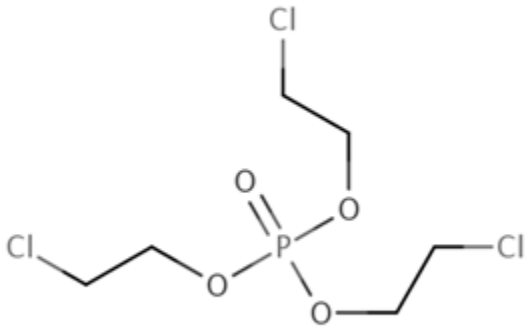
United States
Environmental Protection Agency

December 2023
Office of Chemical Safety and
Pollution Prevention

Draft Risk Evaluation for Tris(2-chloroethyl) Phosphate (TCEP)

Supplemental File:

Systematic Review Protocol for the Draft Risk Evaluation for Tris(2-chloroethyl) Phosphate (TCEP) CASRN: 115-96-8



December 2023

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

The U.S. EPA's Office of Pollution Prevention and Toxics (OPPT) applies systematic review principles in the development of risk evaluations under the amended Toxic Substances Control Act (TSCA) . TSCA section 26(h) requires EPA to use scientific information, technical procedures, measures, methods, protocols, methodologies, and models consistent with the best available science and base decisions under section 6 on the weight of the scientific evidence. Within the TSCA risk evaluation context, the weight of the scientific evidence is defined as "a systematic review method, applied in a manner suited to the nature of the evidence or decision, that uses a pre-established protocol to comprehensively, objectively, transparently, and consistently identify and evaluate each stream of evidence, including strengths, limitations, and relevance of each study and to integrate evidence as necessary and appropriate based upon strengths, limitations, and relevance" (40 CFR 702.33).

To meet the TSCA section 26(h) science standards, EPA used the TSCA systematic review process described in the *Draft Systematic Review Protocol Supporting TSCA Risk Evaluations for Chemical Substances* (U.S. EPA, 2021) (also referred to as "2021 Draft Systematic Review Protocol"). Figure 1-1 Section 3 of the 2021 Draft Systematic Review Protocol depicts the steps in which information is identified and whether it undergoes the formal systematic review process (U.S. EPA, 2021). Information attained via the systematic review process is integrated with information attained from sources of information that do not undergo systematic review (*e.g.*, EPA-generated model outputs) to support a weight of the scientific evidence analysis.

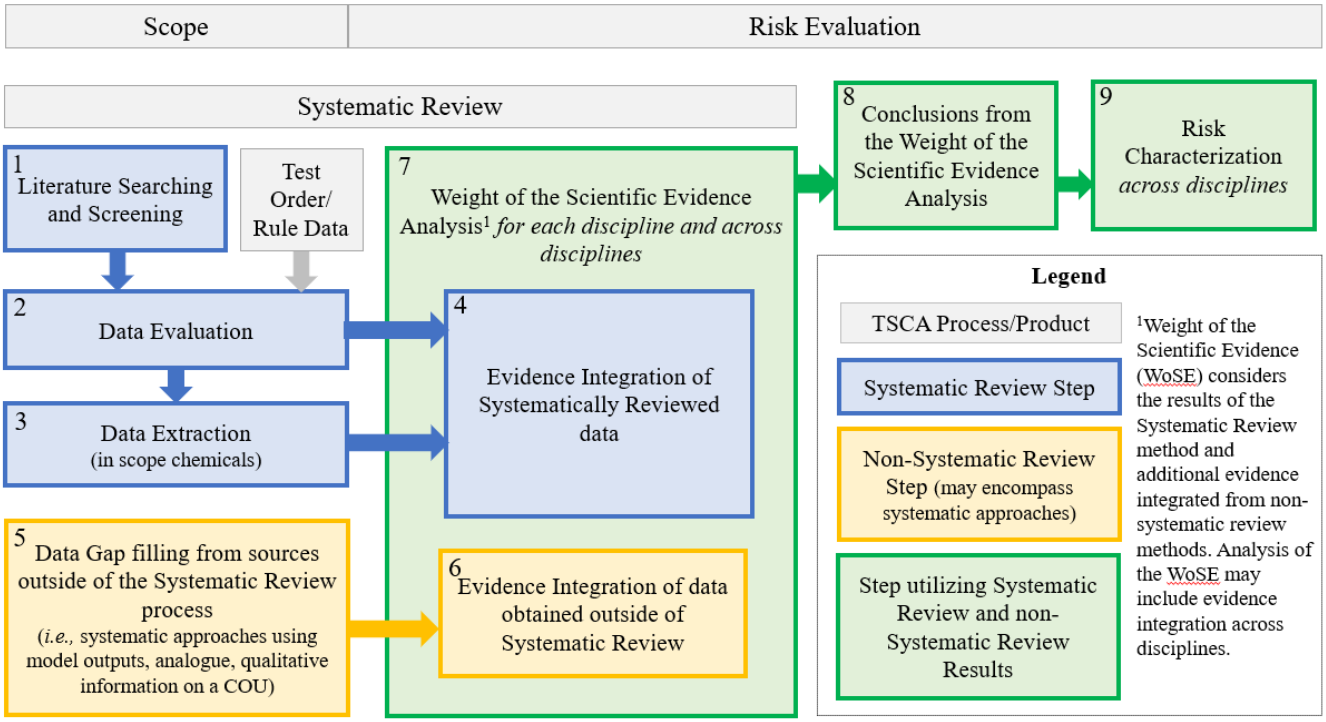


Figure 1-1. Overview of the TSCA Risk Evaluation Process with Identified Systematic Review Steps

The process complements the risk evaluation process in that it is used to develop the exposure and hazard assessments based on reasonably available information. EPA defines "reasonably available information" to mean information that EPA possesses or can reasonably obtain and synthesize for use in risk evaluations, considering the deadlines for completing the evaluation (40 CFR 702.33).

2 CLARIFICATIONS AND UPDATES TO THE 2021 DRAFT SYSTEMATIC REVIEW PROTOCOL

In 2021, EPA released the *Draft Systematic Review Protocol Supporting TSCA Risk Evaluations for Chemical Substances* (U.S. EPA, 2021), a framework of systematic review approaches under TSCA, to address comments received on a precursor systematic review approaches framework, the *Application of Systematic Review in TSCA Risk Evaluations* (U.S. EPA, 2018). In April 2022, the SACC provided comments on the 2021 Draft Systematic Review Protocol and additional comments on OPPT's systematic review approaches were garnered during the public comment period. In lieu of an update to the 2021 Draft Systematic Review Protocol, this systematic review protocol for the *Draft Risk Evaluation for Tris(2-chloroethyl) Phosphate (TCEP)* describes some clarifications and different approaches that were implemented than those described in the 2021 Draft Systematic Review Protocol in response to (1) SACC comments, (2) public comments, or (3) to reflect chemical-specific risk evaluation needs.

2.1 Clarifications

Throughout the 2021 Draft Systematic Review Protocol, there were some terms used that were not explicitly defined, resulting in inconsistent uses within the document (U.S. EPA, 2021). Table 2-1 Table 2-1 lists the terms that were updated to resolve some of the confusion expressed by the public and SACC comments regarding the implementation of the respective systematic review-related step. One main clarification is that *all references that undergo systematic review are considered for use in the risk evaluation*, even those that do not meet the various discipline and sub-discipline screening criteria or those that are categorized as supplemental information at title and abstract (TIAB) or full-text screening.

Section 4.2.5 of the 2021 Draft Systematic Review Protocol describes how data sources (*e.g.*, individual references, databases) may be tagged and linked in epidemiological cohort studies when information is present in multiple studies (U.S. EPA, 2021). References will generally undergo data quality evaluation and extraction if there are data that pass screening criteria; however, to prevent the same data from being represented multiple times and conflating the amount of available information there is on a subject area, EPA selects the reference(s) that most appropriately describes the extractable results (indicated as the parent reference in DistillerSR). For example, if two references portray the same information from the same dataset, only one is counted in the overall dataset (*i.e.*, deduplication). If two references contain information about the same dataset, but only provides additional contextual information or summary statistics (*e.g.*, mean), both data sources are linked but the extractable information from both may be combined in DistillerSR. This allows the capture of key information while avoiding double counting the data of interest, to capture key information while avoiding double counting the data of interest, which may be the case whether one reference contains original or extractable data that passes screening criteria. The linked reference containing most of the data, which are evaluated and extracted, is identified in DistillerSR as the parent reference; the "complementary child reference" in DistillerSR does not undergo data evaluation and extraction. Linking the references in DistillerSR allows the reference with more limited information or only contextual information to be tracked and utilized to evaluate the extracted data in the other related studies. The child reference may undergo data quality evaluation and extraction if there are additional unique and original data that pass screening criteria. One clarification is that this procedure of identifying potential duplicative information *applies to all information that is considered in a risk evaluation under TSCA (not just epidemiological cohort studies)*. Also, this procedure may apply when there is duplicative information in two references even if it is more than just "contextual."

Section 4.5 of the 2021 Draft Systematic Review Protocol describes how data may be obtained using TSCA authorities and test orders. One update to that section is that in addition to requiring data reporting under TSCA sections 4 (test order), 8(a) (Chemical Data Reporting) and 8(d) (Health and Safety Data Reporting), *EPA may also require data reporting under TSCA section 8(c)* (Call-in of Adverse Reactions Records). Appendix 5.3 also describes how information may be submitted to EPA under other TSCA authorities (*e.g.*, TSCA sections 4, 5, 6, 8(d) and 8 (e), as well as FYI submissions).

Section 5 of the 2021 Draft Systematic Review Protocol describes how EPA conducts data quality evaluation of data/information sources considered for a respective chemical risk evaluation, with Section 5.2 specifically explaining the terminology used to describe both metric and overall data/information source quality determinations (U.S. EPA, 2021). To respond to both SACC and public comments regarding the inappropriate use of quantitative methodologies to calculate both "Metric Rankings" and "Overall Study Rankings," *EPA decided to not implement quantitative methodologies to attain either metric and overall data/information source quality determinations* and therefore updated the terminology used for both metric ("Metric Ranking") and overall data/information source ("Overall Study Ranking") quality determinations (Table 2-1). Specifically, metric and overall data/information source quality determination terminology have been updated to "Metric Rating" and "Overall Quality Determination", respectively. The word "level" was also often used synonymously and inconsistently with the word "ranking" in the 2021 Draft Systematic Review Protocol; that inconsistency has been rectified, resulting in the word "level" no longer being used to indicate either metric or overall data/information source quality determinations (U.S. EPA, 2021).

Sections 4.3.2.1.3 and 6 of the 2021 Draft Systematic Review Protocol describe when EPA may reach out to authors of data/information sources to obtain raw data or missing elements that are important to support the data evaluation and data integration steps (U.S. EPA, 2021). In such cases, the request(s) for additional data/information, number of contact attempts, and responses from the authors are documented. EPA's outreach is considered unsuccessful if those contacted do not respond to email or phone requests within 1 month of initial attempt(s) of contact. One important clarification to this guidance is that *EPA may reach out to authors anytime during the systematic review process for a given data/information source or reference, and that contacting authors does not explicitly happen during the data quality evaluation or extraction steps*

Table 2-1. Terminology Clarifications between the 2021 Draft Systematic Review Protocol and the Draft Risk Evaluation for Tris(2-chloroethyl) Phosphate (TCEP)

2021 Draft Systematic Review Protocol Term	TCEP Systematic Review Protocol Term Update	Clarification
"Title and abstract" or "Title and abstract" or "title/abstract"	"Title and abstract"	To increase consistency, the term "title and abstract" will be used to refer to information specific to "title and abstract" screening.
Variations of how "include," "on topic" or "PECO[a]/PESO[b]/RESO[c] relevant" implied a reference was considered for use in the risk evaluation, whereas "exclude," "off topic" or "not	Meets/does not meet PECO/PESO/RESO screening criteria	The term "include" or "exclude" falsely suggests that a reference was or was not, respectively, considered in the risk evaluation. There was also confusion regarding whether "on topic" and "PECO/PESO/RESO relevant" were synonymous and suggested those references were explicitly considered for use in the risk evaluation (and by default, "off topic" and "not PECO/PESO/RESO relevant" references were not). References that meet

2021 Draft Systematic Review Protocol Term	TCEP Systematic Review Protocol Term Update	Clarification
PECO/PESO/RESO relevant" implied a reference was *not* considered for use in the risk evaluation.		the screening criteria proceed to the next systematic review step; however, all references that undergo systematic review at any time are considered in the risk evaluation. Information that is categorized as supplemental or does not meet screening criteria are generally less relevant for quantitative use in the risk evaluation but may be considered if there is a data need identified. For instance, mechanistic studies are generally categorized as supplemental information at either title and abstract or full-text screening steps but may undergo the remaining systematic review steps if there is a relevant data need for the risk evaluation (*e.g.*, dose response, mode of action).
Database source not unique to a chemical	Database	Updated term and definition of "Database": Data obtained from databases that collate information for the chemical of interest using methods that are reasonable and consistent with sound scientific theory and/or accepted approaches and are from sources generally using sound methods and/or approaches (***e.g.,*** state or federal governments, academia). Example databases include STORET and the Massachusetts Energy and Environmental Affairs Data Portal. The term in the 2021 Draft Systematic Review Protocol (Table_Apx N-1) incorrectly suggested that databases that contain information on a singular chemical are not considered (U.S. EPA, 2021). Furthermore, the wording "large" was removed to prevent confusion and the incorrect suggestion that there is a data size requirement for databases that contain information that may be considered for systematic review.
Metric Ranking or Level	Metric Rating	As explained above, EPA is not implementing quantitative methodologies to indicate metric quality determinations, therefore the term "ranking" is inappropriate. The term "level" was inconsistently used to indicate metric quality determinations previously, therefore EPA is removing the use of this term to reduce confusion when referring to metric quality determinations. The term "Rating" is more appropriate to indicate the use of professional judgement to determine a quality level for individual metrics.
Overall Study Ranking or Level	Overall Quality Determination (OQD)	As explained above, EPA is not implementing quantitative methodologies to indicate overall data/information source quality determinations, therefore the term "ranking" is inappropriate. The term

2021 Draft Systematic Review Protocol Term	TCEP Systematic Review Protocol Term Update	Clarification
		"level" was inconsistently used to indicate overall data/information source quality determinations previously, therefore EPA is removing the use of this term to reduce confusion when referring to overall data/information source quality determinations. The term "Rating" is more appropriate to indicate the use of professional judgement to determine a quality level for the overall data/information source quality determination.
Sub-discipline	No change in term	Sub-discipline explicitly indicates the two categories of receptor-based studies relevant to evaluate human health hazard (discipline): epidemiological (human receptor) or human health animal model toxicological studies (non-human animal receptor). Although environmental hazard is a discipline, Appendix T incorrectly suggested that environmental hazard is a sub-discipline in the 2021 Draft Systematic Review Protocol.
Evidence Stream	No change in term	Evidence streams were updated for both environmental and human health hazard disciplines to more appropriately categorize the hazardous endpoints that were considered. Please see additional descriptions of the evidence stream updates in Section 6.5 below.

[a] "PECO" stands for Population, Exposure, Comparator or Scenario, and Outcomes.
[b] "PESO" stands for Pathways or Processes, Exposure, Setting or Scenario, and Outcomes.
[c] "RESO" stands for Receptors, Exposure, Setting or Scenario, and Outcomes.

3 DATA SEARCH

As described in Section 4 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021), EPA conducts a comprehensive search for reasonably available information to support the TSCA risk evaluations. Chemical-specific literature searches are conducted as described in Section 4.2.1 of the 2021 Draft Systematic Review Protocol for all disciplines (*i.e.*, physical and chemical properties, fate, engineering, exposure, environmental hazard, and human health hazard) (U.S. EPA, 2021). Additional details on the chemical verification process, and the methodology used to search for chemical-specific peer-reviewed and gray literature are available in Sections 4.2 and 4.3 of the 2021 Draft Systematic Review Protocol, respectively (U.S. EPA, 2021). The search for peer-reviewed and gray literature relevant references was completed in September and May 2019, respectively. Appendix C.1.12 contains the specific search strings used to identify peer-reviewed literature on TCEP (U.S. EPA, 2021). All reasonably available information submitted to EPA under TSCA authorities was considered.

3.1 Multi-Disciplinary Updates and Clarifications to the Data Search

For the *Draft Risk Evaluation for Tris(2-chloroethyl) Phosphate (TCEP)*, the literature search was conducted as described in Section 4 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021), where the peer-reviewed and gray literature updated search followed the approach outlined in Sections 4.2 and 4.3 of the 2021 Draft Systematic Review Protocol, respectively (U.S. EPA, 2021). Occasionally additional data sources relevant for the risk evaluation may be identified after the initial search for peer-reviewed and gray literature; these data sources will then undergo systematic review for the relevant discipline(s). Additionally, each discipline utilizes different strategies (*e.g.*, search strings) to attain their discipline-specific pools of data sources that undergo systematic review

SWIFT-Review Validation

EPA received comments regarding the lack of detail on the use and validation of SWIFT-Review to determine discipline-specific peer-reviewed reference set considered for use in TSCA risk evaluations. In response to those comments, EPA conducted validation exercises to clarify the search process and build consistency among all the disciplines. The 2021 Draft Systematic Review Protocol contains validation results for the use of SWIFT-Review to determine which peer-reviewed references may be relevant for the characterization of occupational exposure and environmental releases and general population, consumer, and environmental exposure for the respective chemical risk evaluations. However, to expand upon the information provided in the 2021 Draft Systematic Review Protocol, EPA validated references relevant for determining chemical-specific peer-reviewed reference set for the characterization of physical and chemical properties, environmental fate properties, and environmental and human health hazard. EPA manually screened the references that were found in the overall peer-reviewed search results that did not undergo TIAB screening (*i.e.*, references that were not identified using a discipline-specific search string). If a reference that did not undergo further review after TIAB screening was found to meet the screening criteria for a respective discipline (*e.g.*, data needs on physical chemical properties, environmental fate and transport properties, and environmental and human health hazard) and identified for the chemical of interest, it was flagged as a false negative. This analysis validated and verified the use of the search terms in SWIFT-Review, as it showed that less than 5 percent of references were false negatives across all three disciplines. This method was repeated for several of the TSCA High Priority Substances to build confidence in our discipline-specific search strings.

Supplemental Search for Dermal Absorption Data

Dermal absorption studies are needed to accurately assess dermal exposure resulting from different exposure routes associated with specific conditions of use. However, dermal absorption data may not

meet screening criteria for any discipline (Appendix H of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021)). To identify additional studies potentially relevant for characterizing dermal absorption and exposure, EPA searched for data sources that were found in the chemical-specific search (*e.g.*, peer-reviewed, gray and those submitted under various TSCA sections (clarified above in Section 2)) using the search strings in Section 3.7. As described in Section 4.2.2 for supplemental searching and in conjunction with Sections 4.2.4 and 4.2.5 in the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021), search strings (listed below in Section 3.7.1) were used in SWIFT-Review to identify references that are predicted to be the most relevant for evaluating dermal exposure to TCEP.

Additional Gray Literature Sources

Physical and Chemical Properties: In addition to the gray literature sources listed in Appendix E of the 2021 Draft Systematic Review Protocol, an additional database was added to the list of gray literature sources for physical and chemical properties. The National Institutes for Standards and Technology (NIST) Chemistry Webbook was searched in September 2021 to capture spectroscopic data, specifically ultra-violet and visible absorption (UV-Vis) data, if recorded. This source may also provide thermodynamic data that informs chemical stability and behavior under various conditions.

General Population, Consumer, and Environmental Exposure: In addition to the gray literature sources listed in Appendix E of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021), two additional sources were added in January 2023, to capture database outputs from the National Health and Nutrition Examination Survey (NHANES) and Water Quality Portal (WQP).

Because the literature set for many chemicals, including tris(2-chloroethyl) phosphate, includes a record from EPA's retired STOrage and RETrieval (STORET) database, EPA downloaded all the TCEP data from the WQP, the successor database that now contains data from STORET. This data was uploaded into HERO and added to the literature set that is considered for systematic review. In addition, to obtain information on TCEP exposures to the U.S. population, EPA added data from the Centers for Disease Control and Prevention's National Health and Nutrition Examination Survey (NHANES) to its literature set. Although NHANES did not contain relevant information on TCEP, EPA did identify potentially relevant information on its primary metabolite, bis(2-chloroethyl) phosphate (BCEP). After entering the human body, TCEP is typically metabolized into BCEP in various receptors; Section 3.5 further describes the metabolism of TCEP into BCEP. NHANES data on BCEP was also evaluated as part of the systematic review process for data on general population, consumer, and environmental exposure. At the time of download, the three tables available from CDC included "Analysis of Whole Blood, Serum, and Urine Samples, NHANES 1999-2018," "Analysis of Pooled Serum Samples for Select Chemicals, NHANES 2005-2016," and "Analysis of Chemicals Found in Cigarette Smoke in a Special Sample of U.S. Adults, NHANES 2011-2016." Of these, the only dataset containing BCEP data was "Analysis of Whole Blood, Serum, and Urine Samples, NHANES 1999-2018." Similar to the WQP data, the relevant NHANES data were also uploaded into HERO.

3.2 Physical and Chemical Properties

The search for peer-reviewed and gray literature are as described in Sections 4.2 and 4.3, respectively, in the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). Specifically, SWIFT-Review was used to identify peer-reviewed references that are predicted to be the most relevant for evaluating physical and chemical properties for TCEP. The search string used for physical and chemical properties in SWIFT-Review was developed by EPA's Office of Research and Development (ORD) in collaboration with Sciome and is presented in Appendix G-1, Table_Apx G-1 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). As mentioned above in Section 3.1, the search string used to identify potentially relevant peer-reviewed data references for evaluation of the physical and chemical properties of TCEP were

validated. When the search string terms are identified in the title, abstract, or keywords of a given reference in SWIFT-Review, those references proceed with TIAB screening.

3.3 Environmental Fate and Transport Properties

The peer-reviewed and gray literature searches for environmental fate and transport properties are as described in Sections 4.2 and 4.3, respectively, in the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). Specifically, SWIFT-Review was used to identify peer-reviewed references that are predicted to be the most relevant for evaluating environmental fate and transport properties for TCEP. The search string used for environmental fate and transport literature in SWIFT-Review was developed by EPA's ORD in collaboration with Sciome and is presented in Appendix G.2, Table_Apx G2 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). As mentioned above in Section 3.13.1, the search string used to identify potentially relevant peer-reviewed data references for evaluation of the environmental fate and transport properties of TCEP were validated. When the search string terms are identified in the title, abstract, or keywords of a given reference in SWIFT-Review, those references proceed with TIAB screening.

3.4 Environmental Releases and Occupational Exposure

The peer-reviewed and gray literature searches for environmental releases and occupational exposure are as described in Sections 4.2 and 4.3, respectively, in the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). Specifically, SWIFT-Review was used to identify peer-reviewed references that are predicted to be the most relevant for evaluating environmental release and occupational exposure for TCEP. As described in Sections 4.2.4.2 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021), EPA identified on-topic and off-topic references from the broad search results of the TCEP peer-reviewed literature as positive and negative "seeds" to classify which references contained environmental release and occupational exposure to prioritize for further review. When the relevant references were identified in SWIFT Review, those references proceeded with title and abstract screening.

3.5 General Population, Consumer, and Environmental Exposure

The peer-reviewed and gray literature searches for general population, consumer, and environmental exposure are as described in Sections 4.2 and 4.3, respectively, in the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). Specifically, SWIFT-Review was used to identify peer-reviewed references that are predicted to be the most relevant for evaluating general population, consumer, and environmental exposures to TCEP. As described in Sections 4.2.4.2 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021), EPA identified on-topic and off-topic references from the broad search results of the peer-reviewed literature as positive and negative "seeds" to classify which references on general population, consumer, and environmental exposures to prioritize for further review. As noted previously in Section 3.1 3.1, two additional references were added to the literature search protocol to capture database data from the WQP and NHANES. The database data were compared to other database and monitoring data found during the literature search to ensure no duplication of data. A record from a predecessor database to Water Quality Portal, EPA's STORET database, that was found during the literature search was not counted as a separate reference, to avoid double-counting data. In addition, any study summarizing data from NHANES was considered, but not evaluated and extracted, because it contained some of the data that was in the master record for NHANES database data. As noted above, while there were no data for TCEP in the NHANES database, there were data for its metabolite, BCEP. Urinary BCEP is a recommended target for biomonitoring of TCEP (Dodson et al., 2014). To further support the evaluation of TCEP exposure to humans, if there is information that meets the screening criteria for BCEP, a relevant biomarker for human TCEP exposure,

this information was also considered in a similar manner as data for TCEP during the systematic review process. There were no other changes to the process identified in the 2021 Draft Systematic Review Protocol for information considered for the evaluation of general population, consumer, and environmental exposure to TCEP (U.S. EPA, 2021).

3.6 Environmental and Human Health Hazard

The peer-reviewed and gray literature searches for environmental and human health hazard are as described in Sections 4.2 and 4.3, respectively, in the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). Specifically, SWIFT-Review was used to identify peer-reviewed references that are predicted to be the most relevant for evaluating environmental and human health hazard for TCEP. Search strings were developed for the two hazard disciplines by EPA's ORD in collaboration with SWIFT-Review developer, Sciome. As mentioned above in Section 3.1, the search string used to identify potentially relevant peer-reviewed data references for evaluation of the environmental and human health hazard of TCEP were validated. When the search string terms are identified in the title, abstract, or keywords of a given reference in SWIFT-Review, those references proceed with TIAB screening. The environmental and human health hazard search strings are provided online.

3.7 Dermal Absorption

As explained above in Section 3.1, EPA performed a targeted filtering of the references from our existing TCEP literature set to identify potentially relevant information for the characterization of dermal absorption of TCEP. The search string used is listed below in Section 3.7.1.

3.7.1 Dermal Absorption Search String

"Dermal flux" OR "Skin flux" OR "Dermal penetration" OR "Skin penetration" OR "Dermal absorption fraction" OR "Absorption fraction" OR "Neat Kp" OR "Aqueous Kp" OR "Kp" OR "Skin permeability coefficient" OR "Permeability coefficient" OR "Skin permeation coefficient" OR "Permeation coefficient" OR "Skin permeation" OR "Skin absorption" OR "Dermal absorption" OR "Dermal permeation" OR "OECD 427" OR "OECD 428."

4 DATA SCREENING

Sections 4.2.5 and 4.3.2 of the 2021 Draft Systematic Review Protocol describe how title and abstract (TIAB) and full-text screening respectively, are conducted to identify references that may contain relevant information for use in risk evaluations under TSCA using discipline-specific screening criteria (U.S. EPA, 2021). Specifically, TIAB screening efforts may be conducted using the specialized web-based software programs DistillerSR[1] and SWIFT-Active-Screener,[2] [3] and the below sub-sections will describe whether TIAB screening was done manually in DistillerSR or utilized machine learning to help prioritize reference screening in SWIFT-Active-Screener. Additional details on how SWIFT Active-Screener utilizes a machine-learning algorithm to automatically compute which unscreened documents are most likely to be relevant[4] are available in Section 4.2.5 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). During TIAB screening, if it was unclear whether a reference met the screening criteria (*e.g.*, PECO/RESO/PESO statements) without having the full reference to review, or if a reference was determined to meet the screening criteria, that reference advanced to full-text screening if the full reference could be retrieved and generated into a Portable Document Format (PDF).

Literature inventory trees were introduced in the scoping process for the risk evaluations that began systematic review in 2019 in response to comments received from the SACC and public to better illustrate how references underwent various systematic review steps (*e.g.*, TIAB and full-text screening). As explained in Section 2.1.2 of the *Final Scope of the Risk Evaluation for Tris(2-chloroethyl) Phosphate (TCEP)* (U.S. EPA, 2020)), literature inventory trees demonstrate how references that meet screening criteria progress to the next systematic review step. EPA used the Health Assessment Workplace Collaborative (HAWC) tool to develop web-based literature inventory trees that enhance the transparency of the decisions resulting from the screening processes. Additional references that EPA has obtained via public comments and other sources were also considered in the systematic review process and are reflected in the interactive HAWC hyperlinks available in the figure captions below each respective literature inventory tree. The web-based interactive literature inventory trees in HAWC also allow users to directly access the references in the Health & Environmental Research Online (HERO) database (more details available in Section 1 of the 2021 Draft Systematic Review Protocol). Instructions for accessing information about references and data sources in each node via HERO are available in HAWC for each respective literature inventory tree. Each node indicates whether a reference has met screening criteria at different screening steps and/or contains types of content that may be discerned at that respective systematic review step (U.S. EPA, 2021). Furthermore, the sum of the numbers for the various nodes in the literature inventory trees may be smaller or larger than the preceding node because some studies may have unclear relevance or be relevant for many categories of information. The screening process for each discipline varies and the nodes in the literature inventory tree indicate the screening decisions determined for each reference and whether specific content could

[1] As noted on the DistillerSR web page, this systematic review software "automates the management of literature collection, triage, and assessment using AI and intelligent workflows...to produce transparent, audit ready, and compliant literature reviews." EPA uses DistillerSR to manage the workflow related to screening and evaluating references; the literature search is conducted external to DistillerSR.

[2] SWIFT-Active Screener is another systematic review software that EPA is adopting in the TSCA systematic review process. From Sciome's SWIFT-Active Screener web page: "As screening proceeds, reviewers include or exclude articles while an underlying statistical model in SWIFT-Active Screener automatically computes which of the remaining unscreened documents are most likely to be relevant. This 'Active Learning' model is continuously updated during screening, improving its performance with each reference reviewed. Meanwhile, a separate statistical model estimates the number of relevant articles remaining in the unscreened document list."

[3] SWIFT is an acronym for "*Sciome Workbench for Interactive Computer-Facilitated Text-mining.*" SWIFT-Active Screener uses machine learning approaches to save screeners' time and effort.

[4] Description comes from the SWIFT-Active Screener web page.

be determined; if no references had a specific screening decision and/or contained specific content relevant for a respective discipline, a node will not be present on the literature tree to depict this.

Occasionally some references or data sources are identified in the literature search because of the availability of the title and abstract, however EPA may not be able to always locate the entire or original version. Therefore references or data sources that meet TIAB screening criteria may be unattainable for full-text screening. The "PDF not available" node within the literature inventory tree refers to references that were identified in the literature search but which EPA was unable to obtain the entire reference or source of information.

While all information contained in references that enter systematic review is considered for use in the risk evaluation, the references that satisfy the screening criteria are generally deemed to contain the most relevant and useful information for characterizing the uses of, exposure to, and hazard associated with a chemical of interest and are generally utilized in the risk evaluation (and can be used later on to identify further data needs). On the other hand, data or information sources that do not satisfy the screening criteria outlined below may undergo data quality evaluation and extraction should a data need arise for the risk evaluation.

4.1 Multi-Disciplinary Updates and Clarifications to the Data Screening

As stated above in Section 2, all references that are found in the initial chemical-specific searches are considered for use in the respective chemical risk evaluation. Previously Section 4.2.5 of the 2021 Draft Systematic Review Protocol explained that references tagged as potentially having supplemental information may be considered for data quality evaluation and extraction. However, one clarification to that description is that even references that are tagged as not meeting TIAB or full-text screening criteria (*e.g.*, PECO/PESO/RESO) for a respective discipline or sub-discipline may also undergo additional screening to meet information needs that were not stated in the original screening criteria and be considered for data quality evaluation and extraction, should there be additional relevant information that may not have met the original screening criteria.

4.2 Physical and Chemical Properties

During data screening, EPA followed the process described in Appendix H-1 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021) to conduct TIAB and full-text screening for TCEP guided by the data or information needs on various physical and chemical properties or endpoints as listed in Table_Apx H-1 of the protocol. The same screening criteria were used during TIAB and full-text screening of references containing information about the physical and chemical properties of TCEP. TIAB screening was performed using SWIFT Active-Screener. Upon meeting screening criteria during full-text screening, data or information sources then underwent data quality evaluation and extraction. Figure 4-1 presents the number of references reporting general physical and chemical property information that fulfilled the data needs for TCEP and passed these criteria for TIAB and full-text screening.

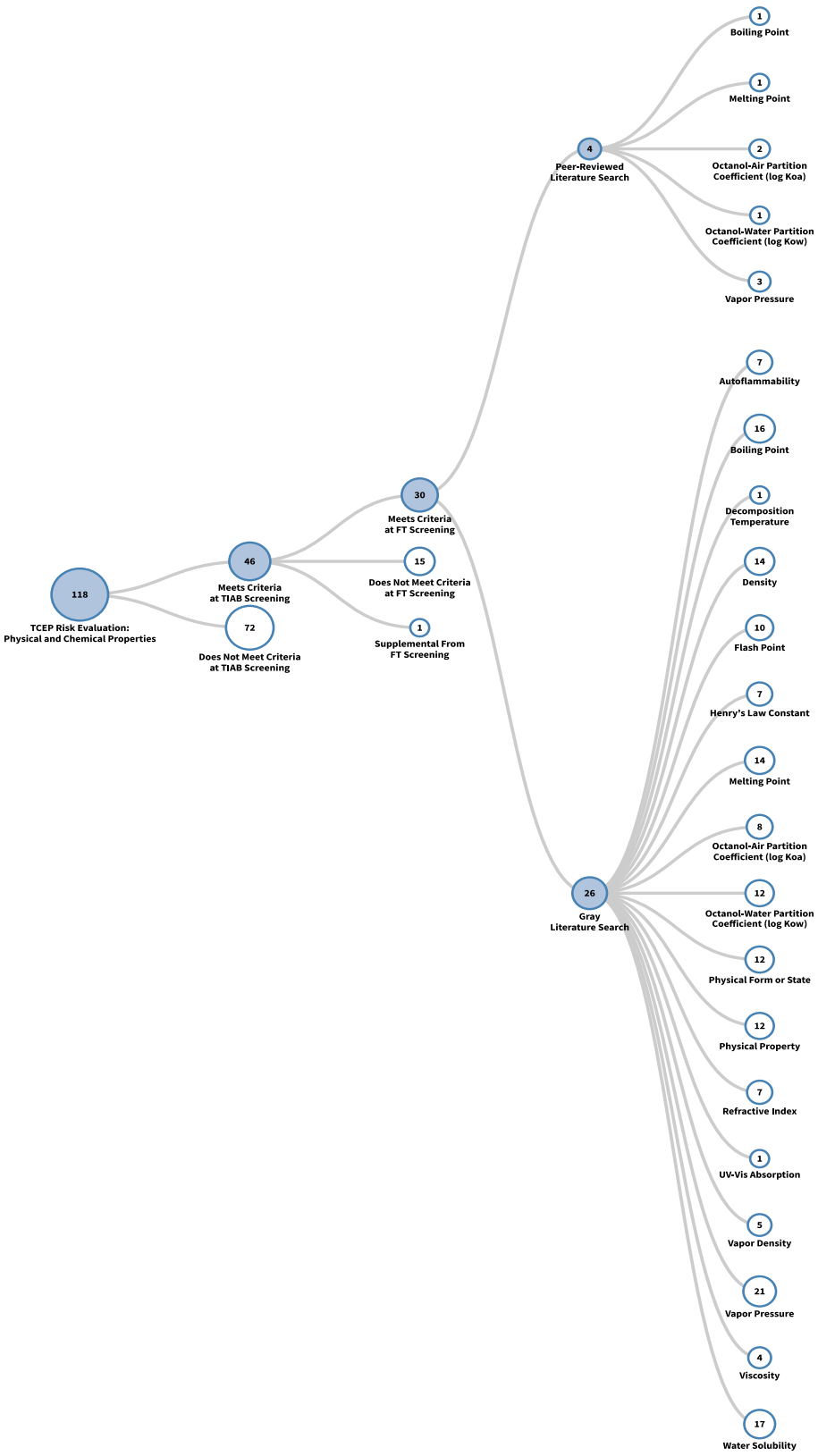


Figure 4-1. Literature Inventory Tree – Physical and Chemical Properties for TCEP
View the interactive literature inventory tree in HAWC. Data in this figure represent all references obtained from the publicly available databases and gray literature references searches that were included in systematic review as of January 27, 2023. Additional data may be added to the interactive version as they become available.

4.3 Environmental Fate and Transport Properties

During data screening, EPA followed the process described in Appendix H.2 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021), to conduct TIAB and full-text screening for TCEP literature search results, as guided by the PESO statement. PESO stands for **P**athways or **P**rocesses, **E**xposure, **S**etting or **S**cenario, and **O**utcomes (see Table_Apx H2 in 2021 Draft Systematic Review Protocol). The same PESO screening criteria was used during TIAB and full-text screening for references considered for the evaluation of environmental fate and transport properties of TCEP. TIAB screening was performed using SWIFT Active-Screener. Data or information sources that complied with the screening criteria specified in the PESO statement then underwent data quality evaluation and extraction. Figure 4-2 presents the number of references that report chemical-specific fate processes and endpoints, or environmental and exposure pathways that passed PESO screening criteria at TIAB and full-text screening.

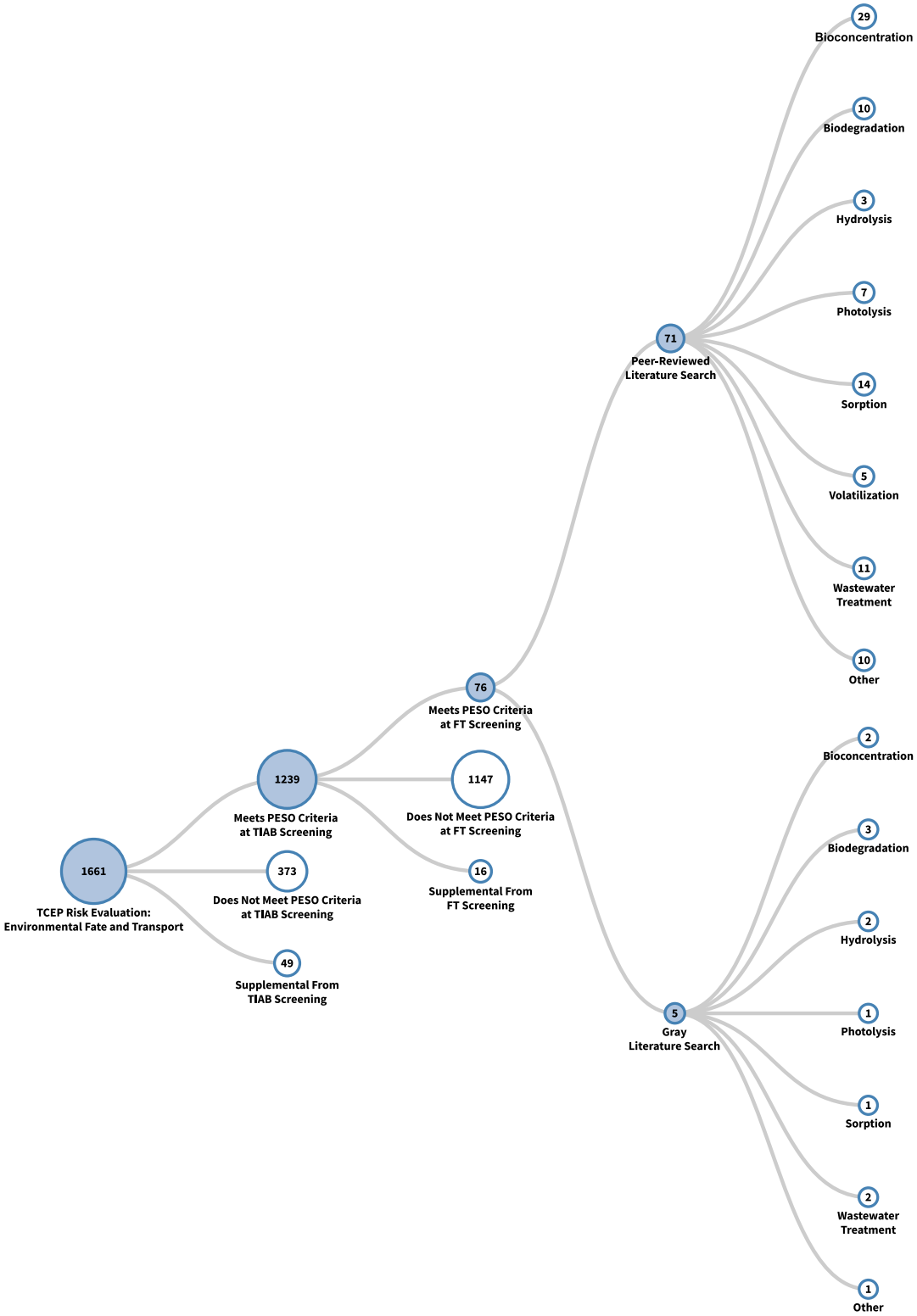


Figure 4-2. Literature Inventory Tree – Environmental Fate and Transport Properties for TCEP
View the interactive literature inventory tree in HAWC. Data in this figure represent all references obtained from the publicly available databases and gray literature references searches that were included in systematic review as of January 27, 2023. Additional data may be added to the interactive version as they become available

4.4 Environmental Release and Occupational Exposure

During data screening, EPA followed the process described in Appendix H.3 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021), to conduct TIAB and full-text screening for TCEP literature search results, as guided by the RESO statement. RESO stands for **R**eceptors, **E**xposure, **S**etting or Scenario, and **O**utcomes. The same RESO statement was used during title and abstract and full-text screening for references considered for the evaluation of environmental release and occupational exposure information for TCEP. Title and abstract screening was performed using SWIFT Active-Screener. Data or information sources that complied with the screening criteria specified in the RESO statement then underwent data quality evaluation and extraction. Figure 4-3 presents the number of references that report general engineering data, environmental release, and occupational exposure data that passed RESO screening criteria at title and abstract and full-text screening.

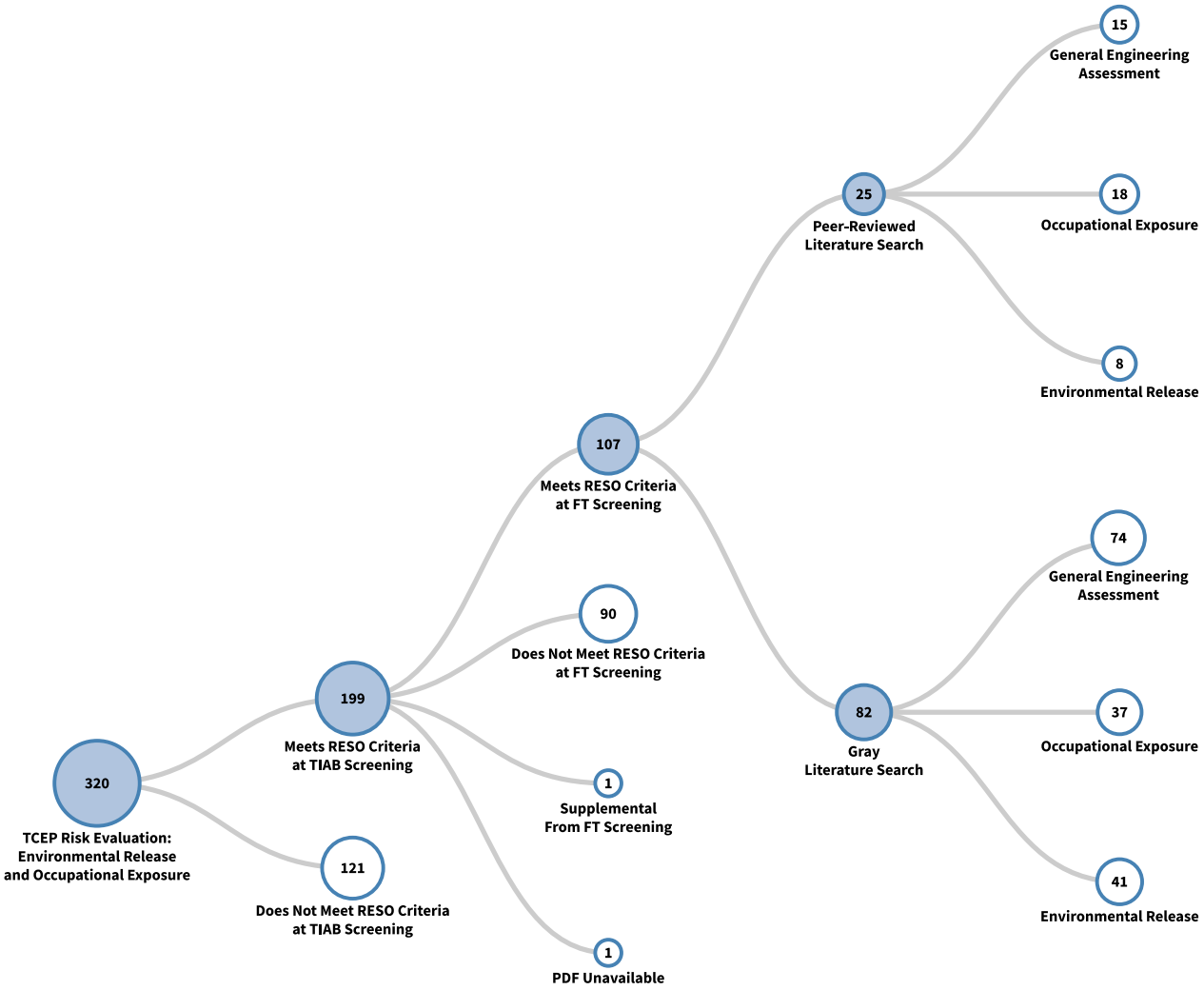


Figure 4-3. Literature Inventory Tree – Environmental Release and Occupational Exposure Search Results for TCEP

View the interactive literature inventory tree in HAWC. Data in this figure represent all references obtained from the publicly available databases and gray literature references searches that were included in systematic review as of April 18, 2023. Additional data may be added to the interactive version as they become available.

4.5 General Population, Consumer, and Environmental Exposure

During data screening, EPA followed the process described in Appendix H.4 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021), to conduct TIAB and full-text screening for TCEP

literature search results, as guided by the PECO statement. PECO stands for **P**opulation, **E**xposure, **C**omparator or Scenario, and **O**utcomes for Exposure Concentration or Dose. The same PECO statement was used during TIAB and full-text screening for references considered for the evaluation of general population, consumer, and environmental exposure information for TCEP. TIAB screening was performed using SWIFT Active-Screener. Figure 4-4 presents the number of references that report general population, consumer, and environmental exposure data that passed PECO screening criteria at TIAB and full-text screening.

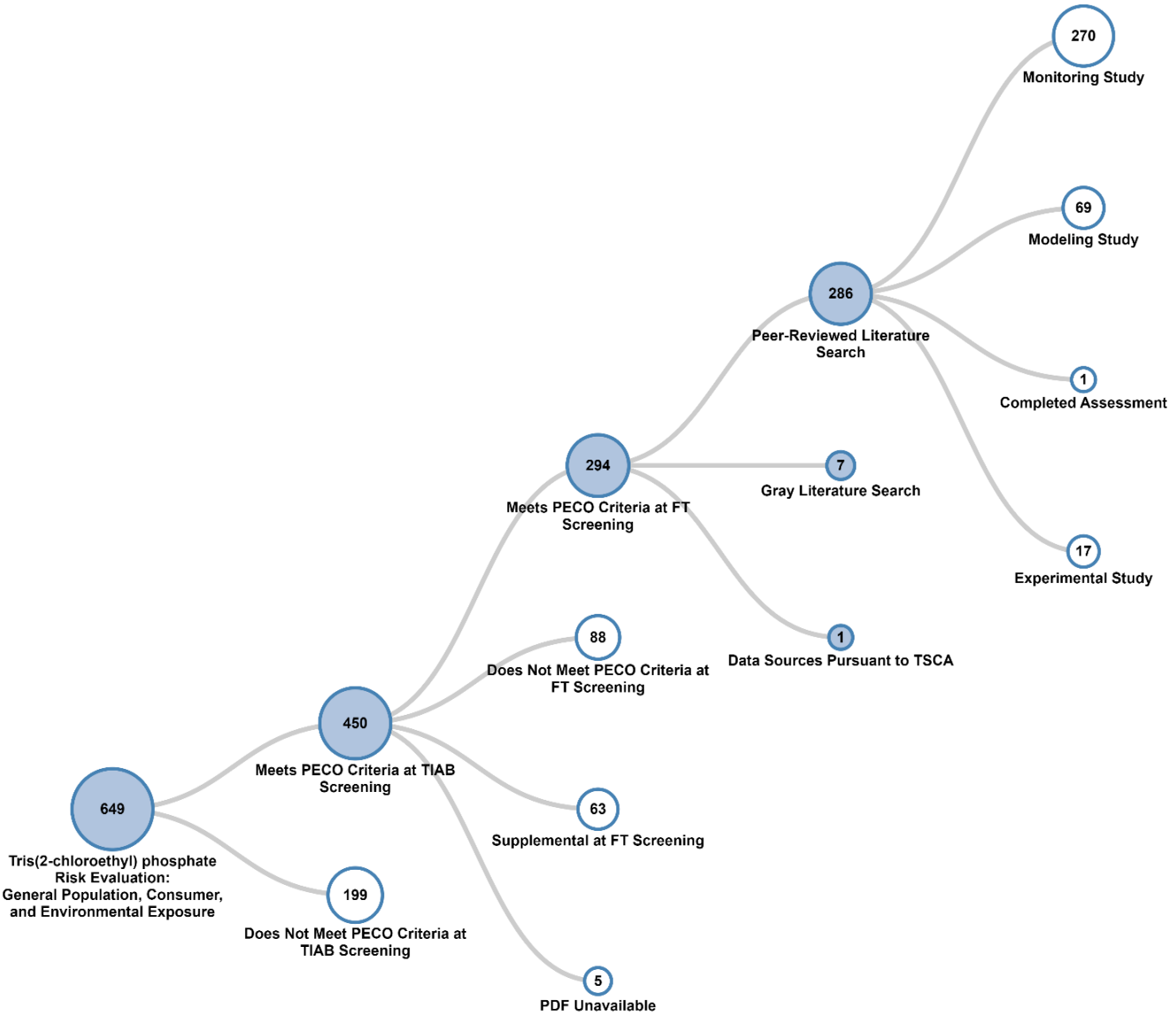


Figure 4-4. Literature Inventory Tree – General Population, Consumer, and Environmental Exposure Search Results for TCEP

View the interactive literature inventory tree in HAWC. Data in this figure represent all references obtained from the publicly available databases and gray literature references searches that were included in systematic review as of May 22, 2023. Additional data may be added to the interactive version as they become available.

4.6 Environmental and Human Health Hazard

During data screening, EPA followed the process described in Appendix H.5.7 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021), to conduct TIAB and full-text screening for TCEP literature search results, as guided by the PECO statement. The same PECO statement was used during

TIAB and full-text screening for references considered for the evaluation of environmental and human health hazard resulting from exposure to TCEP. During TIAB screening, EPA manually screened each reference that was identified as being potentially relevant in DistillerSR. EPA performed full-text screening in DistillerSR on references that met the PECO screening criteria as well as on references that did not clearly fail to meet the PECO screening criteria based on the information available in the title and abstract. Figure 4-5 presents the number of references that report environmental and human health hazard data that passed PECO screening criteria at TIAB and full-text screening.

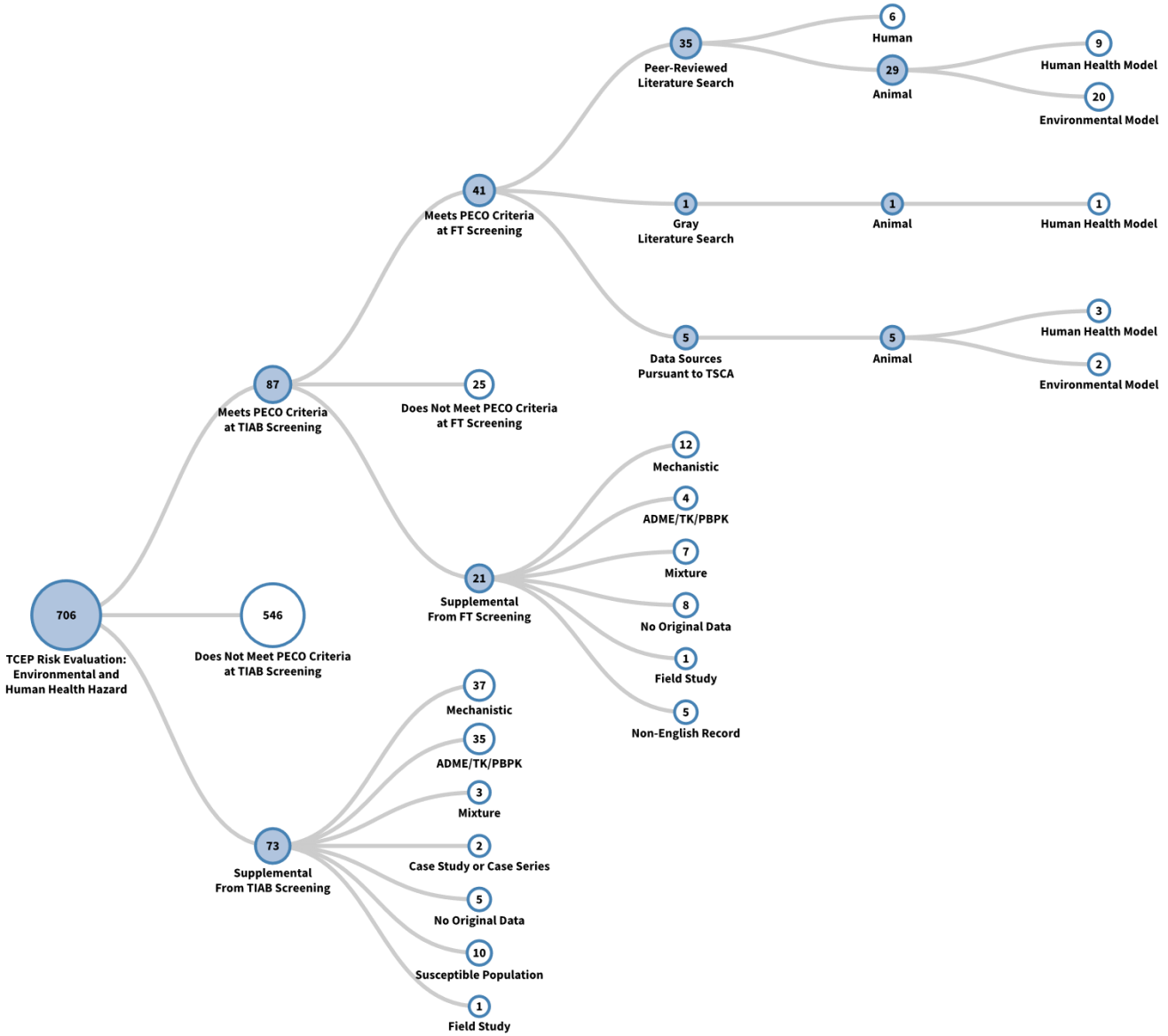

706
TCEP Risk Evaluation: Environmental and Human Health Hazard
87
Meets PECO Criteria at TIAB Screening
546
Does Not Meet PECO Criteria at TIAB Screening
73
Supplemental From TIAB Screening
41
Meets PECO Criteria at FT Screening
25
Does Not Meet PECO Criteria at FT Screening
21
Supplemental From FT Screening
35
Peer-Reviewed Literature Search
1
Gray Literature Search
5
Data Sources Pursuant to TSCA
6
Human
29
Animal
9
Human Health Model
20
Environmental Model
1
Animal
1
Human Health Model
5
Animal
3
Human Health Model
2
Environmental Model
12
Mechanistic
4
ADME/TK/PBPK
7
Mixture
8
No Original Data
1
Field Study
5
Non-English Record
37
Mechanistic
35
ADME/TK/PBPK
3
Mixture
2
Case Study or Case Series
5
No Original Data
10
Susceptible Population
1
Field Study

Figure 4-5. Literature Inventory Tree – Environmental and Human Health Hazard for TCEP

View the interactive literature inventory tree in HAWC. Data in this figure represent all references obtained from the publicly available databases and gray literature references searches that were included in systematic review as of March 6, 2023. Additional data may be added to the interactive version as they become available.

4.7 Dermal Absorption

EPA developed a PECO statement (Table 4-1) to conduct both TIAB and full-text screening of references considered for the evaluation of dermal absorption resulting from TCEP exposure. Each reference was manually screened by two reviewers at both the TIAB and full-text screening steps. Figure 4-6 presents the number of references that report dermal absorption data that passed PECO screening criteria at TIAB and full-text screening. The references considered for dermal absorption (n = 18) were screened multiple times across disciplines. Table 4-2 shows the TIAB and full-text screening results of the references considered to have dermal absorption data, resulting from TCEP exposure, that were also considered by other disciplines; screening results are based on the criteria outlined in the screening criteria for each discipline. Specifically, references that either have supplemental information or do not meet screening criteria for a respective discipline at either TIAB or full-text screening are combined, whereas only the references that met screening criteria at full-text screening for a respective discipline are indicated below in Table 4-2.

Table 4-1. PECO Statement for the Screening of Dermal Absorption References for TCEP

PECO Element	Evidence
P	Tests of the single toxicants on *ex vivo* tissues (including permeation and retention studies) or on live, whole, taxonomically verifiable organisms are included. **Human:** Any population and life stage (occupational or general population, including children and other sensitive populations). **Animal:** All human health models, including (but not limited to) rat, mouse, rabbit, dog, hamster, guinea pig, cat, non-human primate, and pig. **Supplemental:** Tests using 3D human skin equivalent/reconstructed tissue models (*e.g.*, EpiDerm, EPISKIN) or any other *in vitro* systems are considered supplemental.
E	**Human and Animal:** Any quantified dermal exposure to TCEP or in a vehicle, including exposure that occurs *in vivo* or *ex vivo* for any duration. Studies are included only if exposure is intentional and quantified. If exposure is not intentional and is not experimentally controlled, the study is excluded. For example, studies of absorption in workers will be excluded, even if exposure has been quantified. Studies assessing exposures to mixtures (*i.e.*, containing substances other than a vehicle) will be included only if they also contain an exposure or treatment group assessing the chemical of interest alone or in aqueous solution. **Supplemental:** *In vitro* exposures and/or studies in which exposure occurs only to a mixture containing one or more of the chemicals of interest.
C	**Human and Animal:** Any or no comparison group.
O	**Human and Animal**: Any quantitative assessment of the rate or extent of dermal absorption of the substance. Measurements may include the amount of substance that has passed through the skin, or was retained in the skin, distributed within the organism (*e.g.*, blood and tissue concentrations), and/or excreted by the organism (*e.g.*, through urine, feces, or expired air). Absorption may be measured directly (by chemical analysis for the substance and/or its metabolites) or indirectly (*e.g.*, measurement of radioactivity if using a radio-labelled test substance). Absorption may be quantified via determination of percent absorption, dermal/penetrative flux rate, or dermal penetration coefficient (Kp).

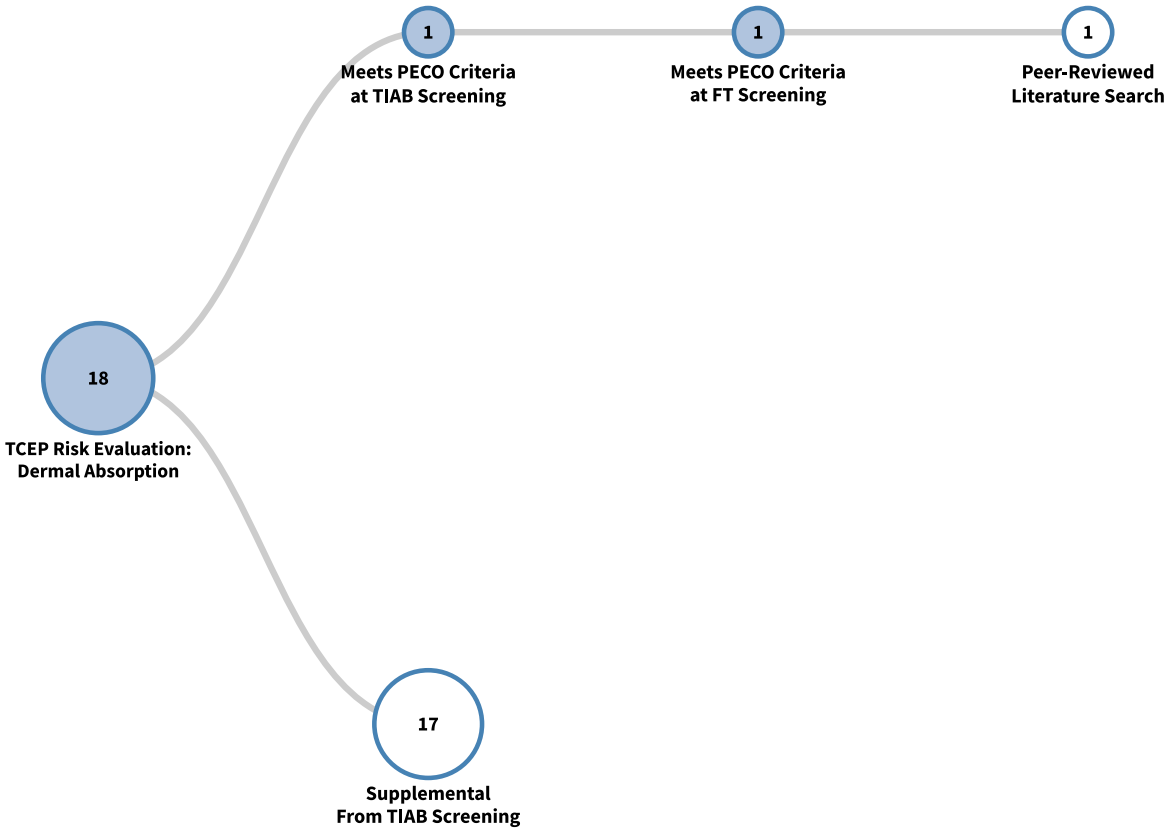


Figure 4-6. Literature Inventory Tree – Dermal Absorption for TCEP
View the interactive literature inventory tree in HAWC. Data in this figure represent all references obtained from the publicly available databases and gray literature references searches that were included in systematic review as of February 28, 2023. Additional data may be added to the interactive version as they become available.

Table 4-2. Dermal Absorption Data Sources Considered by Other Disciplines for TCEP

Screening Result	Discipline				
	Environmental and Human Health Hazard	Environmental Release and Occupational Exposure	General Population, Consumer, and Environmental Exposure	Environmental Fate and Transport	Physical and Chemical Properties
Met Screening Criteria	0	3	9	0	0
Supplemental	5	0	0	0	0
Did Not Meet Screening Criteria	13	2	0	1	0

5 DATA EVALUATION AND DATA EXTRACTION

Data evaluation and extraction were conducted as described in Sections 5 and 6 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). Data evaluation is the systematic review step in which EPA assesses quality of the individual data sources using the evaluation strategies and criteria for each discipline (*e.g.*, physical and chemical property data; fate and transport data; occupational exposure and environmental release data; general population, consumer, and environmental exposure data; environmental hazard; human health hazard) or sub-discipline (*e.g.*, animal toxicity or epidemiology). The data quality evaluation method uses a structured framework with predefined criteria for each type of data/information source. Data extraction is the systematic review step in which EPA uses structured forms or templates to extract quantitative and qualitative data and information from references that meet screening criteria. The overall goal is to provide transparency, consistency, and as much objectivity as possible to the data quality evaluation and extraction processes along with meeting the TSCA scientific standards in section 26(h).

References that meet screening criteria following full-text screening will generally proceed to data quality evaluation and extraction steps, however one clarification to the procedures outlined in Section 6 of the 2021 Draft Systematic Review Protocol is that in situations where EPA is unable to extract data/information from sources that meet screening criteria (*e.g.*, formatting prohibits accurate extraction), that source may not have extracted data to present in the risk evaluation or respective supplemental documents. The systematic review supplemental files that contain results from the data quality evaluation and extraction systematic review steps may use updated templates from those that were provided in the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021) because the purpose of these supplemental documents is to accommodate the data needs for each respective risk evaluation. The following sections describe the data quality and extraction process followed by each discipline or sub-discipline to address various information needs for the Draft Risk Evaluation of TCEP and any clarifications or updates regarding these systematic review steps as described in the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021).

5.1 Physical and Chemical Properties

As described in the 2021 Draft Systematic Review Protocol, data quality evaluation and extraction followed the steps outlined in Sections 5, 6, and 6.1 (U.S. EPA, 2021). The data quality criteria for physical and chemical property data are summarized in Appendix K of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). The *Draft Risk Evaluation for Tris(2-chloroethyl) phosphate (TCEP) – Systematic Review Supplemental File: Data Quality Evaluation and Data Extraction Information for Physical and Chemical Properties* provides details of the data extracted and evaluated, including metric ratings and the overall study quality determination for each data source (U.S. EPA, 2023g).

5.2 Environmental Fate and Transport Properties

As described in Appendix L the 2021 Draft Systematic Review Protocol, data quality evaluation and extraction followed the steps outlined in Sections 5, 6, and 6.2 (U.S. EPA, 2021). Appendix L.4 describes how the overall quality determinations of data sources containing environmental fate and transport data that may be used to quantitatively or qualitatively support the risk evaluations (U.S. EPA, 2021). Table_Apx L4 illustrates the possible quality ratings across the selected metrics for environmental fate data with examples in Table_Apx L5, Table_Apx L6 and Table_Apx L7 (U.S. EPA, 2021). Specific fate data quality rating quality criteria are in Table_Apx L8 (U.S. EPA, 2021).The *Draft Risk Evaluation for Tris(2-chloroethyl) phosphate (TCEP) – Systematic Review Supplemental File: Data*

Quality Evaluation and Data Extraction Information for Environmental Fate and Transport (U.S. EPA, 2023e) provides details of the data extracted and evaluated, including metric rating and the overall study quality determination for each data source.

5.3 Environmental Release and Occupation Exposure

As described in the 2021 Draft Systematic Review Protocol, data quality evaluation and extraction followed the steps outlined in Sections 5, 6, and 6.2 (U.S. EPA, 2021). The data quality criteria for environmental release and occupational exposure data are summarized in Appendix M of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). The *Draft Risk Evaluation for Tris(2-chloroethyl) phosphate (TCEP) – Systematic Review Supplemental File: Data Quality Evaluation and Data Extraction Information for Environmental Release and Occupational Exposure* details the data extracted and evaluated, including metric rating and the overall study quality determination for each data source (U.S. EPA, 2023f).

5.4 General Population, Consumer, and Environmental Exposure

As described in the 2021 Draft Systematic Review Protocol, data quality evaluation and extraction generally followed the steps outlined in Section 5 and 6 (U.S. EPA, 2021). However, one modification is that, since BCEP is formed when TCEP metabolizes in the human body, EPA also considered data with concentrations of BCEP reported because it provides relevant information on human exposures. Please see Section 3.5 for further explanations regarding TCEP metabolism. Therefore, information on BCEP that met PECO screening criteria also underwent data quality evaluation and extraction.

In addition, a few updates were made to the data quality evaluation metrics for some evidence streams (*i.e.*, data types) after the metrics were published in the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). Most of the changes were editorial or minor clarifications, including the standardization of some metrics that apply to multiple evidence streams, where appropriate. For example, in the quality assurance/quality control (QA/QC) metric for evaluating monitoring and experimental evidence streams, the acronym QA/QC was defined and replaced all references to quality assurance and quality control, whether occurring separately or together, and the term "QA/QC techniques" was changed to "QA/QC measures," which already appeared in the metrics.

A few metrics applicable to multiple evidence streams were slightly modified to better fit some of the unique situations that frequently arise for a certain type of evidence stream (*e.g.*, databases). For example, some metrics were updated to clarify the intent of the metric and better account for variation in types of evidence included in one grouping (*e.g.*, experiments involving chamber studies vs. product concentration assessments). The domains did not change; however, see below for the changes and updates made to the data evaluation metrics for the respective evidence streams (*i.e.*, monitoring studies, experimental studies, and databases) as presented in Section 5.4.1. No changes were made to the data evaluation metrics for modeling data, as described in Appendix N.6.2 or to the data evaluation metrics for completed exposure assessments and risk characterizations, as described in Appendix N.6.7 in the 2021 Draft Systematic Review Protocol, respectively (U.S. EPA, 2021). Data quality evaluations for all the references that met PECO screening criteria are included in the *Draft Risk Evaluation for Tris(2-chloroethyl) phosphate (TCEP) – Systematic Review Supplemental File: Data Quality Evaluation Information for General Population, Consumer, and Environmental Exposure* (U.S. EPA, 2023i), referred to hereafter as the "TCEP Data Quality Evaluation Information for General Population, Consumer, and Environmental Exposure."

Data extraction of general population, consumer, and environmental exposure data and information was conducted as described in Section 6 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). However, with respect to information stored within databases, if EPA has access to the data tables, EPA does not conduct a separate data extraction because the data are more accessible and have additional context in the original database format. Data present in the database when the database underwent full-text screening are available in the HERO database (*e.g.*, HERO IDs: 10663361 and 10668533), along with the date the data were downloaded. If a reference (*e.g.*, peer-reviewed reference) presents data from a database that did not undergo systematic review directly (*e.g.*, a foreign database that is not publicly accessible), the data would be extracted from the reference to the extent possible; this did not apply to references that underwent systematic review for this chemical.

As mentioned above in Section 5, references may not undergo data extraction, regardless of the overall quality determination, if they contain no extractable data points (*e.g.*, values are contained in a non-digitizable figure or are representative of unspecified media or treatment processes). In addition, there may be other reasons that EPA decides not to extract all the data from a reference that undergoes data evaluation, depending on the needs of the assessment. While EPA may not extract all the data from all sources, EPA extracted data from studies from the U.S. and other high-income countries that are most relevant for characterizing exposure, use conditions, patterns of use, and product characteristics in the U.S. Decisions about whether to limit extractions to certain timeframes or certain countries were made on an evidence stream by evidence stream basis based on available data and the conditions of use being evaluated to better characterize general population, consumer, and environmental exposure and meet assessment needs. This constitutes an update to Section 6 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). Extraction forms, templates, and decisions are tailored to fit the data extraction needs for each risk evaluation.

The types of fields extracted vary by evidence stream and generally followed Section 6.3 of the 2021 Draft Systematic Review Protocol with regard to the data characteristics captured (U.S. EPA, 2021). Examples of types of data extracted and the extraction formats for the evidence streams for environmental, general population, and consumer exposure data are listed in the extraction tables provided in the *Draft Risk Evaluation for Tris(2-chloroethyl) phosphate (TCEP) – Systematic Review Supplemental File: Data Extraction Information for General Population, Consumer, and Environmental Exposure* (U.S. EPA, 2023c), referred to hereafter as the "TCEP Data Extraction Information for General Population, Consumer, and Environmental Exposure."

5.4.1 Data Quality Evaluation Metric Updates

Shown below are the data evaluation metrics for three evidence streams, showing which data evaluation metrics changed since the publication of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). Other data quality criteria for studies on consumer, general population, and environmental exposure appear in Appendix N of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). For the modeling, and the completed exposure assessments and risk characterizations evidence streams, there were no changes made to the data evaluation metrics since the 2021 Draft Systematic Review Protocol was published. The criteria for modeling studies appear in Table_Apx N-9 of the 2021 Draft Systematic Review Protocol, and criteria for completed exposure assessments and risk characterizations appear in Table_Apx N-19. In some cases, references can meet the criteria for two evidence streams evaluated by exposure assessors, and they can also be reviewed by other disciplines. Upon review, each study is evaluated and extracted using the criteria for the most appropriate and applicable evidence streams given the information therein. In order to make it easier for the reader to see changes made to the data evaluation metrics in Table 5-1, Table 5-2, and Table 5-3, the following convention is used: text inserted is underlined, and text deleted is in ~~strikethrough~~.

5.4.1.1 Data Evaluation Criteria for Monitoring Data, as Revised

Table 5-1. Updated Data Quality Evaluation Criteria for Monitoring Data Sources

Quality Rating	Description
Domain 1. Reliability	
Metric 1. Sampling methodology	
High	Samples were collected according to publicly available SOPs that are scientifically sound and widely accepted (*i.e.*, from a source generally ~~using~~ known to use sound methods and/or approaches) for the chemical and media of interest. Example SOPs include U.S. Geological Survey (USGS') "National Field Manual for the Collection of Water-Quality Data," EPA's "Ambient Air Sampling" (SESDPROC-303-R5), etc. OR The sampling protocol used was not a publicly available SOP from a source generally known to use ~~using~~ sound methods and/or approaches, but the sampling methodology is clear, appropriate (*i.e.*, scientifically sound), and similar to widely accepted protocols for the chemical and media of interest. All pertinent sampling information is provided in the data source or companion source. Examples include: 1. sampling equipment 2. sampling procedures/regime 3. sample storage conditions/duration 4. performance/calibration of sampler 5. study site characteristics 6. matrix characteristics
Medium	Sampling methodology is discussed in the data source or companion source and is generally appropriate (*i.e.*, scientifically sound) for the chemical and media of interest; however, **one or more pieces of sampling information is not described.** The missing information is unlikely to have a substantial impact on results. OR Standards, methods, protocols, or test guidelines may not be widely accepted, but a successful validation study for the new/unconventional procedure was conducted prior to the sampling event and is consistent with sound scientific theory and/or accepted approaches. Or a review of information indicates the methodology is acceptable and differences in methods are not expected to lead to lower quality data.
Low	Sampling methodology is only briefly discussed; therefore, **most sampling information is missing** and likely to have a substantial impact on results. AND/OR The sampling methodology **does not represent best sampling methods, protocols, or guidelines** for the chemical and media of interest (*e.g.*, outdated [but still valid] sampling equipment or procedures, long storage durations). AND/OR There are **some inconsistencies** in the reporting of sampling information (*e.g.*, differences between text and tables in data source, differences between standard method and actual procedures reported to have been used, etc.) that led to a low confidence in the sampling methodology used.

Quality Rating	Description
Critically Deficient	The sampling methodology is not discussed in the data source or companion source. AND/OR Sampling methodology is not scientifically sound or is not consistent with widely accepted methods/approaches for the chemical and media being analyzed (*e.g.*, inappropriate sampling equipment, improper storage conditions). AND/OR There are **numerous inconsistencies** in the reporting of sampling information, resulting in high uncertainty in the sampling methods used.
Not rated/not applicable	
Reviewer's comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 2. Analytical methodology	
High	Samples were analyzed according to publicly available analytical methods that are scientifically sound and widely accepted (*i.e.*, from a source generally ~~using~~ known to use sound methods and/or approaches) and are appropriate for the chemical and media of interest. Examples include EPA SW-846 Methods, NIOSH Manual of Analytical Methods 5th Edition, etc. OR The analytical method used was not a publicly available method from a source generally ~~using~~ known to use sound methods and/or approaches, but the methodology is clear and appropriate (*i.e.*, scientifically sound) and similar to widely accepted protocols for the chemical and media of interest. All pertinent sampling information is provided in the data source or companion source. Examples include: 1. extraction method 2. analytical instrumentation (required) 3. instrument calibration 4. limit of quantitation (LOQ), LOD, detection limits, and/or reporting limits 5. recovery samples 6. biomarker used (if applicable) 7. matrix-adjustment method (*i.e.*, creatinine, lipid, moisture)
Medium	Analytical methodology is discussed in detail and is clear and appropriate (*i.e.*, scientifically sound) for the chemical and media of interest; however, **one or more pieces of analytical information is not described**. The missing information is unlikely to have a substantial impact on results. AND/OR The analytical **method may not be standard/widely accepted, but a method validation study was conducted** prior to sample analysis and is expected to be consistent with sound scientific theory and/or accepted approaches. AND/OR Samples were collected at a site and immediately analyzed using an on-site mobile laboratory, rather than shipped to a stationary laboratory.

Quality Rating	Description
Low	Analytical methodology is only briefly discussed. Analytical instrumentation is provided and consistent with accepted analytical instrumentation/methods. However, **most analytical information is missing** and likely to have a substantial impact on results. AND/OR Analytical method **is not s**tandard/widely accepted, and method validation is limited or not available. AND/OR Samples were analyzed using field screening techniques. AND/OR LOQ, LOD, detection limits, and/or reporting limits not reported. AND/OR There are **some inconsistencies or possible errors** in the reporting of analytical information (*e.g.*, differences between text and tables in data source, differences between standard method and actual procedures reported to have been used, etc.) which leads to a lower confidence in the method used.
Critically Deficient	Analytical methodology is not described, **including analytical instrumentation** (*i.e.*, HPLC, GC). AND/OR Analytical methodology is not scientifically appropriate for the chemical and media being analyzed (*e.g.*, method not sensitive enough, not specific to the chemical, out of date). AND/OR There are numerous inconsistencies in the reporting of analytical information, resulting in high uncertainty in the analytical methods used.
Not rated/ Not applicable	
Reviewer's comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 3. Selection of biomarker of exposure	
High	Biomarker in a specified matrix is known to have an accurate and precise quantitative relationship with external exposure, internal dose, or target dose (*e.g.*, previous studies (or the current study) have indicated the biomarker of interest reflects external exposures). AND Biomarker (parent chemical or metabolite) is derived from exposure to the chemical of interest.
Medium	Biomarker in a specified matrix has accurate and precise quantitative relationship with external exposure, internal dose, or target dose. AND Biomarker is derived from multiple parent chemicals, not only the chemical of interest, **but** there is a stated method to apportion the estimate to only the chemical of interest

Quality Rating	Description
Low	Biomarker in a specified matrix has accurate and precise quantitative relationship with external exposure, internal dose, or target dose. AND Biomarker is derived from multiple parent chemicals, not only the chemical of interest, **and** there is NOT an accurate method to apportion the estimate to only the chemical of interest. OR Biomarker in a specified matrix is a poor surrogate (low accuracy and precision) for exposure/dose.
Critically Deficient	Not applicable. A study will not be deemed critically deficient based on the use of biomarker of exposure.
Not rated/ applicable	Metric is not applicable to the data source.
Reviewer's comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Domain 2. Representative	
Metric 4. Geographic area	
High	Geographic location(s) is reported, discussed, or referenced.
Medium	Not applicable. This metric is dichotomous (*i.e.*, high vs. critically deficient).
Low	Not applicable. This metric is dichotomous (*i.e.*, high vs. critically deficient).
Critically Deficient	Geographic location is not reported, discussed, or referenced.
Not rated/ not applicable	
Reviewer's comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 5. Temporality	
High	Timing of sample collection for monitoring data is consistent with current or recent exposures **(within 5 years)** may be expected.
Medium	Timing of sample collection for monitoring data **is less consistent** with current or recent exposures **(>5 to 15 years)** may be expected.
Low	Timing of sample collection for monitoring data is not consistent with when current exposures **(>15 years old)** may be expected and likely to have a substantial impact on results.
Critically Deficient	Timing of sample collection for monitoring data is **not reported, discussed, or referenced.**

Quality Rating	Description
Not rated/ Not applicable	
Reviewer's comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 6. Spatial and temporal variability	
High	Sampling approach accurately captures variability of environmental contamination in population/scenario/media of interest based on the heterogeneity/homogeneity and dynamic/static state of the environmental system. For example: 1. **Large sample size** (*i.e.*, ≥10 samples for a single scenario). 2. Use of replicate samples. 3. Use of systematic or continuous monitoring methods. 4. Sampling over a sufficient period of time to characterize trends. 5. For urine, 24-hour samples are collected (vs. first morning voids or spot). 6. For biomonitoring studies, the timing of sample collected is appropriate based on chemical properties (*e.g.*, half-life), the pharmacokinetics of the chemical (*e.g.*, rate of uptake and elimination), and when the exposure event occurred.
Medium	Sampling approach likely captures variability of environmental contamination in population/scenario/media of interest based on the heterogeneity/homogeneity and dynamic/static state of the environmental system. Some uncertainty may exist, but it is unlikely to have a substantial impact on results. For example: 1. **Moderate sample size** (*i.e.*, 5–10 samples for a single scenario), or 2. Use of judgmental (non-statistical) sampling approach, or 3. No replicate samples. 4. For urine, first morning voids or pooled spot samples.
Low	Sampling approach poorly captures variability of environmental contamination in population/scenario/media of interest. For example: 1. **Small sample size** (*i.e.*, <5 samples), or 2. Use of haphazard sampling approach, or 3. No replicate samples, or 4. Grab or spot samples in single space or time, or 5. Random sampling that does not include all periods of time or locations, or 6. For urine, un-pooled spot samples.
Critically Deficient	Sample size is not reported. **Single sample** collected per data set. For biomonitoring studies, the timing of sample collected is not appropriate based on chemical properties (*e.g.*, half-life), the pharmacokinetics of the chemical (*e.g.*, rate of uptake and elimination), and when the exposure event occurred.
Not rated/not applicable	

Quality Rating	Description
Reviewer's comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 7. Exposure scenario	
High	The data closely represent relevant exposure scenario (*i.e.*, the population/scenario/media of interest). Examples include: 1. amount and type of chemical/product used 2. source of exposure 3. method of application or by-stander exposure 4. use of exposure controls 5. microenvironment (location, time, climate)
Medium	The data likely represent the relevant exposure scenario (*i.e.*, population/scenario/media of interest). **One or more key pieces of information may not be described** but the deficiencies are unlikely to have a substantial impact on the characterization of the exposure scenario. AND/OR If surrogate data, activities seem similar to the activities within scope.
Low	The data lack multiple key pieces of information, and the deficiencies are likely to have a substantial impact on the characterization of the exposure scenario. AND/OR There are **some inconsistencies or possible errors** in the reporting of scenario information (*e.g.*, differences between text and tables in data source, differences between standard method and actual procedures reported to have been used, etc.) which leads to a lower confidence in the scenario assessed. AND/OR If surrogate data, activities have lesser similarity but are still potentially applicable to the activities within scope.
Critically Deficient	If reported, the exposure scenario discussed in the monitored study does not represent the exposure scenario of interest for the chemical.
Not rated/ Not applicable	
Reviewer's comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Domain 3. Accessibility/clarity	
Metric 8. Reporting of results	
High	Supplementary or raw data (*i.e.*, individual data points) are reported, allowing summary statistics to be calculated or reproduced. AND Summary statistics are detailed and complete. Example parameters include: 1. Description of data set summarized (*i.e.*, location, population, dates, etc.)

Quality Rating	Description
	2. Range of concentrations or percentiles 3. Number of samples in data set 4. Frequency of detection 5. Measure of variation (coefficient of variation [CV], standard deviation) 6. Measure of central tendency (mean, geometric mean, median) 7. Test for outliers (if applicable) AND Both adjusted and unadjusted results are provided (*i.e.*, correction for void completeness in urine biomonitoring, whole-volume or lipid adjusted for blood biomonitoring, wet or dry weight for environmental tissue samples or soil samples) [only if applicable].
Medium	Supplementary or raw data (*i.e.*, individual data points) are not reported, and therefore summary statistics cannot be reproduced. AND/OR Summary statistics are reported but are missing one or more parameters (see description for high). AND/OR Only adjusted or unadjusted results are provided, but not both [only if applicable].
Low	Supplementary data are not provided, and summary statistics are missing most parameters (see description for high). AND/OR There are some inconsistencies or errors in the results reported, resulting in low confidence in the results reported (*e.g.*, differences between text and tables in data source, less appropriate statistical methods).
Critically Deficient	There are numerous inconsistencies or errors in the calculation and/or reporting of results, resulting in highly uncertain reported results.
Not Rated/ Not Applicable	
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 9. Quality assurance	
High	The study quality assurance/quality control (QA/QC) measures and all pertinent ~~quality assurance~~ QA/QC information is provided in the data source or companion source. Examples include: 1. Field, laboratory, and/or storage recoveries. 2. Field and laboratory control samples. 3. Baseline (pre-exposure) samples. 4. Biomarker stability 5. Completeness of sample (*i.e.*, creatinine, specific gravity, osmolality for urine samples) AND No QA/QC ~~quality control~~ issues were identified, or any identified issues were minor and adequately addressed (*i.e.*, correction for low recoveries, correction for completeness).

Quality Rating	Description
Medium	The study applied and documented ~~quality assurance/quality control~~ QA/QC measures; however, **one or more pieces of QA/QC information is not described.** Missing information is unlikely to have a substantial impact on results. AND No QA/QC ~~quality control~~ issues were identified, or any identified issues were minor and addressed (*i.e.*, correction for low recoveries, correction for completeness).
Low	QA/QC measures ~~Quality assurance/quality control techniques~~ and results were not directly discussed but are implied through the study's use of standard field and laboratory protocols. AND/OR Deficiencies were noted in ~~quality assurance/quality control~~ QA/QC measures that are likely to have a substantial impact on results. AND/OR There are some inconsistencies in the ~~quality assurance~~ QA/QC measures reported, resulting in low confidence in the QA/QC ~~quality assurance/control~~ measures taken and results (*e.g.*, differences between text and tables in data source).
Critically Deficient	QA/QC issues have been identified which significantly interfere with the overall reliability of the study.
Not Rated/ Not Applicable	
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Domain 4. Variability and uncertainty	
Metric 10. Variability and uncertainty	
High	The study characterizes variability in the population/media studied. AND Key uncertainties, limitations, and data gaps have been identified. AND The uncertainties are minimal and have been characterized.
Medium	The study has limited characterization of variability in the population/media studied. AND/OR The study has limited discussion of key uncertainties, limitations, and data gaps. AND/OR Multiple uncertainties have been identified but are unlikely to have a substantial impact on results.
Low	The characterization of variability is absent. AND/OR Key uncertainties, limitations, and data gaps are not discussed. AND/OR

Quality Rating	Description
	Uncertainties identified may have a substantial impact on the exposure the exposure assessment
Critically Deficient	Estimates are highly uncertain based on characterization of variability and uncertainty.
Not Rated/ Not Applicable	
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*

5.4.1.2 Data Evaluation Criteria for Experimental Data, as Revised

Table 5-2. Updated Evaluation Criteria for Experimental Data Sources

Data Quality Rating	Metric Description	
Domain 1. Reliability		
Metric 1. Sampling Methodology and Conditions		
High	Samples were collected according to publicly available SOPs, methods, protocols, or test guidelines that are scientifically sound and widely accepted from a source generally known to use sound methods and/or approaches such as EPA, NIST, American Society for Testing and Materials, ISO, and ACGIH. OR The sampling protocol used was not a publicly available SOP from a source generally known to use sound methods and/or approaches, but the sampling methodology is clear, appropriate (***i.e.,*** scientifically sound), and similar to widely accepted protocols for the chemical and media of interest. All pertinent sampling information is provided in the data source or companion source. Examples include: sampling conditions (***e.g.,*** temperature, humidity) sampling equipment and procedures sample storage conditions/duration performance/calibration of sampler	
Medium	Sampling methodology is discussed in the data source or companion source and is generally appropriate (***i.e.,*** scientifically sound) for the chemical and media of interest, however, one or more pieces of sampling information is not described. The missing information is unlikely to have a substantial impact on results. OR Standards, methods, protocols, or test guidelines may not be widely accepted, but a successful validation study for the new/unconventional procedure was conducted prior to the sampling event and is consistent with sound scientific theory and/or accepted approaches.	

Data Quality Rating	Metric Description	
Low	Sampling methodology is only briefly discussed. Therefore, most sampling information is missing and likely to have a substantial impact on results. AND/OR The sampling methodology does not represent best sampling methods, protocols, or guidelines for the chemical and media of interest (***e.g.,*** outdated (but still valid) sampling equipment or procedures, long storage durations). AND/OR There are some inconsistencies in the reporting of sampling information (***e.g.,*** differences between text and tables in data source, differences between standard method and actual procedures reported to have been used, etc.) which lead to a low confidence in the sampling methodology used.	
Critically Deficient	The sampling methodology is not discussed in the data source or companion source. AND/OR Sampling methodology is not scientifically sound or is not consistent with widely accepted methods/approaches for the chemical and media being analyzed (***e.g.,*** inappropriate sampling equipment, improper storage conditions). AND/OR There are numerous inconsistencies in the reporting of sampling information, resulting in high uncertainty in the sampling methods used.	
Not Rated/Not Applicable		
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*	
Metric 2. Analytical methodology		
High	Samples were analyzed according to publicly available analytical methods that are scientifically sound and widely accepted (**i.e.,** from a source generally using sound methods and/or approaches) and are appropriate for the chemical and media of interest. Examples include EPA SW-846 Methods, NIOSH Manual of Analytical Methods 5th Edition, etc. OR The analytical method used was not a publicly available method from a source generally known to use sound methods and/or approaches, but the methodology is clear and appropriate (***i.e.,*** scientifically sound) and similar to widely accepted protocols for the chemical and media of interest. All pertinent analytical ~~sampling~~ information is provided in the data source or companion source. Examples include: extraction method analytical instrumentation (required) instrument calibration LOQ, LOD, detection limits, and/or reporting limits recovery samples biomarker used (if applicable)	

Data Quality Rating	Metric Description	
	matrix-adjustment method (***i.e.,*** creatinine, lipid, moisture)	
Medium	Analytical methodology is discussed in detail and is clear and appropriate (***i.e.,*** scientifically sound) for the chemical and media of interest; however, one or more pieces of analytical information is not described. The missing information is unlikely to have a substantial impact on results. AND/OR The analytical method may not be standard/widely accepted, but a method validation study was conducted prior to sample analysis and is expected to be consistent with sound scientific theory and/or accepted approaches. AND/OR Samples were collected at a site and immediately analyzed using an on-site mobile laboratory, rather than shipped to a stationary laboratory.	
Low	Analytical methodology is only briefly discussed. Analytical instrumentation is provided and consistent with accepted analytical instrumentation/methods. However, most analytical information is missing and likely to have a substantial impact on results. AND/OR Analytical method is not standard/widely accepted, and method validation is limited or not available. AND/OR Samples were analyzed using field screening techniques. AND/OR LOQ, LOD, detection limits, and/or reporting limits not reported. AND/OR There are some inconsistencies or possible errors in the reporting of analytical information (***e.g.,*** differences between text and tables in data source, differences between standard method and actual procedures reported to have been used, etc.) which leads to a lower confidence in the method used.	
Critically Deficient	Analytical methodology is not described, including analytical instrumentation (***i.e.,*** HPLC, GC). AND/OR Analytical methodology is not scientifically appropriate for the chemical and media being analyzed (***e.g.,*** method not sensitive enough, not specific to the chemical, out of date). AND/OR There are numerous inconsistencies in the reporting of analytical information, resulting in high uncertainty in the analytical methods used.	
Not Rated/Not Applicable		
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*	

Data Quality Rating	Metric Description	
Metric 3. Selection of biomarker of exposure		
High	Biomarker in a specified matrix is known to have an accurate and precise quantitative relationship with external exposure, internal dose, or target dose (***e.g.,*** previous studies (or the current study) have indicated the biomarker of interest reflects external exposures). AND Biomarker (parent chemical or metabolite) is derived from exposure to the chemical of interest.	
Medium	Biomarker in a specified matrix has accurate and precise quantitative relationship with external exposure, internal dose, or target dose. AND Biomarker is derived from multiple parent chemicals, not only the chemical of interest, **but** there is a stated method to apportion the estimate to only the chemical of interest	
Low	Biomarker in a specified matrix has accurate and precise quantitative relationship with external exposure, internal dose, or target dose. AND Biomarker is derived from multiple parent chemicals, not only the chemical of interest, **and** there is NOT a stated method to apportion the estimate to only the chemical of interest. OR Biomarker in a specified matrix is a poor surrogate (low accuracy and precision) for exposure/dose.	
Critically Deficient	Not applicable. A study will not be deemed critically deficient based on the use of biomarker of exposure. ~~Biomarker in a specified matrix is a poor surrogate (low accuracy and precision) for exposure/dose.~~	
Not Rated/Not Applicable	Metric is not applicable to the data source.	
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*	
Domain 2. Representative		
Metric 4. Testing scenario		
High	Testing conditions closely represent relevant exposure scenarios (*i.e.*, population/scenario/media of interest). Examples include: 1. amount and type of chemical/product used 2. source of exposure/test substance 3. method of application or by-stander exposure 4. use of exposure controls	

Data Quality Rating	Metric Description	
	5. microenvironment (location, time, climate, temperature, humidity, pressure, airflow) AND Testing conducted under a broad range of conditions for factors such as temperature, humidity, pressure, airflow, and chemical mass/weight fraction (if appropriate).	
Medium	The data likely represent the relevant exposure scenario (*i.e.,* population/scenario/media of interest). One or more key pieces of information may not be described but the deficiencies are unlikely to have a substantial impact on the characterization of the exposure scenario. AND/OR If surrogate data, activities seem similar to the activities within scope.	
Low	The data lack multiple key pieces of information and the deficiencies are likely to have a substantial impact on the characterization of the exposure scenario. AND/OR There are some inconsistencies or possible errors in the reporting of scenario information (*e.g.*, differences between text and tables in data source, differences between standard method and actual procedures reported to have been used, etc.) which leads to a lower confidence in the scenario assessed. AND/OR If surrogate data, activities have lesser similarity but are still potentially applicable to the activities within scope. AND/OR Testing conducted under a single set of conditions, except for experiments to determine a weight fraction or concentration in a product.	
Critically Deficient	Testing conditions are not relevant to the exposure scenario of interest for the chemical.	
Not Rated/Not Applicable		
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*	
Metric 5. Sample size and variability		
High	Sample size is reported and large enough (*i.e.*, ≥ 10 samples) to be reasonably assured that the samples represent the scenario of interest. AND Replicate tests performed and variability across tests is characterized (if appropriate).	
Medium	Sample size is moderate (*i.e.*, 5 to ~~10~~ <10 samples), thus the data are likely to represent the scenario of interest.	

Data Quality Rating	Metric Description	
	AND Replicate tests performed and variability across tests is characterized (if appropriate).	
Low	Sample size is small (*i.e.*, <5 samples), thus the data are likely to poorly represent the scenario of interest. AND/OR Replicate tests were not performed.	
Critically Deficient	Sample size is not reported. AND/OR Single sample collected per data set, except for experiments to determine a weight fraction or concentration in a product. AND/OR For biomonitoring studies, the timing of sample collected is not appropriate based on chemical properties (*e.g.*, half-life), the pharmacokinetics of the chemical (*e.g.*, rate of uptake and elimination), and when the exposure event occurred.	
Not Rated/Not Applicable		
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*	
Metric 6. Temporality		
High	Source(s) of tested items appears to be current (within 5 years).	
Medium	Source(s) of tested items is less consistent with when current or recent exposures (>5 to 15 years) are expected.	
Low	Source(s) of tested items is not consistent with when current or recent exposures (>15 years) are expected or is not identified.	
Critically Deficient	Temporality of tested items is not reported, discussed, or referenced.	
Not Rated/Not Applicable		
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*	
Domain 3. Accessibility/clarity		
Metric 7. Reporting of results		

Data Quality Rating	Metric Description	
High	Supplementary or raw data (***i.e.,*** individual data points) are reported, allowing summary statistics to be calculated or reproduced. AND Summary statistics are detailed and complete. Example parameters include: Description of data set summarized (***i.e.,*** location, population, dates, etc.) Range of concentrations or percentiles Number of samples in data set Frequency of detection Measure of variation (CV, standard deviation) Measure of central tendency (mean, geometric mean, median) Test for outliers (if applicable) AND Both adjusted and unadjusted results are provided (**i.e.,** correction for void completeness in urine biomonitoring, whole-volume or lipid adjusted for blood biomonitoring) [only if applicable].	
Medium	Supplementary or raw data (***i.e.,*** individual data points) are not reported, and therefore summary statistics cannot be reproduced. AND/OR Summary statistics are reported but are missing one or more parameters (see description for high). AND/OR Only adjusted or unadjusted results are provided, but not both [only if applicable].	
Low	Supplementary data are not provided, and summary statistics are missing most parameters (see description for high). AND/OR There are some inconsistencies or errors in the results reported, resulting in low confidence in the results reported (***e.g.,*** differences between text and tables in data source, less appropriate statistical methods).	
Critically Deficient	There are numerous inconsistencies or errors in the calculation and/or reporting of results, resulting in highly uncertain reported results.	
Not Rated/Not Applicable		
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*	
Metric 8. Quality assurance		

Data Quality Rating	Metric Description	
High	The study applied quality assurance/quality control (QA/QC) measures and all pertinent QA/QC ~~quality assurance~~ information is provided in the data source or companion source. Examples include: 1. Laboratory, and/or storage recoveries. 2. Laboratory control samples. 3. Baseline (pre-exposure) samples. 4. Biomarker stability 5. Completeness of sample (*i.e.*, creatinine, specific gravity, osmolality for urine samples) AND No QA/QC ~~quality control~~ issues were identified, or any identified issues were minor and adequately addressed (*i.e.*, correction for low recoveries, correction for completeness).	
Medium	The study applied and documented ~~quality assurance/quality control~~ QA/QC measures; however, one or more pieces of QA/QC information is not described. Missing information is unlikely to have a substantial impact on results. AND No QA/QC ~~quality control~~ issues were identified, or any identified issues were minor and addressed (*i.e.*, correction for low recoveries, correction for completeness).	
Low	QA/QC ~~Quality assurance/quality control techniques~~ measures and results were not directly discussed but are ~~can be~~ implied through the study's use of standard field and laboratory protocols. AND/OR Deficiencies were noted in QA/QC ~~quality assurance/quality control~~ measures that are likely to have a substantial impact on results. AND/OR There are some inconsistencies in the QA/QC ~~quality assurance~~ measures reported, resulting in low confidence in the ~~quality assurance/control~~ QA/QC measures taken and results (*e.g.*, differences between text and tables in data source).	
Critically Deficient	QA/QC issues have been identified which significantly interfere with the overall reliability of the study.	
Not Rated/Not Applicable		
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*	
Domain 4. Variability and uncertainty		
Metric 9. Variability and uncertainty		

Data Quality Rating	Metric Description	
High	The study characterizes variability in the population/media studied. AND Key uncertainties, limitations, and data gaps have been identified. AND The uncertainties are minimal and have been characterized.	
Medium	The study has limited characterization of variability in the population/media studied. AND/OR The study has limited discussion of key uncertainties, limitations, and data gaps. AND/OR Multiple uncertainties have been identified but are unlikely to have a substantial impact on results.	
Low	The characterization of variability is absent. AND/OR Key uncertainties, limitations, and data gaps are not discussed. AND/OR Uncertainties identified may have a substantial impact on ~~the exposure~~ the exposure assessment	
Critically Deficient	Estimates are highly uncertain based on characterization of variability and uncertainty.	
Not Rated/Not Applicable		
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*	

5.4.1.3 Data Evaluation Criteria for Databases, as Revised

Table 5-3. Updated Data Evaluation Criteria for Database Data

Data Quality Rating	Description
Domain 1. Reliability	
Metric 1. Sampling methodology	
High	Widely accepted sampling methodologies (*i.e.*, from a source generally known to use ~~using~~ sound methods and/or approaches) were used to generate the data presented in the database. Example SOPs include USGS's "National Field Manual for the Collection of Water-Quality Data," EPA's "Ambient Air Sampling" (SESDPROC-303-R5), etc.

Data Quality Rating	Description
Medium	One or more pieces of sampling methodology information is not described, but missing information is unlikely to have a substantial impact on results. OR The sampling methodologies were consistent with sound scientific theory and/or accepted approaches based on the reported sampling information but may not have followed published procedures from a source generally known to use sound methods and/or approaches.
Low	The sampling methodology was not reported in data source or readily available companion data source.
Critically Deficient	The sampling methodologies used were not appropriate for the chemical/media of interest in the database (*e.g.*, inappropriate sampling equipment, improper storage conditions).
Not Rated/Not Applicable	
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 2. Analytical methodology	
High	Widely accepted analytical methodologies (*i.e.*, from a source generally using sound methods and/or approaches) were used to generate the data presented in the database. Example SOPs include EPA SW-846 Methods, NIOSH Manual of Analytical Methods 5th Edition, etc.
Medium	The analytical methodologies were consistent with sound scientific theory and/or accepted approaches based on the reported analytical information but may not have followed published procedures from a source generally known to use sound methods and/or approaches.
Low	The analytical methodology was not reported in data source or companion data source.
Critically Deficient	The analytical methodologies used were not appropriate for the chemical/media of interest in the database (*e.g.*, method not sensitive enough, not specific to the chemical, out of date).
Not Rated/Not Applicable	
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Domain 2. Representative	
Metric 3. Geographic area	
High	Geographic location(s) is reported, discussed, or referenced.
Medium	Not applicable. This metric is dichotomous (*i.e.*, high vs. critically deficient).
Low	Not applicable. This metric is dichotomous (*i.e.*, high vs. critically deficient).

Data Quality Rating	Description
Critically Deficient	Geographic location is not reported, discussed, or referenced.
Not Rated/Not Applicable	
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 4. Temporal	
High	The data reflect current conditions (within 5 years) AND/OR Database contains robust historical data for spatial and temporal analyses (if applicable).
Medium	The data are less consistent with current or recent exposures (>5 to 15 years) AND/OR Database contains sufficient historical data for spatial and temporal analyses (if applicable).
Low	Data are not consistent with when current exposures (>15 years old) may be expected AND/OR Database does not contain enough historical data for spatial and temporal analyses (if applicable).
Critically Deficient	Timing of sample data is not reported, discussed, or referenced.
Not Rated/Not Applicable	
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 5. Exposure scenario	
High	The data closely represent relevant exposure scenario (*i.e.*, the population/scenario/media of interest). Examples include: 1. Amount and type of chemical/product used 2. Source of exposure 3. Method of application or by-stander exposure 4. Use of exposure controls 5. Microenvironment (location, time, climate)
Medium	The data likely represent the relevant exposure scenario (*i.e.*, population/scenario/media of interest). **One or more key pieces of information may not be described** but the deficiencies are unlikely to have a substantial impact on the characterization of the exposure scenario. AND/OR If surrogate data, activities seem similar to the activities within scope.

Data Quality Rating	Description
Low	The data lack multiple key pieces of information and the deficiencies are likely to have a substantial impact on the characterization of the exposure scenario. AND/OR There are **some inconsistencies or possible errors** in the reporting of scenario information (*e.g.*, differences between text and tables in data source, differences between standard method and actual procedures reported to have been used, etc.) which leads to a lower confidence in the scenario assessed. AND/OR If surrogate data, activities have lesser similarity but are still potentially applicable to the activities within scope.
Critically Deficient	If reported, the exposure scenario discussed in the monitored study does not represent the exposure scenario of interest for the chemical.
Not Rated/Not Applicable	
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Domain 3. Accessibility/clarity	
Metric 6. Availability of database and supporting documents	
High	Database is widely accepted and/or from a source generally known to use sound methods and/or approaches (*e.g.*, raw data from NHANES, STORET).
Medium	The database may not be widely known or accepted (*e.g.*, state-maintained databases), but the database is adequately documented with most or all of the following information: 1. Within the database, metadata is present (sample identifiers, annotations, flags, units, matrix descriptions, etc.) and-data fields are generally clear and defined. 2. A user manual and other supporting documentation is available, or there is sufficient documentation in the data source or companion source. Database quality assurance and data quality control measures are defined and/or a QA/QC protocol was followed.
Low	The database may not be widely known or accepted, and only limited database documentation is available (see the medium rating).
Critically Deficient	No information is provided on the database source or availability to the public.
Not Rated/ Applicable	
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*

Data Quality Rating	Description
Metric 7. Reporting of results	
High	The database or information source reporting the analysis of the database data is well organized and understandable by the target audience. AND Summary statistics in the data source are detailed and complete. Example parameters include: 1. Description of data set summarized (*i.e.*, location, population, dates, etc.) 2. Range of concentrations or percentiles 3. Number of samples in data set 4. Frequency of detection 5. Measure of variation (CV, standard deviation) 6. Measure of central tendency (mean, geometric mean, median) 7. Test for outliers (if applicable)
Medium	The database or information source reporting the analysis of the database data is well organized and understandable by the target audience. AND/OR Summary statistics are missing one or more parameters (see description for high).
Low	The database or information source reporting the analysis of the database data is unclear or not well organized. AND/OR Summary statistics are missing most parameters (see description for high) AND/OR There are some inconsistencies or errors in the results reported, resulting in low confidence in the results reported (*e.g.*, differences between text and tables in data source, less appropriate statistical methods).
Critically Deficient	There are numerous inconsistencies or errors in the calculation and/or reporting of results, resulting in highly uncertain reported results. AND/OR The information source reporting the analysis of the database data is missing key sections or lacks enough organization and clarity to locate and extract necessary information.
Not Rated/Not Applicable	
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Domain 4. Variability and uncertainty	
Metric 8. Variability and uncertainty	
High	Variability, key uncertainties, limitations, and/or data gaps have been identified. AND/OR The uncertainties are minimal and have been characterized.

Data Quality Rating	Description
Medium	The study has limited discussion of variability, key uncertainties, limitations, and/or data gaps. AND/OR Multiple uncertainties have been identified but are unlikely to have a substantial impact on results.
Low	Variability, key uncertainties, limitations, and data gaps are not discussed. AND/OR Uncertainties identified may have a substantial impact on the exposure the exposure assessment
Critically Deficient	Estimates are highly uncertain based on characterization of variability and uncertainty.
Not Rated/Not Applicable	
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*

5.5 Environmental and Human Health Hazard

Details regarding the data quality evaluation and extraction of environmental and human health hazard information from references that passed PECO screening criteria are available in Sections 5 and 6.4 of the 2021 Draft Systematic Review Protocol. Data quality criteria for environmental studies, animal and *in vitro* toxicity studies and epidemiological studies are available in Appendices P, Q, and R in the 2021 Draft Systematic Review Protocol, respectively (U.S. EPA, 2021). The below-listed supplemental documents provide details of the data evaluated and extracted. Data evaluation information for each discipline (*i.e.*, environmental and human health hazard) is contained in separate supplemental documents and includes metric rating and the overall study quality determination for each data source. On the other hand, data extraction information for both disciplines are contained in a single supplemental document to increase the ease of accessing hazard data that may be relevant for both environmental- and human health-related receptors. One clarification that applies to the data extraction of human health hazard data is that all the data extraction was conducted in DistillerSR. Regarding environmental hazard references that meet full-text PECO screening criteria, the available environmental hazard data were extracted from those references in the ECOTOXicology Knowledgebase (ECOTOX) database and then imported into DistillerSR.

- *Draft Risk Evaluation for Tris(2-chloroethyl) phosphate (TCEP) – Systematic Review Supplemental File: Data Quality Evaluation Information for Environmental Hazard* (U.S. EPA, 2023h)
- *Draft Risk Evaluation for Tris(2-chloroethyl) phosphate (TCEP) – Systematic Review Supplemental File: Data Quality Evaluation Information for Human Health Hazard Animal Toxicology* (U.S. EPA, 2023j)
- *Draft Risk Evaluation for Tris(2-chloroethyl) phosphate (TCEP) – Systematic Review Supplemental File: Data Quality Evaluation Information for Human Health Hazard Epidemiology* (U.S. EPA, 2023k)

- *Draft Risk Evaluation for Tris(2-chloroethyl) phosphate (TCEP) – Systematic Review Supplemental File: Data Extraction Information for Environmental Hazard and Human Health Hazard Animal Toxicology and Epidemiology* (U.S. EPA, 2023b)

5.5.1 Environmental Hazard

As described in Appendix R of the 2021 Draft Systematic Review Protocol, references that met PECO screening criteria at full text screening underwent data quality evaluation (U.S. EPA, 2021). Likewise, for references that met PECO screening criteria at full text screening underwent data extraction as described in Section 6.4.1 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). One clarification regarding the extraction of environmental hazard data is that all of the extracted data, except those with confidential business information claims, will be available in the ECOTOX database, which is publicly available.

5.5.2 Human Health Hazard

As described in Appendices Q and R of the 2021 Draft Systematic Review Protocol, references that met PECO screening criteria at full text screening underwent data quality evaluation (U.S. EPA, 2021). Likewise, references that met PECO screening criteria at full text screening underwent data extraction as described in Section 6.4.1 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). Any clarifications or updates regarding the data quality evaluation or extraction of data from references that met PECO screening criteria at full text screening will be discussed further below for epidemiological and animal toxicity studies.

5.5.2.1 Epidemiology Studies

For TCEP, all references that met PECO screening criteria and were categorized as having potentially epidemiological information for the evaluation of human health hazard underwent data quality evaluation and data extraction as described in Appendix R and Section 6.4.1 of the 2021 Draft Systematic Review Protocol, respectively (U.S. EPA, 2021). There were no changes to the data evaluation domains and metrics or data extraction methodologies since the 2021 Draft Systematic Review Protocol was published.

5.5.2.2 Animal Toxicity Studies

Although there were no updates made to the data extraction methodologies described in the 2021 Draft Systematic Review Protocol for references with potentially relevant animal toxicity studies for the evaluation of human health hazard, EPA did update language in some of the metrics used to conduct data quality evaluation for those references. Below are updates to the data evaluation metrics from the versions published in Appendix Q.4.2 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). Language that was inserted since the draft protocol was published is **bolded**, and language removed is shown in ~~strikethrough~~. Language was removed from metric 12 to ensure the metric rating was not discounted due to assessment of liver metabolism via an injection pathway. The description for the rating of high for metric 12 was updated to clarify that the caveat of 10 or more air changes/hour applies only to dynamic whole-body chambers. For metrics not listed below, no changes were made since the 2021 was published (U.S. EPA, 2021). Draft Systematic Review Protocol was published.

Table 5-4. Updated Data Quality Evaluation Criteria for Animal Toxicity Studies

Data Quality Rating	Description
Metric 12. Exposure route and method Were the route and method of exposure reported and suited to the test substance (*e.g.*, accounting for volatility, ~~injection was not used for assays of liver metabolism~~, an appropriate vehicle was used when necessary)? For nose-only or head-only inhalation studies, were the animals appropriately acclimated or was the lack of acclimation controlled for?	
High	The route and method of exposure were reported and were suited to the test substance (see above) For inhalation studies, a dynamic, nose-only or head-only chamber was used~~. with greater than 10 or more air changes/hour~~. While dynamic nose-only (or head-only) studies are generally preferred, dynamic whole-body chambers are acceptable for gases **as long as there were 10 or more air changes/hour**.
Medium	There were minor limitations regarding the route and method of exposure, but the researchers took appropriate steps to mitigate the problem (*e.g.*, attempted to minimize headspace for volatile compounds in drinking water). These limitations are unlikely to have a substantial impact on results. For inhalation studies, a dynamic whole-body chamber was used for vapors that may condense (assume most will condense at high concentrations unless otherwise stated) or for aerosols, having 10 or more air changes/hour. A medium rating can also be assigned if the study indicates a dynamic chamber but not the number of air changes.
Low	There were deficiencies regarding the route and method of exposure that are likely to have a substantial effect on results. Researchers may have attempted to correct the problem, but the success of the mitigating action was unclear. For inhalation studies, there are significant flaws in the design or operation of the inhalation chamber, such as uneven distribution of test substance in a whole-body chamber, having less than 10 air changes/hour in a whole-body chamber, or using a whole-body chamber that is too small for the number and volume of animals exposed. OR Only very minimal if any details about the methods for inhalation exposure administration (as described above) were reported, resulting in significant uncertainty about the true exposure parameters.
Critically Deficient	The route or method of exposure was not reported OR An inappropriate route or method (*e.g.*, administration of a volatile organic compound via the diet) was used for the test substance without taking steps to correct the problem (*e.g.*, mixing fresh diet). These are serious flaws that makes the study unusable. For inhalation studies, either a static chamber was used, there is no description of the inhalation chamber, or an atypical exposure method was used, such as allowing a container of test substance to evaporate in a room.
Not Rated/Not Applicable	Do not select for this metric.

Data Quality Rating	Description
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*

5.6 Dermal Absorption

EPA conducts data quality evaluation and extraction for references that contain potentially relevant information considered to inform the human dermal absorption for TSCA risk evaluations using DistillerSR. To evaluate the studies, data evaluation metrics were developed as described in Appendix S in the publication of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). The *in vitro/ex vivo* criteria are based on *in vitro* criteria used for mechanistic studies. As described in Section 5.3 of the 2021 Draft Systematic Review Protocol, both an initial and QC review will be conducted per references for data quality evaluation (U.S. EPA, 2021). The sections below identify updates to the criteria made since publication of the 2021 Draft Systematic Review Protocol.

Data extraction involves the cataloguing of experimental methodologies as well as the results from references that met the PECO screening criteria described above in Section 4.7. Some examples of the information that will be extracted include skin material/species was used, type and thickness of skin preparation, diffusion cell exposure set up, occlusion used, vehicle in the donor chamber, and concentration of the test substance. Because of complexities with the dermal absorption information, both Kp/flux information as well as fraction absorption information will be extracted. EPA included this relevant information in extraction forms developed for DistillerSR. The data entered into DistillerSR will then be coded for output into tables that are included in the published risk evaluations.

The 3D human skin equivalent models are not recommended by OECD Guidance (OECD Series on Testing and Assessment No. 156 (September 2022)) (OECD, 2022) for use in evaluating risks. Therefore, if adequate data are available from *in vivo* or *in vitro/ex vivo* (excised skin) studies, EPA will not (1) evaluate the studies using 3D human skin for data quality; (2) extract information from the studies; or (3) consider the studies when developing quantitative dermal absorption estimates for the TSCA risk evaluations. However, EPA may discuss the 3D models when integrating evidence and may consider evaluating them if no other experimental dermal absorption information is available. *Draft Risk Evaluation for Tris(2-chloroethyl) phosphate (TCEP) – Systematic Review Supplemental File: Data Quality Evaluation and Data Extraction Information for Dermal Absorption* provides details of the data extracted and evaluated, including metric rating and the overall study quality determination for each data source (U.S. EPA, 2023d).

5.6.1 Data Quality Metrics – *In Vitro/Ex Vivo*

Shown below is the dermal absorption *in vitro/ex vivo* data evaluation table that has been modified from Appendix S of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021) (Table 5-5). The scope of the data evaluation was expanded to include the evaluation of *ex vivo* dermal absorption references. Language that was inserted is **bolded**, and language removed is shown in ~~strikethrough~~. For metrics 1, 3, 5, 6, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, and 21 changes were made to the original description to provide context and/or clarity to the evaluation question and/or ratings for the metric. Because the data evaluation form was not finalized prior to the publication of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021), language was added to metrics 4, 5, 7, 10 in the places that were marked as "TBD" in Appendix S. For metric 4, the description for the original rating of "medium" was used instead to indicate a high rating for TCEP, and the description for a medium rating has been modified.

Clarifications were also made to the descriptions for the low and "critically deficient" ratings. For metric 8, the rating of high was removed, and that description was incorporated into the medium rating.

Table 5-5. Updated Data Evaluation Criteria for *In Vitro/Ex Vivo* Dermal Absorption Studies

Data Quality Rating	Description
Domain 1. Test substance	
Metric 1. Test substance identity Was the test substance identified definitively (*i.e.*, established nomenclature, CASRN, physical nature, physical and chemical properties, and/or structure reported, including information on the specific form tested [*e.g.*, salt or base, valence state, isomer, if applicable] for materials that may vary in form)? If test substance was a mixture, were mixture components and ratios characterized?	
High	The test substance (*i.e.*, chemical of interest) was identified definitively (*i.e.*, nomenclature, CASRN, structure) and where applicable the specific form (*e.g.*, particle characteristics for solid state materials, salt or base, valence state, hydration state, isomer, radiolabel, etc.) was definitively and completely characterized. For mixtures, the components and ratios were characterized (*i.e.*, provided as concentration, ratio of percentage of the mixture or product). Additionally, for radiolabeled substances, the location of the radiolabel within the substance should be indicated, ideally with **the radiolabel** ^{14}~~C~~ in a metabolically stable position.
Medium	The test substance (*i.e.*, chemical of interest) was identified and the specific form was characterized (where applicable). For mixtures, some components and components and ratios were identified and characterized but at least the chemical of interest has a percentage/concentration reported. There were minor uncertainties (*e.g.*, minor characterization details were omitted **such as about the** radiolabel ~~details~~) that were unlikely to have a substantial impact on results.
Low	The test substance and form (if applicable) were identified, and the components and ratios of mixtures were characterized, but there were uncertainties regarding test substance identification or characterization that are likely to have a substantial impact on the results (*e.g.*, no information on isomer (or enantiomer) composition of differences could affect toxicokinetic properties, limited particle size information, omitted details regarding branched or straight chain structure).
Critically Deficient	The test substance identity and form (the latter if applicable) could not be determined from the information provided (*e.g.*, nomenclature was unclear and CASRN or structure were not reported) OR For mixtures, the components and ratios were not characterized.
Not Rated/Not Applicable	Do not select for this metric
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*

Data Quality Rating	Description
Metric 2. Test substance source Was the source of the test substance reported, including manufacturer and batch/lot number for materials that may vary in composition? If synthesized or extracted, was test substance identity verified by analytical methods?	
High	The source of the test substance was reported as a manufacturer or the production process was specifically identified. The batch/lot number was identified (for materials that may vary in composition), and the chemical identity was either certified by the source in the publication or could be verified on a manufacturer's website. OR The test substance identity was analytically verified by the laboratory that performed the toxicity study.
Low	The test substance was synthesized or extracted by a source other than the manufacturer [and no production process was identified]. OR The source was not reported. AND The test substance identity was NOT analytically verified by the performing laboratory.
Not Rated/Not Applicable	Do not select for this metric
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 3. Test substance purity Was the purity or grade (*i.e.*, analytical, technical) of the test substance (including the radiolabeled substance) reported and adequate? Were impurities identified? Were impurities present in quantities that could influence the results?	
High	For discrete substances, the test substance **(including radiolabel)** purity and composition were such that any observed effects were highly likely to be due to the nominal test substance itself (*e.g.*, highly pure at >98% or analytical grade test substance or a formulation of lower purity that contains ingredients considered to be inert, such as water). All components, including impurities and residual chemicals, were identified and the chemical of interest was the main component.
Medium	The nature and quantity of reported impurities are such that study results were not likely to be substantially impacted by the impurities (impurities not known to induce outcome of interest at low levels, impurities are inert or GRAS, etc.). Regardless of the nature and purity, for discrete chemicals, the purity of the chemical of interest should be >70%, unless water is the only impurity.

Data Quality Rating	Description
Low	Purity and/or grade of test substance were not reported
Critically Deficient	The nature and quantity of reported impurities were such that study results were likely to be due to one or more of the impurities. ~~This is a serious flaw that makes the study unusable.~~ **AND/OR** **For discrete chemicals, purity was <70% with an impurity other than water.**
Not Rated/Not Applicable	Do not select for this metric
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Domain 2. Test design	
Metric 4. Reference compounds Were the results of a reference compound (*e.g.*, caffeine, testosterone, benzoic acid) run concurrently or separately and recently by the same laboratory and reported in the study? Was the absorption response appropriate? **Alternately, has the performing lab demonstrated previous technical sufficiency in dermal absorption studies?** ~~[TBD: need to decide how important it is to have reference compounds]~~	
High	**An appropriate concurrent reference compound was tested or data from a historical reference compound was provided, and an appropriate response was observed. Any uncertainties (*e.g.*, omission of minor details regarding exposure or response) are minor.**
Medium	~~When applicable, an appropriate concurrent or historical reference compound was used, and an appropriate response was observed. Any uncertainties (*e.g.*, omission of minor details regarding exposure or response) are minor.~~ **An appropriate concurrent or historical reference compound was used, but there were some deficiencies regarding the reference compound exposure or response (*e.g.*, the response was not well described, it is unclear whether the response was acceptable).**
Low	~~When applicable, an appropriate concurrent or historical reference compound was used, but there were deficiencies regarding the reference compound exposure or response (*e.g.*, the response was not described).~~ ~~OR~~ ~~No reference compound was used or reported.~~ **No appropriate reference compound was used or reported AND there is no established history of test performance in the performing laboratory.**

Data Quality Rating	Description
Critically Deficient	Reference compounds were run but ~~an inadequate response for the reference compounds (outside historical controls results~~) indicates ~~that the assay would not accurately measure absorption.~~ **the response was unacceptable (*e.g.*, outside historical control results), raising concerns about the validity of the assay.**
Not Rated/Not Applicable	Do not select for this metric.
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 5. Assay procedures Were assay methods and procedures (*e.g.*, diffusion cell set up, temperature, humidity, physiological ~~conductivity~~ **compatibility** of receptor fluid, volumes applied ~~and surface area of skin~~, **amount of test substance per surface area of skin**, use/measurement of occlusion or carbon trap, materials and procedures used for tape stripping, capture of volatile compounds if required) described in detail and applicable/justified? ~~See other metrics for additional assay procedures (*e.g.*, metrics 1–3 for test substance information; metric 11 for exposure duration; metric 15 for replicates per group).~~ **Do the study methods describe how they ensure that quantification of the receptor fluid is adequately sensitive (*e.g.*, sufficient signal-to-noise ratio, high enough specific activity of radiolabel, sufficient amount of time or number of scintillations detected).** **Diffusion cell setup should indicate static vs flow-through, and for flow-through the flow rate should be indicated.** ~~OECD 428, OECD GD28 and OECD GD156 should be consulted and used to consider quality ratings.~~	
High	Study authors described the methods and procedures (*e.g.*, diffusion cell set up, temperature, humidity, physiological ~~conductivity~~ **compatibility** of receptor fluid, volumes applied and surface area of skin, use/measurement of occlusion or carbon trap, **specific activity of radiolabel**, materials and procedures used for tape stripping, capture of volatile compounds if required) used for the test in detail **and justified any relevant choices**. Either a static cell or flow-through system was used, with either constant stirring (static cell) or an appropriate flow- rate (flow-through). **These methods were appropriate based on the TGs and GDs above.**
Medium	Methods and procedures were partially described **(*e.g.*, all but temperature and humidity are described)** but appeared to be appropriate ~~(*e.g.*, TBD)~~, so the omission of details is unlikely to have a substantial impact on results.
Low	The methods and procedures were not well described or deviated from customary practices **(*e.g.*,** ~~TBD~~ **absence of occlusion or carbon trap for volatile test substance)** and this is likely to have a substantial impact on results, **however conservative statistical adjustments could possibly account for these deviations.**
Critically Deficient	Assay methods and procedures were not appropriate **and would result in unusable data that cannot be statistically accounted for** (*e.g.*, ~~TBD~~ **failure to use a diffusion cell with sufficient seal, too low volume/mass of test substance applied per surface area, tape stripping and wash fractions combined and not measured independently).**

Data Quality Rating	Description
Not Rated/Not Applicable	Do not select for this metric
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 6. Standards for tests For assays with established criteria, were the test validity, acceptability, reliability, and/or QC criteria reported and consistent with current standards and guidelines? **Were sufficient data provided to determine that the standards/guidelines have been met?** Example criteria: *Percent recovery:* 100±10% of the radioactivity as stated in OECD TG 428; 100±20% for volatile and unlabeled compounds as stated in OECD GD 28. ***Coefficient of Variation*: Variance across replicates should be measured and indicated when standard deviation exceeds 25%.** ***Skin integrity:* (1) Tritiated water – a.) a 'limit value' for a maximum Kp of 4.5 x10 $^{-3}$ cm/h (Guth et al., 2015; Meidan and Roper, 2008) and mean Kp of 2.5 x 10-3 cm/h (Bronaugh and Franz, 1986) for human *ex vivo* skin and b.) percent absorption (≤ 0.6% of applied dose in 1 hr) (Learn et al., 2018).** **(2) Electrical conductance - minimal threshold of 17 kilo-ohms (Fasano et al., 2002). (3) Trans-epidermal water loss - Less than 10 grams/m^2/hr (Zhang et al., 2018) (4) Other internal reference standard methods (*e.g.*, 3H-labeled compounds, methylene blue) as cited in (Guth et al., 2015)** ~~*Skin integrity*: (1) Tritiated water – minimal flux threshold TBD (2) Electrical conductance - minimal threshold of 17 kilo-ohms (Fasano et al., 2002).~~ ~~OECD 428, OECD GD28, and OECD GD156 should be consulted; deviations should be explained.~~	
Medium	**Criteria used to determine the** ~~The test~~ validity acceptability, reliability, and/or **quality of the experiment** ~~QC criteria~~ (*e.g.*, threshold for skin integrity, percent recovery considered acceptable) were reported and consistent with current standards and guidelines, as/if applicable **and authors stated that results met those criteria or the results provided enough detail to compare with the criteria**
Low	**Few or no QC criteria were reported, however, the reported results provided enough information to evaluate how the study compared against the criteria stated in the study and/or external criteria and standards.** ~~Some QC criteria were not reported~~.
Critically Deficient	Inadequate **information was provided on the standards used to evaluate the study results AND 1) the authors did not report whether the test met pre-established criteria,** **OR** 2**) inadequate** data **on results** were **presented** ~~provided~~ to demonstrate **the** validity, acceptability, and reliability of the test when compared with current standards and guidelines **or the pre-established standards/criteria identified by the authors. In this case, adequate QC cannot be performed.**

Data Quality Rating	Description
Not Rated/Not Applicable	Do not select for this metric
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Domain 3. Exposure characterization	
Metric 7. Preparation and storage of test substance (chemical) Did the study characterize preparation of the test substance and storage conditions? Were the frequency of preparation and/or storage conditions appropriate to the test substance stability and solubility (if applicable)?	
High	The test substance preparation and/or storage conditions (*e.g.*, test substance stability, homogeneity, mixing temperature, stock concentration, stirring methods, storage conditions) were reported and appropriate for the test substance (*e.g.*, stability and solubility in diluents or solvents confirmed especially if they differ from what is used commercially; volatile test substances prepared and stored in sealed containers; same stock solution for all exposure concentrations).
Medium	The test substance preparation and storage conditions were reported, but minor limitations in the test substance preparation and/or storage conditions were identified (*e.g.*, **test substance formulations were stirred instead of centrifuged for a specific number of rotations per minute** ~~TBD~~). OR There is an omission of details that are unlikely to have a substantial impact on results (*e.g.*, preparation/administration of test substance is described, but storage is not reported; however, storage is unlikely to affect results based on likely stability over the time frame of the test or the physical and chemical properties of the chemical make concerns about volatility or solubility unlikely).
Low	Deficiencies in reporting of test substance preparation, and/or storage conditions are likely to have a substantial impact on results (*e.g.*, available information on physical and chemical properties suggests that stability and/or solubility of test substance in diluent/solvent may be poor). OR Information on preparation and storage was *not* reported and lack of details could substantially impact results (*e.g.*, preparation for volatile or low-solubility chemicals).
Critically Deficient	Serious flaws reported regarding test substance preparation and/or storage conditions will have critical impacts on dose/concentration estimates and make the study unusable (*e.g.*, instability of test substance, test substance volatilized rapidly from storage containers).
Not Rated/Not Applicable	Do not select for this metric

Data Quality Rating	Description
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 8. Consistency of exposure administration Were exposures administered consistently across study groups (*e.g.*, consistent volumes and area of skin surface for application)?	
~~High~~	~~Details of exposure administration were reported and exposures were administered consistently across study groups in a scientifically sound manner (*e.g.*, consistent volumes, thickness and area of skin surface for application,).~~
Medium	Details of exposure administration were reported or inferred from the text, **and** ~~but the minor limitations in administration of~~ exposures **were administered consistently across study groups in a scientifically sound manner** (*e.g.*, **consistent volumes** ~~slight variation in volume~~, thickness and area **of** ~~or~~ skin surface ~~used~~ for application). **Any minor deviations/limitations are considered)** ~~that were identified~~ are unlikely to have a substantial impact on results. OR Details of exposure administration are incompletely reported, but the missing information is unlikely to have a substantial impact on results.
Low	Details of exposure administration were reported, but deficiencies in administration of exposures (*e.g.*, moderate differences in volume, thickness, and area of skin surface used for application) that were reported or inferred from the text are likely to have a substantial impact on results. OR Details of exposure administration are insufficiently reported and the missing information is likely to have a substantial impact on results
Critically Deficient	Exposures were not administered consistently across and/or within study groups (*e.g.*, large differences in volume, thickness, and area of skin surface used for application) resulting in serious flaws that make the study unusable.
Not Rated/Not Applicable	Do not select for this metric.
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 9. Reporting of concentrations Were exposure doses/concentrations or amounts of test substance reported without ambiguity (*e.g.*, point estimate instead of range, analytical instead of nominal)? Note: Ambiguity also applies to doses/concentrations if values were only reported as points on a figure without numerical values.	

Data Quality Rating	Description
High	The exposure doses/concentrations or amounts of test substance were reported without ambiguity (*e.g.*, point estimate instead of range, analytical/measured instead of nominal).
Medium	The exposure doses/concentrations or amounts of test substance were reported with some ambiguity (*e.g.*, range instead of point estimate OR nominal instead of analytical/measured).
Low	The exposure doses/concentrations or amounts of test substance were reported but with substantial ambiguity about precision (*e.g.*, only an estimated range AND only nominal instead of analytical measurements).
Critically Deficient	The exposure doses/concentrations or amounts of test substance were not reported, resulting in serious flaws that make the study unusable.
Not Rated/Not Applicable	Do not select for this metric.
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 10. Exposure duration Was the exposure duration (*e.g.*, hours) reported and was it appropriate for this study type and/or outcome(s) of interest? **Was the duration of exposure relevant to conditions of use and physical-chemical properties of the test substance? Did measurements continue post-exposure to account for retained dose in skin?** ~~[TBD: add text about human exposure relevancy]~~.	
High	The exposure duration (*e.g.*, hours) was reported and was appropriate for the study type and/or outcome(s) of interest (*e.g.*, at least 6 to 10 hours prior to washing and up to at least 24 hours **total including** post-washing). A shorter exposure duration may also be included but is less useful unless the substance is demonstrated to be volatile, **the results demonstrate that absorption approached completion (*e.g.*, nothing left in the skin wash or tape strip samples),** or the timepoint is used only for Kp/flux measurements.
Low	The duration(s) of exposure differed slightly from current standards and guidelines for studies of this type (*e.g.*, <6 to 10 hours prior to washing and less than 24 hours **total including** post-washing), **and** ~~but~~ the differences **may** ~~are unlikely to~~ have a substantial impact on results.
Critically Deficient	No information on exposure duration(s) was reported OR the exposure duration was not appropriate OR Duration(s) differed significantly from studies of the same or similar types **and these differences (most likely shorter duration)** . ~~These deficiencies~~ are likely to have a substantial impact on **interpretation of** results.

Data Quality Rating	Description
Not Rated/Not Applicable	Do not select for this metric.
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 11. Number of exposure groups and concentrations spacing Were the number of exposure groups/**tested concentrations** and dose/concentration spacing **appropriate and** justified by study authors (*e.g.*, to mimic a specific type of human exposure) and adequate for addressing the purpose of the study **across a wide range of conditions of use (COUs) (*e.g.*, dilute, concentrated, and neat)?(***e.g.*, to evaluate dermal absorption)?	
High	**There were three or more dose** ~~The number of exposure~~ groups **tested** and dose/concentration spacing were justified by study authors (*e.g.*, to mimic a specific type of human exposure) and **were** ~~was~~ adequate for addressing the purpose of the study.
Low	There were ~~minor~~ limitations regarding the number of exposure groups and/or applied dose/concentration spacing (*e.g.*, unclear if lowest dose was low enough or the highest dose was high enough, **or less than three doses/concentrations tested), restricting the applicability of the results to only a subset of COUs and weight fractions**~~.), but the number of exposure groups and spacing of exposure levels were adequate and are unlikely to have a substantial impact on results.~~
Critically Deficient	The number of exposure groups and dose/concentration spacing were not reported ~~OR~~ ~~the number of exposure groups and dose/concentration spacing were not adequate and did not mimic expected human exposures.~~
Not Rated/Not Applicable	Do not select for this metric.
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Domain 4. Test model	
Metric 12. Test model **(skin)** Were the test models (*e.g.*, viable skin, cadaver/cosmetic surgery skin, **animal skin**) and descriptive information (*e.g.*, tissue origin, anatomical site, tissue storage, **initial** integrity or viability) reported? What was the source of the test model? Was the model routinely used for the outcome of interest? For example, for human skin, split thickness (200–400µm), dermatomed skin is preferred.	

Data Quality Rating	Description
High	The test model (*e.g.*, viable skin, cadaver skin, cosmetic surgery **skin, animal** skin) and descriptive information (*e.g.*, tissue origin, anatomical site, tissue storage, integrity or viability, **lot/batch used**) were reported and the test model was routinely used for the outcome of interest.
Low	The test model was **insufficiently** reported **and** ~~reporting along with limited descriptive information.~~ ~~OR~~ ~~The test model was routinely used for the outcome of interest.~~ **r**eporting limitations **may** ~~are unlikely to~~ have a substantial impact on results.
Critically Deficient	The test model and necessary descriptive information were not **at all** reported OR the test model was not appropriate for evaluation of **dermal absorption** ~~the specific outcome of interest~~
Not Rated/Not Applicable	Do not select for this metric.
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 13. Number/Replicates per group Was the number of replicates per dose/concentration group appropriate for the study type and outcome analysis?	
Medium	The number of replicates per dose/concentration were reported and was appropriate (*e.g.*, acceptable data from a minimum of four replicates per test preparation).
Low	The number of replicates per dose/concentration **and timepoint** was reported but was less than recommended by current standards and guidelines **(*i.e.*, less than four replicates for each test preparation according to OECD TG 428)**. This is likely to have an impact on results. OR The number of replicates per dose/concentration was not reported.
Critically Deficient	The number of organisms or tissues per study group and/or replicates per study group was insufficient to characterize dermal absorption (*e.g.*, less than four replicates per test preparation produced acceptable data).
Not Rated/Not Applicable	**Do not select for this metric.** ~~Not applicable for qualitative studies not requiring any statistics.~~

Data Quality Rating	Description
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Domain 5. Outcome assessment	
Metric 14. Outcome assessment methodology Did the outcome assessment methodology address or report the intended outcome(s) of interest? Was the outcome assessment methodology (including ~~nature of endpoints evaluated~~, measurement technique and timing of measurement[s]) **appropriate** ~~sensitive~~ for the **associated conditions of use (COUs)**~~outcome(s) of interest (*e.g.*, measured endpoints that are able to detect a true effect~~)? OECD 428, OECD GD28 and **the dosing scenario?** ~~OECD GD156 should be consulted, and deviations should be documented and explained.~~	
High	The outcome assessment methodology addressed the intended outcome(s) of interest AND was sensitive for the outcome(s) of interest and followed OECD guidance documents. The **selected formulations are reasonable for the chemical of interest and would result in** ~~dosing reflected~~ a **sufficiently conservative estimate representative** ~~range~~ of conditions of use **for the chemical of interest (*e.g.*, use of IPM diluent).** ~~(COUs) to which humans are exposed~~. The infinite dose scenario **should be used** ~~is optimum~~ for Kp determinations while finite dosing is **required** ~~optimal~~ for **percent**~~%~~ absorption calculations. **For finite**~~The~~ dose **conditions, normally 1-5 mg/cm^2 of** ~~in the~~ **skin for a solid, and up to 10 µL/cm^2 for liquids of test material should be loaded, unless otherwise justified. For dilutions (*i.e.*, not neat test material), finite** ~~should be considered to be the potentially absorbable~~ dose **testing for each concentration of should ideally be conducted with application of 10 µL/cm^2 test material. For infinite dose testing of solids, it is required that at least 10 mg/cm^2 of pure substance be used to establish an undepletable dose, regardless of concentration. For infinite dose testing of liquids, at least 100 µL/cm^2 of pure substance should be used** to **establish an undepletable dose, regardless of concentration.** ~~calculate the final % absorption. Recovery is 90±10% or 80±20% for volatile substances.~~
Medium	The outcome assessment methodology used partially addressed the intended outcomes(s) of interest and deviations were explained, ~~(*e.g.*, mutation frequency evaluated in the absence of cytotoxicity in a gene mutation test)~~, but minor uncertainties **(*e.g.*, dosing was slightly below or above the recommendations for finite or infinite scenarios)** are unlikely to have a substantial impact on results.
Low	Significant deficiencies in the implementation of the reported outcome assessment methodology were identified (*e.g.*, **a volatile diluent was used with a volatile test substance** ~~matrix/assay interference, assay yielded anomalous results~~, etc.) **OR** The outcome assessment methodology was not clearly reported and it was unclear whether methods were sensitive for the outcome of interest. This is likely to have a substantial impact on results.
Critically Deficient	The reported assessment methodology was not sensitive to the outcome(s) of interest. For example, **percentage absorption was determined only from an infinite dose, and**/~~the reported measurement endpoint(s)~~ or **Kp/flux was derived from a finite dose, and statistics could** ~~timing were~~ not **easily be calculated independently.** ~~sensitive for~~

Data Quality Rating	Description
	~~the outcome(s) of interest (*e.g.*, cells were evaluated for chromosomal aberrations immediately after exposure to the test substance instead of after post-exposure incubation period).~~ These are serious flaws that make the study unusable.
Not Rated/Not Applicable	Do not select for this metric.
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 15. Consistency of outcome assessment Was the outcome assessment carried out consistently (*i.e.*, using the same protocol) across study groups (*e.g.*, assessment at the same time after initial exposure in all study groups)?	
High	Details of the outcome assessment protocol were reported and outcomes were assessed consistently across study groups (*e.g.*, at the same time after initial exposure) using the same protocol in all study groups. All study groups utilized the same **vehicle for the** blank formulation as **for the study concentration groups** ~~a vehicle~~, the duration of exposure was the same across groups, the same receptor fluid composition was **used** ~~utilized~~ for each group, the sampling period was consistent across groups, etc.
Medium	There were minor differences in the timing of outcome assessment across study groups, or incomplete reporting of minor details of outcome assessment protocol execution were explained, but these uncertainties or limitations are unlikely to have substantial impact on results.
Low	Details regarding the execution of the study protocol for outcome assessment (*e.g.*, timing of assessment across groups) were confusing, limited, or not reported nor deviations explained (or cited to another publication with no description in the paper itself), and these deficiencies are likely to have a substantial impact on results.
Critically Deficient	There were large inconsistencies in the execution of study protocols for outcome assessment across study groups. These are serious flaws that make the study unusable.
Not Rated/Not Applicable	Do not select for this metric.
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 16. Sampling adequacy **and sensitivity** Was the reported sampling size adequate for the outcome(s) of interest, including number of evaluations per exposure group, and endpoint (*e.g.*, **scintillation counts/sample)?**~~number of slides/cells/metaphases evaluated per~~	

Data Quality Rating	Description
~~test concentration)? OECD 428, OECD GD28, and OECD GD156 should be consulted, deviations should be explained.~~	
High	The study reported adequate sampling for the outcome(s) of interest including number of evaluations per exposure group, and **measurement sensitivity** ~~endpoint~~ (*e.g.*, scintillation counts/sample **and/or duration of radioactivity detection, adequate signal to noise [*i.e.*, background] ratio for detection [*e.g.*, signal 3x noise**]). The sampling intervals should be adequate to allow accurately graphically representing the results of the receptor fluid content of the test article versus time.
Medium	Details regarding sampling for the outcome(s) of interest were reported, but minor limitations were identified in the reported sampling of the outcome(s) of interest and were explained. However, those limitations are unlikely to have a substantial impact on results.
Low	Details regarding sampling of outcomes were not fully reported nor explained and the omissions are likely to have a substantial impact on results.
Critically Deficient	Reported sampling was not adequate for the outcome(s) of interest and/or serious uncertainties or limitations were identified in how the study carried out the sampling of the outcome(s) of interest (*e.g.*, replicates from control and test concentrations were evaluated at different times).
Not Rated/Not Applicable	**N/A** ~~NA~~ should be used for assays/studies that do not require a certain number of slides/cells/metaphases etc. be sampled for scoring (*i.e.*, mutagenicity assays, mechanistic studies).
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Domain 6. Confounding/variable control	
Metric 17. Confounding variables in test design and procedures Were there confounding differences among the study groups in the size, and/or quality of tissues exposed that could influence the outcome assessment, **(*e.g.*, skin integrity)**?	
High	There were no differences reported among study group parameters (*e.g.*, test substance lot or batch, strain/batch/ lot number of ~~organisms~~ ~~or models used per group or size~~ **skin samples used per group or size**, and/or quality of tissues exposed) that could influence the outcome assessment. Skin integrity was ~~acceptable~~ **measured by preferable methods (*e.g.*, electrical resistance and TEWL). Results of skin integrity testing were acceptable for all replicates and exposure groups (*e.g.*, > 17 kilo-ohms based on electrical resistance, less than 10 grams/m2/hr)**

Data Quality Rating	Description
Medium	Minor differences were reported and explained in initial conditions that are unlikely to have a substantial impact on results (*e.g.*, tissues from two different lots were used and QC data were similar for both lots). Skin integrity ~~had variability but were acceptable~~ **was measured by a less desirable method (*e.g.*, tritiated water) , but results were acceptable (*e.g.*, a 'limit value' for Kp of 4.5 x10 $^{-3}$ cm/h or percent absorption of $\leq$ 0.6% of applied dose in 1 hr).** Outliers were statistically evaluated. **Most results of skin integrity testing were acceptable, and the number of replicates/donors was adequate after excluding any unacceptable results.**
Low	Initial strain/batch/lot number skin samples used per group, size, and/or quality of tissues exposed was not reported. These deficiencies are likely to have a substantial impact on results.
Critically Deficient	There were significant differences among the study groups with respect to the strain/batch/lot number of ~~organisms or models used per group or size and/or quality of tissues exposed (*e.g.*, initial number of viable bacterial cells were different for each replicate [105 cells in replicate 1, 108 cell in replicate 2, and 103 cells in replicate 3], tissues from two different lots were used for in vitro skin corrosion test, but the control batch quality for one lot was outside of the acceptability range). Skin integrity results were below thresholds. Recovery was below guidance limits or not quantified. Exposures did not reflect worker COUs.~~ **skin samples used per group or size and/or quality of tissues exposed (*e.g.*, several replicates demonstrated integrity issues). Recovery varied greatly among replicates (*i.e.* >10%). In this situation, results are not reliable for estimating actual absorption.**
Not Rated/Not Applicable	Do not select for this metric.
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 18. Confounding variables in outcomes unrelated to exposure Were there differences among the study groups unrelated to exposure to test substance (*e.g.*, **solubility in receptor fluid** ~~contamination~~) that could influence the outcome assessment? Did the test material interfere in the assay (*e.g.*, altering fluorescence or absorbance, signal quenching by heavy metals, altering pH, solubility, or stability issues)?	
High	There were no reported differences among the study replicates or groups in test model unrelated to exposure (*e.g.*, **solubility in receptor fluid** ~~contamination~~) and the test substance did not interfere with the assay (*e.g.*, signal quenching by heavy metals). The test substance was demonstrated to be soluble in the receptor fluid.
Medium	Authors reported that one or more replicates or groups experienced disproportionate outcomes unrelated to exposure (*e.g.*, **solubility issues** ~~contamination~~), but data from the remaining exposure replicates or groups were valid and is unlikely to have a substantial impact on results. OR

Data Quality Rating	Description
	The test material interfered in the assay, but the interference did not cause substantial differences among the groups. OR Solubility in the receptor fluid was not demonstrated, but solubility is not likely to be an issue based on the expected concentration relative to the receptor fluid formulation.
Low	Data on outcome differences unrelated to exposure (including receptor fluid formulation) were not reported for each study replicate or group and the missing information is likely to have a substantial impact on results. OR Assay interference was present or inferred resulting in large variabilities among the groups.
Critically Deficient	There were indications of assay interference several replicates or groups or there is evidence of insolubility in the receptor fluid such that no outcomes could be assessed.
Not Rated/Not Applicable	Do not select for this metric.
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Domain 7. Data presentation and analysis	
Metric 19. Data analysis Were statistical methods, calculations methods, and/or data manipulation clearly described and appropriate for dataset(s)? **Were absorption estimates presented across a time series for each compartment of the test system?**	
High	Statistical methods (including any calculations or data transformations) were clearly described or had only minor omissions and were appropriate for the dataset(s). Percentage absorption estimates were presented across a time series for each compartment of the test system, and Kp/flux measurements were based on the linear/steady-state part of the absorption curve. **Calculated absorption estimates properly accounted for outliers consistently across replicates/timepoints** ~~Any selection of outliers was justified.~~
Low	Statistical analysis was performed but not described adequately to understand what was performed or whether it was properly applied (*e.g.*, determination of outliers). OR Statistical analysis was inconsistently/inappropriately applied across replicates and datasets (*e.g.*, absorption not measured across time series, inconsistent exclusion of outliers **{perhaps due to integrity failure}** across measurements, **but** coefficient of variation for several replicates (SD relative to mean) was <~~>~~ 25%).

Data Quality Rating	Description
	OR **Absorption estimates were not presented across a time series for each component.**
Critically Deficient	Statistical analysis was performed using an inappropriate method (*e.g.*, parametric test for non-normally distributed data), and/or coefficient of variation for several replicates (SD relative to mean) was >25%. OR Statistical analysis was not performed. AND Data enabling an independent statistical analysis were not provided. These are serious flaws that make the study unusable.
Not Rated/Not Applicable	Statistical analysis was not possible (n = 1−2) or not necessary (clearly negative findings across all groups; Ames assay using 2-fold increase as benchmark).
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 20. Data interpretation **Is**~~Were the evaluation criteria reported and is~~ the interpretation of results consistent with standards and guidelines? **For example, did** reported absorption estimates account for sufficient recovery? Was the combined amount of test substance in the skin and receptor fluid counted in the overall estimate? **Was derivation of Kp vs fractional absorption applied to the appropriate exposure conditions (infinite dose vs finite dose, respectively)?**	
High	~~Study authors followed evaluation criteria for the test, and these were consistent with established practices~~ [a]. Recovery of applied test substance was adequate (90% for occluded or non-volatile substance, 80% for non-occluded, volatile ~~substance~~ **or unlabeled substance) or the absorption estimate was normalized to account for any reduction below these levels. Both the skin compartment and any tape-stripping washes after the first two were included in the absorption estimate**. **AND** Assay results were correctly interpreted relative to the properties of the test substance and the assay setup (sufficient duration to capture all absorption if not evaporated, proper interpretation of finite vs infinite dose).
Medium	Absorption estimates were reported improperly **or incompletely** (*e.g.*, skin compartment not included, values not normalized if recovery less than adequate), however simple independent data analysis is possible to overcome these issues.
Low	**Measurements of permeability flux/Kp were evaluated from a finite dose or fractional absorption estimates were included from a single and/or infinite dose.**~~Complex reanalysis of the data is required in order to obtain usable interpretations (*e.g.*, external outlier analysis may be required, Kp determination must be recalculated from the time series).~~

Data Quality Rating	Description
Critically Deficient	The reported **scoring**~~rating~~ and/or evaluation criteria were very inconsistent with established practices, resulting in the interpretation of data results that are seriously flawed and highly misleading relative to the properly interpreted results (*e.g.*, study author claims 5% absorption but correct analysis results in 40% absorption, only percentage absorption is reported from a finite dose) **and therefore not usable for any scenarios.**
Not Rated/Not Applicable	Do not select for this metric.
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*
Metric 21. Reporting of data Were the data for all outcomes presented? Were data reported by exposure group?	
High	Data for exposure-related findings were presented for all outcomes by exposure group **(*e.g.*, all timepoints, formulations, concentrations, finite vs infinite dose)**. Negative findings were reported qualitatively or quantitatively.
Medium	Data for exposure-related findings were reported for most, but not all, outcomes by exposure group (*e.g.*, both short and long-term exposures). The minor uncertainties in outcome reporting are unlikely to have substantial impact on results **(*e.g.*, intermediate timepoints not included in the data tables but the full curve is included).**
Low	Data for exposure-related findings were not shown for each study group, but results were described in the text. OR Data were only reported for some outcomes. OR Continuous data were presented without measures of variability or n/group.
Critically Deficient	Data presentation was inadequate (*e.g.*, the report does not differentiate among findings in multiple exposure groups) **OR** Major inconsistencies were present in reporting of results that render the findings uncertain regarding hazard identification or dose- response.
Not Rated/Not Applicable	Do not use for this metric.
Reviewer's Comments	*[Document concerns, uncertainties, limitations, and deficiencies and any additional comments that may highlight study strengths or important elements such as relevance]*

6 EVIDENCE INTEGRATION

As described in Section 7 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021), evidence integration refers to the consideration of evidence obtained from systematic review and scientific information obtained from sources that did not undergo systematic review to implement a weight of the scientific evidence approach. The weight of the scientific evidence is defined as "a systematic review method, applied in a manner suited to the nature of the evidence or decision, that uses a pre-established protocol to comprehensively, objectively, transparently, and consistently identify and evaluate each stream of evidence, including strengths, limitations, and relevance of each study and to integrate evidence as necessary and appropriate based upon strengths, limitations, and relevance" (40 CFR 702.33). The consideration of the quality and relevance of the data, while taking into account the strengths and limitations of the data, to appropriately evaluate the evidence for this supplement, is described in Section 7 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021).

6.1 Physical and Chemical Properties

Section 7.1 in the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021) describes how information from data sources that undergo systematic review are integrated for use in risk evaluations under TSCA for physical and chemical property data. Appendix E.1 in the Draft Risk Evaluation for TCEP provides the rationale for selecting data values from systematic review (U.S. EPA, 2023a).

6.2 Environmental Fate and Transport

Sections 7.2 – 7.2.3.1 in the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021) describes how information from data sources that undergo systematic review are integrated for use in risk evaluations under TSCA for environmental fate and transport data. In some cases, multiple high-quality data values or a range of values may be given. Including multiple data values or a range of values provides some transparency on how TCEP occurs in real world scenarios and to highlight the variability and/or potential uncertainties in any individual value. A determination of confidence in the range of fate endpoint(s) are also made based on the study quality of contributing data values. The main purpose of this determination is to evaluate how consistent the conclusions are for studies of congruent ratings. Interpretations regarding the strength of a study, model, or data point contribute to how these are individually judged and then considered together. This process culminates in a final judgment about the extent to which an endpoint is supported by the available evidence.

6.3 Environmental Release and Occupational Exposure

TCEP did not have any data available from TRI, DMR, NEI, etc., therefore, EPA evaluated environmental releases based on modeling approaches and industry sector information from standard engineering sources such as Generic Scenarios and Emission Scenario Documents. As described in Section 3.1 of the Draft Risk Evaluation for TCEP (U.S. EPA, 2023a), EPA estimated COU-specific releases using modelled estimates from a generic site using a theoretical production volume based on CDR thresholds and loss fractions to the most likely media of release.

TCEP does not have an OSHA PEL, therefore there was very little monitoring data available through systematic review. Where available, EPA evaluated occupational exposures based on monitoring data, surrogate monitoring data, modeling approaches, and worker activity information from standard engineering sources and systematic review as described in Section 5.1.1 of Draft Risk Evaluation for TCEP (U.S. EPA, 2023a).

6.4 General Population, Consumer, and Environmental Exposure

6.4.1 General Population and Environmental Exposure Assessment

No TRI or DMR data were available for TCEP. As such, EPA did not have site-specific information for TCEP releases. Section 3.2 of the Draft Risk Evaluation for TCEP describes the approach and methodology used to estimate environmental releases (U.S. EPA, 2023a).

General population exposures were evaluated for the inhalation, dermal and ingestion exposure pathways based on environmental release data identified as described above in Section 6.3. Section 5.1.3.1 of the Draft Risk Evaluation for TCEP summarizes the approach and methodology used to assess general population exposures (U.S. EPA, 2023a). In summary, modeled environmental release estimates were used as inputs for the general population exposure modeling. Air release estimates were used as inputs for estimating ambient air concentrations and deposition fluxes via IIOAC and AERMOD. Surface water release estimates were used as inputs for estimating surface water concentrations via EFAST 2014 and VVWM-PSC. Modeled ambient air concentrations were used to estimated inhalation exposure. Modeled deposition fluxes were used to estimate soil concentrations. Soil concentrations were used to estimate ingestion and dermal exposure. Modeled surface water concentrations were used for dermal and ingestion estimates through various scenarios (*e.g.*, drinking water, dermal via swimming, incidental ingestion via swimming, and fish ingestion).

For the environmental exposure assessment, EPA used modeled surface water concentrations via EFAST 2014 and VVWM-PSC, and surface water concentrations, benthic pore water concentrations and sediment concentrations modeled via VVWM-PSC.

Where available, EPA compared reported environmental monitoring data and reported environmental modeling data (see also the supplemental file with TCEP Data Extraction Information for General Population, Consumer, and Environmental Exposure (U.S. EPA, 2023c)) with EPA modeled media concentrations. Section 3.3 of the Draft Risk Evaluation for TCEP summarizes the EPA estimated environmental concentrations and provides tornado plots for the aggregated environmental monitoring data for various media (U.S. EPA, 2023a). Section 4.1.2 and 4.1.3 of the Draft Risk Evaluation for TCEP includes measured and modeled concentrations of TCEP in aquatic and terrestrial species (U.S. EPA, 2023a). Furthermore, EPA gathered available information on TCEP in surface water and groundwater from the WQP database as presented in Section 3.3.2 of the TCEP Draft Risk Evaluation (U.S. EPA, 2023a). Appendix S 2.1 of the TCEP Draft Risk Evaluation describes the approach taken to retrieve and process the WQP data (U.S. EPA, 2023a).

6.4.1 General Population Exposure: Dietary, Biomonitoring and Exposure Reconstruction

Dietary data from the systematic review monitoring literature is summarized in Section 5.1.3.4.7 of the Draft Risk Evaluation for TCEP (U.S. EPA, 2023a). EPA did not estimate dietary exposures to TCEP but did summarize the monitoring data to describe the contribution of TCEP from food sources.

Biomonitoring data from the systematic review monitoring literature is summarized in Section 5.1.3.4.6 (Human Milk Exposure) and 5.1.3.5 (Exposure Reconstruction using Human Biomonitoring Data and Reverse Dosimetry) of the Draft Risk Evaluation for TCEP (U.S. EPA, 2023a); see also the TCEP Data Extraction Information for General Population, Consumer, and Environmental Exposure (U.S. EPA, 2023c).

EPA estimated a range of concentrations of TCEP in human milk based on maternal doses using a readily available multi-compartment physiologically based pharmacokinetic model, referred to as the Verner model. To ground truth the results, EPA compared modeled estimates with human milk biomonitoring concentrations from three available studies from the systematic review.

In addition to human milk, EPA summarized biomonitoring data (*e.g.*, urine, surface wipes, silicone wristbands) in tornado plots and accompanying tables in Section 5.1.3.5 of the Draft Risk Evaluation for TCEP (U.S. EPA, 2023a). EPA selected Urinary BCEP as a biomarker of exposure for TCEP and conducted an exposure reconstruction and reverse dosimetry using urinary BCEP data reported in NHANES.

6.4.1 Consumer Exposure Assessment

Section 3.4 of the Draft Risk Evaluation for TCEP summarizes the monitoring data and reported modeled estimates of TCEP in indoor air, personal air and indoor dust (U.S. EPA, 2023a); see also TCEP Data Extraction Information for General Population, Consumer, and Environmental Exposure (U.S. EPA, 2023c).

EPA conducted consumer modeling for TCEP-containing articles via CEM 3.0. EPA relied on the experimental data from the systematic review (see experimental data shown in the TCEP Data Extraction Information for General Population, Consumer, and Environmental Exposure (U.S. EPA, 2023c)) to inform input parameters for CEM 3.0, namely information related to consumer article weight fractions and dermal parameters (*e.g.*, fraction absorbed).

6.4.1 Other data sources

The exposure models relied heavily on the physical chemical and fate properties as input parameters. Sections 5.1 and 5.2 describe how the physical chemical and fate properties were selected. Where applicable, EPA relied on model defaults, exposure factors and activity patterns available from EPA's Exposure Factors Handbook (U.S. EPA, 2017).

6.5 Environmental and Human Health Hazard

Sections 7.4 and 7.5 of the 2021 Draft Systematic Review Protocol explain how information from data sources that undergo systematic review and those that do not undergo systematic review are considered for use in risk evaluations under TSCA for evaluating environmental and human health hazard, respectively (U.S. EPA, 2021). The sections below identify updates to the evidence integration process and the evidence streams used in the evidence integration step for environmental and human health hazard data in the Draft Risk Evaluation for TCEP. Specifically, as Table 2-1 indicated, the evidence streams were updated for both environmental and human health hazard disciplines as described below in Sections 6.5.1 and 6.5.2.1, respectively. The evidence streams for both environmental and human health hazard are high level categorizations of how apical and mechanistic endpoints are binned. When relevant for a respective chemical risk evaluation, data across evidence streams are integrated for hazard characterization.

6.5.1 Environmental Hazard

Section 7.1 of the 2021 Draft Systematic Review Protocol describes how environmental hazard integration is organized into different evidence streams. The environmental hazard evidence streams, as described in Table 7-8 of the 2021 Draft Systematic Review Protocol, have been updated to increase the level of clarity and consistency of granularity (U.S. EPA, 2021). Specifically, for risk evaluations conducted under TSCA, the environmental hazard evidence streams were updated (Table 6-1) to more

clearly reflect how apical and mechanistic hazardous endpoints (as defined by the screening PECO statement) resulting from either controlled field, laboratory, or uncontrolled exposure field studies are binned to better consider the relevancy of the data for the respective risk evaluation. Bold text in Table 6-1 indicates updates made to the evidence stream and the respective questions and considerations.

Table 6-1. Querying the Evidence to Organize Integration for Environmental Data and Information

Evidence Stream	Questions
Apical endpoints (controlled field/laboratory conditions)	Of the available data, are there endpoints that could have population level effects such as reproduction, growth, and/or mortality?
Mechanistic **data (controlled field/laboratory conditions**)	**Is the mechanistic endpoint linked to an apical endpoint? Is it part of an AOP? If not, can you instead use it qualitatively? If a transcriptomic point of departure (tPOD) is available, is it appropriate to use quantitatively?**
Apical endpoints (uncontrolled exposure field conditions)	Are there any field studies available showing adverse effects? How does exposure to the chemical of interest affect the community of organisms? Are there any co-occurring adverse environmental conditions other than exposure to the chemical of interest that should be taken into consideration?
Mechanistic endpoints (uncontrolled exposure field conditions)	**Is the mechanistic endpoint linked to an apical endpoint? Is it part of an AOP? If not, can you instead use it qualitatively? If a transcriptomic point of departure (tPOD) is available, is it appropriate to use quantitatively? Are there any co-occurring adverse environmental conditions other than exposure to the chemical of interest that should be taken into consideration?**

Evidence streams for environmental hazard included empirical data with apical endpoints and mechanistic data from controlled laboratory experiments for aquatic and terrestrial organisms. Predictive models represented within the body of evidence included the EPA's Web-based interspecies Correlation Estimation (Web-ICE) application and the Ecological Structure Activity Relationships (ECOSAR) Predictive Model. Modeled data served as evidence streams outside of SR and were integrated with evidence streams within SR. Web-ICE and ECOSAR both predict hazard values for apical endpoints; therefore, the modeled data and empirical data are both part of the apical endpoint data represented from controlled laboratory experiments evidence stream.

In the environmental hazard characterization for TCEP, for aquatic organisms EPA integrated environmental hazard data from empirical data with modeled data within the apical endpoint evidence stream. Specifically, EPA used empirical data and modeled data from Web-ICE to create a Species Sensitivity Distribution (SSD) and calculate an HC_{05} which was then used to calculate a concentration of concern (COC) (See Appendices Q.2.1.1, Web-based Interspecies Correlation Estimation (Web-ICE) and Q.2.1.2 Species Sensitivity Distribution (SSD). ECOSAR predictions for aquatic species were available for green algae and daphnid. ECOSAR results are presented within Section 4.2.2, Aquatic Species Hazard and were used qualitatively to support empirical apical data represented within the quantitative analysis presented above. Terrestrial hazard values are represented with apical endpoints as

described within Section 4.2.3, Terrestrial Species Hazard. Additional representation of mechanistic endpoints for Earth worm (*Eisenia fetida*) and American kestrel (*Falco sparverius*) were represented within the screening level trophic transfer analysis for TCEP detailed within Section 4.3.1.1, Risk Characterization Approach for Trophic Transfer (U.S. EPA, 2023a).

Evaluations of the strength of evidence and weight of scientific evidence for environmental hazard was conducted as described within Section 7.4.2 of the 2021 Draft Systematic Review Protocol (U.S. EPA, 2021). For additional details on the application of this methodology, please see Section 4.2.6 of the Draft Risk Evaluation for TCEP and Appendix Q.2.3.1 (U.S. EPA, 2023a).

6.5.2 Human Health Hazard

6.5.2.1 Updates to the Systematic Review Protocol

Section 7.5 of the 2021 Draft Systematic Review Protocol described how EPA considers individual evidence streams (human, animal toxicity, and mechanistic/supplemental studies) when integrating evidence (U.S. EPA, 2021). In the Draft Risk Evaluation for TCEP, the evidence integration tables have been updated to clarify that both the human and animal evidence streams reflect studies that were considered in deriving toxicity values (U.S. EPA, 2023a). This clarification was made to distinguish these streams from data that are more supportive in nature (*e.g.,* animal studies that only examined mechanistic data). The third stream is unchanged, as shown below.

These updates (bolded text) are included in the headings in relevant rows within the evidence integration tables in Appendix K of the Draft Risk Evaluation for TCEP:

- Evidence in Studies of Exposed Humans **Considered for Deriving Toxicity Values**
- Evidence from *in vivo* Mammalian Animal Studies **Considered for Deriving Toxicity Values**
- Evidence in Mechanistic Studies and Supplemental Information

EPA also implemented a second update to the 2021 Draft Systematic Review Protocol within the TCEP risk evaluation (U.S. EPA, 2021). For human health outcomes with limited data, EPA presented evidence integration conclusions as short narratives because the data did not warrant full tables. See Appendix K.2 of the Draft Risk Evaluation for TCEP (U.S. EPA, 2023a).

6.5.2.2 Data Available for Human Health Hazard Evidence Integration

EPA conducted data quality evaluation of human epidemiological and animal toxicity studies that were considered for deriving toxicity values. EPA integrated these evidence streams with mechanistic data for several human health hazard outcomes (*e.g.,* neurological/behavioral, reproductive/ developmental, renal, hepatic, cancer). Human epidemiological studies were available for cancer and immune effects from the 2019 literature search and animal toxicity studies were available for all human health outcomes considered. *In vivo* and *in vitro* mechanistic data were available for most human health hazard outcomes.

7 Bibliography

Bronaugh, RL; Franz, TJ. (1986). Vehicle effects on percutaneous absorption: in vivo and in vitro comparisons with human skin. Br J Dermatol 115: 1-11. http://dx.doi.org/10.1111/j.1365-2133.1986.tb06214.x

Dodson, RE; Van den Eede, N; Covaci, A; Perovich, LJ; Brody, JG; Rudel, RA. (2014). Urinary Biomonitoring of Phosphate Flame Retardants: Levels in California Adults and Recommendations for Future Studies. Environ Sci Technol 48: 13625-13633. http://dx.doi.org/10.1021/es503445c

Fasano, WJ; Manning, LA; Green, JW. (2002). Rapid integrity assessment of rat and human epidermal membranes for in vitro dermal regulatory testing: correlation of electrical resistance with tritiated water permeability. Toxicol In Vitro 16: 731-740. http://dx.doi.org/10.1016/s0887-2333(02)00084-x

Guth, K; Schäfer-Korting, M; Fabian, E; Landsiedel, R; van Ravenzwaay, B. (2015). Suitability of skin integrity tests for dermal absorption studies in vitro. Toxicol In Vitro 29: 113-123. http://dx.doi.org/10.1016/j.tiv.2014.09.007

Learn, DB; Roper, CS; Craig, S; Toner, FS. (2018). An assessment of the intra-lab variability of the in vitro percutaneous absorption of testosterone through human skin [Poster]. In 57th Annual Meeting and ToxExpo (SOT 18). Retrieved from https://www.criver.com/sites/default/files/resource-files/SP-SOT-18-assessment-intra-lab-variability-in-vitro-percutaneous-absorpt.pdf

Meidan, VM; Roper, CS. (2008). Inter- and intra-individual variability in human skin barrier function: a large scale retrospective study. Toxicol In Vitro 22: 1062-1069. http://dx.doi.org/10.1016/j.tiv.2008.01.009

OECD. (2022). Series on Testing & Assessment, No. 156: Guidance notes on dermal absorption studies (Second edition). (ENV/JM/MONO(2011)36/REV1). Paris, France: Organisation for Economic Co-operation and Development (OECD). https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV-JM-MONO(2011)36%20&doclanguage=en

U.S. EPA. (2017). Update for Chapter 5 of the Exposure Factors Handbook: Soil and dust ingestion [EPA Report]. (EPA/600R-17/384F). Washington, DC: National Center for Environmental Assessment, Office of Research and Development. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P100TTX4.txt

U.S. EPA. (2018). Application of Systematic Review in TSCA Risk Evaluaitons. (EPA 740-P1-8001).

U.S. EPA. (2020). Final Scope of the Risk Evaluation for Tris(2-chloroethyl) Phosphate (TCEP); CASRN 115-96-8. (EPA 740-R-20-009). Washington, DC: Office of Chemical Safety and Pollution Prevention. https://www.epa.gov/sites/default/files/2020-09/documents/casrn_115-96-8_tris2-chloroethyl_phosphate_tcep_final_scope.pdf

U.S. EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances, Version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies. (EPA Document #EPA-D-20-031). Washington, DC: Office of Chemical Safety and Pollution Prevention. https://www.regulations.gov/document/EPA-HQ-OPPT-2021-0414-0005

U.S. EPA. (2023a). Draft Risk Evaluation for Tris(2-chloroethyl) Phosphate (TCEP). Washington, DC: Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention.

U.S. EPA. (2023b). Draft Risk Evaluation for Tris(2-chloroethyl) Phosphate (TCEP) – Systematic Review Supplemental File: Data Extraction Information for Environmental Hazard and Human Health Hazard Animal Toxicology and Epidemiology. Washington, DC: Office of Pollution

Prevention and Toxics, Office of Chemical Safety and Pollution Prevention. https://www.regulations.gov/docket/EPA-HQ-OPPT-2018-0476

U.S. EPA. (2023c). Draft Risk Evaluation for Tris(2-chloroethyl) Phosphate (TCEP) – Systematic Review Supplemental File: Data Extraction Information for General Population, Consumer, and Environmental Exposure. Washington, DC: Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention.

U.S. EPA. (2023d). Draft Risk Evaluation for Tris(2-chloroethyl) Phosphate (TCEP) – Systematic Review Supplemental File: Data Quality Evaluation and Data Extraction Information for Dermal Absorption. Washington, DC: Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention.

U.S. EPA. (2023e). Draft Risk Evaluation for Tris(2-chloroethyl) Phosphate (TCEP) – Systematic Review Supplemental File: Data Quality Evaluation and Data Extraction Information for Environmental Fate and Transport. Washington, DC: Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention.

U.S. EPA. (2023f). Draft Risk Evaluation for Tris(2-chloroethyl) Phosphate (TCEP) – Systematic Review Supplemental File: Data Quality Evaluation and Data Extraction Information for Environmental Release and Occupational Exposure. Washington, DC: Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention.

U.S. EPA. (2023g). Draft Risk Evaluation for Tris(2-chloroethyl) Phosphate (TCEP) – Systematic Review Supplemental File: Data Quality Evaluation and Data Extraction Information for Physical and Chemical Properties. Washington, DC: Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention.

U.S. EPA. (2023h). Draft Risk Evaluation for Tris(2-chloroethyl) Phosphate (TCEP) – Systematic Review Supplemental File: Data Quality Evaluation Information for Environmental Hazard. Washington, DC: Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention.

U.S. EPA. (2023i). Draft Risk Evaluation for Tris(2-chloroethyl) Phosphate (TCEP) – Systematic Review Supplemental File: Data Quality Evaluation Information for General Population, Consumer, and Environmental Exposure. Washington, DC: Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention.

U.S. EPA. (2023j). Draft Risk Evaluation for Tris(2-chloroethyl) Phosphate (TCEP) – Systematic Review Supplemental File: Data Quality Evaluation Information for Human Health Hazard Animal Toxicology. Washington, DC: Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention. https://www.regulations.gov/docket/EPA-HQ-OPPT-2018-0476

U.S. EPA. (2023k). Draft Risk Evaluation for Tris(2-chloroethyl) Phosphate (TCEP) – Systematic Review Supplemental File: Data Quality Evaluation Information for Human Health Hazard Epidemiology. Washington, DC: Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention.

Zhang, Q; Murawsky, M; Lacount, T; Kasting, GB; Li, SK. (2018). Transepidermal water loss and skin conductance as barrier integrity tests. Toxicol In Vitro 51: 129-135. http://dx.doi.org/10.1016/j.tiv.2018.04.009