



Science Approach Document

Chemical Screening and Prioritization: Health Canada's Automated Workflow for Prioritization (HAWPr)

Health Canada

December 2025

Cat. No.: En84-392/2025E-PDF
ISBN: 978-0-660-79652-9

Unless otherwise specified, you may not reproduce materials in this publication, in whole or in part, for the purposes of commercial redistribution without prior written permission from Environment and Climate Change Canada's copyright administrator. To obtain permission to reproduce Government of Canada materials for commercial purposes, apply for Crown Copyright Clearance by contacting:

Environment and Climate Change Canada
Public Inquiries Centre
Place Vincent Massey Building
351 Saint-Joseph Boulevard
Gatineau QC K1A 0H3
Telephone: 819-938-3860
Toll Free: 1-800-668-6767 (in Canada only)
Email: enviroinfo@ec.gc.ca

Cover photo: © Environment and Climate Change Canada

© His Majesty the King in Right of Canada, as represented by the Minister of Environment and Climate Change, 2025

Aussi disponible en français

Synopsis

The Health Canada Automated Workflow for Prioritization (HAWPr) has been developed to more efficiently collect, organize and process chemical data to further expand on the methods used for identification of risk assessment priorities under the *Canadian Environmental Protection Act, 1999* (CEPA). HAWPr is a computational tool that integrates inputs from across various information sources and scripting languages, as well as conducts analytics on large data sets. HAWPr was built to automate 4 categories of sequential prioritization tasks:

- 1) chemical data collection;
- 2) data gap filling and predictive modelling;
- 3) evidence evaluation and confidence scoring; and
- 4) hazard and exposure-based prioritization.

This science approach document (SciAD) presents the key elements of HAWPr including:

- the information sources where data is collected;
- predictive modelling approach for data-poor chemicals;
- the human health hazards considered and the criteria for determining if a substance has a hazard indicator;
- substance use patterns and volume considerations used in a weight-of-evidence evaluation to determine exposure potential;
- the integration of hazard and exposure information for risk-based prioritization;
- validation of the workflow by comparing prioritization decisions against known hazardous substances and those with known human exposure;
- proposed outcomes for substances on the Domestic Substances List (DSL); and
- limitations and uncertainties of the tool.

The SciAD demonstrates that HAWPr is a robust tool that will improve how substances are prioritized for assessment work under CEPA to continue to protect the health of people living in Canada. Automation of these tasks helps to improve transparency and thoroughness of information review by simultaneously enabling access to over a million records from curated toxicity and exposure datasets while increasing efficiencies and reproducibility in the overall process. HAWPr was designed to allow flexibility of components within the tool as appropriate to keep pace with evolving science. As a result, future developments are not limited by the confines of any one piece of software, scripting language, or individual expertise and decision flows can be refined and expanded as information presents itself. Implementation of the HAWPr approach described in this document will assist the Government of Canada in identifying substances with a potential human health concern in a more efficient manner. Preliminary results for the use of HAWPr on the substances on the DSL are available as a supporting document to this approach. These results were taken into consideration

along with other sources of information in the subsequent steps for the identification of priorities for risk assessments.

Table of contents

Synopsis.....	i
1 Introduction.....	1
1.1 Background	1
1.2 HAWPr general overview.....	2
2 HAWPr modules	5
2.1 Hazard.....	5
2.2 Exposure	20
2.3 Integrating hazard and exposure for risk-based outcomes	29
3 Validation.....	30
4 HAWPr outcomes for substances on the DSL	32
4.1 Identifying emerging areas of concern to people in Canada under an amended Canadian Environmental Protection Act	33
4.2 Limitations and uncertainties	34
5 Future work.....	37
5.1 Automated collection and relevance ranking of open scientific literature.....	37
5.2 Clustering for data-gap filling and group assessment.....	38
5.3 <i>In silico</i> modeling	38
6 Conclusions	40
7 References	41
8 Appendix.....	44
8.1 List of sources	44
8.2 Validation of HAWPr.....	48
8.3 Detailed results of HAWPr	56

Table of figures

Figure 1-1. HAWPr integrates emerging science tools and datasets, all of which are a piece of the overall IRAP process.	2
Figure 2-1. General stepwise process for collecting and analyzing chemical hazard information.	5
Figure 2-2. Product Count scoring matrix.....	23
Figure 2-3. Tonnage band scoring matrix.....	25
Figure 2-4. Presence-absence scoring matrix.....	26
Figure 2-5. Conceptual prioritization matrix.....	29
Figure 4-1 Prioritization matrix for the DSL	33
Figure 8-1. Multi-class confusion matrix for IRAP equivalent CTP flag versus HAWPr CTP flag	52

Table of tables

Table 2-1. Sources of classification information for carcinogenicity	7
Table 2-2. Assigning confidence levels associated with hazard indicator determination for carcinogenicity.	8
Table 2-3. Models used in the genotoxicity screen within HAWPr	10
Table 2-4. Assigning confidence levels associated with hazard indicator determination for genotoxicity	10
Table 2-5. Effect level thresholds used for hazard indicators in HAWPr	12
Table 2-6. Confidence levels for repeated dose toxicity	12
Table 2-7. Models used in the EDC screen.....	16
Table 2-8. Confidence level criteria for the EDC screen.....	17
Table 2-9. Criteria for overall hazard indicator level of a substance.....	18
Table 2-10. Example designation of overall hazard score for Substance X	19
Table 2-11. Product ruleset score for Substance X.....	23
Table 2-12. Calculation of the total exposure indicator score for Substance X	27
Table 2-13. Thresholds for exposure prioritization	28
Table 4-2. Overall HAWPr results showing the number of DSL substances that have been identified for each hazard flag	34
Table 8-1. Sources of hazard information.	44
Table 8-2. Data sources and assigned relevance for the Product Count ruleset.....	45
Table 8-3. Data sources and assigned relevance for the Tonnage Band ruleset	46
Table 8-4. Data sources and assigned relevance for the Presence-Absence ruleset ...	46
Table 8-5. Precision, recall, and F1-score for each flag, as well as overall and weighted average, assuming the IRAP 2019 equivalent CTP flag is the true-positive case.	53

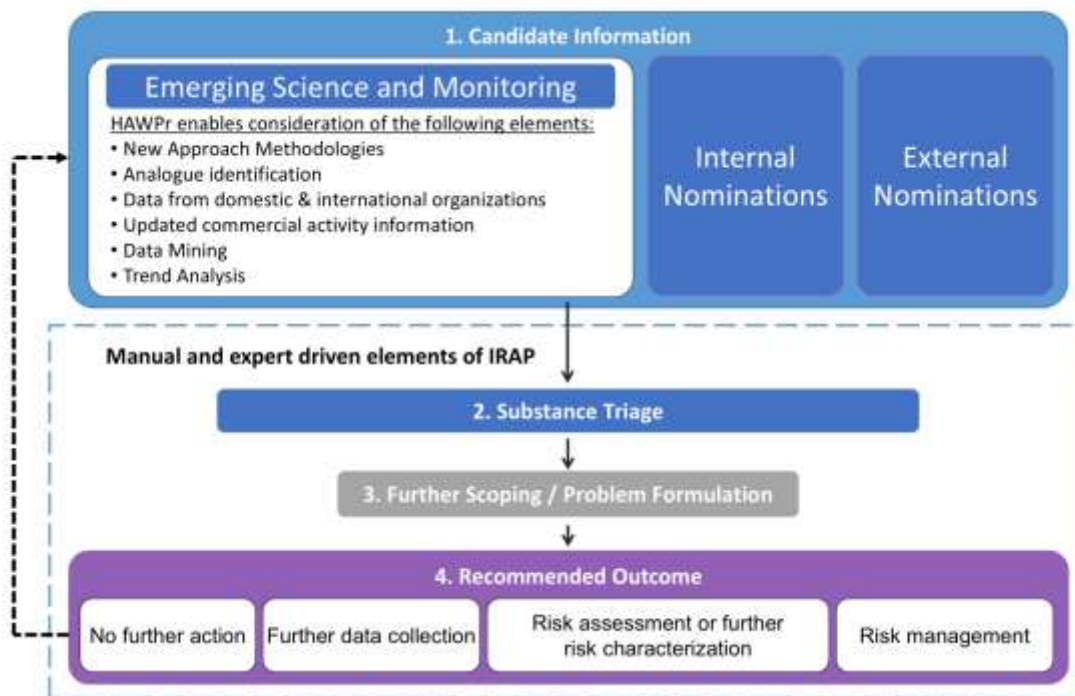
1 Introduction

1.1 Background

The Identification of chemicals and polymers as Risk Assessment Priorities (IRAP) approach was first published in 2014 (Environment Canada, Health Canada 2014) in response to a need to consider new and emerging science to identify substances that may have the potential to cause harm to the environment or human health. This approach is a manual systematic compilation and review of information from a large variety of sources, which allows the Government of Canada to identify and prioritize substances requiring further work on a cyclical basis.

The overall IRAP process is comprised of multiple steps as outlined in [Figure 1-1](#). The first step is to collect chemical information to help inform prioritization outcomes. This includes internal and external nominations (for example, public requests or international decisions), as well as the consideration of emerging science and monitoring data. The Health Canada Automated Workflow for Prioritization (HAWPr) is a decision support tool which focuses on integrating advances in the emerging science and monitoring streams of chemical information with traditionally available data to evolve the process for identification of priorities for further work. This offers a significant opportunity to modernize and expand on the data considerations for the identification of substances with the greatest potential for human concern, including improvements to previous methods for how data and information is gathered, how it is reviewed and considered, as well as how it is weighted and compared. Moreover, this progressive tool brings together the various pieces of data to create a coherent, robust, transparent, and reproducible approach for identifying candidates for prioritization that can also be coupled with internal and public nominations. As illustrated in [Figure 1-1](#), HAWPr is designed to inform prioritization activities and considers and integrates a large amount of both hazard and exposure information. The role of HAWPr is captured within the Emerging Science and Monitoring step of the overall IRAP approach. The results of HAWPr are then considered, along with other streams of information. The overall recommended outcomes from IRAP are made following a manual review and consideration of a breadth of inputs (for example, ongoing or previous outcomes of risk assessment activity by the program), one of which includes the HAWPr tool.

Figure 1-1. HAWPr integrates emerging science tools and datasets, all of which are a piece of the overall IRAP process.



[Long description: This figure depicts the overall IRAP process. Starting at the top is step 1, where sources of candidate information are collected to help inform the prioritization approach. While both internal and external nominations are sources of candidate information, the focus in this graph is on the emerging science and monitoring information sources, which represent the place where HAWPr fits into this overall process. The emerging science and monitoring information includes elements such as new approach methodologies, analogue identification, data from domestic and international organizations, updated commercial activity information, data mining, as well as trend analysis. After collection of data and information in the first step, the process moves to a substance triage step and from there to a third step of further scoping or problem formulation. Finally, the last step is where a decision on a recommended outcome of the prioritization process is made. Recommended outcomes can include further data collection, risk assessment or further risk characterization, risk management, as well as no further action. There is also a process to have substances return to the first step, for example, if further data collection is required.]

This document was prepared by the staff of the *Canadian Environmental Protection Act, 1999* (CEPA) Risk Assessment Program at Health Canada and has undergone external written peer review. Comments on the technical portions relevant to human health were received from Theresa Lopez, Jennifer Flippin and Joan Garey (TetraTech Inc.). Comments received were taken into consideration, noting that the final content and outcome of the report remains the responsibility of Health Canada.

1.2 HAWPr general overview

There were 2 key considerations for developing HAWPr. First, it was recognized that collecting, organizing, and processing chemical data in previous cycles of IRAP to identify health priorities was a manual and labour-intensive process. Second, during the

[2016 Chemicals Management Plan \(CMP\) Science Committee](#) meeting on incorporating New Approach Methodologies (NAMs) into multiple aspects of chemical risk assessment, Health Canada sought input on modernizing future cycles of prioritization to include: (1) computational approaches to mitigate the labour-intensive nature of the activity and explore the incorporation of an approach for ranking hazard, exposure and risk; and (2) the integration of NAMs to better address chemicals that lack traditional data sources and to harness emerging science, technologies and non-animal toxicity testing methods. Further, [incorporating NAMs](#) supports the efforts to promote the replacement, reduction or refinement of vertebrate animal testing while ensuring that non-animal approaches provide equivalent or better protection to the health of people in Canada. Since the 2016 meeting, Health Canada has undertaken a review of the existing IRAP approach to identify areas of improvement related to human health hazards and exposures. To this end, Health Canada has developed HAWPr to aid in the implementation of this vision.

At its foundation, HAWPr is an integrated computational tool built using the [KNIME analytics platform](#). KNIME is a versatile software which can integrate inputs across various sources and scripting languages (such as [Python](#) and [R](#)), as well as conduct analytics on large data sets. HAWPr was built to automate 4 categories of sequential tasks:

- 1) chemical data collection;
- 2) data gap filling and predictive modelling;
- 3) evidence evaluation and confidence scoring; and
- 4) hazard and exposure-based prioritization.

Together, automation of these tasks helps to improve transparency and thoroughness of information review by simultaneously enabling access to over a million records from curated toxicity and exposure datasets while increasing efficiencies and reproducibility in the overall process. HAWPr was designed to allow substitution, addition, or removal of components within the tool as appropriate to keep pace with evolving science. As a result, future developments are not limited by the confines of any piece of software, scripting language, or individual expertise and decision flows can be refined and expanded as information becomes available.

1.2.1 Chemical data collection

HAWPr collects substance specific toxicity and exposure data across numerous sources (see the list of sources in section 8.1), enabling users to consider the type and quantity of data that is available for each substance on the Domestic Substances List (DSL) simultaneously. This is accomplished through various data retrieval methods such as interacting with the [Organisation for Economic Co-Operation and Development \(OECD\) Quantitative Structure-Activity Relationship \(QSAR\) Toolbox](#) and the [European Chemicals Agency's \(ECHA's\) International Uniform Chemical Information Database \(IUCLID\)](#) for Registration, Evaluation, Authorisation and Restriction of Chemicals

(REACH) dossiers via an application programming interface (API); querying Health Canada structured database (via PostgreSQL6); and custom scripting for web data extraction. HAWPr also incorporates alternative data such as the output of NAMs from the United States (US) Environmental Protection Agency (EPA) [ToxCast Program](#).

1.2.2 Data gap filling and predictive modelling for hazard

Following data collection, HAWPr then identifies data gaps that are apparent for each substance on the DSL and leverages various novel consensus modelling techniques and a rapid read-across approach to address the data gaps for hazard identification. The aim of this task is to ensure that substances are not only prioritized based on the availability of traditional toxicity data which is often limited for the diverse chemical space of the DSL, but that information provided by non-animal methods such as predictive models and read-across approaches are used to capture data-poor chemicals that merit further consideration.

1.2.3 Evidence evaluation and confidence levels of hazard and exposure indicators

After data collection and data gap filling techniques have been applied, HAWPr makes use of consistent rule-based criteria to identify both hazard and exposure indicators on a substance-by-substance basis, and reconciles any differences between *in silico*, *in vitro* and *in vivo* hazard data. Decision rules are also used to determine the relative confidence level for the identified data; this concept is discussed in more detail later in the document for each respective indicator (see section 2.1.1). Application of these criteria allows the identification of substances where there is the highest confidence in potential hazard and exposure indicators, as well as more comprehensive understanding of emerging concerns based on alternative methods employed in the approach. The ability to refine and revise the rules used to identify specific indicators provides improved flexibility for targeting the identification of substances relevant to specific issues, such as vulnerable populations, and chemical classes of concern (for example, endocrine disrupting chemicals), which must be considered under recent amendments to CEPA (see section 4 for more details).

1.2.4 Hazard and exposure-based prioritization

Using the data identified above and the associated confidence levels, HAWPr applies criteria to triage substances for both hazard and exposure potential and place substances into overall priority levels. As discussed previously and illustrated in Figure 1-1, the results from HAWPr are then considered, along with additional streams of information, as well as manual substance triaging and further scoping steps, within the broader IRAP framework to determine what, if any, further actions may be required to address potential concerns with a substance(s).

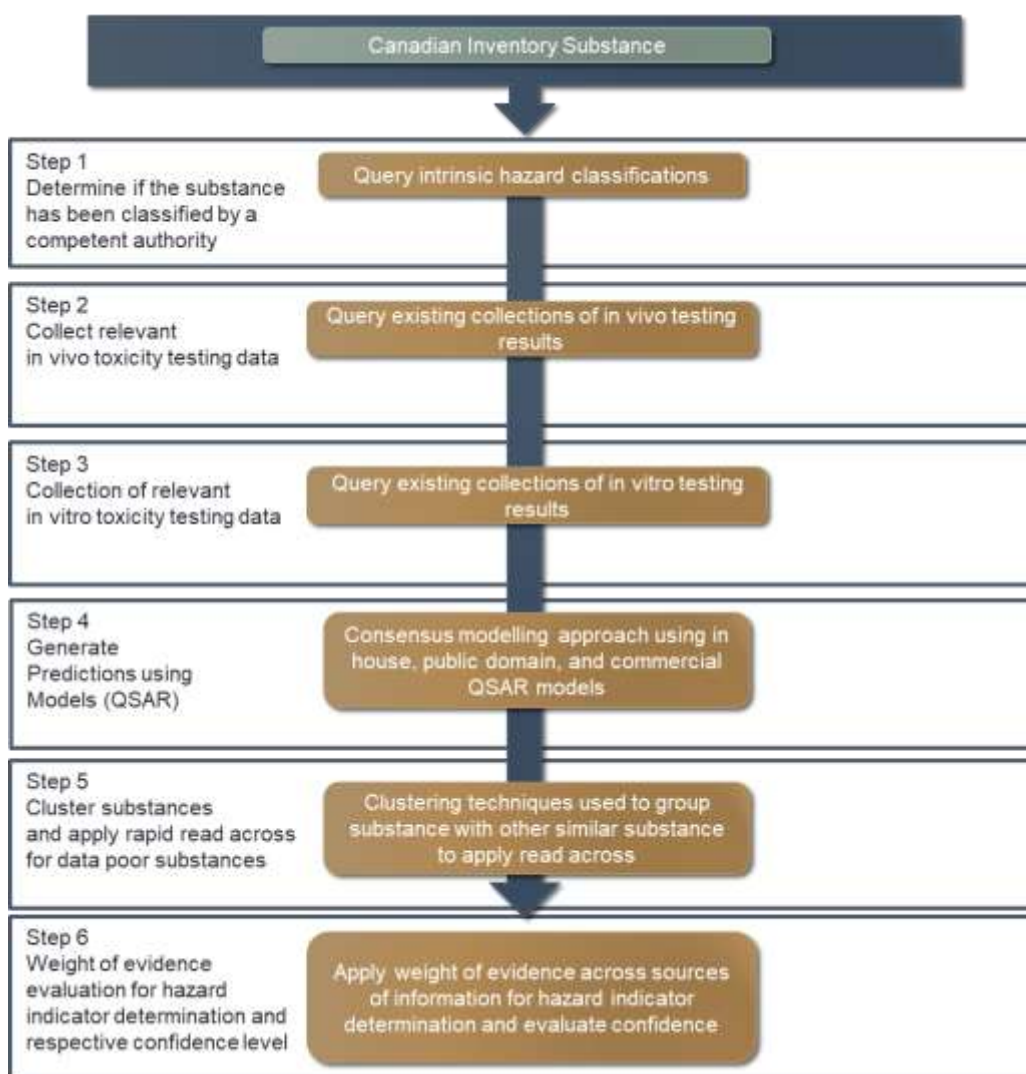
Greater details on the above-described HAWPr tasks are outlined below in the respective hazard and exposure module descriptions (section 2).

2 HAWPr modules

2.1 Hazard

The general process steps for collecting and analyzing hazard data is presented in Figure 2-1. The hazard module currently addresses critical risk assessment toxicity endpoints including carcinogenicity, genotoxicity, developmental and reproductive toxicity, repeated dose toxicity, as well as certain endocrine pathways. It is intended that the coverage of toxicological effects will continue to be expanded as NAMs and data become available. Specific details for each of the human health endpoints are provided below.

Figure 2-1. General stepwise process for collecting and analyzing chemical hazard information.



[Long description: This figure illustrates a step-wise process for collecting and analyzing hazard data in HAWPr. The first 3 steps reflect the collection of empirical hazard data from various sources, starting from intrinsic hazard classification, through *in vivo* testing results, and then *in vitro* testing results. Steps 4 and

5 represent the collection, generation and application of predicted results from models, for example, from QSARs, as well as approaches to cluster, group and apply read-across to inform the hazard for data-poor substances. The final step is the application of weight of evidence across sources of information for hazard indicator determination and an evaluation of confidence.]

The first step for each toxicity endpoint is to identify if the substance has been classified by a competent authority based on its intrinsic hazard properties. Competent authorities are identified as other domestic or international regulatory agencies that have classified a substance after a robust scientific review and are outlined with the endpoint specific considerations in section 2.1.1. The second and third step of the process involves querying various databases for *in vivo* and *in vitro* hazard data. To gain efficiencies in screening chemicals for available data, multiple data sources from numerous international regulators (for example, US EPA ToxVal) and other collections (for example, Carcinogenic Potency Database) have been compiled into one common database maintained and updated yearly by Health Canada. Moreover, other datasets that are maintained externally are also queried where possible via APIs (for example, ECHA's IUCLID for REACH) or by directly interacting with a downloadable database (for example, ToxCast) (detailed in Table 8-1).

To facilitate the prioritization of data-poor substances, predictive models are used to address data gaps. There are a variety of models available which can be used in regulatory toxicology (Madden et al. 2020). For certain endpoints, multiple models are available to predict toxicity; these may include structural alert-based or expert systems, QSARs, as well as machine-learning techniques. For example, there are QSAR models developed by MultiCase and Advanced Chemistry Development (ACD) that predict estrogen receptor binding activity, while other models in the OECD QSAR Toolbox might predict developmental and reproductive toxicity, acute toxicity, or mutagenicity. While each model has its own biases and errors, using an informed combination of the models on an endpoint-by-endpoint basis aims to minimize those effects. Additionally, combining models allows for better coverage of the broad range of chemicals on the DSL. During Step 4, the results of multiple *in silico* models that predict the same toxicological endpoint are combined into a single consensus prediction. Detailed methodology and discussion on the development of optimal consensus predictions are described in Collins et al. (2024). Where QSAR models are not available, [read-across](#) is used to address data gaps for prioritization in Step 5. A common method in regulatory assessment to fill gaps for toxicological effects data, read-across makes use of the hazard data from a structural analogue when no substance-specific data is available (OECD 2017). Read-across typically requires manual expert judgement, though recently there has been work to systematize and automate the approach (Low et al. 2013; Shah et al. 2016; Helman 2019). In HAWPr, rapid read-across (RRA) is applied based on structural similarity using [PubChem fingerprints](#) and Tanimoto scoring (Bajusz et al. 2015). Selection of the read-across analogue is based on the closest structural neighbour on the DSL that has toxicological data available to inform a risk assessment. Manual consideration of the appropriateness of the analogue will be done in later steps of IRAP if the substance is selected for further work.

Finally, in Step 6, the information gathered from Step 1 through Step 5 is processed to determine if the substance has a given hazard indicator for a particular endpoint. The process for the evaluation of data is endpoint specific and described in section 2.1.1. In general, existing classifications from domestic or international regulatory agencies carry the greatest weight when determining if a substance has a hazard indicator. When a substance does not have an existing classification, *in vivo* data is considered before *in vitro* data. Where *in vivo* and *in vitro* data are both available, but the results are conflicting, *in vivo* data takes precedence for determining the presence of a hazard indicator. In the absence of adequate substance-specific *in vivo* or *in vitro* data, read-across and QSAR are used to identify potential hazard indicators. Confidence is assigned based on the relevancy and strength of the underlying data or prediction. In general, there is higher confidence in the determination of a hazard indicator if the information underpinning the decision comes from existing classifications or *in vivo* data. Confidence in the hazard indicator is moderate when only *in vitro* data are available, whereas confidence is considered low when the hazard indicator determination is based only on *in silico* predictions or the RRA approach.

2.1.1 Endpoint specific considerations for hazard indicators

2.1.1.1 Carcinogenicity

The first line of evidence used as a hazard indicator for carcinogenicity is examining if a substance has previously been classified by a competent authority. If a substance has been classified as outlined in Table 2-1, this serves as a high confidence hazard indicator which prioritizes a substance for further consideration.

Table 2-1. Sources of classification information for carcinogenicity

Source	Jurisdiction	Indicator Classifications
International Agency for Research on Cancer (IARC)	International	Group 1: carcinogenic to humans Group 2A: probably carcinogenic to humans Group 2B: possible carcinogenic to humans
ECHA Harmonized Classifications	Europe	Carc. 1A: known to cause cancer to humans Carc. 1B: presumed to cause cancer to humans Carc. 2: suspected of causing cancer to humans
US EPA Classifications	US	Group A: human carcinogens Group B1 and B2: probable human carcinogens Group C: possible human carcinogen OR Carcinogenic to humans Likely to be carcinogenic to humans Suggestive evidence for carcinogenicity in humans
National Toxicology Program Report on Carcinogens	US	Known to be a human carcinogen Reasonably anticipated to be a human carcinogen

Source	Jurisdiction	Indicator Classifications
National Institute for Occupational Safety and Health Occupational Carcinogens	US	Potential occupational carcinogen

If a substance has not been classified as outlined above, data collected from sources listed in Table 8-1 are evaluated. Substances with “positive” summary data, available carcinogenic potency information (for example, Tumourigenic Dose (TD) or Benchmark Dose (BMD)), and/or cancer slope factors / unit risk estimates are considered as hazard indicators for carcinogenicity.

Carcinogens act through complex biological processes and involve multiple different toxicity mechanisms. As a result, carcinogenicity is a challenging hazard to model using *in silico* tools as “carcinogenicity” is not a singularly defined endpoint. At present, *in silico*-based methods for predicting carcinogenicity are being evaluated at Health Canada but are not currently part of the automated prioritization process outlined here. Typically, *in silico* approaches intended to inform carcinogenicity assessment examine specific genotoxic mechanisms as the 2 hazards are related; the application of the genotoxicity mechanistic models is captured further below.

Assignment of confidence level for hazard indicators pertaining to carcinogenicity based on source of information and rationale are outlined in Table 2-2.

Table 2-2. Assigning confidence levels associated with hazard indicator determination for carcinogenicity.

Relative level of confidence in hazard indicator	Lowest	Moderate	Highest
Information source underpinning hazard indicator	RRA (only structural features considered)	Summary evidence from a cancer study <i>and/or</i> Availability of BMD, TD, or other metric for carcinogenicity	International classification <i>and/or</i> Cancer slope factor or similar from competent authority
Rationale	RRA based solely on structural similarity is used to fill the data gap for	Effect levels were extracted from summary databases. There has been no	The substance has been classified as a carcinogen after an extensive review by

Relative level of confidence in hazard indicator	Lowest	Moderate	Highest
	the purposes of prioritization. The RRA approach does not include a consideration of similarity across physical/chemical properties, metabolism, mechanism of action or other elements used in a comprehensive risk assessment justification. Therefore, there is higher uncertainty associated with the hazard indicator determination.	expert review of the studies used by HAWPr and, as such, there is less confidence associated with the indicator.	qualified experts (for example, IARC classifications) or a cancer risk estimate has been established by a competent authority indicating that an assessment was conducted (for example, US EPA's Integrated Risk Information System (IRIS) program).

2.1.1.2 Genotoxicity

Existing classifications for genotoxicity are first examined when determining if a substance has a hazard indicator related to genotoxicity. At present, only 1 source of classification information is available and considered reliable for genotoxicity. This is the ECHA Harmonized Classifications for mutagenicity (Muta 1A/1B and Muta 2) which is based on the [Globally Harmonized System of Classification and Labelling of Chemicals \(GHS\)](#).

Where classification information is not available, data collected from sources outlined in Table 8-1 are then considered along with the predictions from QSAR models.

For the purposes of the workflow, a substance is considered potentially genotoxic if the data show:

- a) positive results in *in vivo* mammalian studies for gene mutation or clastogenicity/aneugenicity; and/or

- b) positive results in an *in vitro* test for gene mutation or clastogenicity/aneugenicity and absence of conflicting data from *in vivo* mammalian tests that cover the same genetic endpoint that gave a positive response in the *in vitro* test.

Currently, when no *in vitro* or *in vivo* data are available, Health Canada is using a series of models that predict outcomes for various OECD guideline assays that can detect the effects outlined in points (a) and (b) above. To integrate available data with *in silico* predictions, Health Canada has automated (with modification) an *in silico* protocol published by an international consortium of experts from over 50 organizations including those from academia and government (Hasselgren et al. 2019). For the *in silico* predictions for genotoxicity, a combination of commercial, free, and in-house developed models are used (Table 2-3). Consensus models were developed for bacterial mutation, *in vitro* chromosome aberration, and *in vivo* micronucleus that resolve predictions from multiple models into 1 value. The approach to building the consensus models is presented in Collins et al. (2024). Confidence scoring of genotoxicity flags is outlined in Table 2-4.

Table 2-3. Models used in the genotoxicity screen within HAWPr

Model	Assay	Availability	Reference
Advanced Chemistry Development (ACD) Percepta	Bacterial mutation	Licence required	ACD/ Percepta 2018
VEGA – QSAR models	<i>In vivo</i> micronucleus	Free	Benfenati et al. 2013
SimulationPlus – ADMET Predictor	<i>In vitro</i> chromosome aberration	Licence required	SimulationsPlus 2022
OASIS TIMES	Bacterial mutation	Licence required	TIMES 2016
LeadScope Model Applier – Expert Alert System	Bacterial mutation	Licence required	Myatt et al. 2022
LeadScope Model Applier – QSAR prediction	Bacterial mutation, <i>in vitro</i> chromosome aberration, <i>in vivo</i> micronucleus	Licence required	Landry et al. 2019

Table 2-4. Assigning confidence levels associated with hazard indicator determination for genotoxicity

Relative level of confidence in hazard indicator	Lowest	Moderate	Highest
Information source underpinning prioritization decision	QSAR prediction showing <i>in vivo</i> or <i>in vitro</i> evidence of mutation, chromosome aberration and/or micronuclei.	Studies showing <i>in vivo</i> or <i>in vitro</i> evidence of mutation, chromosome aberration and/or micronuclei (study quality unknown).	International classification; <i>and/or</i> Guideline studies showing <i>in vivo</i> or <i>in vitro</i> * evidence of mutation, chromosome aberration and/or micronuclei.
Rationale	QSAR models have been validated and are considered adequate for prioritization. However, since the decision is not based on empirical data, there is less relative confidence than using <i>in vitro</i> or <i>in vivo</i> data.	Results extracted from databases where study details could not be determined, as such, there is less confidence associated with the indicator.	The substance has been classified as a mutagen after an extensive review by qualified experts (for example, ECHA Harmonized Classification, Labelling, and Packaging Regulations (CLP)) or results were extracted from databases where study details could be determined (for example, OECD guideline); as such, more confidence is associated with the indicator.

* absence of conflicting data from *in vivo* mammalian tests that covers the same genetic endpoint

2.1.1.3 Repeated dose toxicity

Where available, ECHA Harmonized Classifications for specific target organ toxicity for repeated exposure (Specific Target Organ Toxicity RE 1, 2), are considered as affirmative hazard indicators for repeated dose toxicity. Where classification information is not available, available data collected from sources outlined in Table 8-1 are then considered. A substance is considered to have a hazard indicator for repeated dose toxicity if a point of departure (POD) based on a No-Observed-Effect-Level (NOAEL) and/or Lowest-Observed-Effect-Level (LOAEL) from a repeated dose toxicity study falls below the thresholds as outlined in Table 2-5. The thresholds were developed by examining GHS criteria for specific target organ toxicity (STOT) categories. The STOT

category 2 threshold for observable adverse effects in a toxicity study for the oral route is 100 mg/kg bw/day. Generally, STOT classification is based on a 90-day study conducted in rodents. Shorter duration studies can be used where the threshold is modified by a factor of 3 (to 300 mg/kg bw/day) to account for the additional uncertainty with a study of shorter duration. There are also GHS STOT thresholds for dermal and inhalation studies. When using summary level information from databases of studies, it is often not possible to determine the target organ, the severity of the effect or the exact duration of the study. The thresholds outlined in Table 2-5 were selected to be conservative while aligning with GHS criteria to the extent possible for use in a high-throughput prioritization process within HAWPr. An identified effect level of less than the specified threshold serves as an indicator that requires further follow up under IRAP. Confidence scoring criteria for repeated dose tests are available in Table 2-6.

Table 2-5. Effect level thresholds used for hazard indicators in HAWPr

Route	Hazard indicator thresholds (NOAEL or LOAEL)
Oral	≤ 300 mg/kg bw/day
Dermal	≤ 600 mg/kg bw/day
Inhalation (vapour)	≤ 3 mg/L/6h/day
Inhalation (gas)	≤ 750 ppmV/6h/day
Inhalation (dust/mist/fume)	≤ 0.6mg/L/6h/day

Table 2-6. Confidence levels for repeated dose toxicity

Relative level of confidence in hazard indicator	Lowest	Moderate	Highest
Information source underpinning prioritization decision	RRA of POD value	Study with a POD below the specified hazard indicator threshold (guideline quality unknown).	International classification; <i>and/or</i> Guideline study with a POD below the specified hazard indicator threshold.

Relative level of confidence in hazard indicator	Lowest	Moderate	Highest
Rationale	RRA is based solely on structural similarity without a full justification leading to lower confidence.	Results extracted from databases where study details could not be determined; as such, there is less confidence associated with the indicator.	The substance has been classified after an extensive review by qualified experts (for example, ECHA Harmonized CLP). Results extracted from databases where study details could be determined (for example, OECD guideline); as such, more confidence is associated with the indicator.

2.1.1.4 Reproductive and developmental toxicity

Classification information is first considered when determining if a substance has a hazard indicator for reproductive or developmental toxicity. The ECHA Harmonized Classifications for reproductive and developmental toxicity (Repr 1A/1B and Repr 2), where available, are considered as affirmative hazard indicators.

Where classification information is not available, available data collected from sources outlined in Table 8-1 are then considered. A substance is considered to have a potential indicator for reproductive or developmental toxicity if a POD NOAEL and/or LOAEL from a study examining reproductive and/or developmental toxicity falls below the thresholds outlined in Table 2-5. The selected thresholds align with those described for repeated dose toxicity. Since the hazard screen makes use of summary level data extracted from the various information sources, it is not possible to evaluate effect or severity to establish if the substance of interest is conclusively a reproductive or developmental toxicant. This determination requires examining the identified adverse effects from the corresponding study which, at present, cannot be done in an automated fashion. Thus, an identified effect level below the specified thresholds serves as an indicator that requires further follow-up, but itself should not be taken as a confirmatory indication that the substance is a reproductive or developmental toxicant. Confidence scoring of reproductive and developmental flags is consistent with that for repeated dose toxicity and outlined in Table 2-6.

2.1.1.5 Endocrine disrupting chemicals screen

As defined by the World Health Organization (WHO) and adopted by the OECD, “an endocrine disruptor is an exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes adverse health effects in an intact organism, or its progeny, or (sub)populations” while “a potential endocrine disruptor is an exogenous substance or mixture that possesses properties that might be expected to lead to endocrine disruption in an intact organism, or its progeny, or (sub)populations” (OECD 2018).

The HAWPr endocrine disrupting chemicals (EDC) screen examines available evidence across *in vivo*, *in vitro*, and *in silico*-based mechanistic information. At present, the process is focused on a subset of possible endocrine-related pathways, namely chemicals that may interact with the estrogen and androgen systems. Chemicals may disrupt other endocrine mediated pathways (for example, thyroid and steroidogenesis); however, datasets for screening assays and *in silico* models beyond the Estrogen Receptor (ER) and Androgen Receptor (AR) system are limited and have not been evaluated for performance and confidence across the chemical space of interest -- this is an area of ongoing research and development. As scientific confidence and program experience evolves, other endocrine based pathways will be integrated in the workflow.

For certain endocrine pathways, [existing test methods](#) can detect substances that may be potential EDCs, and some are also sufficient for the detection and characterization of apical adverse effects and dose-response assessments necessary for risk characterization. However, individual Test Guidelines (TGs) have limitations and often identification of an EDC or potential EDC may require a complement of these available tests (or supplementation by appropriate NAMs) to screen out, or conversely, show a potential endocrine mode of action and related apical adverse effects. During a risk assessment, a detailed weight-of-evidence evaluation across multiple studies and multiple levels of the conceptual framework is often required to conclude that a substance is an endocrine disruptor (OECD 2018).

For prioritizing a substance as a potential EDC, *in vivo* mechanistic data is considered first, where available. The Uterotrophic and Hershberger bioassays are mechanistic assays that examine for the ER and AR pathways, respectively. The primary source of this information for chemicals is derived from reference datasets compiled by the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). The NICEATM Guideline-Like Uterotrophic Database contains 458 studies on 118 chemicals demonstrating the potential for *in vivo* estrogenic bioactivity (Kleinstreuer et al. 2016). The Hershberger dataset is smaller with 49 reference chemicals for AR pathway responses (Browne et al. 2018). If a chemical shows activity in either the Uterotrophic or Hershberger assays, it is used as an affirmative hazard indicator.

HAWPr also collects *in vitro* assay data pertaining to the ER and AR pathways including data and models developed for the US Endocrine Disrupting Screening Program (EDSP). The EDSP has developed high throughput screening assays for some important ED-related molecular initiating events including ER activation and AR activation, with the intention of providing alternatives to the existing methods used in the tiered testing strategy to accelerate the screening process (US EPA 2017a). The ToxCast-based ER Bioactivity model (Area Under the Curve (AUC) model) integrates the dose-response curves across 18 ToxCast ER *in vitro* assays into a computational model that can discriminate bioactivity from assay-specific interference and responses related to cytotoxicity (Browne et al. 2015). The model provides a score that is relative to the activity seen for the endogenous hormone, 17 β -estradiol. The model has been evaluated against a set of reference chemicals for which guideline-type studies were available for the *in vivo* uterotrophic assay, and it was shown to be highly predictive (Browne et al. 2015). [Under the EDSP](#), the US EPA will allow a test order recipient to satisfy Tier 1 requirements (ER binding, ER transactivation, and uterotrophic assays) by citing existing data for the ToxCast ER Model or generate new data relying on the 18 ER high-throughput assays for bioactivity. A similar model for the androgen pathway is also available using 11 ToxCast-based assays (Kleinstreuer et al. 2017). Chemicals are considered to have ER or AR activity when their respective AUC scores in the ER or AR ToxCast models are greater than or equal to 0.1, which are considered affirmative hazard indicators in HAWPr. Since the AUC score represents the response across multiple assays covering different portions of the adverse outcome pathway (receptor binding, transcriptional activation, and cell proliferation), it is considered more reliable than any single assay response alone (for example, receptor binding). The results of the AUC models take precedence over responses found in single assays from other sources of information (for example, the Collective Estrogen Receptor Activity Prediction Project (CERAPP) literature data). There is a significant portion of chemicals in the ToxCast database that do not have full testing results for all 18 ER and 11 AR assays. As a result of this partial testing, these chemicals do not have a computed AUC score available. In these situations, substances with partial ToxCast/Tox21 testing were considered ER or AR “active” in ToxCast where the majority of available ER or AR assays had a positive hit call. If the majority of available ER assays were not considered active, the substance was considered “inactive”. Care was taken to consider agonist and antagonist responses separately using this approach.

Beyond ToxCast, *in vitro* data for ER-related activity is also used from CERAPP, an international collaborative effort to predict estrogenic potential for tens of thousands of chemicals found in the environment (Mansouri et al. 2016). As part of the project, validation sets were developed containing *in vitro* data on numerous substances. These data included an activity outcome (active/inactive) as well as a potency value for active compounds. Experimental data were collected from the US Food and Drug Administration (FDA) Estrogenic Activity Database, Japan’s Ministry of Economy, Trade and Industry, Japan database and ChEMBL database. The collected data is available

for download on over 7000 substances on the US EPA ToxCast website. All data entries were categorized into 3 assay classes:

- a) binding,
- b) reporter gene/transactivation, or
- c) cell proliferation and potency levels for binding, where agonist and antagonist activity were assigned based on the results (Mansouri et al. 2016).

ER binding, agonist, or antagonist activity observed in the CERAPP training set are used as hazard indicators in HAWPr.

In the absence of *in vivo* or *in vitro* data, HAWPr makes use of predictive models for the ER and AR pathways. A combination of commercial, free, and in-house developed models is used to predict ER and AR related activities. The consensus models developed under the CERAPP and the Collaborative Modeling Project for Androgen Receptor Activity (CoMPARA) (Mansouri et al. 2020) projects are used for ER and AR activity predictions, respectively. In addition, the commercial/free models outlined in Table 2-7. Models used in the EDC screen are also used. Finally, in-house models were developed to predict ER and AR activity using a machine learning approach, specifically Random Forest (RF) algorithms (Collins and Barton-Maclaren 2022). To determine an overall prediction in HAWPr for AR and ER binding, agonism and antagonism are considered separately and are predicted using a consensus model approach. Each consensus model was constructed to maximize the predictive power of the *in silico* predictions as well as have predictions for as many substances as possible. Explicit definition of what models were used in each consensus model is beyond the scope of this work (details in Collins et al. 2024); however, the models used are given in Table 2-7.

Table 2-7. Models used in the EDC screen

Model	Pathway	Availability	Reference
Health Canada RF	Estrogen/Androgen	Free	Manuscript Submitted
CERAPP	Estrogen	Results for some for free	Mansouri et al. 2016
CoMPARA	Androgen	Results for some for free	Mansouri et al. 2020
CaseUltra binding	Estrogen	Licence required	Klopman 1992; Chakravarti et al. 2012; Saiakhov et al. 2013
CaseUltra agonism	Estrogen/Androgen	Licence required	Klopman 1992; Chakravarti et al. 2012; Saiakhov et al. 2013

Model	Pathway	Availability	Reference
CaseUltra antagonism	Estrogen/Androgen	Licence required	Klopman 1992; Chakravarti et al. 2012; Saiakhov et al. 2013

The last step in the ER and AR screen includes an automated process to reconcile differences in observed activities across *in vivo*, *in vitro*, and *in silico* sources of information. Highest confidence is placed in the results obtained from *in vivo* mechanistic studies (where available). If the results from the *in vivo* studies conflict with what is observed *in vitro* (or predicted *in silico*), the *in vivo* results prevail for the hazard indicator decision. Likewise, greater confidence is placed on *in vitro* results over the *in silico* predictions alone and, where conflicts arise, *in vitro* results take precedence over *in silico* predictions. Upon completion of the EDC screen, confidence criteria are applied in accordance with Table 2-8.

Table 2-8. Confidence level criteria for the EDC screen

Relative level of confidence in hazard indicator	Lowest	Moderate	Highest
Information source underpinning prioritization decision	QSAR prediction for binding, agonism or antagonism for ER or AR pathways.	Non-Guideline (or where guideline status is unknown) <i>in vitro</i> or <i>in vivo</i> mechanistic study <i>and/or</i> US EPA ToxCast data (No AUC model score available)	Guideline <i>in vivo</i> mechanistic study (for example, uterotrophic assay) <i>and/or</i> Guideline <i>in vitro</i> mechanistic study <i>and/or</i> US EPA ToxCast data (AUC models)
Rationale	QSAR models have been validated and are considered adequate for prioritization. However,	Results extracted from databases where study details could not be determined; as such, there is less confidence associated with the indicator.	Results extracted from databases where study details could be determined (for example, OECD guideline); as such, more confidence is

Relative level of confidence in hazard indicator	Lowest	Moderate	Highest
	since the decision is not based on empirical data there is less relative confidence than using <i>in vitro</i> or <i>in vivo</i> data.		associated with the indicator. US EPA ToxCast data (AUC models) have been validated against guideline mechanistic studies.

2.1.2 Overall hazard prioritization

HAWPr currently considers 7 endpoints within the hazard module, namely carcinogenicity, genotoxicity, reproductive toxicity, developmental toxicity, repeated dose toxicity as well as ER and AR activity. An overall hazard indicator level is assigned based on the positive hazard indicator with the highest confidence (from any endpoint) (Table 2-9). Substances with a high overall hazard indicator level are those where a positive hazard indicator was found and where there is the highest confidence in the data underpinning the indicator.

Table 2-9. Criteria for overall hazard indicator level of a substance

Overall hazard indicator level	Criteria
High	Any substance with at least 1 positive hazard indicator where the data underpinning the indicator has a high confidence level (for example, an international classification and/or a guideline <i>in vivo</i> study).
Moderate	Any substance with at least 1 positive hazard indicator where the data underpinning the indicator has a moderate confidence level (for example, <i>in vitro</i> studies or where a study guideline could not be determined).
Low	Any substance with at least 1 positive hazard indicator where the data underpinning the indicator has a lower confidence level (for example, RRA or QSAR predictions).
Non-priority	Any substance with data or predictions showing no positive hazard indicators.

Overall hazard indicator level	Criteria
Unknown	Any substance with no data to inform presence of hazard indicators and where read-across or QSAR was not possible (for example, substances of unknown or variable composition, complex reaction products or biological materials (UVCBs)).

Table 2-10 presents an example substance with positive hazard indicators for carcinogenicity, reproductive toxicity, developmental toxicity and for endocrine activity, while there are no hazard indicators for repeated dose toxicity. For genotoxicity, there is a positive finding for chromosomal aberrations in an *in vitro* assay, but the results are superseded by the negative findings of chromosomal aberrations *in vivo*. Since this substance has positive hazard indicators for several endpoints and the confidence level for at least 1 indicator is high (in this case, carcinogenicity), the overall hazard indicator level is high.

Table 2-10. Example designation of overall hazard score for Substance X

Endpoint	Indicator	Evidence	Confidence level
Carcinogenicity	Positive	International Classification (IARC 2B)	High
Genotoxicity	Negative ames Positive <i>in vitro</i> chromosome aberration Negative <i>in vivo</i> chromosome aberration	Guideline studies for <i>in vitro</i> evidence of mutation, chromosome aberration and/or micronuclei	High
Reproductive toxicity	Positive LOAEL < 300 mg/kg-bw/day	RRA from structural analogue	Low
Developmental toxicity	Positive LOAEL < 300 mg/kg-bw/day	Non-guideline <i>in vivo</i> study	Moderate
Endocrine activity	Positive	EPA ER AUC models using <i>in vitro</i> ToxCast data	High
Repeated dose toxicity	Negative ≥300 mg/kg-bw/day	Non-guideline <i>in vivo</i> study	Moderate
Overall hazard indicator level	N/A	N/A	High

2.2 Exposure

In general, it has been acknowledged that the lack of exposure data is one of the main challenges to chemicals regulation (European Commission, Joint Research Centre 2022). In most respects, hazard indicators have the same value internationally. For example, another jurisdiction's classification of a substance (for example, a classification from the ECHA) is just as relevant to people in Canada. Exposure indicators, on the other hand, are much more variable as substances are not necessarily used the same way or in the same amounts globally. Identifying Canadian-specific use and exposure information for all substances is difficult and therefore it is often necessary to consider non-Canadian sources of information. In these cases, additional considerations are required to determine if a given exposure indicator from a non-Canadian source would be anticipated to be applicable in Canada. This could involve a multitude of considerations, including, but not limited to, the ability to order and have a given product shipped to Canada from another jurisdiction (for example, purchase from a US retailer) or evidence of availability from a retailer's operations in another jurisdiction that also have operations in Canada. Temporal factors may also have a greater impact on the applicability of an exposure indicator than for hazard indicators. Given the rapid change in formulations of products, an older Safety Data Sheet (SDS) may no longer be truly reflective of the current formulation. Consequently, the approach for identifying indicators to evaluate the potential for exposure to people in Canada involves the consideration of a wide and varied set of sources, and the relevance of an indicator to exposure of people in Canada is a key determinant in this approach.

While approaches to deal with exposure information were developed as part of previous IRAP reviews to streamline and address the evaluation of exposure as efficiently as possible, it was recognized that there was considerable room for improvement and automation in this area. Most notable is the implementation of a rules-based weight-of-evidence approach (see section 2.2.1) to score indicators with both a relevance and occurrence metric. HAWPr allows the program to rapidly review a multitude of sources, such as databases and publicly available SDSs, in addition to the implementation of metrics to weigh confidence and relevance of a source. While many of these sources had been identified and used in previous IRAP reviews, the ability to systematically consider and use them to inform prioritization across thousands of substances was limited by the requirement to manually identify and consider them on an as needed, ad hoc basis. Table 8-2,

Table 8-3, and Table 8-4 outline the relative relevance of the exposure indicator data sources being considered in HAWPr.

2.2.1 Direct exposure

The term "direct exposure" refers to exposure to substances available to people in Canada for their use, either directly or as part of a mixture, product, or manufactured item. The user in this context is any consumer that has access to a product advertised, imported, or sold in Canada (including those sold online), and does not include direct

uses which may result from chemicals used by workers in an industrial or occupational setting. Different sources of data can provide different exposure indicators, which can be referred to generally as “occurrence”. Three rulesets were developed to partition exposure indicator data according to count (for example, how many products the substance is found in), tonnage (for example, how much of the substance is used), or presence/absence (for example, other indicators of exposure such as reported use in food packaging materials).

Conceptually, these rulesets provide a means to assign a relative relevance and occurrence that serves as an indicator for the potential for direct human exposure to people in Canada for a given substance. Exposure indicator data with the highest relevance and highest occurrence represent the substances with the highest potential for direct human exposure. Conversely, data with the lowest relevance and lowest occurrence represent the substances that HAWPr predicts to be least likely to have the potential for direct human exposure. Determinations of relevance are based on several considerations including previous program experience, origin of the information (for example, data from Canada is more relevant than international data), temporal relevance, and the overall quality/comprehensiveness of the data source. In addition, some sources can be linked more directly with a use that is likely to lead to, or represent, uses with direct exposure (for example, notification of use in a cosmetic). For other sources, however, the potential for direct exposure is not as intuitively clear, and it is less straightforward to determine the potential for direct exposure from the exposure information. In such cases, the information is used conservatively as a potential surrogate for evidence of direct exposure, but it is assigned a lower confidence ranking in the approach. It is important to note that the determination of potential for exposure in HAWPr only indicates the likelihood that there could be exposure of people in Canada and should not be used to infer the extent of exposure. This potential, coupled with binning of occurrence, low to high, will partition substances from those which have the highest potential for direct human exposure to those with the lowest, and all the combinations of relevance and occurrence in between.

To facilitate combining outcomes from multiple sources of information within a ruleset, each data point was assigned a relevance score and occurrence score (for example, 1, 2, and 3 for low, moderate, and high, respectively). The relevance and occurrence scores are summed together. For each substance, a total score is calculated as the sum of scores from all data sources. The total scores are then rescaled to a percentage using Equation 2-1, where the total score is divided by the maximum possible score (calculated by summing all maximum scores from each individual indicator in the ruleset).

Equation 2-1.

$$\text{rescaled score} = \frac{\text{score}}{\text{maximum possible score}} \times 100$$

[Long description: This figure shows an equation used to rescale the total scores from all sources. The equation depicts that the rescaled score is calculated by taking the total of the relevance and occurrence scores, divided by the maximum possible score from each individual indicator, then multiplying that value by 100.]

To emphasize high-relevance data sources in the product count and tonnage band rulesets, their relevance scores were increased in increments of 2; for example, low (1), moderate (3), high (5), and very high relevance (7). This approach allows a high relevance-low occurrence pair to score higher than a low relevance-high occurrence pair, placing greater importance on the relevance, or confidence, of the source.

2.2.1.1 Products available to consumers – product count ruleset

Indicators of direct consumer exposure can arise from a wide variety of sources and include indicators where the potential for direct exposure is clear, such as from use in cosmetics, as well as indicators where exposure potential is not as readily apparent. Canadian sources of exposure are used directly as indicators with a high relevance score. Other sources that are not explicitly from Canada (for example, Consumer Product Information Database (CPID)) are used as potential surrogate indicators, based on potential for availability in the Canadian market or potential access by people in Canada via inter-jurisdictional shipping (for example, online ordering from non-Canadian based retailer). The selection of surrogate data has been done using program experience on which of these indicators have proven to be the most reliable for the purpose of prioritization in Canada. To account for the use of surrogate data, the relevance score is lower thereby lowering the overall score.

For the product count ruleset (Figure 2-2), the occurrence frequency of a substance is binned according to the count of the number of products that substance occurs in from each given source. The occurrence frequency was partitioned into logarithmic bins of 1 to 10, 10 to 100, 100 to 1000, 1000 to 10000, and greater than 10000 products. These bins were assigned an occurrence score of 1 (very low), 2 (low), 3 (moderate), 4 (high), and 5 (very high), respectively. Data source relevance was assigned as follows and outlined further in Table 8-2: low relevance (for example, data on product counts not directly attributable to specific jurisdictions), moderate relevance (for example, product data or cosmetic notifications from international jurisdictions), high relevance (data on general products in Canada), and very high relevance (data on cosmetic products in Canada).

Figure 2-2. Product Count scoring matrix

		Product Count					
		0	1 to 10	10 to 100	100 to 1000	1000 to 10000	> 10000
		None	Very Low	Low	Moderate	High	Very High
Relevance	Score	0	1	2	3	4	5
Low	1	0	2	3	4	5	6
Moderate	3	0	4	5	6	7	8
High	5	0	6	7	8	9	10
Very High	7	0	8	9	10	11	12

[Long description: This figure shows the scoring matrix for the product count ruleset. On the left side is the listing of the 4 possible relevance levels and the associated scores. For example, low has a score of 1, moderate a score of 3, high a score of 5, and very high a score of 7. Along the top of the graphic is the 6 different product count bins that were developed for this approach. Going from 0 or no product count, followed by a 1-10 product count bin receiving of a very low occurrence, the product count bins and occurrence scores increase incrementally by an order of magnitude moving to the right, up to the final product count bin in the far right of greater than 10000 products, or an assignment of very high occurrence score of 5. The matrix therefore reflects a combination of the possible scores from combining the product count and relevance scores. For example, a substance with data from a indicator with very high relevance (score of 7), and a very high product count (score of 5), would result in an overall score of 12.]

For the product count ruleset, the relevance and occurrence scores from each data source are summed, then the results from all available data sources are summed to calculate an overall score. To illustrate this, consider the results shown in Table 2-11 for an example substance from the product count ruleset and the individual contributions to this score from the various product count indicators. The maximum indicator score for each data source (refer to the left-most column of Figure 2-2 and Table 8-2 to determine this) reflects the highest possible relevance and occurrence scores from a given product count indicator. Once calculated, the summed indicator score for a given substance is converted to a percentage using Equation 2-1. In this example, Substance X had a summed score of 33 out of a maximum indicator score of 52 from the ruleset, representing an overall score of 63%.

Table 2-11. Product ruleset score for Substance X

Data source	Relevance level (score)	Occurrence level (score)	Indicator score	Max indicator score
Cosmetic notifications (S.30)	Very High (7)	Moderate (3)	10	12
Canadian SDS	High (5)	Low (2)	7	10
VCRP + CalEPA (US Cosmetic notifications)	Moderate (3)	Very Low (1)	4	8
US SDS	Moderate (3)	Low (2)	5	8
CPID	Moderate (3)	Very Low (1)	4	8

Data source	Relevance level (score)	Occurrence level (score)	Indicator score	Max indicator score
US EPA Chemical & Product Categories (CPCat)	Low (1)	Low (2)	3	6
Sum	N/A	N/A	33	52

At present, the data sources used as exposure indicators within HAWPr do not allow for the granularity to make additional determinations on relevance based on product type. For example, the system considers a count of SDSs from retailers, but this count is not influenced by the specific type of product (for example, motor oil versus children's paint). However, this approach is highly amenable to consideration of this type of information, where available, as well as to the ongoing addition of new sources of product count indicators. Work will continue to identify and integrate new data sources moving forward.

2.2.1.2 Regulatory information on commercial status of substances - tonnage band ruleset

Exposure indicators based on tonnage commonly reflect information collected by regulatory authorities on the commercial activity of a substance in their jurisdiction. In many cases, this includes the volumes of a substance reported by industry/users, either manufactured or imported, within a particular period (for example, tonnes per year). This may also include details on intended use or function, and the sector of use. Within HAWPr, an approach has been developed to leverage the information that is available and currently amenable to evaluation in an automated fashion. The tonnage band ruleset (see Figure 2-3) was developed to score various sources of commercial information on substances based on the relevance of the source and the occurrence (reported tonnage). Where available, the reported use of substances in products available to consumers or use in products intended for children that is associated with reported tonnages was used to inform the relevance of this data as an indicator of direct exposure. For example, a substance reported with volumes in Canada in combination with a reported use in products would be considered to have higher relevance (higher likelihood of direct exposure to people in Canada) than a substance with volumes reported, but no reported use in products. The same decision framework would apply for commercial information available from other jurisdictions, such as in the US or Europe.

In this ruleset, the occurrence frequency of a substance is indicated by its yearly import and manufacture tonnage. A survey of the available data sources showed that bins spanning 2 orders of magnitude would adequately capture the range of occurrence, namely less than 10^1 tonnes (very low, score = 1), 10^1 to 10^3 tonnes (low, score = 2), 10^3 to 10^5 tonnes (moderate, score = 3), 10^5 to 10^7 tonnes (high, score = 4), 10^7 to 10^9 tonnes (very high, score = 5), and greater than 10^9 tonnes (extremely high, score = 6). Figure 2-3 shows the tonnage bands developed for this ruleset and the corresponding

score associated with an incremental increase in volumes by 2 orders of magnitude from a very low band to extremely high.

Figure 2-3. Tonnage band scoring matrix

		Tonnage						
		0	$\leq 10^1$	10^1 to 10^3	10^3 to 10^5	10^5 to 10^7	10^7 to 10^9	$> 10^9$
		None	Very Low	Low	Moderate	High	Very High	Extremely High
Relevance	Score	0	1	2	3	4	5	6
Low	1	0	2	3	4	5	6	7
Moderate	3	0	4	5	6	7	8	9
High	5	0	6	7	8	9	10	11

[Long description: This figure shows the scoring matrix for the tonnage band ruleset. On the left side is the listing of the 3 possible relevance levels and the associated scores. For example, low has a score of 1, moderate a score of 3, and high a score of 5. Along the top of the graphic is the 7 different tonnage bands that were developed for this approach. Going from 0 or no tonnage occurrence, followed by greater than 0 but less than 1 tonne for an assignment of a very low tonnage occurrence, the tonnage bands then increase incrementally by 2 orders of magnitude moving to the right, up to the final tonnage band in the upper right of greater than 10^9 tonnes, or an assignment of an extremely high occurrence score. The matrix therefore reflects a combination of the score assigned to the tonnage band and the relevance scores. For example, data from a indicator with high relevance (score of 5) and an extremely high tonnage occurrence (score of 6) would result in an overall score of 11.]

The selection of these bands is consistent with ones used both by Canada and by other jurisdictions internationally. In Canada, mandatory notices under section 71 of CEPA often collect quantities in order-of-magnitude ranges. Similarly, registrations in Europe under REACH are determined by registered volumes falling within 1 of 4 tonnage bands per year; specifically, 1-10 tonnes, 10-100 tonnes, 100-1000 tonnes and more than 1000 tonnes.

Similar with other rulesets, the confidence of direct exposure to people in Canada would also be reliant on the source of the information (for example, Canadian commercial information is assigned a higher confidence than US or European). Confidence in this context is reflected in the relevance level and score assigned to a given source used to inform the tonnage band ruleset (see

Table 8-3 for detailed relevance scoring). The further removed an indicator is from being an indicator of direct exposure to people in Canada, the lower the relevance. For example, tonnage information that is not associated with reported use in products is assigned a lower relevance than information from a source that is associated with both tonnage and reported use in products. While this approach allows for further interpretation of commercial data available, additional evaluation during the substance triage step of IRAP would be needed to confirm that direct exposure is anticipated. Similar to the product count ruleset, the score is calculated as the sum of relevance and occurrence scores multiplied by the occurrence call, summed over all data sources and then rescaled to a percentage using the maximum possible score of 52 (see Equation 2-1).

2.2.1.3 Presence-absence ruleset

The information considered as exposure indicators in this ruleset do not provide a clear link to specific products, but rather serve as a pointer to potential uses which may result in direct exposure (for example, listing in an international cosmetic ingredient database) or include surveillance data that may indicate current or past exposures in the population (for example, detected in a survey population in a biomonitoring database). For these data sources, occurrence frequency was not applicable or not readily available. Consequently, for this ruleset, occurrence is binary (yes or no) and assigned a score of 1 or 0, respectively. Relevance was partitioned into 3 levels (low, moderate, high) and assigned scores of 1, 2, and 3, respectively (see Figure 2-4). Considerations for the relevance assigned to a given indicator source are consistent with methods used in the other rulesets. For example, presence indicated in a biomonitoring study conducted in Canada is assigned a higher relevance score than presence on a list of substances provided by an international industry association (for example, the International Fragrance Association (IFRA) Fragrance Ingredient List).

The presence/absence score for a substance is the sum of its relevance and occurrence scores, summed over all data sources. Once calculated, the indicator score is converted to a percentage by dividing the summed score by the maximum possible score of 141 (calculated by summing all maximum scores from each individual indicator in the ruleset).

Figure 2-4. Presence-absence scoring matrix

Relevance	Occurrence Score	Absent	Present
		0	1
Low	1	0	2
Moderate	2	0	3
High	3	0	4

[Long description: This figure represents the scoring matrix for the presence-absence ruleset. On the left are the 3 levels for relevance, which are low, moderate and high and assigned scores of 1, 2 and 3, respectively. On the top is the occurrence, or presence, of an indicator. In this ruleset, there was only a score assigned of 1 or 0, based on a presence or absence of an occurrence of a substance from a given source. Therefore, the matrix scoring for the ruleset represents a combination of the presence score and the relevance score assigned to a given indicator source. For example, all substances that were absent from a source were assigned a score of 0 regardless of the relevance score of the source, whereas a combination of low relevance and presence received a score of 2, or a high relevance and presence received a score of 4.]

Many of the indicators in this ruleset have been used in previous IRAP reviews to help inform the potential for direct exposure. These include the identification of substances with releases reported in the National Pollutant Release Inventory (NPRI), those measured in the US EPA's Suspect Screening Analysis of Chemicals in Consumer Products, the Household and Commercial Products Association (HCPA) Consumer

Product Ingredient Database, as well as substances with potential, probable or very probable direct exposure from the Substances in Preparations in Nordic Countries (SPIN) database, amongst others (see Table 8-4).

The use of biomonitoring information as indicators of exposure has also been a part of previous reviews. Due to the nature of the information, manual review of the data is required to interpret the usefulness of this data in prioritization -- for example, determining the detection rate, trends over time, or what populations are included in the study. For the purpose of the automated approach, the biomonitoring data available will only be considered in the presence-absence ruleset. Further interpretation of the data will be required if the substance is flagged for further consideration.

Work will continue to develop approaches and automated tools that can better interpret and incorporate the data from these sources into more granular rulesets.

2.2.2 Overall exposure prioritization

For prioritization purposes, substances are given a quantitative score based on combining the scores from the individual rulesets. To combine the scores, the 3 rulesets are assigned a weight that incorporates the relative importance of the different types of information used to inform each ruleset, and their role in driving the prioritization of substances based on direct potential for exposure decisions in previous program experience. For instance, program experience has demonstrated that information such as domestic cosmetic notifications or the identification of SDSs for products are key indicators of a substance's potential for direct exposure based on anticipated use and availability to people in Canada. Consequently, these sources of information are weighted more highly than others. Likewise, program experience has shown that information on tonnage alone is not a strong predictor of the potential for direct exposure for people in Canada. For example, substances with high reported volumes in commerce may have uses that are considered unlikely to lead to direct exposure to the general population, such as industrial site-limited uses or use of substances as intermediates in product manufacture. Given these considerations, the weights assigned to each ruleset are: 70% to the product count, 20% to the tonnage band, and 10% to presence-absence.

An example showing how the scores from the individual rulesets are converted to a percentage and then to a weighted score is provided in Table 2-12. The scores from all 3 rulesets are then summed to get an overall exposure indicator score for the substance.

Table 2-12. Calculation of the total exposure indicator score for Substance X

Ruleset	Score	Score (%)	Assigned weight (%)	Weighted score
Product count	33/52	63.46	70	44.42
Tonnage	18/52	34.62	20	6.92
Presence - absence	25/141	17.73	10	1.77

Ruleset	Score	Score (%)	Assigned weight (%)	Weighted score
Exposure indicator score	N/A	N/A	N/A	53.11

For comparison with the qualitative hazard prioritization outcomes, the quantitative exposure indicator score is also converted to a qualitative level based on the thresholds shown in Table 2-13. As many of the exposure data sources used within HAWPr employ voluntary data reporting, and acknowledging that data collection using HAWPr is not exhaustive, a substance with an exposure score of 0 is categorized as having "unknown" exposure, rather than no exposure.

Table 2-13. Thresholds for exposure prioritization

Exposure indicator score	Exposure indicator level
40-100	High
20-40	Moderate
>0-20	Low
0	Unknown

2.2.3 Direct exposures from use in foods, natural health products, non-prescription drugs, and pesticides

Use of substances as food additives in Canada, or as active (or medicinal) ingredients in natural health products, non-prescription drugs, and pesticides all represent sources of potential direct exposure to people in Canada. However, for the purposes of prioritization under CEPA, these uses are considered regulated under other acts in Canada; and therefore, these specific uses do not represent a basis for prioritization within this approach. As part of the broader IRAP process, further consideration of these types of uses will be part of the substance triage and problem formulation/further scoping steps.

While not considered within HAWPr to inform prioritization for potential direct exposure, a brief description of the types of uses and their rationale for exclusion from this approach can be found below to provide more clarity on the extent that this information is, or is not, being used to inform outcomes from HAWPr.

These types of exclusions commonly reflect uses of substances which, while representing a potential for direct exposure to people in Canada, are addressed by programs, acts and/or regulations other than CEPA. The [Pest Control Products Act](#) regulates both active ingredients in pesticides as well as pesticide formulants. Additional exclusions include the use of substances in drugs including natural health products that are addressed and regulated in Canada under the [Food and Drugs Act](#) as well as under the [Controlled Drugs and Substances Act](#). Use in foods in Canada, including as a food additive, in food packaging materials or as food contact substances, are also regulated under the *Food and Drugs Act*.

It is important to note that, while exposures from these uses are not used explicitly for prioritization in HAWPr, the identification of these uses is valuable information within the broader IRAP prioritization framework. These indicators of exposure and use are valuable in assisting with the determination of the full breadth of exposures anticipated for the general population, such as in cases where there is co-occurrence of exposures from other sources.

2.3 Integrating hazard and exposure for risk-based outcomes

As IRAP is a risk-based approach, the decision to recommend a substance for further consideration depends on both the potential for hazard and exposure. The overall prioritization matrix applied within HAWPr to combine the results from both hazard binning (shown in Table 2-9) and exposure binning (shown in Table 2-13) is depicted in Figure 2-5.

Figure 2-5. Conceptual prioritization matrix

		Exposure Indicator			
		Unknown	Low	Moderate	High
Hazard Indicator	Unknown	Unknown			
	Non-priority	Non-priority			
	Low	Low	Low	Moderate	
	Moderate	Low	Moderate		High
	High	Moderate	High		High

[Long description: This figure shows the overall prioritization bins resulting from the overlaying of the hazard and exposure indicators bins. On the left side, from top to bottom, are listed the various hazard indicator bins, including unknown hazard, non-priority, low, moderate and high priority. On the top of the graphic, listed from left to right, are the exposure indicators unknown, low, moderate and high. Various combinations of these indicator bins result in different prioritization bins. For example, a high hazard bin along with a high exposure bin results in a high prioritization bin, a high exposure bin and low hazard bin result in a moderate prioritization bin, a low exposure and low hazard bin result in a low prioritization bin, and so on.]

The outcomes resulting from HAWPr will be considered in conjunction with other input streams and used to inform the next steps in the overall IRAP process (Figure 1-1), with emphasis placed on the substances falling in the high priority bins and the associated scores within those bins. Further, the modular design of HAWPr allows for ongoing refinements and the flexibility to target areas of interest or emerging concern such as chemical classes of concern (for example, EDCs), substances with the potential to cause cancer, or potentially disproportionately impacted populations (for example, pregnant people, children). HAWPr output will be coupled with other information including, but not limited to, whether the substance has already been assessed or managed under CEPA, whether there is ongoing international work, or whether the substance should be considered as an individual or as part of a larger group.

3 Validation

Prior to the implementation of HAWPr to screen and score the entire DSL, subsets of substances that had expected outcomes were selected and used to validate that both the hazard and exposure arms of the workflow were generating results in line with what would be expected. In general, each approach was not expected to have a 1:1 alignment with the retrospective test set as long as differences could be understood and described as a function of new information, new approaches, advanced decision rules or legislative considerations. Below is a brief overview of the results of the validation tests; for detailed results and analysis, please refer to section 8.2 Validation of HAWPr.

For hazard, a curated list of substances was created using substances that had been [previously concluded as meeting criteria under section 64c of CEPA](#), were included on the [ECHA Authorisation List](#) consisting of substances of very high concern (SVHC), or were listed on the [EPA Toxic Substances Control Act \(TSCA\) Low Priority list](#). This design allowed for the testing of both a 'positive control' with 2 lists of substances with expected high hazard outcomes, and a 'negative control' with a list of substances with expected low hazard outcomes. The concordance between the list of 226 substances that had been previously concluded as meeting the criteria under section 64c of CEPA and HAWPr was 94% for the low indicator level and above, with 90% concordance at the moderate indicator level and above. The 111 ECHA SVHC substances had a concordance level of 79% at the low indicator level and above, dropping to 46% when looking at moderate indicator level and above. More substances are captured at the low indicator level as these substances make use of QSAR and read-across in the absence of substance specific data. Looking at the TSCA Low Priority list, 42% of substances were found in the moderate indicator level or above, which is higher than expected. However, as HAWPr is one component of the overall IRAP approach, the system is designed to make conservative decisions rather than to miss substances during the automated triage. TSCA Low Priority substances are typically captured by HAWPr for repeat dose toxicity indicators, as the HAWPr triggering threshold for this module is a NOAEL or LOAEL at 300mg/kg-bw/day which is considered conservative.

For exposure, [outcomes for substances that underwent rapid screening under the CMP](#) were compared to the results generated by HAWPr. Rapid screening approaches were applied to substances which reportedly had low or no use in Canada. As such, comparison against the tonnage band ruleset would not be useful. When comparing the results of the product count ruleset from HAWPr with the rapid screening results, it was found that 87% of the substances had agreement between both approaches in their identification of direct exposure indicators. The 13% of substances for which direct exposure was identified via the rapid screening process and not HAWPr were determined to be a result of intensive manual searching using disperse sources unavailable to HAWPr at this time. Although the sources included in HAWPr will continue to be expanded, it should be noted that not all sources will be amenable to automated collection or review in HAWPr, and these sources will instead continue to be considered as part of the broader IRAP review and prioritization framework.

Finally, the [2019 IRAP priorities](#) were used to support validation for both hazard and exposure systems. It was expected that there would be good concordance between the substances flagged for hazard between the 2 approaches given that HAWPr incorporates the same hazard indicators previously used by IRAP. This was reaffirmed by observing that 85% of the substances flagged by IRAP as high hazard were assigned high priority based on hazard flags in HAWPr. In addition, it was observed that HAWPr has greatly improved the amount of hazard information considered over previous IRAP reviews. A powerful illustration of this was the identification of almost 2900 substances assigned either moderate or high priority based on hazard flag by HAWPr that did not have high hazard flags from the IRAP 2019 review.

The process to flag the potential for direct exposure in IRAP 2019 was binary (Yes/No), and therefore did not allow for a comparable evaluation to HAWPr. To address this issue and allow comparison between the 2 approaches, data from the IRAP 2019 exposure review was manually reviewed and assigned exposure flags, following the logic used in HAWPr. For instance, the use of cosmetic notifications in IRAP was aligned with HAWPr and assigned an appropriate flag to facilitate cross-comparison between the 2 prioritization methods. Using this approach, the 2 systems showed high levels of concordance. Exact agreement between IRAP and HAWPr was found for 82% of substances and HAWPr identified more exposure indicators than had been captured by IRAP 2019 in 9% of cases. An additional 8% of substances were found by HAWPr to have exposure indicators where IRAP 2019 had not identified exposure data. In total, HAWPr performed as well or better than IRAP 2019 for 99% of substances.

The IRAP 2019 overall results reflect decisions made as a result of subsequent triaging and scoping work done as part of that initiative as described previously for [Figure 1-1](#). As a result, it is recognized that the validation of overall outcomes from IRAP 2019 versus HAWPr requires some additional considerations. However, as HAWPr is one important feeder into the broader IRAP process, the expectation is that the system should provide a more conservative outcome that can be further refined as a substance or group of substances progresses through subsequent expert-guided stages of the IRAP process. With that in mind, it was found that there was still good alignment observed between the substances with outcomes of no further action at this time from IRAP 2019 and the results from HAWPr; specifically, over 91% of IRAP 2019's 'no further action' substances were assigned overall priority scores of moderate-low or below by HAWPr. The remaining ~9% of substances were those where the difference between the overall outcomes of the 2 systems was driven primarily by new hazard indicators used within HAWPr but not captured previously by IRAP. With respect to the 67 substances on the DSL identified for problem formulation and further scoping from the IRAP 2019 review, approximately 65% were also identified by HAWPr as having an overall priority of moderate or higher. All of these substances (43) were identified by Health Canada as human health priorities. The 24 remaining substances were the result of adding group members based on read-across, expert judgement during the subsequent steps of IRAP, or were originally identified as ecological priorities. Overall, the results of HAWPr are consistent with IRAP 2019, and the majority of discrepancies

identified reflect follow-up work undertaken in IRAP 2019 to further scope, triage, cluster and group substances.

These validation exercises tested the automated system, including advanced rulesets, against sets of substances with known outcomes for both hazard and exposure. As discussed above and detailed in section 8.2, the results indicate that HAWPr performs exceptionally well, capturing high hazard substances and substances with potential for exposure with greater efficacy and efficiency than previous manual approaches. The automated rule-based data integration and decision workflow also provides the added strength of assigning a defined level of confidence to the overall priority ranking. Further, the discrepancies identified are understood, and as such do not represent critical faults with the system; rather, discrepancies can often be explained by general improvements to the process and, in some cases, the use of additional datasets and emerging science.

4 HAWPr outcomes for substances on the DSL

As highlighted, the primary function of HAWPr is to gather, organize and score available data from existing and emerging science and monitoring approaches (Figure 1-1) as a feeder into the overall IRAP process. Through the validation exercises described above, it was shown that HAWPr reliably delivers on these critical aspects of the IRAP process, providing assurance that similar success will occur following application to the DSL. Substances on the DSL with valid CAS RNs (n=25,286) have been processed by the HAWPr workflow, with the general outcomes described below. Detailed results for each substance are available as a supporting document.¹ It is recognized that many DSL substances have been previously assessed and, in some cases, listed on Schedule 1 to CEPA. As such, this initial prioritization will be further curated to take into account previous outcomes and actions as appropriate.

The overall prioritization outcomes for the DSL utilizing the binning system as described in section 2.3 is illustrated in Figure 4-1. The values in each cell indicate the number of substances in that priority bin. The colour-coding is described in the “Overall Prioritization Outcomes” section and illustrated by the conceptual prioritization matrix in Figure 2-5. Approximately 39% of substances were classified as priorities (low to high). Substances classified as non-priorities make up 12% of the total, with 2233 substances identified as non-priority hazard and low exposure. The remaining 49% of substances lack enough information for prioritization, with the high confidence hazard-unknown exposure (n=42) and high exposure-unknown hazard substances (n=42) potentially being candidates for further investigation.

¹ Detailed results for the 25286 substances processed by the HAWPr workflow contain: hazard data and their associated confidence; exposure scores; and the overall hazard, exposure, and priority levels.

Figure 4-1 Prioritization matrix for the DSL

		Exposure Indicator			
		Unknown	Low	Moderate	High
Hazard Indicator	Unknown	4468	2544	379	42
	Non-priority	975	1704	342	114
	Low	4757	5428	615	154
	Moderate	191	579	168	93
	High	42	1925	535	231

[Long description: Similar to Figure 2.5, this graphic shows the prioritization matrix based on the results from the exposure and hazard binning. However, in this case, rather than the conceptual binning of prioritization bins, this graphic includes the results of HAWPr for the DSL. On the left side, the various hazard indicator bins are listed from top to bottom, including unknown hazard, non-priority, low, moderate and high priority. On the top of the graphic, listed from left to right, are the exposure indicators unknown, low, moderate and high. The number of substances that fall into each priority bin are then shown in the matrix. 231 substances were identified as having high hazard and high exposure indicators.]

The qualitative binning allowed easy distinction of which substances should be identified for further consideration in IRAP, where further curation of the list and consideration of other data sources will be conducted before recommending priorities for future work. The unknown bins in the HAWPr outcomes also revealed that insufficient data is available to prioritize a significant portion of the DSL substances even with the additional incorporation of emerging science and automated read-across approaches as modernized elements of the evolving system. This finding emphasizes the need to continue to innovate and advance the development of NAMs to support ongoing data-driven prioritization and assessment activities to address key data gaps. Section 5 below outlines work underway to continue to evolve and expand the use of emerging science and technologies to enhance HAWPr incrementally as data and tools become available.

4.1 Identifying emerging areas of concern to people in Canada under an amended CEPA

In addition to using the general binning approach as described above, HAWPr also allows for the probing of specific human health-related concerns that can be considered during the overall IRAP process. In 2017, the House of Commons Standing Committee on Environment and Sustainable Development released a report with recommendations on strengthening CEPA as part of its 5 year review cycle (Parliament of Canada 2017). In the report were several recommendations for amendments to CEPA that pertain to a substance's intrinsic hazard properties including the identification of substances which have carcinogenic, mutagenic or toxic to reproduction (CMR) properties or other

substances of equivalent concern such as EDCs. More recently, [Bill S-5](#) received Royal Assent and included several amendments to CEPA, which have impacted the focus of priority-setting activities.

Given the modular nature of HAWPr, it is possible to adapt the process to further prioritize substances based on specific CMR and EDC hazard indicators ([Table 4-1](#)) or other identified concerns. Moreover, by collecting hazard information across the entire DSL and combining this information with structural identifiers for substances, clusters can be formed to gain efficiencies in the assessment process as well as explore the potential for cumulative effects based on specific identified hazards (see [section 5 Future work](#)).

Table 4-1. Overall HAWPr results showing the number of DSL substances that have been identified for each hazard flag

Hazard flag	Number of DSL substances
Carcinogenic	2473
Mutagenic	5172
Reproductive	4638
Developmental	4300
Endocrine (ER)	3664
Endocrine (AR)	3503

In addition, by coupling hazard data to available exposure data, impacts to disproportionately impacted populations can be considered and given priority going forward. For example, using HAWPr collected information, 395 DSL substances were found in children's products that also have either a flag for developmental effects or flags for potential endocrine activity. As additional exposure data becomes available, linking hazard to exposure becomes increasingly feasible and thus the potential for identifying additional impacted populations can be realized.

4.2 Limitations and uncertainties

While validation has demonstrated that HAWPr is a robust and reliable tool to aid in prioritization within IRAP, it is important to understand the limitations and uncertainties of HAWPr when reviewing the output during the substance triage phase of IRAP.

For the hazard module, the first limitation relates to using summary level information from toxicity studies without confirming the endpoint of effect of concern. The hazard data sources within HAWPr cover tens of thousands of toxicity studies. However, depending on the source of the information and endpoint considered, not all study details are available and what can be currently processed in an automated manner are limited to certain study details, such as study type, dosing route and duration, species, and POD information like the identified NOAEL. When working with summary level information, it may not be possible to identify specific adverse effects without examining the full study reports. HAWPr is intended to facilitate prioritization, and using POD

values to identify potential repeated dose toxicants or reproductive/developmental toxicants is appropriate given that further follow-up on a substance happens during the substance triage phase in IRAP. During this phase, specific study reports can be reviewed to examine specific adverse effects. Using POD values as a basis for hazard indicators likely overestimates the true number of reproductive or developmental toxicants. International classification information is also used to identify specific hazards for CMR endpoints (for example, IARC classifications for carcinogens) which ensures that chemicals reviewed and classified by experts are considered for prioritization. For other endpoints, such as genotoxicity and the endocrine pathways examined within HAWPr, this limitation does not exist as the prioritization decision makes use of assays that provide more mechanistic information, and hence the prioritization decisions are more robust for these endpoints.

An additional limitation in the hazard module of HAWPr relates to the application of read-across for data-poor chemicals. As currently implemented, the clustering-based read-across is entirely dependent on Tanimoto scores using chemical structural fingerprints. More robust read-across justifications rely on mechanism of toxicity information, physicochemical property similarity and metabolic similarity. Moreover, an examination of similarity across hazard endpoints within a cluster of structurally related substances also provides additional scientific support and greater confidence for applying read-across. Incorporating these additional elements for analogue identification and read-across within HAWPr is planned for future versions of the tool.

Mixtures and UVCBs also present challenges for the read-across method deployed in this workflow. Many mixtures and UVCBs do not have adequately defined structural representations and, since the read-across method in HAWPr relies on comparing structural similarity for analogue identification, this read-across method cannot be used to fill data gaps for these substances. This lack of structural representation also precludes the use of predictive modelling, which also requires structure information as a prerequisite. If there is substance-specific information available for a UVCB or mixture (based on CAS RN as the query), then this data is used for prioritization. For those without substance-specific data, HAWPr cannot be used to facilitate prioritization through the identification of hazard flags and different strategies must be implemented outside of HAWPr for such substances.

A current uncertainty in the hazard module relates to not capturing all potential toxicological effects important for human health risk assessment during the prioritization process. While there may be certain toxicological effects that are not currently well identified with HAWPr, more specific hazard identification, beyond carcinogenicity, mutagenicity, reproductive and developmental toxicity, and certain endocrine pathways, is envisioned in the future as more NAMs are incorporated into the hazard module. This can be accomplished as more NAM-based assays are mapped to adverse outcome pathways (AOP). Health Canada participates in the development of the [AOP Knowledge Base](#) maintained by the OECD, which provides a platform where

mechanistic events measured in *in vitro* assays are tied to specific adversities observed *in vivo*. Once available, these *in vitro* assays will be used to further expand the hazard module to identify potential hazards using NAM (for example, developmental neurotoxicity).

The exposure module has its own limitations. The module is not intended to provide a definitive determination of exposure, but rather identifies the potential for exposure. The presence of exposure indicators provides an increasing weight of evidence that direct human exposure may occur without specifically affirming or refuting it. Related to this, the system does not provide information on possible routes of exposure for comparison with potential hazards and no quantitative estimates of exposure are currently generated. Such information can be discerned at the substance triage step within IRAP, if warranted.

In addition to this, the system shares the same limitations other approaches must deal with when attempting to identify and capture exposure information relevant for not only prioritization, but also risk assessment more generally. This includes, but is not limited to, the availability of exposure information amenable to systematic evaluation and limitations of the indicator sources themselves. For example, while concerted efforts can be made to extract a comprehensive dataset of SDSs from Canadian and international retailers, there are recognized limitations with respect to the types of products that may or may not have SDSs commonly available for them. Similarly, manufactured items present additional limitations, with difficulties identifying substances present in manufactured items (for example, an SDS does not exist for a manufactured product such as an article of clothing) and potential for release from those items. Proprietary information on ingredients also impacts the availability of information on product composition. Fragrances are an example of this issue, where evidence of potential use of substances as fragrance may have been identified, but specific listing of the substance as an ingredient in an SDS is not available to corroborate this use information as many are listed under Fragrance/Parfum in an SDS.

At present, more detailed information in a number of areas, which is valuable for determining potential exposure, is not evaluated or currently considered as part of the HAWPr system. For instance, information on specific function/use or product type is not systematically considered nor reflected in HAWPr, and therefore prioritization of exposure relies primarily on metrics such as the frequency of indicators identified (for example, number of Canadian SDSs). The current system lacks the ability to distinguish and systematically consider variability in potential exposures based on product type (for example, high exposures anticipated from use in body lotion versus lower from use in nail polish). Enhancements in this area would improve HAWPr's ability to not only qualitatively identify potential indicators for direct human exposure, but also allow for more quantitative measures of this potential. Future iterations of the system will consider ways to improve the evaluation of this valuable exposure information. However, it is recognized that there are issues with respect to both the availability and

ability to extract this information comprehensively, systematically, and consistently from these sources. This type of information, if available and extracted more consistently, would allow for further refinements in the exposure workflow, allowing for additional weighting based on factors such as concentration limits or age of an SDS. The tools to better extract this information, and therefore allow for these types of further refinements in the exposure workflow, will continue to be addressed and improved as data and methods for natural language processing (NLP) and text extraction permits. Until this can be automated, this supplemental information is considered in the overall IRAP process to help further inform selection of priorities.

While there are current limitations and uncertainties pertaining to the use of HAWPr, the validation exercises conducted in section 3 have demonstrated that the tool is accurately identifying risk assessment priorities while simultaneously being much faster to deploy than previous cycles of IRAP. HAWPr will continue to undergo improvements to further enhance the prioritization capabilities while addressing the current known limitations and uncertainties, which are largely a function of the current state of the science and information that is available to process in an automated manner.

5 Future work

Several projects are currently underway to ensure continual innovation to keep the pace with technological and scientific developments. This section presents an overview of the tools and approaches currently in development ranging from the conceptual to the near complete. As new tools are developed and datasets emerge, stakeholders can expect that these will be incorporated into HAWPr.

5.1 Automated collection and relevance ranking of open scientific literature

One of the largest challenges in prioritization has been the identification of chemical data, such as hazard data, derived from unstructured sources like open scientific literature. Research groups across the world continually publish new chemical toxicity information in peer-reviewed scientific journals and, as a result, there are thousands of available papers for DSL chemicals from a wide variety of journals. A simple search of a chemical name or [CAS registry number](#) using a journal abstract and citation service, such as [Scopus](#), can yield thousands of results per chemical; however, only a selection of the returned results may be useful for hazard assessment. As part of efforts to modernize chemical prioritization and assessment and to ensure that all available relevant information is considered in the process, NLP and machine learning techniques were used to develop a pilot model to screen journal abstracts for relevancy for hazard assessment and prioritization. This process presents all abstracts captured through a search query and sorts most potentially relevant abstracts to the top of the user's list. The development and testing of the NLP model are currently underway. The NLP neural

network will help prioritize articles for evaluators to read as well as serve as an overall indicator of the volume of PubMed literature available for a given chemical.

5.2 Clustering for data-gap filling and group assessment

As presented in Table 4-1, many chemicals on the DSL are data poor despite the new and varied data sources included in this workflow, and are expected to remain as such given the resources required to generate experimental data or perform large-scale surveys. This lack of data makes the assignment of hazard or exposure potential difficult. Therefore, it is proposed that substances be clustered into groups based on structural similarity. Substances with common chemical structure have often been shown to exhibit common activity – this is the basis for qualitative and quantitative structure-activity relationship (SAR and QSAR) models, as well as the rationale for applying read-across (OECD 2017). With this relationship in mind, it may be possible to fill some of the data gaps for data-poor substances from those similar substances that are data-rich. Additionally, grouping similar chemicals increases efficiency and allows the consideration of clustered chemicals for a group risk assessment.

Structural features of a chemical can be expressed as a unique chemical fingerprint. Rudimentary read-across using structural features assessed via a Tanimoto metric provides a high throughput systematic method of comparison of chemicals to cluster similarities. Preliminary work has indicated that a Tanimoto similarity threshold of 0.85 yields clusters of high structural homogeneity while maintaining a reasonably large average cluster size. High throughput clustering allows potential to overlay hazard flags from data-rich chemicals to data-poor chemicals and to explore if chemicals have functional similarities in products. This information can help to identify co-exposures and aid in cumulative risk assessments.

5.3 *In silico* modeling

5.3.1 Hazard

The DSL has many substances that are data poor, specifically with the availability of hazard information. As shown with this work, *in silico* methods can be a useful tool to supplement the information about a substance when other data is not available. *In silico* models for endocrine and genotoxic endpoints are incorporated into HAWPr. Currently, *in silico* models pertaining reproductive, developmental and neurotoxicity are being evaluated for use in HAWPr. To develop additional models, specific well-defined toxicological endpoints need to be determined. The well-defined toxicological endpoint needs to meet 2 requirements: 1) multiple *in silico* models which predict a given endpoint and 2) availability of high-quality datasets containing information about the toxicology endpoint of interest to compare against the individual *in silico* models. Further work for *in silico* models will be focused on developing additional in-house QSAR models to predict toxicological endpoints as data allows, as was previously done with ER and AR activity. This process does require a large high-quality dataset for each toxicological endpoint of interest, which is the roadblock in continuing the process.

5.3.2 Exposure

The greatest advances within the program to date have focused on *in silico* models for predictions of the hazardous properties of substances. However, *in silico* approaches can also help to predict information on potential exposures to substances of concern. This can be done by using models known as quantitative structure use relationships (QSUR). QSURs work like QSAR models; using machine learning software, the structure of a substance can be related to a potential use. There are online datasets available which have a variety of substances and have associated use data.²

Future work will involve the exploration, development, and implementation of QSUR models within HAWPr. The QSUR models will use the data from available datasets and have multiple models developed to predict if a substance is likely to have a particular function or use. This information can also aid in the potential grouping of possible substitutes. Work is continuing to evaluate ways to incorporate other approaches and tools, such as ExpoCast (Ring et al. 2019) and PROTEX-HT (Li et al. 2021) for use in the prioritization, screening and ranking of exposure potential for large substance data sets (for example, DSL). Next steps in this area will not only consider how best to consider the use and outputs of these individual *in silico* tools and approaches, but also how, and if, they may be used in tandem with existing approaches to provide and inform a more holistic approach to prioritization and assessment.

5.3.3 Indirect exposure

Although indirect exposures are commonly considered in CMP assessments if relevant, results of previous assessments have shown that concern for human health is more commonly driven by direct exposures from products available to consumers (Bonnell et al. 2018). For this reason, indirect exposures have not been considered systematically in previous iterations of IRAP. In the future, Health Canada will move to integrate approaches to indirect exposure as part of the overall process of determining exposure flags and prioritization in HAWPr. The intention is to develop a process for flagging indirect exposures, leveraging and building on the process developed and utilized in Environment and Climate Change Canada's Ecological Risk Classification 2 (ERC2) (Environment and Climate Change Canada 2022), with a focus on exposures via air, water, soil, and incidental exposure from food. Potential for indirect exposure will add to the overall exposure prioritization scores and could be used to further help with differentiation amongst priorities.

² [US EPA Functional Use Database](#), [ECHA REACH Registered Substances](#), [US FDA Voluntary Cosmetic Registration Program \(VCRP\)](#), [California EPA cosmetics list](#), [UL Prospector](#), [Consumer Product Ingredients Database](#)

6 Conclusions

HAWPr is a tool that provides an expansive scope, enhanced efficiency and flexibility to modernize the Health Canada approach to prioritize chemicals. Developed as a high-throughput automated system to collect, organize, and extract various types and levels of information, this decision support tool modernizes the data considerations to identify key hazard and exposure indicators and inform prioritization outcomes. The analysis of the DSL resulted in the collection of over one million data points which the program can use to prioritize chemicals and identify emerging issues, such as substances with CMR or EDC flags.

Validation of the results from the use of HAWPr compared to previous prioritization exercises showed a high level of concordance. In addition, it was observed that HAWPr has greatly improved the amount of hazard information available to inform decisions compared to previous reviews, including the integration of advances in the emerging science with traditionally available animal data. To ensure that key priorities are not missed in the process, the system has been designed to make conservative decisions during the automated triage phase. It is recognized that additional manual review and scoping will be required for any priorities identified using this tool. This ensures the automated interpretation of the data was accurate and allows for further scoping of the priorities as needed, for example to identify appropriate assessment groups.

HAWPr was designed to allow flexibility of components within the tool, ensuring that future developments are not limited by the confines of any one piece of software, scripting language, or individual expertise, and decision flows can be refined and expanded as information becomes available. HAWPr will be improved through continual innovation to keep pace with technological and scientific developments. It is one important component of the overall IRAP approach which can greatly improve the efficiency, thoroughness and consistency of the process of identifying substances with a potential human health concern.

7 References

- Bajusz D, Rácz A, Héberger K. 2015. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J Cheminformatics*. 7(1):20.
- Benfenati E, Manganaro A, Gini G. 2013. [VEGA-QSAR: AI Inside a Platform for Predictive Toxicology](#). In: [Proceedings of the Workshop Popularize Artificial Intelligence co-located with the 13th Conference of the Italian Association for Artificial Intelligence \(AIXIA 2013\)](#). Vol. 1107. Turin, Italy. p. 21–28.
- Bonnell MA, Zidek A, Griffiths A, Gutzman D. 2018. Fate and exposure modeling in regulatory chemical evaluation: new directions from retrospection. *Environ Sci Process Impacts*. 20(1):20–31.
- Browne P, Kleinstreuer NC, Ceger P, Deisenroth C, Baker N, Markey K, Thomas RS, Judson RJ, Casey W. 2018. Development of a curated Hershberger database. *Reprod Toxicol*. 81:259–271.
- Chakravarti SK, Saiakhov RD, Klopman G. 2012. Optimizing Predictive Performance of CASE Ultra Expert System Models Using the Applicability Domains of Individual Toxicity Alerts. *J Chem Inf Model*. 52(10):2609–2618.
- Collins SP, Barton-Maclaren TS. 2022. Novel machine learning models to predict endocrine disruption activity for high-throughput chemical screening. *Front Toxicol*. 4:981928.
- Collins SP, Mailloux B, Kulkarni SA, Long A, Barton-Maclaren TS. 2023. Development of Consensus in Silico Models for Toxicological Predictions - In preparation.
- Environment and Climate Change Canada. 2022. [Science approach document - ecological risk classification of organic substances version 2.0 \(ERC2\)](#).
- 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. [accessed yr mon date].
- European Commission. Joint Research Centre. 2022. [Addressing evidence needs in chemicals policy and regulation](#). LU: Publications Office.
- Hasselgren C, Ahlberg E, Akahori Y, Amberg A, Anger LT, Atienzar F, Auerbach S, Beilke L, Bellion P, Benigni R, et al. 2019. Genetic toxicology in silico protocol. *Regul Toxicol Pharmacol*. 107:104403.
- Helman G. 2019. [Generalized Read-Across \(GenRA\): A workflow implemented into the EPA CompTox Chemicals Dashboard](#). ALTEX.

Kleinstreuer NC, Ceger PC, Allen DG, Strickland J, Chang Xx, Hamm JT, Casey WM. 2016. A Curated Database of Rodent Uterotrophic Bioactivity. *Environ Health Perspect.* 124(5):556–562.

Klopman G. 1992. MULTICASE 1. A Hierarchical Computer Automated Structure Evaluation Program. *Quant Struct-Act Relatsh.* 11(2):176–184.

Landry C, Kim MT, Kruhlak NL, Cross KP, Saiakhov R, Chakravarti S, Stavitskaya L. 2019. Transitioning to composite bacterial mutagenicity models in ICH M7 (Q)SAR analyses. *Regul Toxicol Pharmacol.* 109:104488.

Li L, Sangion A, Wania F, Armitage JM, Toose L, Hughes L, Arnot JA. 2021. Development and Evaluation of a Holistic and Mechanistic Modeling Framework for Chemical Emissions, Fate, Exposure, and Risk. *Environ Health Perspect.* 129(12):127006.

Low Y, Sedykh A, Fourches D, Golbraikh A, Whelan M, Rusyn I, Tropsha A. 2013. Integrative Chemical–Biological Read-Across Approach for Chemical Hazard Classification. *Chem Res Toxicol.* 26(8):1199–1208.

Mansouri K, Abdelaziz A, Rybacka A, Roncaglioni A, Tropsha A, Varnek A, Zakharov A, Worth A, Richard AM, Grulke CM, et al. 2016. CERAPP: Collaborative Estrogen Receptor Activity Prediction Project. *Environ Health Perspect.* 124(7):1023–1033.

Mansouri K, Kleinstreuer N, Abdelaziz AM, Alberga D, Alves VM, Andersson PL, Andrade CH, Bai F, Balabin I, Ballabio D, et al. 2020. CoMPARA: Collaborative Modeling Project for Androgen Receptor Activity. *Environ Health Perspect.* 128(2):027002.

Myatt GJ, Bassan A, Bower D, Johnson C, Miller S, Pavan M, Cross KP. 2022. Implementation of in silico toxicology protocols within a visual and interactive hazard assessment platform. *Comput Toxicol.* 21:100201.

[OECD] Organisation for Economic Cooperation and Development. 2017. [Guidance on Grouping of Chemicals, Second Edition](#). OECD (OECD Series on Testing and Assessment).

[OECD] Organisation for Economic Cooperation and Development. 2018. [Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption](#). OECD (OECD Series on Testing and Assessment).

Parliament of Canada. 2017. [Healthy Environment, Healthy Canadians, Healthy Economy: Strengthening the Canadian Environmental Protection Act, 1999](#). 42ndParl., 1st sess.

Ring CL, Arnot JA, Bennett DH, Egeghy PP, Fantke P, Huang L, Isaacs KK, Jolliet O, Phillips KA, Price PS, et al. 2019. Consensus Modeling of Median Chemical Intake for

the U.S. Population Based on Predictions of Exposure Pathways. Environ Sci Technol. 53(2):719–732.

Saiakhov R, Chakravarti S, Klopman G. 2013. Effectiveness of CASE Ultra Expert System in Evaluating Adverse Effects of Drugs. Mol Inform. 32(1):87–97.

Senate of Canada. 2022. [Bill S-5, Strengthening Environmental Protection for a Healthier Canada Act - Summary of Amendments](#).

Shah I, Liu J, Judson RS, Thomas RS, Patlewicz G. 2016. Systematically evaluating read-across prediction and performance using a local validity approach characterized by chemical structure and bioactivity information. Regul Toxicol Pharmacol. 79:12–24.

8 Appendix

8.1 List of sources

Table 8-1. Sources of hazard information.

Indicator/Source	C	G	R	D	RD	NAM	EDC	TK
California Office of Environmental Health Hazard Assessment	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Carcinogenicity & Mutagenicity (ISSCAN)	Y	Y	N/A	N/A	N/A	N/A	N/A	N/A
Carcinogenic Potency Database (CPDB)	Y	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Carcinogenesis Research Information System (CCRIS)	Y	Y	N/A	N/A	N/A	N/A	N/A	N/A
Procter & Gamble Developmental & Reproductive Database [via OECD QSAR Toolbox]	N/A	N/A	Y	Y	N/A	N/A	N/A	N/A
International Life Science Institute (ILSI) Developmental Toxicity Database [via OECD QSAR Toolbox]	N/A	N/A	N/A	Y	N/A	N/A	N/A	N/A
ECHA REACH Dossiers	Y	Y	Y	Y	Y	N/A	Y	N/A
European Food Safety Authority Open Food Tox Database [via OECD QSAR Toolbox]	N/A	N/A	Y	Y	Y	N/A	Y	N/A
US EPA Health Effects Assessment Summary Tables (HEAST) [via US EPA ToxVal DB]	Y	N/A	Y	Y	Y	N/A	N/A	N/A
US EPA High Production Volume Information System (HPVIS) [via US EPA ToxVal DB]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
US EPA Human Health Benchmarks for Pesticides [via US EPA ToxVal DB]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
US EPA IRIS Assessments [via US EPA ToxVal DB]	Y	N/A	Y	Y	Y	N/A	N/A	N/A
US EPA OPP Assessments [via US EPA ToxVal DB]	Y	N/A	Y	Y	Y	N/A	N/A	N/A
US EPA Provisional Peer-Reviewed Toxicity Values (PPRTVs) [via US EPA ToxVal DB]	Y	N/A	Y	Y	Y	N/A	N/A	N/A
US EPA Regional Screening Levels (RSL)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Indicator/Source	C	G	R	D	RD	NAM	EDC	TK
[via US EPA ToxVal DB]								
US EPA ToxRefDB [via US EPA ToxVal DB]	Y	N/A	Y	Y	Y	N/A	N/A	N/A
US EPA ToxVal Database	Y	N/A	Y	Y	Y	N/A	N/A	N/A
EU COSMOS [via US EPA ToxVal DB]	Y	Y	Y	Y	Y	Y	Y	N/A
Genotoxicity & Carcinogenicity (ECVAM)	Y	Y	N/A	N/A	N/A	N/A	N/A	N/A
Genotoxicity (OASIS) [via OECD QSAR Toolbox]	N/A	Y	N/A	N/A	N/A	N/A	N/A	N/A
Genotoxicity pesticides (EFSA) [via OECD QSAR Toolbox]	N/A	Y	N/A	N/A	N/A	N/A	N/A	N/A
Health Assessment Workplace Collaborative (HAWC) Public Assessments [via US EPA ToxVal DB]	Y	Y	Y	Y	Y	Y	Y	N/A
Health Canada Toxicological Reference Values (TRVs) for Contaminated Sites [PDF]	Y	N/A	Y	Y	Y	N/A	N/A	N/A
Hazard Evaluation Support System (HESS) [via OECD QSAR Toolbox]	N/A	N/A	N/A	N/A	Y	N/A	N/A	N/A
NTP Technical Reports for Carcinogens	Y	N/A	N/A	N/A	N/A	N/A	N/A	N/A
OECD QSAR Toolbox	Y	Y	Y	Y	Y	Y	Y	Y
ToxCast / Tox21 Database	N/A	N/A	N/A	N/A	N/A	Y	N/A	N/A

C- Carcinogenicity; G- Genotoxicity, R- Reproductive Toxicity; D- Developmental Toxicity; RD- Repeated Dose Toxicity; NAM- New Approach Methodology Data; EDC- Endocrine Pathway related Data; TK- Toxicokinetics Data; Y – Yes (included in data set); N/A – not applicable

Table 8-2. Data sources and assigned relevance for the Product Count ruleset

Source	Relevance (score)
Health Canada Cosmetic Notifications	Very High (7)
SDS for products known to be available in Canada	High (5)
SDS for products likely to be available in Canada	Moderate (3)
US FDA Voluntary Cosmetic Registration Program	Moderate (3)
California EPA Cosmetic Registration Program	Moderate (3)
Consumer Product Information Database (CPID)	Moderate (3)
CPCat Database	Low (1)

Table 8-3. Data sources and assigned relevance for the Tonnage Band ruleset

Source	Use in children's products	Use in products	Relevance (score)
Health Canada Section 71 Notice	Yes	Yes	High (5)
Health Canada Section 71 Notice	Yes	No	High (5)
Health Canada Section 71 Notice	No	Yes	High (5)
Health Canada Section 71 Notice	No	No	Moderate (3)
US EPA Chemical Data Reporting (CDR)	Yes	N/A	Moderate (3)
US EPA CDR	No	N/A	Low (1)
ECHA REACH Registered Substances	N/A	Yes	Moderate (3)
ECHA REACH Registered Substances	N/A	No	Low (1)

Table 8-4. Data sources and assigned relevance for the Presence-Absence ruleset

Source	Relevance (score)
Alberta Biomonitoring Phase 1 - Serum of Pregnant Women Data	High (3)
Alberta Biomonitoring Phase 2 - Serum of Pregnant women and children (2004-2006)	High (3)
Asian/Pacific Islander Community Exposures (ACE) Project - ACE 1/ACE 2	Moderate (2)
Biomonitoring Exposures Study (BEST) - Pilot/Expanded study	Moderate (2)
California Childhood Leukemia Study (CCLS)	Moderate (2)
California Regional Exposure Study, Los Angeles County (CARE-LA)	Moderate (2)
California Teachers Study (CTS)	Moderate (2)
CCSPA List of interest from CMP3	Moderate (2)
Center for the Health Assessment of Mothers and Children of Salinas (CHAMACOS)	Moderate (2)
Children's Safe Products Act (Washington and Oregon State). High Priority Chemicals	Moderate (2)
CHMS Cycle 1 (2007–2009)	High (3)
CHMS Cycle 2 (2009-2011)	High (3)
CHMS Cycle 3 (2012–2013)	High (3)
CHMS Cycle 4 (2014-2015)	High (3)

Source	Relevance (score)
CHMS Cycle 5 (2016-2017)	High (3)
Cleaning Product Ingredient Inventory (Sept 2012)	Low (1)
Cleanright list of substances commonly used in soaps, detergents and maintenance products	Low (1)
Color Pigments Manufacturers Association (CPMA) DSL Categorization List	Low (1)
Consumer Product Survey Measured substances	Moderate (2)
Cosmetic Ingredient Review (CIR)	Moderate (2)
CPPDB Likely or Possibly in Plastics list	Low (1)
US EPA Suspect Screening Analysis of Chemicals in Consumer Products	Moderate (2)
EU Detergent Ingredient Database (DID)	Low (1)
EU Export Notices to Canada in accordance with Rotterdam convention	Low (1)
European Inventory of Cosmetic ingredients (COSING)	Moderate (2)
Firefighter Occupational Exposures (FOX) Project	Moderate (2)
First Nations Biomonitoring Initiative - National Results (2011)	High (3)
GerES IV (2003-2006)	Moderate (2)
IFRA Fragrance Ingredient List from 2022 Transparency List (Member volume/use survey)	Moderate (2)
Investigation of the composition and use of permanent make-up (PMU) inks in Australia	Low (1)
IPCHEM Human biomonitoring record in database	Low (1)
Kirk-Othmer List of substances used in cosmetics	Low (1)
List of Fragrance ingredients in Clorox products	Low (1)
Man-Made Chemicals in Human Blood	Moderate (2)
Markers of Autism Risk in Babies–Learning Early Signs (MARBLES)	Moderate (2)
Maternal and early life exposure to phthalates: The Plastics and Personal-care Products use in Pregnancy (P4) study	High (3)
Maternal and Infant Environmental Exposure Project (MIEEP)	Moderate (2)
Measuring Analytes in Maternal Archived Samples (MAMAS)	Moderate (2)
MEEC Study	High (3)
MIREC Study	High (3)
NHANES (Fourth Report)	Moderate (2)
NPRI reported releases >100 kg in one or more years 2015-2020	Moderate (2)
RAPEX- Substances detected in consumer products alert system	Moderate (2)
SPIN DB (Potential human consumer exposure)- Highest Potential level- Very Probable	Moderate (2)
SPIN DB (Potential human consumer exposure)- Mid Potential level - Probable and Possible	Low (1)

Source	Relevance (score)
Study of Per- and Polyfluoroalkyl Substances (PFAS) in Clothing, Apparel, and Selected Children's Items	Moderate (2)
US Export Notifications to Canada (TSCA 12b)	Moderate (2)

8.2 Validation of HAWPr

8.2.1 Ability to identify substances with hazard potential

To assess the ability of HAWPr to identify substances with known hazard potential, a curated list of 226 substances that had been previously concluded as toxic under section 64c of CEPA was run through the system. The results indicated that 212 (94%) of previously identified toxic substances had flags at the low indicator level and above, with 204 (90%) being assigned flags at the moderate indicator level or higher. The workflow classified 5 (2%) and 9 (5%) of the toxic substances as non-priority and unknown, respectively. Upon review, it was noted that the unknown classified substances mainly consisted of petroleum-like substances, which are challenging to capture using automated data collection methods due to lack of representative structures or lack of knowledge regarding the key components driving the risk. Non-priority substances were generally identified as organometallic substances which are data poor and not amenable to the QSAR methods used in HAWPr. Moiety-based assessments (for example, lead-based compounds) were excluded from the validation exercise because the toxic call for many of the substances in the moiety are based on information from the parent substance (for example, lead salts are considered toxic due to possible release of lead). As a result, many substances are relatively data poor and would likely not rank highly in the workflow even though they are considered to meet the toxic definition as part of the moiety.

To further assess the ability to capture known hazardous substances beyond those under CEPA, 111 chemicals in ECHA's Annex XIV of REACH ("Authorisation List") which consists of Substances of Very High Concern (SVHC) were evaluated by the workflow. This exercise resulted in 88 (79%) SVHC substances being assigned a hazard flags at the low indicator level and above, dropping to 51 substances (46%) when looking at the moderate level and above. HAWPr classified 21 (19%) and 2 (2%) SVHC substances as unknown and low priority, respectively. Of the substances with the unknown classification, 13 (62%) were data-poor substances without defined structures making read-across and QSAR not possible.

When considering the 19 substances on the TSCA Low Priority list, the intent was to look at substances with fewer expected hazard flags to see if the system would return fewer captures in the hazard indicator bins. It was found that 8 (42%) substances were allocated in the moderate confidence bin or higher. This is higher than expected but indicates that the workflow may trend toward more conservative outcomes. As HAWPr

is only one component of the overall IRAP approach, it is preferred for the system to make conservative decisions rather than to omit substances of potential interest during the automated triage phase.

8.2.2 Ability to identify substances with potential direct exposure

The exposure validation process relied on validation sets of substances which had been previously reviewed for exposure. In keeping with the approach to the hazard validation, the principal metric in determining success of the approach was concordance between the HAWPr outcome and the known or expected outcomes.

8.2.2.1 Comparison of direct potential for exposure (DPE) evaluation results from rapid screening conducted under CMP with HAWPr results

Under the CMP, a number of substances were part of various rapid screening assessment initiatives. Rapid screening was an approach to identify possible sources of direct exposure to the people in Canada from substances that were not reported in use via mandatory notices under section 71 of CEPA or were reported at tonnages of less than 1000 kg/yr, and thereby not expected to pose risks from potential indirect exposures. If there were no sources of direct exposure identified as part of this assessment, the substances were considered as candidates for rapid screening. If a potential for DPE was identified, the substance was not considered a suitable candidate for rapid screening and was identified for further assessment.

Based on the context of rapid screening, the absence of a direct exposure indicator would lead to a conclusion of non-toxic under section 64 of CEPA. As a consequence, in rapid screening, the identification of direct exposure from even a single product that was considered to have the potential to be available to consumers in Canada, often via exhaustive manual searching for products and SDSs online, was sufficient to flag the substance as having the potential for direct exposure and for it to be disqualified from the rapid screening approach. The expectation is that the battery of exposure sources used by HAWPr is sufficiently large and comprehensive that it will provide comparable results to the manual DPE evaluations done under rapid screening in a much more efficient, transparent, and reproducible manner.

To adequately compare how HAWPr and DPE evaluations identified indicators of potential direct exposure, the focus was placed on comparing the results from the product count ruleset from HAWPr and the DPE results for identifying product and cosmetic uses. The results of the other 2 direct exposure indicator rulesets from HAWPr were not considered to represent an evaluation of comparable information between the 2 approaches for this validation exercise. Further, as rapid screening was a subset of CMP substances with no or low volumes, tonnage was not considered to be a meaningful point of comparison.

Any product count from HAWPr of 1 or greater was compared with the substances identified as having DPE indicators in products or cosmetics from rapid screening. Similarly, a product count occurrence of 0 was seen as agreeing with DPE results from

rapid screening if no direct exposures from products and/or cosmetics were identified via rapid screening assessments.

The results of this evaluation showed a high level of concordance between the 2 approaches. The results of HAWPr were compared with the DPE results for 1143 substances evaluated as part of rapid screening initiatives and demonstrated that 997 (87%) substances had agreement between both approaches in their identification of direct exposure indicators. Indicators for direct exposure were identified for 108 (9%) of all 1143 substances, with 889 (78%) substances demonstrating concordance in not identifying an indicator for DPE. The larger proportion of substances not having evidence of the potential for direct exposure is consistent with this subset of substances identified as candidates for rapid screening - that being substances presumed to have no or minimal commercial use based on responses received from mandatory surveys under CEPA.

For the approximately 13% (146) of substances that did not show concordance between HAWPr and rapid screening DPE, these were substances where rapid screening did identify a potential direct exposure via product or cosmetic use, but HAWPr did not (at least with respect to an exposure indicator from the product count ruleset). The main drivers of this difference were related to the substance-by-substance search for function and use data as part of rapid screening that identified potential DPE indicators from relatively obscure sources, or informed by uses and manufactures associated with the substances via Section 71 submissions; these information sources are not currently accessible within HAWPr. These DPE indicators from rapid screening accounted for 71 of the 119 substances found to lack concordance between the 2 approaches. The remaining 48 outliers were identified in rapid screening as being used as fragrances in cosmetics and products. That information is not readily available in HAWPr in the product count ruleset, as fragrance ingredients are normally not explicitly listed in SDSs or in cosmetic notifications received due to concerns with proprietary information on formulations. Frequently, ingredient listings in SDSs or cosmetic notifications simply state "Fragrance", "Parfum" etc., with corresponding product concentration, without substance-specific information related to the fragrance composition. This is a limitation of HAWPr that will need to be considered for future improvements.

There were also 27 substances that were flagged with exposure indicators for direct exposure by HAWPr, that were not identified as part of their rapid screening DPE evaluation. These primarily represent changes in the data and use over time, or situations in which both approaches identified possible indicators of DPE but the further manual evaluation under rapid screening may have resulted in a determination that exposure to people in Canada was unlikely. This level of in-depth, substance-by-substance evaluation of the exposure information is not within the scope of HAWPr but will be considered if the substance is brought forward in IRAP.

As noted previously, the results of the validation exercises show a high level of concordance between the results of the 2 approaches. The 13% of substances that had

results that were not well aligned were primarily related to differences in the context and conservatism of the 2 approaches, as well as the in-depth substance-by-substance searching for additional indicators of DPE that was undertaken as part of rapid screening assessments.

8.2.3 Overall system validation using IRAP 2019

An approach was developed to compare the results of both the hazard and exposure components of the HAWPr system, with the corresponding results from the IRAP 2019 review cycle. The IRAP 2019 results were conducive to an in-depth evaluation, as the process to prioritize substances based on exposure and hazard indicators share many common data sources with HAWPr, coupled with more recent data.

8.2.3.1 System validation using IRAP 2019 - exposure

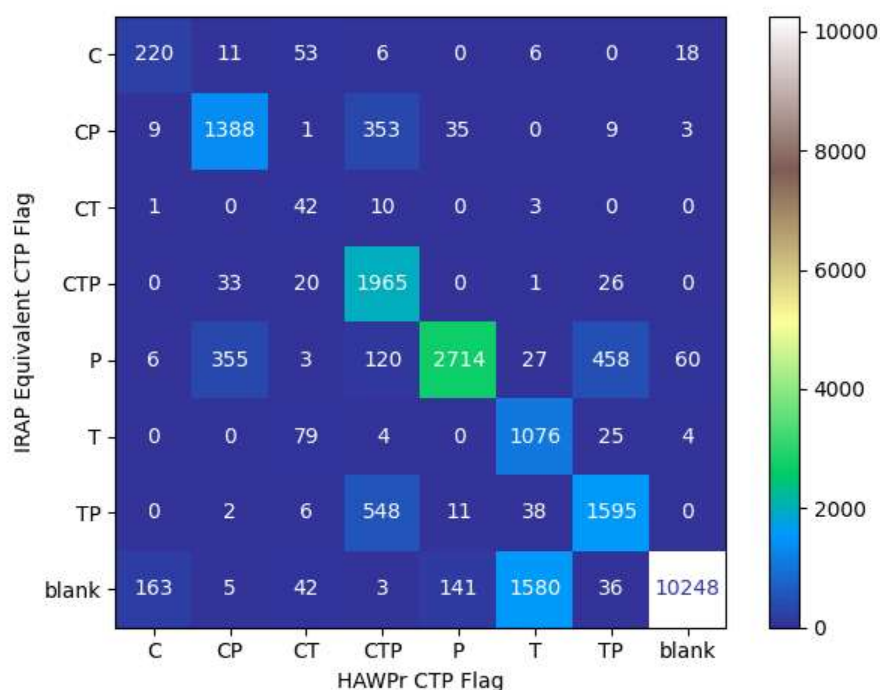
The process in which direct exposure was flagged for IRAP 2019 was binary (Yes/No), and therefore did not allow for a method to evaluate the extent of potential direct exposure in the way HAWPr allows. For example, a higher number of SDSs or cosmetic notifications would result in a higher score from the product count ruleset in HAWPr, which could lead to assignment in a Priority 1 (high) exposure bin. The equivalent did not exist from IRAP results, where the actual count of products did not impact the potential for direct exposure. The substance either had a potential for direct exposure based on indicators from these product count sources, or it did not. Consequently, this limited the comparison of the IRAP exposure outcomes directly to results from HAWPr.

To address this issue and allow comparison between the 2 approaches, data from the IRAP 2019 exposure review was manually reviewed and assigned CTP flags, following the logic used in HAWPr for the Product Count (C), Tonnage (T) and Presence-Absence (P) rulesets. For instance, the use of cosmetic notifications in IRAP was aligned with the use of this information to inform the product count ruleset for HAWPr, and therefore, if present, counted as a C flag in the IRAP review. Similarly, quantity information reported via Section 71 that was used in IRAP was treated as a potential T flag (tonnage ruleset in HAWPr). This allowed direct quantitative comparison between flags identified in the original IRAP approach and those identified using the HAWPr approach. Of particular interest was the identification of indicators of potential direct exposure that IRAP 2019 may have captured that may not be well captured by HAWPr.

The results of the comparison of CTP flags between approaches depicted as a multi-class confusion matrix is illustrated in Figure 8-1. The on-diagonal elements indicate agreement between the 2 methods, while non-zero off-diagonal elements indicate some disparity. Further investigation into the composition of these disparities found that the majority of these off-diagonals are not of concern as they merely reflect differences in the data sources used. For example, HAWPr improved on the ability to utilize SDSs to inform the product count ruleset, a source not readily available in IRAP 2019. A key finding in this validation can be seen in the bottom row of Figure 8-1. This row indicates HAWPr returned flags, whereas IRAP 2019 did not. This demonstrates that HAWPr has increased coverage of the available data allowing for a comprehensive review of

indicators of direct exposure. The bottom right white box in Figure 8-1 reflects where there was an absence of any flags for a substance(s) from all of the indicators considered from both approaches. Effectively, this is the subset of substances that are data poor with respect to information on exposure from the sources identified by both approaches. Previous program experience has found that, in many cases, this scenario is indicative of negligible to no commercial use of a substance; however, there are situations where certain types of uses, such as presence in manufactured items or plastics, are not well captured by current exposure indicators readily available for use in approaches such as HAWPr.

Figure 8-1. Multi-class confusion matrix for IRAP equivalent CTP flag versus HAWPr CTP flag



[Long description: This figure depicts the results of the comparison of CTP flags between approaches as a multi-class confusion matrix. The x-axis is for the various iterations of CTP flags from HAWPr, for example, all 3 CTP flags, or only C, only C and T, and so on. The y-axis shows the iterations of the equivalent CTP flags from IRAP. The on-diagonal elements indicate agreement between the 2 methods, while non-zero off-diagonal elements indicate some disparity. The white box in the bottom right of the matrix reflects where there was an absence of any flags for a substance(s) from all of the indicators considered from both approaches.]

Assuming the IRAP equivalent CTP flag represents the true-positive case and HAWPr CTP flags represent the prediction, then the precision, recall, and F1-score may be determined for each flag as well as for the overall statistics (see Table 8-5). Low precision for CT and T is due to differences in data sources between the 2 methods, especially the introduction of new data in the case of the tonnage ruleset. The overall accuracy of 0.82, combined with the weighted average precision, recall, and F1-score being 0.82 or greater indicate good concordance between the methods.

Table 8-5. Precision, recall, and F1-score for each flag, as well as overall and weighted average, assuming the IRAP 2019 equivalent CTP flag is the true-positive case.

	Precision^a	Recall^b	F1-score^c	Support^d
C	0.55	0.7	0.62	314
CP	0.77	0.77	0.77	1798
CT	0.17	0.75	0.28	56
CTP	0.65	0.96	0.78	2045
P	0.94	0.73	0.82	3743
T	0.39	0.91	0.55	1188
TP	0.74	0.72	0.73	2200
blank	0.99	0.84	0.91	12218
Accuracy^e	N/A	N/A	0.82	23562
Average	0.65	0.8	0.68	23562
Weighted average	0.88	0.82	0.83	23562

^a Precision: the ability of the classifier not to label a negative sample as positive. This is calculated by dividing the number of true positives (the number of predictions where the classifier correctly predicts the positive class as positive) by the sum of true positives and false positives (the number of predictions where the classifier incorrectly predicts the negative class as positive).

^b Recall: the ability of the classifier to find all the positive samples. This is calculated by dividing the number of true positives by the sum of true positives and false negatives (the number of predictions where the classifier incorrectly predicts the positive class as negative).

^c F1-score: the harmonic mean of the precision and recall. This is equal to 2 times the precision x recall divided by the precision plus recall.

^d Support: the number of occurrences of each class for the true case

^e Accuracy: the fraction of true positives and true negatives (the number of predictions where the classifier correctly predicts the negative class as negative) found amongst all classifications. This is calculated as the sum of true positives and true negatives divided by the total number of predictions.

8.2.3.2 System validation using IRAP 2019 - hazard

Like exposure, the hazard prioritization of substances in IRAP 2019 was ultimately a binary (Yes/No) decision on high hazard potential. This was primarily because the main sources of data used to inform hazard indicators for IRAP were international classifications for CMR, EDC and repeated dose toxicity. While it is possible to compare the substances across the DSL to identify if the hazard identified by HAWPr aligns with a high hazard flag from IRAP 2019, it was expected that there would be good alignment between the substances flagged for hazard between the 2 systems, given that HAWPr incorporates the hazard indicators used previously by IRAP. This was reaffirmed by observing that 85% of the substances IRAP 2019 flagged as high hazard were assigned high confidence based on hazard flags in HAWPr. This represents a good alignment between approaches for substances with high hazard indicators from IRAP 2019, also having high hazard scores from HAWPr. In addition, it was observed that HAWPr has greatly improved the amount of hazard information considered over previous IRAP reviews. A powerful illustration of this was the identification of approximately 2900

substances assigned either moderate or high hazard scores by HAWPr that did not have high hazard flags from the IRAP 2019 review.

There were over 22,000 substances on the DSL that had hazard prioritization results from both IRAP 2019 and HAWPr to compare. From that number, 571 substances had a high hazard flag from the IRAP 2019 review. By comparison, HAWPr identified 12,839 substances that had sufficient information to fall within hazard indicator bins (low-moderate-high). Given the considerable advancements by HAWPr to capture more data beyond the more traditional high hazard classification data that IRAP 2019 relied on, this high number of substances with information upon which to bin hazard is not unexpected. It also does provide clear evidence of the magnitude of information that is now being brought to bear in the prioritization decision making workflow for hazard relative to previous prioritization exercise(s) undertaken by Health Canada.

8.2.3.3 System validation using IRAP 2019 - overall outcomes

As IRAP 2019 results reflect decisions made as a result of subsequent triaging and scoping work done as part of that initiative, the validation of overall outcomes from IRAP 2019 versus HAWPr has some limitations. For example, the IRAP determinations of No Further Work incorporate the outcomes of expert judgement at the further triaging and scoping phase of IRAP, whereas other elements such as the program status (for example, considered in previous CMP assessment and concluded non-toxic) may have impacted the final decision on a prioritization path forward. As a result, the respective outputs of each systems' respective hazard and exposure evaluation provide a better comparison on a substance-by-substance basis of the 2 approaches to screen for priorities. As HAWPr is just one feeder into the IRAP process, the expectation is that HAWPr should provide a more conservative outcome that can be refined further as a substance progresses through subsequent stages of the IRAP process. The comparison of overall outcomes from IRAP 2019 and HAWPr is limited to the identification of any critical areas of improvement for HAWPr.

Rather than a global comparison of the alignment of outcomes, the analysis has been divided by outcome type. This better facilitates a consideration of the nuances of the various IRAP outcomes, as well as an appreciation of where HAWPr is identifying substances as potential priorities that were not captured in the previous IRAP review. This review will focus specifically on the No Further Work and Problem Formulation outcomes from IRAP 2019. Other outcomes, such as exposure or hazard Data Gathering or International Activity, represent such a large breadth of scenarios and contexts, frequently relying on input from information outside the HAWPr system (for example, consideration of ongoing evaluation of substances/groups by international jurisdictions), that comparison between the outputs of both systems is not appropriate.

8.2.3.3.1 IRAP 2019 no further work (NFW) outcomes

The IRAP 2019 review identified approximately 20,500 substances that did not have sufficient evidence for further work at this time (hereafter referred to as NFW). The breakdown of the comparison of results of HAWPr with those substances with NFW

outcomes from IRAP 2019. Overall, these results show that both approaches show good alignment in the identification of low priority substances for approximately 91% of substances. For the 91% (approximately 18,800) of substances with good alignment in results, these had overall priority results from HAWPr of moderate-low or below that is considered consistent with an overall low priority outcome from IRAP 2019.

A further approximately 930 substances (5%) that were determined to have insufficient evidence for prioritization and assigned an outcome of NFW from the IRAP 2019 review had results from HAWPr that were seen as being in general agreement with IRAP. For these substances, the overall outcomes are the same, but the combinations of hazard and exposure indicators may have differed. This subset of approximately 930 substances comprised of those that were found to not have exposure indicators from both systems, but HAWPr had identified a potential hazard. For an additional approximately 330 substances (2%), these can be regarded as ones which HAWPr identified with exposure indicators as priorities that IRAP 2019 did not capture because hazard flags were not identified by that process.

The remaining approximately 450 substances were ones for which the outcomes of HAWPr and IRAP 2019 did not align. This final subset had a mix of rationales for the differences between the IRAP 2019 and HAWPr results. The reasons for the differences ultimately were found to come down to additional evaluation of the available flags that occurred as part of the triaging and further scoping stage of IRAP 2019 (for example, risk management measures already in place).

8.2.3.3.2 IRAP 2019 problem formulation outcomes

A comparison was also undertaken as part of this validation exercise of the results of HAWPr for the subset of 67 substances from IRAP 2019 that had Problem Formulation and further scoping outcomes. A little under 65% (43 of 67) of these substances were also identified by HAWPr as having moderate or higher overall priority. All these substances were identified by Health Canada as human health priorities and represent an area where the results of the 2 approaches aligned. The reasons why the remaining 24 of 67 substances were not identified as priorities by HAWPr was investigated to better understand areas HAWPr could be improved, or where the role of additional triaging in IRAP, post-HAWPr, may have been a factor. Of the 24 remaining substances, 20 were the result of adding group members based on read-across or expert judgement during the further scoping/problem formulation step of IRAP. The remaining 4 substances were identified as ecological priorities for problem formulation by Environment and Climate Change Canada. Health Canada identified these substances as having direct exposure indicators in the IRAP 2019 review but had no human health hazard indicators to warrant prioritization. Overall, the results of HAWPr are consistent with IRAP 2019, and the majority of discrepancies identified reflect follow-up work undertaken in IRAP 2019 to further scope, triage, cluster and group substances post-screening for the presence of hazard and exposure indicators.

8.3 Detailed results of HAWPr

Detailed results from the HAWPr workflow are available in a spreadsheet provided as a supporting document to this Science Approach Document.