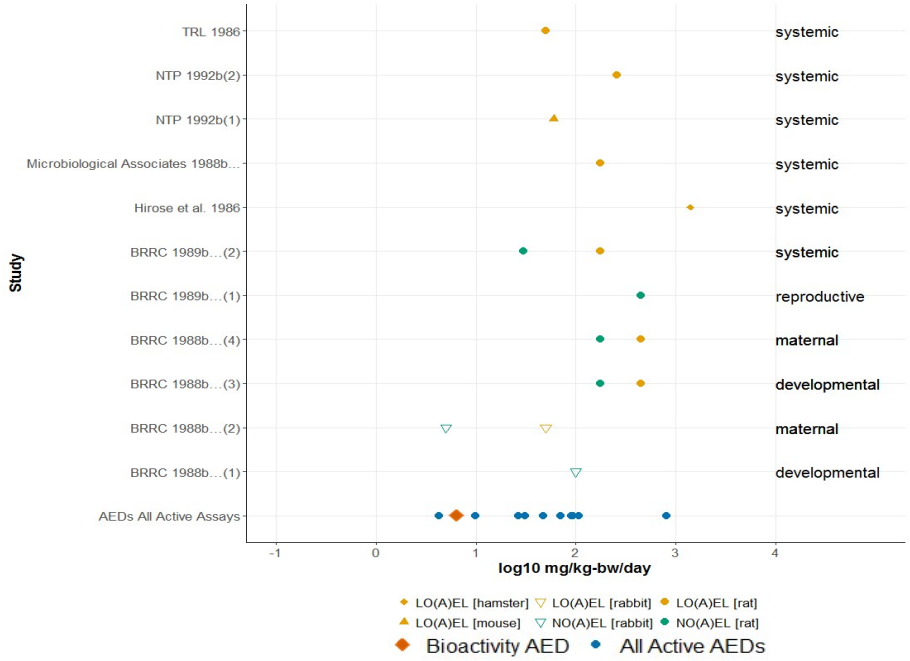

Figure A3-1. Comparison of ToxCast derived $POD_{Bioactivity}$ with $POD_{Traditional}$ for p-cresol. The AC_{50} values across the active assays (post-filtering) were converted to their respective administered equivalent dose (AED) using reverse dosimetry and high-throughput toxicokinetic information in the HTTK package in R (blue points). The $POD_{Bioactivity}$ is represented by the red point which corresponds to the 5th percentile of the AC_{50} values converted to an AED.

Bisphenol A (BPA)

Bioactivity

There are 954 assay endpoints in the ToxCast database for bisphenol A (BPA) and after applying the filtering criteria there are 202 assays that are considered active indicating a broad range of bioactivity. BPA has been tested in all 79 available cytotoxicity related assays and was considered to be active in 15. The lower bound cytotoxicity limit as estimated by the tcpl package using the available "burst assays" for invitrodb version 3.0 is ~13.95 µM. There are 73 active assay endpoints with AC_{50} values below this lower bound estimate. An overview with these assays grouped based on their intended target family along with a total sum of active assays in each subgroup is presented in Table A3-2. **Error! Reference source not found.**Much of the activity seen by BPA is likely confounded by cytotoxicity. Assay endpoints related to nuclear receptors (majority being related to the estrogen receptor pathway) are the family target groups with the most hits. The 5th percentile of distribution of AC_{50} values from active assays was estimated to be 0.351 µM, which is below the cytotoxicity limit. This 5th percentile converts to an AED of 0.01 mg/kg bw/day, which is the basis for the $POD_{Bioactivity}$. The range of AEDs across all active assays is 1.05×10^{-4} to 4.96 mg/kg bw/day.

Table A3-2. Overview of 73 active assays endpoints with AC50 values below the cytotoxicity estimate for bisphenol A grouped based on their intended target family

Intended Target Family	Sub Group	Number of Active Assays
nuclear receptor	steroidal	20
cell morphology	organelle conformation	1
nuclear receptor	non-steroidal	7
cell cycle	cytotoxicity	3
cell adhesion molecules	Immunoglobulin CAM	2
cell adhesion molecules	collagen	1
cytokine	chemotactic factor	1
gpcr	rhodopsin-like receptor	6
cytokine	inflammatory factor	2
cyp	xenobiotic metabolism	17
protease	matrix metalloproteinase	1
transporter	neurotransmitter transporter	2
transporter	vesicular transporter	1
background measurement	baseline control	4
steroid hormone	androgens	2

Intended Target Family	Sub Group	Number of Active Assays
malformation	NA	2
oxidoreductase	peroxidase	1

In vivo animal studies

The final screening assessment for BPA was published in October 2008 under the CMP (ECCC, HC 2008).

The toxicity literature reviewed at the time of the assessment for BPA was extensive and a complete summary of studies examined is not presented here. Rather, the considerations made during risk characterization of this substance are available in the published assessment (ECCC, HC 2008). The results of the studies are summarized in Figure A3-2 to provide context to the comparison of the $POD_{Bioactivity}$ to $POD_{Traditional}$.

For the risk characterization of BPA, the NO(A)ELs of 5 and 50 mg/kg bw/day for adult systemic toxicity and developmental/reproductive effects, respectively, were considered in the derivation of a POD for the MOE (Tyl et al. 2002; 2007). However, as part of the overall weight of evidence, the findings of effects below these thresholds (e.g. developmental neurotoxicity) from low dose studies played a notable role in the risk characterization. At the time of assessment, although it was determined that there was considerable uncertainty related to these low dose studies, they were suggestive of the potential for effects at much lower doses that reflected by the more traditional PODs. The low dose developmental neurotoxicity effects were used to support a toxic conclusion under CEPA using a precautionary approach.

Comparison of $POD_{Bioactivity}$ with $POD_{Traditional}$

The derived $POD_{Bioactivity}$ value of 0.01 mg/kg bw/day (10 µg/kg-bw/day) is within the range of the $POD_{Traditional}$ values from low dose studies considered during the assessment of BPA (2 to 10 µg/kg-bw/day). The $POD_{Bioactivity}$ is ~500 to 5000 fold lower than the NO(A)ELs found in the multi-generational reproductive and developmental toxicity studies used for derivation of an MOE in the assessment (Tyl et al. 2002; 2007).

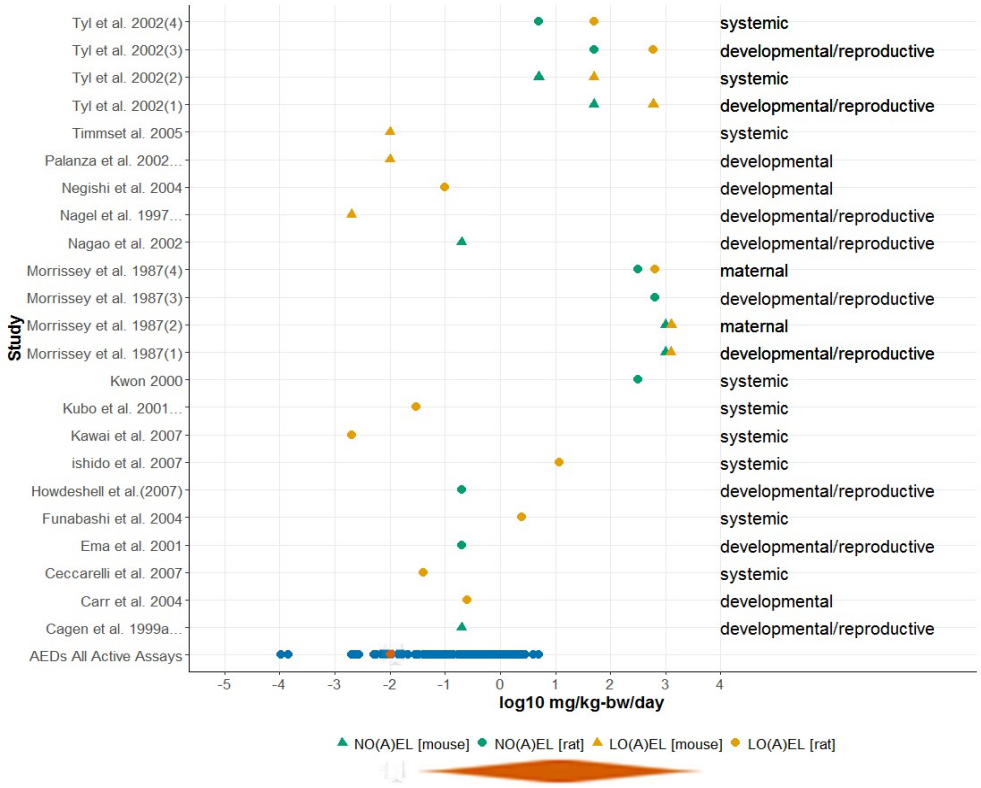

Figure A3-2. **Comparison of ToxCast derived $POD_{Bioactivity}$ with $POD_{Traditional}$ for BPA**. The AC_{50} values across the active assays (post-filtering) were converted to their respective administered equivalent dose (AED) using reverse dosimetry and high-throughput toxicokinetic information in the HTTK package in R (blue points). The $POD_{Bioactivity}$ is represented by the red point which corresponds to the 5th percentile of the AC_{50} values converted to an AED.

Allyl Chloride

Bioactivity

After the assay filtering criteria was applied, allyl chloride was considered active in 6 assay endpoints from a total of 379 tested. Allyl chloride was tested in 62 assays related to cytotoxicity and showed no activity; thus, cytotoxicity is not expected to confound the bioactivity observed. The active assays are listed in Table A3-3 along with their intended family target. Allyl chloride shows some limited activity related to interactions with nuclear receptors such as PXR and ER and other targets that regulate transcriptional activity. Allyl chloride is a volatile, low molecular weight compound, and these types of chemicals may not be well screened in the current ToxCast battery. It is possible that allyl chloride required different assays, conditions, or chemical management to observe bioactivity based on its physico-chemical properties. These properties will be considered as potential criteria for inclusion into the domain of applicability and will serve as flags for closer consideration moving forward. The AC_{50}

values across all active assays were converted to their respective AEDs and ranged from 18.01 to 2412.34 mg/kg bw/day. The AED for the 5th percentile of AC_{50} values from the active assays, which is the basis for the $POD_{Bioactivity}$, was estimated to be 214.36 mg/kg bw/day.

Table A3-3. Overview of 6 active assays endpoints for allyl chloride grouped based on their intended target family

Intended Target Family	Sub Group	Number of Active Assays
nuclear receptor	non-steroidal	1
kinase	receptor tyrosine kinase	1
background measurement	baseline control	2
nuclear receptor	Steroidal	2

In vivo animal studies

The final screening assessment for allyl chloride was published in November 2009 under the “Challenge” phase of the CMP (ECCC, HC, 2009).

Allyl chloride was classified on the basis of carcinogenicity by other national and international agencies (i.e., European Commission and US EPA) but the evidence for genotoxicity and carcinogenicity was considered weak at the time of the Canadian assessment. Thus, the risk characterization for allyl chloride was based on non-cancer effects observed in toxicity studies.

Of the limited oral studies found, the lowest LO(A)EL identified was 45 mg/kg bw/day based on congestion and contained dystrophic changes in unspecified organs (Al’meev and Karmazin 1969). The characterization of effects in the study was considered poor and of limited utility. The lowest oral LO(A)EL in a study for which sufficient information was available to characterize effects was determined to be 73 mg/kg bw/day based on a dose-related decrease in body weight in a chronic gavage study conducted in rats and was carried forward for risk characterization to derive a MOE (National Cancer Institute 1978). At higher doses, hind limb weakness in mice has also been reported (He et al. 1981).

Comparison of $POD_{Bioactivity}$ with $POD_{Traditional}$

The derived $POD_{Bioactivity}$ value of 214.36 mg/kg bw/day is higher than the lowest $POD_{Traditional}$ and the value used for MOE calculation, which were 45 and 73 mg/kg bw/day, respectively (Figure A3-3).

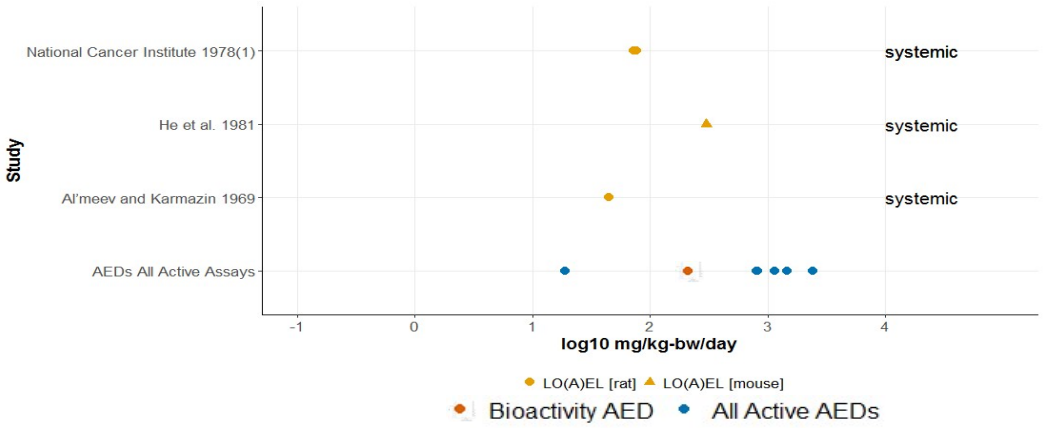

Figure A3-3. Comparison of ToxCast derived $POD_{Bioactivity}$ with $POD_{Traditional}$ for allyl chloride. The AC_{50} values across the active assays (post-filtering) were converted to their respective AED using reverse dosimetry and high-throughput toxicokinetic information in the HTTK package in R (blue points). The $POD_{Bioactivity}$ is represented by the red point which corresponds to the 5th percentile of the AC_{50} values converted to an AED.

Appendix A: Substance list and results for CMP comparative case study

Table A-1: CMP chemicals for comparison of $POD_{Bioactivity}$ to $POD_{Traditional}$

CAS RN	Substance Common Name	CMP Phase	$POD_{Bioactivity}$ (log10 mg/kg bw/day)	Minimum $POD_{Traditional}$ (log10 mg/kg bw/day)	log10 POD ratio	$POD_{Traditional}$ to $POD_{Bioactivity}$ Ratio
102-06-7	1,3-Diphenylguanidine	CMP1	-0.51	0.6	1.11	12.88
103-23-1	Bis(2-ethylhexyl)hexanedioate	CMP1	0.13	1.71	1.58	38.02
106-44-5	p-Cresol	CMP2	0.8	0.7	-0.10	0.79
106-89-8	Epichlorohydrin	CMP1	-2.82	0.3	3.12	1318.26
107-05-1	Allyl chloride	CMP1	2.33	1.65	-0.68	0.21
107-51-7	Octamethyltrisiloxane	CMP1	0.57	2.4	1.83	67.61
108-46-3	Resorcinol	CMP3	-0.13	2.37	2.5	316.23
112-38-9	10-Undecenoic acid	CMP3	-1.16	1.72	2.88	758.58
119-61-9	Benzophenone	CMP3	-1.75	0.49	2.24	173.78
123-91-1	1,4-Dioxane	CMP1	-5.36	0.98	6.34	2187761.62
126-73-8	Tributyl phosphate	CMP1	-2.7	1	3.70	5011.87
127-19-5	N,N-Dimethylacetamide	CMP1	0.09	1.7	1.61	40.74
131-11-3	Dimethyl phthalate	CMP2	-4.03	2.4	6.43	2691534.80
1330-78-5	Tricresyl phosphate	CMP2	-1.78	0.6	2.38	239.88
13674-84-5	Tris(2-chloroisopropyl)phosphate	CMP2	-2.24	1.6	3.84	6918.31
13674-87-8	Tris(1,3-dichloro-2-propyl) phosphate	CMP2	-1.42	0.7	2.12	131.83
149-57-5	2-Ethylhexanoic acid	CMP1	-0.54	0.95	1.49	30.90
1763-23-1	Perfluorooctanesulfonic acid	Pre-CMP	-2.95	-0.3	2.65	446.68
25013-16-5	Butylated hydroxyanisole	CMP1	-4.63	-0.6	4.03	10715.19
27178-16-1	Diisodecyl hexanedioate	CMP3	1.45	3	1.55	35.48
2795-39-3	Potassium perfluorooctanesulfonate	Pre-CMP	-2.5	-1.52	0.98	9.55
28553-12-0	Diisononyl phthalate	CMP2	-0.19	0.3	0.49	3.09
330-54-1	Diuron	CMP1	-0.35	0	0.35	2.24
3380-34-5	Triclosan	CMP2	-2.87	0.48	3.35	2238.72
3825-26-1	Ammonium perfluorooctanoate	CMP2	-2.74	-1.22	1.52	33.11
53-19-0	o,p'-DDD	CMP1	-0.15	0.6	0.75	5.62

534-52-1	2-Methyl-4,6-dinitrophenol	CMP2	-3.66	-0.23	3.43	2691.53
548-62-9	Gentian Violet	CMP3	-2.26	-0.3	1.96	91.20
60-09-3	4-Aminoazobenzene	CMP2	-1.23	1.46	2.69	489.78
69-72-7	Salicylic acid	CMP3	-3.97	1.3	5.27	186208.71
80-05-7	Bisphenol A	CMP1	-1.97	-2.7	-0.73	0.19
87-61-6	1,2,3-Trichlorobenzene	Pre-CMP	-2.85	0.89	3.74	5495.41
88-12-0	N-Vinyl-2-pyrrolidone	CMP1	-2.15	0.56	2.71	512.86
88-72-2	2-Nitrotoluene	CMP1	0.01	1.26	1.25	17.78
88-85-7	Dinoseb	CMP3	-3.78	-0.11	3.67	4677.35
90-04-0	2-Anisidine	CMP2	-1.97	1.2	3.17	1479.11
91-20-3	Naphthalene	CMP1	-4.59	1.7	6.29	1949844.60
91-22-5	Quinoline	CMP2	1.09	1.4	0.31	2.04
92-52-4	Biphenyl	CMP2	-4.23	1.4	5.63	426579.52
93-15-2	Methyleugenol	CMP1	-2.72	1	3.72	5248.07
95-48-7	o-Cresol	CMP2	-0.71	1.48	2.19	154.88
95-94-3	1,2,4,5-Tetrachlorobenzene	Pre-CMP	-1.94	0.32	2.26	181.97
96-23-1	1,3-Dichloro-2-propanol	CMP3	-0.32	0	0.32	2.09
96-29-7	2-Butanone oxime	CMP1	-1.64	0.6	2.24	173.78
97-53-0	Eugenol	CMP3	1.15	2	0.85	7.08
98-01-1	Furfural	CMP1	0.94	1.04	0.10	1.26

Appendix B: Comparison of $POD_{Traditional}$ for developmental and reproductive effects to $POD_{Bioactivity}$

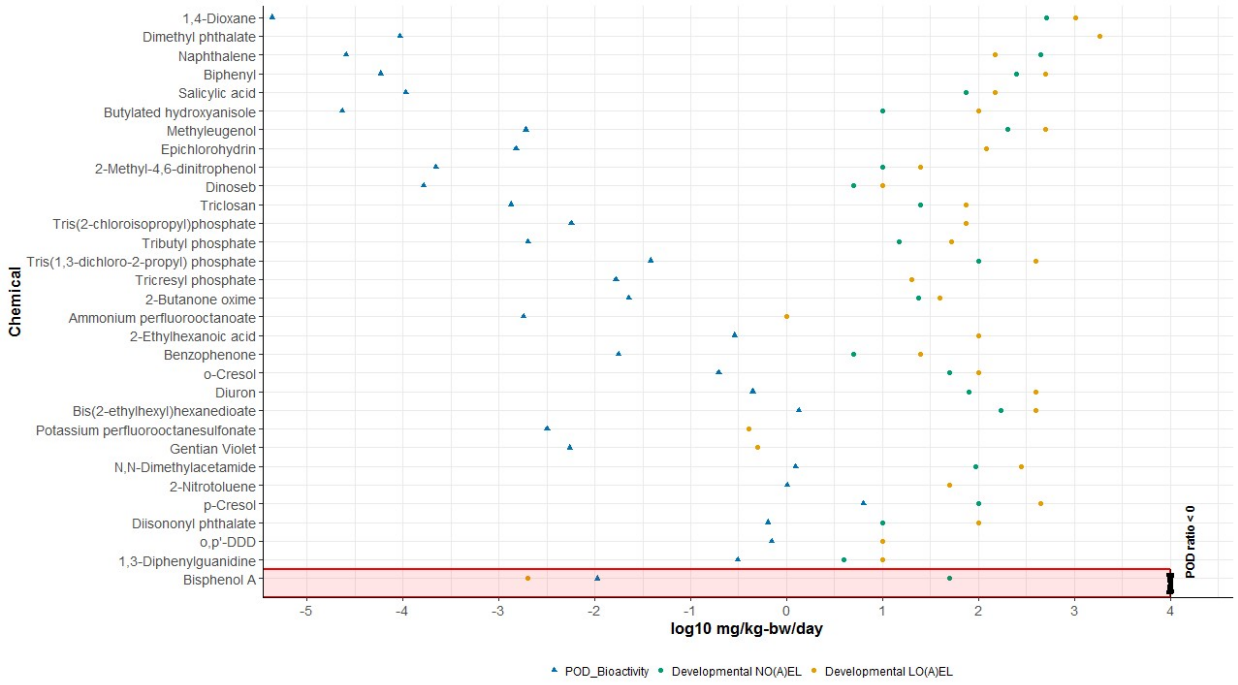

Figure B-1: Comparison of ToxCast derived $POD_{Bioactivity}$ with developmental $POD_{Traditional}$. $POD_{Bioactivity}$ represents the 5th percentile of AC_{50} values from ToxCast assays with an active hit-call converted to an AED using reverse dosimetry and high-throughput toxicokinetic information in the HTTK package in R. The NO(A)EL and LO(A)EL points represent the lowest effect levels across all studies collected in the respective assessment.

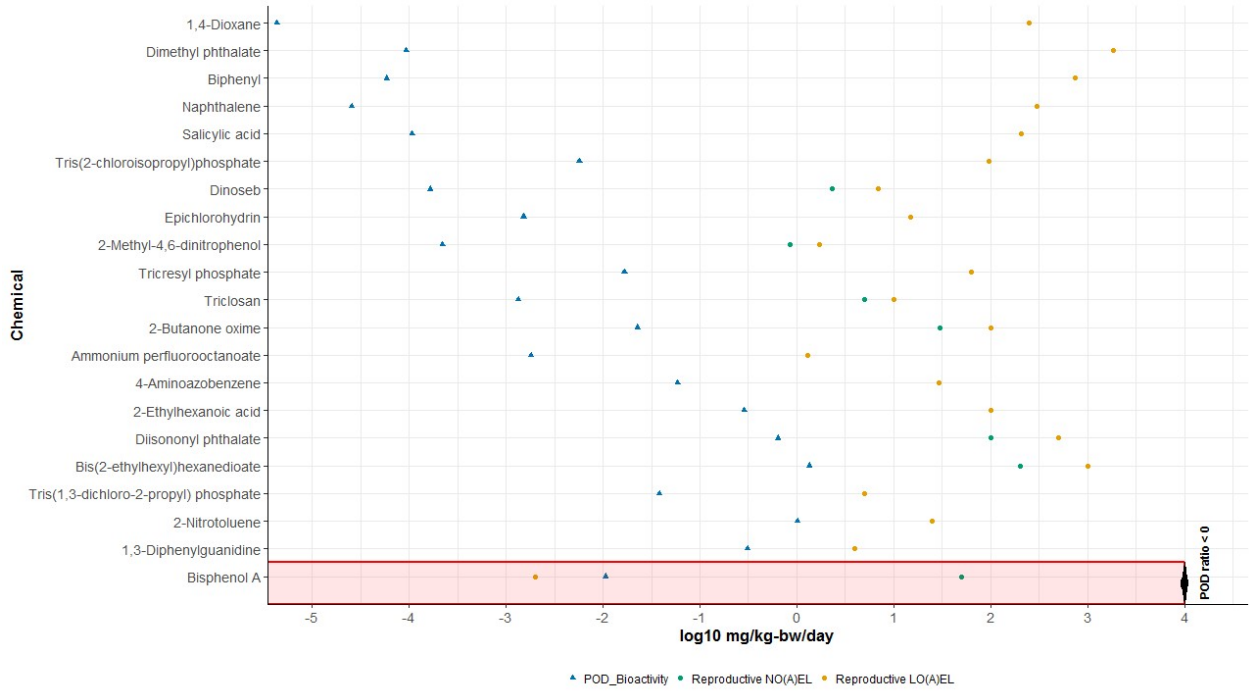

Figure B-2: Comparison of ToxCast derived $POD_{Bioactivity}$ with reproductive $POD_{Traditional}$. $POD_{Bioactivity}$ represents the 5th percentile of AC_{50} values from ToxCast assays with a positive hit call converted to an AED using reverse dosimetry and high-throughput toxicokinetic information in the HTTK package in R. The NO(A)EL and LO(A)EL points represent the lowest effect levels across all studies collected in the respective assessment.