TCEQ Guidelines for Systematic Review and Evidence Integration

Prepared by the

Toxicology Division

White Paper

December 20, 2017

TEXAS COMMISSION ON ENVIRONMENTAL QUALITY

Document Description and Intended Use

A systematic review is defined as a high-level review of the available, relevant information in order to extract and analyze all data to address a specific research question. Systematic reviews are becoming an integral part of risk assessments since key steps of the process include using explicit, reproducible methods to identify, select and critically evaluate all quality research in order to minimize bias and provide reliable findings (Cochrane Collaboration 2011). This document provides guidance on how to conduct a systematic literature review and integrate evidence when developing chemical-specific reference values (ReVs) and unit risk factors (URFs). However, this process can also be modified or expanded to address other questions that would benefit from systematic review practices. These guidelines supplement the Texas Commission on Environmental Quality (TCEQ) Regulatory Guidance-442 (RG-442), *TCEQ Guidelines to Develop Toxicity Factors* (TCEQ 2015).

Since the TCEQ published RG-442, systematic review guidelines were needed, which include explicit criteria for determining study quality prior to identifying a key study (e.g., study inclusion and exclusion criteria). Since data are collected from diverse evidence streams (e.g., human clinical data, epidemiological data, animal toxicological studies, mechanistic data), there is a need to evaluate and integrate information from multiple streams to improve the decision-making process, increase transparency, minimize bias, and improve consistency between different risk assessments. The systematic review and evidence integration framework can improve regulatory decision-making processes, increase transparency, minimize bias, improve consistency between different risk assessments, and further improve confidence in toxicity factor development. This document is not intended to be an explicit instruction manual, but rather a guide to use for any chemical evaluation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Acronyms and Abbreviations

Acronyms and Abbreviations	Definitions
ADME	absorption, distribution, metabolism, and excretion
AOP	Adverse Outcome Pathway
BMD	benchmark dose
BMDL	benchmark dose lower confidence limit
CSAF	Chemical specific adjustment factor
d	day(s)
DSD	development support document
EA	experimental animal
ESL	effects screening level
GD	gestational day
GLP	good laboratory practice
h	hour(s)
HAWC	Health Assessment Workspace Collaboration
HE	human epidemiologic
HEC	human equivalent concentration
HED	human equivalent dose
i.p.	intraperitoneal
i.v.	intravenous
IOM	Institute of Medicine
IPCS	International Programme on Chemical Safety
IRIS	Integrated Risk Information System
LOAEL	lowest observed adverse effect level
MECH	mechanistic
MeSH	medical subject headings
µg	microgram(s)
µg/m^3	micrograms per cubic meter
mg	milligram(s)
mg/m^3	milligrams per cubic meter
min	minute(s)
MOA	mode of action
NAS	National Academy of Sciences
NIEHS	National Institute of Environmental Health Sciences
NRC	National Resource Council
NOAEL	no observed adverse effect level
NTP	National Toxicology Program
OHAT	Office of Health Assessment and Translation
ORD	Office of Research and Development
PBPK	physiologically based pharmacokinetic model
PECO	Populations, Exposure, Comparator/Control, and Outcomes

Acronyms and Abbreviations	Definitions
POD	point of departure
POD_{ADJ}	point of departure adjusted for exposure duration
POD_{HEC}	point of departure adjusted for human equivalent concentration
ppb	parts per billion
ppm	parts per million
REACH	Registration, Evaluation, Authorization, and Restriction of Chemicals
ReV	reference value
RfD	reference dose
RG	regulatory guidance
ROB	risk of bias
SF_o	slope factor
TCEQ	Texas Commission on Environmental Quality
TD	Toxicology Division
UF	uncertainty factor
URF	unit risk factor
USEPA	United States Environmental Protection Agency
wk	week(s)
WOE	weight of evidence
yr	year(s)

Introduction

A systematic review involves a comprehensive plan and search strategy with the intention of reducing bias by "identifying, appraising, and synthesizing all relevant studies on a particular topic" (Uman 2011). Several recent publications have proposed best practices for conducting systematic reviews (Rhomberg et al. 2013, NRC 2014, Rooney et al. 2014).The Office of Health Assessment and Translation (OHAT) Division of the National Toxicology Program (NTP), in the National Institute of Environmental Health Services (NIEHS), recently published their method for conducting systematic reviews and evidence integration for reaching hazard identification conclusions (Rooney et al. 2014).

The overall objective of this guidance is to provide information on conducting a systematic review during the development of chemical-specific toxicity factors based on evidence from human, animal, and mechanistic studies. The following systematic review guidelines supplement TCEQ's 2015 published regulatory guidelines on deriving toxicity factors (RG-442). Figure 1 depicts the TCEQ systematic review and evidence integration process. In general, derivation of chemical reference values (ReVs) or unit risk factors (URFs) begins with a toxicity assessment involving hazard identification, dose-response assessment, and a chemical's mode of action. The toxicity factors developed by the TCEQ are derived to protect potentially sensitive populations, such as children, pregnant women and the elderly; thus, all available health endpoints and various types of studies are considered in order to determine the most sensitive health endpoint (i.e., critical effect) in the most [relevant or] sensitive species. This guidance, in principle, must also be applicable for chemicals for which limited toxicity data are available. Therefore, the TCEQ used the available existing methodologies to develop guidelines for conducting systematic reviews and integrating evidence when developing chemical-specific reference values (ReVs) and unit risk factors (URFs).

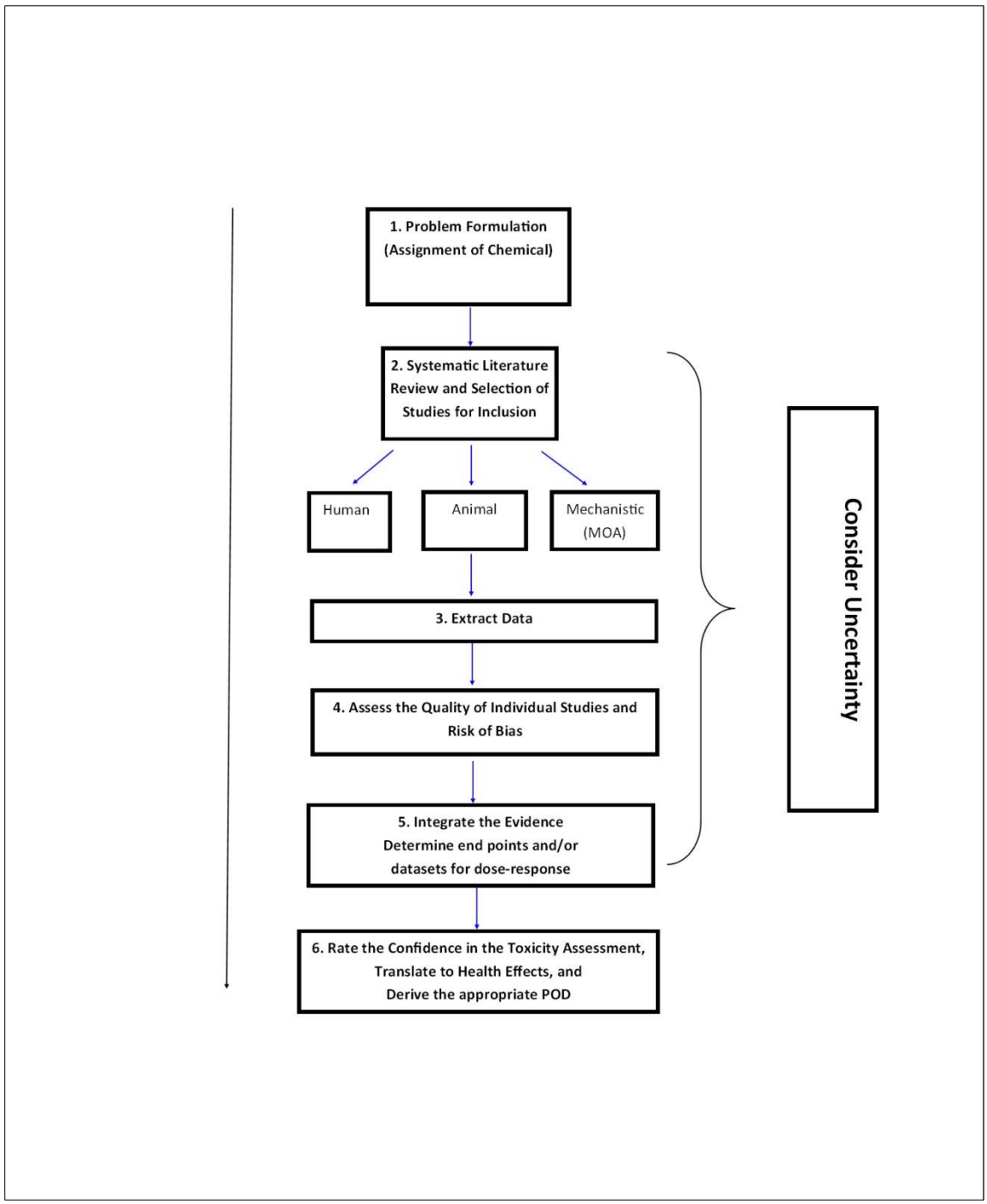

Figure 1. Steps in Systematic Review and Evidence Integration

Step 1: Problem Formulation and Protocol Development

The first step in the systematic review and evidence integration process is problem formulation (Figure 1). This step identifies and specifically states the research question and describes the extent of the evaluation. Problem formulation contains elements that promote transparency and consistency, and can accommodate different biologically plausible hypotheses (Rhomberg et al. 2013).

For the derivation of toxicity factors, the TCEQ reviews all available data to identify the critical effect that occurs at the lowest human equivalent concentration or dose. The TCEQ's Guidelines to Develop Toxicity Factors (TCEQ 2015) is a peer-reviewed publication that outlines the process of critically evaluating a variety of health outcomes and focusing resources on human-relevant adverse health endpoints. The process begins with the selection of a chemical, followed by the review of the physical and chemical properties and a critical review of dose-response data for all of the available health endpoints. The empirical evidence is examined thoroughly to determine the no observed adverse effect level (NOAEL) and/or the lowest observed adverse effect level (LOAEL). When data are available, Benchmark Dose Software (BMDS) is used to develop dose-reponse curves and to establish a point of departure (POD). To the extent possible, determination of the most appropriate mode of action (MOA) for the most sensitive (i.e., critical) adverse endpoint is also included in the analysis. An MOA analysis is important in understanding the potenital for toxicity and the most scientifically-definsible extrapolation to lower exposures (USEPA 2005). Assessing the confidence in the hypothesized or previously propounded MOAs for a critical effect is addressed and described in more detail in Becker et al., 2017.

The problem formulation should be articulated clearly to prevent the systematic review and evidence integration process from becoming unduly resource intensive and to ensure that chemical specific toxicity values are protective of human health and welfare. The output of the problem formulation step is a statement that includes specific questions pertinent to all of the steps of the systematic review process, including the literature search, study selection, data extraction, and synthesis. Examples of specific questions to structure the probem formulation are included below:

- What is the chemical for which the Development Support Document (DSD) is being developed?
- What are the physical and chemical properties of the chemical?
- What is/are the critical effect(s)?
- Are there potentially sensitive subpopulations?
- What is the MOA? (described in detail in Becker et al. 2017)
    - Is the hypothesized MOA biologically plausible?
    - Are adverse effects reversible if dosing is stopped?
    - Is there consistency across sexes, strains, organs and species?

    - Is the MOA consistent across structurally related chemicals?
    - Does the MOA suggest a threshold for departure?
    - How does the dose-response assessment inform the MOA?
    - Is the chemical carcinogenic? If so, is the chemical carcinogenic only by a specific route-of-exposure or when a biologically-plausible threshold is exceeded?
        - What is the MOA and dose response for each tumor type? What is the strength of the MOA for each tumor type? What is the anticipated relevance of each tumor type to humans?
    - Do existing MOA data inform the choice to use linear or non-linear extrapolation in the toxicity factor development?
- Does route-of-exposure play a role in toxicity?
- Is the chemical carcinogenic? If so, is the chemical carcinogenic only by a specific route of exposure or when a biologically-plausible threshold is exceeded?
- Is the chemical a reproductive or developmental toxicant?

Protocol development is another important aspect in the initial step of the systematic review process. A protocol is typically developed around a PECO (Populations, Exposure, Comparator/Control, and Outcomes) statement (Rooney et al. 2014). These identifiers are used to lay out the framework for the literature search and inclusion/exclusion criteria. The PECO statement is particularly helpful if specific aspects of the review have already been identified prior to the literature search, such as species of interest, critical health endpoint, route-of-exposure, or MOA. For example, most chemical assessments conducted by the TCEQ meet these criteria:

Table 1. PECO Statement Used by the TCEQ to Develop Toxicity Factors

Population(s)	General human population and any potentially sensitive human subpopulations, animals, and vegetation
Exposure	Exposure to the selected chemical or any identified metabolites or surrogates with similar MOAs
Comparator/ Control	Populations exposed to concentrations below the concentration that causes the most sensitive (i.e., critical) effect
Outcome(s)	The first adverse effect (i.e. critical effect) that occurred in the most [relevant or] sensitive species caused by the exposure

The TCEQ defines an adverse effect as a biochemical change, functional impairment, or pathologic lesion that affects the performance of the whole organism, or reduces an organism's ability to respond to an additional environmental challenge (TCEQ 2015). Consistent with the goal of protecting public health, the TCEQ calculates conservative health-based toxicity factors to protect against adverse health effects. More information is available in Section 3.6.1 Determination of Adverse Effect (TCEQ 2015).

In this framework, the problem formulation and the protocol are written in a general manner because they must be applicable to a wide array of chemicals, data sets, and endpoints. The problem formulation and protocol development steps may be changed as new information becomes available; however, any changes made to the review questions and protocol should be documented accordingly. For the purpose of conducting systematic reviews and integrating evidence for determining toxicity factors, the TCEQ uses the following protocol as a guideline. Detailed descriptions of the protocol used by the TCEQ to develop toxicity factors can be found in TCEQ (2015) and the steps to the overall protocol are summarized here.

1. Identify the chemical of interest and define the research question(s)
2. Conduct a systematic review:
    a. Identify the study inclusion/exclusion criteria
    b. Conduct a systematic literature search
    c. Identify the study inclusion/exclusion criteria
    d. Extract the relevant data from each data stream (human, animal, mechanistic)
    e. Assess study quality and conduct a risk of bias analysis
    f. Weigh the evidence in each data stream and then integrate the evidence across the data streams
    g. Rate the confidence in the evidence
3. Assuming a potential hazard has been identified, develop toxicity factor (as detailed in TCEQ 2015 guidelines):
    a. Review the essential data and selected key studies from the systematic review
    b. Conduct an MOA analysis
    c. Choose the most appropriate dose metric available (e.g., tissue dose, air concentration)
    d. Select critical adverse effect based on human equivalent exposure, considering each potential key study
    e. Extrapolate from the point of departure (e.g., POD_{HEC}) to lower exposures considering existing MOA data (e.g., for a mutagenic MOA for carcinogenicity)

Step 2: Systematic Literature Review and Selecting Studies for Inclusion

2.1 Systematic Literature Review

The general objective of the literature search strategy for a specific chemical risk assessment is to identify all relevant studies, which may include both published and unpublished studies.

The TCEQ conducts thorough literature searches of relevant databases and takes other prudent steps to identify relevant studies during the literature review. The TCEQ Toxicology Division (TD) trains its toxicology staff to conduct their own systematic literature searches. For example, in addition to relevant guidance (e.g., Section 3.3.2 of TCEQ 2015), TCEQ staff utilize the

National Library of Medicine's resources for training on advanced uses of the various databases (PubMed, TOXNET, etc.), and/or train in person with an Instructional Services Librarian. TCEQ staff also utilizes other resources such as webinars and/or in-person training (as available).

Several months prior to the start of work on a DSD, the TD of the TCEQ conducts a scoping exercise to identify all available toxicity information for the chemical. The TD announces this process using its email listserve to solicit information for a particular chemical or class of chemicals; interested parties are encouraged to provide citations or toxicological information. Chapter 1 of the TCEQ (2015) Guidelines provides more detailed information on the selection of chemicals and data solicitation for DSDs. The literature review may be updated as new information becomes available or additional supplemental literature searches are warranted. Changes made to the initial literature review should be documented accordingly.

2.1.1 Selecting Databases and Sources

Initially, publically available databases (Table 2) are searched using explicitly stated search criteria. Additionally, several governmental and private sector organizations can be consulted for previously published scientific literature and toxicity values for chemicals. This checklist (Table 2) is a dynamic document, and other sources and databases may be added if deemed necessary or advantageous for the toxicity factor derivation process.

Table 2. List of Available Databases

TOXNET is supported by the National Library of Medicine and includes several databases:	
	ChemIDplus
	Chemical Carcinogenesis Research Information System (CCRIS)
	Developmental and Reproductive Toxicology (DART) Database
	Genetic Toxicology Data Bank (GENETOX)
	Hazardous Substances Data Bank (HSDB)
	Integrated Risk Information System (IRIS)
	International Toxicity Estimates for Risk (ITER)
	Toxicology Literature Online (TOXLINE)
Searchable databases from the USEPA:	
	Acute Exposure Guideline Levels (AEGLs)
	Health and Environmental Research Online (HERO)
	National Ambient Air Quality Standards (NAAQS)
	Provisional Peer Reviewed Toxicity Values for Superfund (PPRTV)
	Toxicity and Exposure Assessment for Children's Health (TEACH)
Other searchable databases:	
	Defense Technical Information Center
	National Cancer Institute
	Public Medicine (PubMed)
	Registry of Toxic Effects of Chemical Substances (RTECS)
	National Technical Information Service (NTIS)
Published documents from the public and private sectors	
	Agency for Toxic Substances and Disease Registry (ATSDR) Toxicological Profiles
	American Conference of Government and Industrial Hygienists (ACGIH)
	American Industrial Hygiene Association (AIHA)
	California Environmental Protection Agency (CalEPA)
	CalEPA Office of Environmental Health Hazard Assessment (OEHHA)
	Centers for Disease Control and Prevention (CDC)
	Health Canada
	International Agency for Research on Cancer (IARC)
	International Programme on Chemical Safety (IPCS)
	National Institute for Occupational Safety and Health (NIOSH)
	National Toxicology Program (NTP)
	Occupational Safety and Health Administration (OSHA)
	Organization for Economic Co-operation and Development (OECD)

2.1.2 Selecting Search Terms

Adequate searching of the scientific literature is a vital part of the systematic review process. To the extent possible, search terms should be thoughtfully selected to appropriately narrow

down search results for data-rich chemicals that would otherwise produce an exhaustive amount of literature, much of which is irrelevant to toxicity factor development. The use of Boolean operators is recommended while conducting a systematic literature search:

- "AND" is used to group *keywords or ideas* together in the search (e.g., benzene AND cancer);
- "OR" is used to search for multiple *synonyms* (e.g., inhalation OR air OR aerosol);
- "NOT" is used to exclude keywords (e.g., ethylene NOT diethylene);
- Quotation marks (" ") are used when multiple keywords are searched together (e.g., "ethylene glycol");
- Asterisks (*) are used to search all of the forms of a root word (truncation) to get all derivatives of the term (e.g., a search for carcinogenic effects can include the term carc*, which will search carcinogen, carcinogenic, carcinoma, etc.; and.);
- Medical subject headings [mesh] are used in PubMed to look for the search term in a specified heading group rather than just key words to get more relevant results.

These terms can be grouped together to narrow down a literature search that otherwise may produce an excess of irrelevant results. For example, the ethylene glycol search string may look like this:

"ethylene glycol" [mesh] NOT "ethylene oxide" AND (inhal* OR air OR carc* OR onco*)

This search string identifies studies with the keywords ethylene and glycol together in a medical subject heading, excludes studies referring to ethylene oxide, and only includes the studies that use a form of inhal* (inhale, inhalation), air, carc* (carcinogenic, carcinogen) or onco* (oncogenesis, oncogenicity). Documenting the search criteria and search cutoff dates used in a systematic literature review is important; an example of how this search criterion can be recorded in a DSD is provided in Table 13 in the Appendix.

2.1.3 Maintain a Record of Searches

Currently, there are several available to help inform decisions and transparently document the systematic literature review process. These tools, which can help maintain references in one place and group them based on the selection criteria, can be powerful because they allow query of the databases, and improve transparency of the inclusion/exclusion process.

The TCEQ utilizes the HAWC (Health Assessment Workspace Collaboration) software to conduct the literature search, compile references from PubMed and other sources, tag literature for inclusion or exclusion, and analyze the available literature. HAWC is an open source, modular, content management system designed to synthesize multiple data sources into overall human health assessments of chemicals. The system integrates and documents workflow from the literature search to data extraction, synthesis, and interpretation. This software tool is used to

manage the systematic review and data display. Human health assessments of chemicals are best documented with a systematic review of the scientific literature, and depending on the chemical may require a large amount of data extraction, synthesis, and interpretation by teams of experts across multiple fields. HAWC creates a workspace for interested parties, including reviewers and stakeholders, to have dynamic access to on-going and completed assessments. Additionally, HAWC creates a clear and concise summary of the results of these assessments, enables online access to the literature review, and tracks primary data and/or tabulated study summaries and visual aids (e.g., Forest plots) that constitute the scientific justification for the assumptions and conclusions made by the reviewer(s). TCEQ staff will be formally trained on how to conduct the literature search, compile references, tag literature for inclusion or exclusion, and analyze the available literature using the HAWC database.

2.2 Inclusion and Exclusion Criteria

A strength of the systematic review approach is the documentation of clear study inclusion/exclusion criteria. This step is useful in documenting why particular studies were chosen as potential key studies and the reasons for excluding other studies (i.e., excluding them as potential key studies or completely excluding studies from the review). These criteria improve transparency and subsequently help improve risk communication to a wide range of stakeholders. Clear and direct inclusion/exclusion criteria need to be specified to identify the initial study database from which key and supporting studies are selected. The inclusion and exclusion criteria are formulated based on the specific questions that are established during the problem formulation step. For example, the criteria are based on adverse health outcomes, exposures, durations, and the types of studies relevant to the toxicity factor being developed. Studies that contribute to identifying the relevant critical effect(s) are selected for further review. Developing explicit criteria *a priori* to select or omit studies helps to balance scientific judgment by providing clear and transparent documentation. This documentation allows the search to be easily reproduced by other researchers if needed, which in turn can improve confidence in the TCEQ's derivation of toxicity factors.

Several study-specific questions can be asked to determine whether a study should be included or excluded (examples are included in Table 3). Defining one set of inclusion and exclusion criteria for all chemicals is difficult since often the criteria will be chemical and/or purpose-specific. Therefore, inclusion and exclusion criteria may be modified as needed and should be documented appropriately. For example, if the purpose is to develop an inhalation reference value, oral studies may be excluded. However, if the inhalation database is lacking and the effects are not route dependent, oral studies may be included. More stringent exclusion criteria may be required for data-rich chemicals to narrow down the pool of available literature to only those studies relevant to the specific assessment being conducted. For a thorough review, two or more individuals should review each piece of literature identified from the scientific literature search and classify the journal articles based on the specific inclusion/exclusion criteria utilized.

Table 3. Examples of Study Inclusion and Exclusion Criteria

Study Type	Inclusion Criteria	Exclusion Criteria
General	Complete study available for review	- Only abstract is available - Study in a language other than English - Unpublished report/unable to retrieve
	Exposure concentration is relevant to developing toxicity factors	- Significantly high concentrations used - Study focused on overdose/poisoning or mortality - Exposure concentration unknown
	Study contains original data	- Study is a review article
	Study examines effects caused by chemical exposure	- Study measures concentration in products, etc. - Study does not examine health effects
	Study focused on the chemical of concern or active metabolites	- Study examined multiple chemicals not of interest - Study on treatment following chemical exposure
Mode of Action	Weight of evidence is sufficient to establish MOA in humans.	- Human relevance of the MOA can be reasonably excluded based on fundamental, qualitative or quantitative differences in key events between animals and humans.
Animal	Route of exposure is relevant to environmental exposure and to toxicity factor development	- Exposure through i.v., i.p., or subcutaneous injection - Study examining dermal exposure - Study examining route other than that of interest
	Relevant animal model and endpoints examined	- Study used non-mammalian animal models - Endpoint not relevant to human health - Endpoint not applicable to toxicity factor development
Human/ Epidemiology	Route of exposure is relevant to toxicity factor development	- Study examining exposure route other than that of interest (e.g., dermal) - Multiple routes possible/unknown route of exposure
	Relevant endpoints examined	- Endpoint not clearly defined or measured
Mechanistic	Concentration is relevant to human exposure	- Study did not use a biologically and/or environmentally relevant exposure concentration
	Dose is applicable to ReV development	- Dose cannot be converted into an appropriate POD
	Adverse effect showed a significant positive (e.g., strong, monotonic) dose-response curve	- Adverse effect showed an overall negative dose-response curve or insufficient doses were tested
	Study evaluated effects in tissue (cells) of interest	- Evaluated tissue/cells were not relevant to ReV development

Step 3: Data Extraction

Data extraction is the third step in the systematic review process (Figure 1). During the data extraction step, studies that meet the inclusion criteria are further critically reviewed and adverse health endpoint data are summarized into evidence tables. These tables can be simple and created using Microsoft Word or Excel, or can be created in an open source content management system such as HAWC. Table 4 is a very simple example of an evidence table, and Tables 16-18 in the Appendix are examples of how these tables can be used in a DSD. More extensive data extraction tables may be required for data-rich chemicals to fully characterize the available data, including columns for study design, study size, exposure characterization and/or tested levels, the type of statistical analyses performed, and results. The purpose of these tables or databases is to display an overall view of the available data in the literature, identify potential trends in PODs, and to provide a basis to use the data as evidence.

Table 4. Example Data Extraction Table

Reference	Species/ n/sex	Exposure Concentration	Exposure Duration	NOAEL	LOAEL	Notes
Smith et al. (1973)	Humans/ 10 males	0, 50, 100 ppm	6 hours	50 ppm	100 ppm	Respiratory irritation in 9/10 volunteers

Data extraction will differ for each data stream because of differences in study design, methodologies, and data quality. Epidemiology studies include experimental and observational (analytical and descriptive) studies. Animal toxicity studies are conducted to determine dose-response, and are usually conducted for particular durations (i.e., acute, subacute, subchronic, chronic), or to study a specific effect (e.g., carcinogencity, reproductive/developmental, neurological). Mechanistic or *in vitro* studies are often conducted to determine genotoxic potential, cell transformation, cytotoxicity, or to understand the MOA, but they are often difficult to extrapolate to human-relevant exposures. Toxicity factors are based on a database of the most reliable information available (see Step 4 below) so that the values reflect the most scientifically-supported information on the potential hazards of the chemical and dose-response.

Step 4: Assessing the Quality of Individual Studies and Risk of Bias

Assessing data quality is a critical step in risk assessment (Figure 1). Studies that meet the inclusion criteria should be critically evaluated for study quality and risk of bias (ROB). Studies that were excluded based on previously stated criteria are not assessed for study quality and ROB. Section 3.3.3.1 of the TCEQ (2015) guidance briefly discusses that data quality evaluations should consider method validity, reproducibility, study reliability, dose-response relationships, temporal associations between exposures and adverse health effects, and whether critical effects are relevant to humans. ROB is a concept that was defined by the Institute of Medicine (IOM) as the “extent to which flaws in the design and execution of a collection of studies could

bias the estimate of effect for each outcome under the study" (IOM 2001 as described in NRC 2014). According to the National Academy of Sciences (NRC 2014), bias is defined as an error that decreases validity, and ROB refers to the potential for bias to occur.

Although study quality and ROB are interrelated to some extent, the NRC review of the United States Environmental Protection Agency's (USEPA) Integrated Risk Information System (IRIS) assessment recommends treating the terms separately. However, the NTP OHAT review defines study quality broadly with three main elements, including ROB: 1) reporting quality, which relates to the way the study was reported; 2) internal validity or ROB, which refers to how plausible the results of the study are and depends on how the study was designed and conducted; and 3) external validity or directness of applicability, which refers to evaluating whether the study is pertinent and applicable for the particular issue being considered (Rooney et al. 2014). The Rooney et al. (2014) review provided a comprehensive set of questions to address ROB for the different streams of data including experimental animal studies, human chamber studies, and epidemiology studies. These questions are part of a framework that underwent extensive peer-review and are pertinent to the TCEQ's chemical risk assessment program. The TCEQ uses the Rooney et al. (2014) recommendations for ROB as a guide in the development of study quality criteria. The TCEQ considers the evaluation of study quality and ROB as a single step in the systematic review process in order to efficiently review the included human, animal, and mechanistic studies.

4.1 Determining Study Quality and ROB

Risk assessments often include information from different streams of data (e.g., animal studies, human inhalation chamber studies, epidemiology studies). Each of these categories is different from the other in study design, study protocol, exposure, and species examined. While study quality is a critical component of risk assessment, there are no specific guidelines on how to collectively assess the overall study quality for all of the available data from different data streams. Additionally, defining a distinct set of rules across the different types of studies can be difficult.

The TCEQ's guidance defines study type score criteria (Tables 5, 6, 8, 9, and 10) to determine study quality for individual studies when deriving toxicity factors. Each of the selected studies is evaluated for study quality and ROB based on a number of attributes. The attributes are scored on a scale of 1 to -1, with 1 meaning the study possessed the specific attribute, 0 meaning the study did not examine the attribute, and -1 meaning the study lacked the attribute (Table 5). The total scores are then summed as a guide to compare studies within each evidence group; however, because each evidence group has a different number of scoring criteria, totals cannot be compared across the different data streams. Using scientific judgment, studies are then labeled as key, supporting or informative in nature.

The general guidelines for scoring criteria (Table 5) provide a means to evaluate all studies, regardless of type, to determine the overall quality of the study, not whether a study will be used or selected as a key study. In addition, the scoring criteria for reproductive and developmental studies, which could include data from animal, human, or mechanistic studies, are provided in Table 6.

Assigning study quality tiers can be a useful tool when evaluating data-rich chemicals, especially when the data are primarily from a single stream (e.g., animal studies, human inhalation chamber studies, epidemiology studies). Total quality scores within each stream can be divided into two tiers, with Tier 1 studies having higher overall scores, suggesting more positive attributes, while Tier 2 studies with lower overall scores suggesting more limitations. These tiers would not be used to exclude studies, but rather to present a better idea of the overall quality of the study in relation to other studies in the data stream. Text should be added along with the data tables to explain how the tiers were chosen and what role the aspects of study quality played in the overall selection of the key studies, especially when lower scoring studies are chosen. Studies can be identified at this step as key, supporting, and informative based on their ability to be used in the derivation of a toxicity factor. Since the end goal of the review is the derivation of a toxicity factor, studies that have low quality scores but are amenable to this process may be selected over studies that score higher but that lack the necessary detailed to derive a POD. Supporting studies may be used to support the use of an MOA, a route of exposure, or an exposure concentration, while an informative study may have information on MOA or the critical effect, but lacks any exposure information.

Table 5. General Guidelines for Study Quality and ROB Analysis for General Studies

Score Criteria	1	0	-1
Original data	Authors generated primary data	Authors used data from another source to draw their own conclusions	Review study, data from other sources mentioned but not further analyzed
Applicable route of exposure	Study looks at specific route of exposure relevant to ReV development	Unknown what the exact route of exposure was	Study states that a different route of exposure was studied
Single route	Study looks at a single route of exposure relevant to ReV development	Unknown if multiple routes were accounted for during exposure	Study states that multiple routes were examined
Range of doses/ exposures	Study examines >2 exposure concentrations	Study examines one or two exposure concentrations	Exposure concentration unknown
Exposure concentration known/measured	Study measures the exposure concentration (analytical)	Exposure concentration assumed but not measured/tested (nominal)	Exposure concentration unknown
Blinded study	Study specifically states that blind testing was used	Unclear whether blind testing was used	Study specifically states that blind testing was not used
Health effects relevant to ReV development	Measured health effects relevant to ReV development	Measured effects not relevant to ReV development (e.g., measured changes in protein expression, cellular changes, or other effects that may not be biologically significant)	No health effects were measured (e.g., measured air or mixture concentrations)
Single chemical exposure	Single chemical of interest or activate metabolite was used	Unknown whether additional chemicals may have been present	Study used multiple chemicals not of interest/mixture
Appropriate endpoints measured	Study examines target organ or relevance of adverse effects known or suspected based on the MOA	Study lacks information about certain relevant endpoints (e.g., measure urinary excretion but not irritation or cellular dysfunction)	Appropriate endpoints not measured (study did not examine relevance of adverse effects or adverse effects not related to MOA)
Measured outcomes reported	All measured outcomes were reported in a consistent manner	Some outcomes were reported, but not consistently	Measured outcomes were not reported
Study design sufficient/clearly defined	Study designed clearly defined and detailed in methods	Study design not defined, detailed information not provided	Study design contains an obvious flaw or problem
Calculation of sample size	Study conducts calculation to determine appropriate sample size	Study does not calculate sample size but sample size appears to be appropriate	Study does not calculate sample size and sample size does not appear to be sufficient
Confounding factors	Study eliminates or controls for any possible confounding factors	Confounding factors not identified or addressed	Study has confounding factors (e.g., smoking, behavioral patterns)
Appropriate research practices	Study provides enough detail to assume quality, uniformity, consistency, and reproducibility	Study qualities not clearly or specifically stated	Study lacks a specific aspect of quality, uniformity, consistency, or reproducibility

Table 6. Study Quality and ROB Scoring Criteria for Reproductive/Developmental Studies

Score Criteria	1	0	-1
Critical window for effects	Exposure model based on appropriate critical window (e.g., GD 6-15 for rodents)	Study uses alternate exposure window than would be expected for the measured effect	Exposure window not described or detailed
Maternal and fetal toxicity	Study examines both maternal and fetal toxicity	Study examines either maternal or fetal toxicity	Study fails to appropriately measure maternal or fetal toxicity

4.1.1 Human Studies

There is an increased interest in incorporating human data into chemical risk assessments due to various initiatives such as the World Health Organization's International Programme on Chemical Safety (IPCS) and European Union's Regulation on Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) initiative. Human studies are preferred over animal studies when developing toxicity factors, as the need to conduct animal-to-human extrapolation (e.g., dose, effect) is unnecessary, and uncertainty is decreased. However, while there is guidance on how to conduct human epidemiology studies, there is limited guidance on evaluating the integrity of the study designs and interpretation of the findings.

As mentioned in Section 3.3.3.3 of the TCEQ (2015) guidance, epidemiology studies provide data regarding associations between exposure and health effects that are useful in hazard identification, and if accompanied by sufficient, accurate and reliable exposure data, may be useful in the dose-response assessment for a toxicant. Epidemiological studies may be descriptive, analytical, or experimental in design. Descriptive studies can involve populations (ecological studies) or individuals (case reports and cross-sectional studies). Analytical study designs, where individuals are also the units of observation, include observational studies (cross-sectional, case-control, and cohort studies), and experimental designs include randomized clinical trials, field or community trials, challenge tests (i.e., human inhalation chamber studies), and interventions (Figure 2). Typically, observational study designs are the most common human studies used when determining environmental impacts on health outcomes (Rushton and Elliot 2003). Section 3.3.3.3 of the TCEQ (2015) guidelines provides a brief summary of the different study designs. The following information is provided as supplemental information to complement staff expertise with epidemiology data.

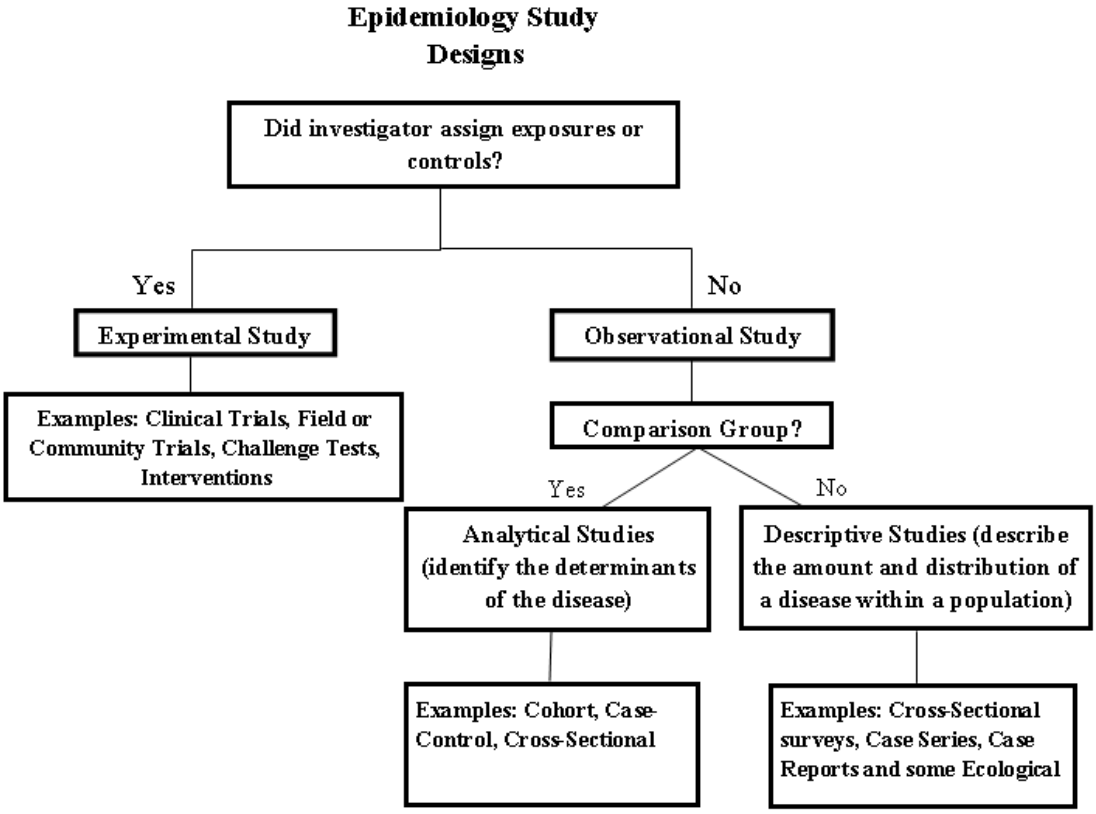

Figure 2. Epidemiology Study Designs (adapted from Rushton and Elliot 2003, and Grimes and Schulz 2002)

Epidemiology studies indirectly evaluate causality through varying exposures; therefore, one must select useful, well-designed studies for derivation of toxicity factors (Künzli and Tager 1997). Study designs can differ based on sample size and availability of subjects, units of observation, data collection methods, and directionality of exposure. Table 7 below provides a general sequence of research efforts in epidemiology, and a hierarchy based on the overall strengths, limitations, and validity of study designs. The table is adapted from Table S2 of the OHAT Approach (Rooney et al. 2014), and study types are listed from strongest to weakest (Künzli and Tager 1997). For example, ecological studies and case-reports are in the lowest tier of the hierarchy because they lack controlled exposure, there is less confidence that exposure occurs prior to the outcome, individual data may or may not be available, and they are of little use for etiologic inference (Künzli and Tager 1997, Rooney et al. 2014).

Table 7. General Sequence of Research Efforts in Epidemiology

Study Strength	Type of Study	Definition	Controlled Exposure	Exposure Prior To Outcome	Individual Outcome data	Comparison Group Used
Strongest	Experimental (Clinical Trials, Human Controlled Studies)	Investigator intentionally alters one or more exposures to study outcome effects.	Likely	Likely	Likely	Likely
	Cohort (Observational)	Two or more groups of people, who are free of disease and differ according to extent of exposure to a potential cause of disease, are compared with respect to incidence of disease in each group. The objective of a cohort study is to investigate whether the incidence of an event is related to a suspected exposure. Cohort studies can be prospective and retrospective in nature. (Szklo and Nieto 2007).	Unlikely	May or May not	Likely	Likely
	Case-Control (Observational)	A case-control study compares diseased individuals-cases and non-diseased individuals-controls with respect to their level of exposure to a suspected risk factor (Szklo and Nieto 2007).	Unlikely	May or May not	Likely	Likely
	Cross Sectional (Observational)	A cross-sectional study design examines the relationship between disease and other variables of interest as they exist in a sample of (or the total) reference population at a given point in time (Szklo and Nieto 2007).	Unlikely	Unlikely	Likely	Likely
	Ecological (Observational)	In an ecologic study, correlations are obtained between exposure rates and disease rates among different groups or populations (Szklo and Nieto 2007).	Unlikely	Unlikely	Unlikely	Unlikely
Weakest	Case Report/Series (Observational)	A case report is a descriptive study that describes and interprets single individual (case report) or small group (case series) cases based on detailed clinical evaluations and histories of the individual(s) (Szklo and Nieto 2007).	Unlikely	Unlikely	May or May not	Unlikely

Adapted from the OHAT Approach, Rooney et al. 2014

Epidemiology data can complement and enhance the evidence from toxicological studies. However, epidemiological data often lacks exposure information and may have confounding issues and bias. Critical issues relevant to exposure data include the type of assessment method used, patterns of exposure over time, and the metric used to represent exposure data (Rushton and Elliot 2003). These issues can reduce confidence due to more uncertainty. Also, controlled experimental exposures rarely occur in epidemiology studies; therefore, reliable exposure data is often limited (e.g., occupational area data as opposed to personal sample data). Controlled exposures that occur in experimental human studies can be extremely useful and are preferred over observational epidemiology studies as they provide evidence of exposure and effect (i.e., cause-and-effect), while potential confounding can be identified and controlled; however, there are also limitations. For example, human controlled exposure studies generally involve small sample sizes. Also, due to the nature of noninvasive methods and ethical considerations, exposures are limited to low exposure levels and only minor and reversible effects are studied (Rushton and Elliot 2003).

Strengths, weaknesses, and ROB should be weighed prior to making a causal association based on epidemiology studies. Further, statistically significant results should not be automatically deemed as evidence of a causal association (e.g., adequate controls or adjustments for confounders may not have been made). Thus, a positive association does not necessarily imply causation (Phillips and Goodman 2004). However, if sufficient exposure data are available and the quality is high, epidemiology studies should be used for dose-response assessment for a number of reasons: human studies are preferred over animal studies when developing toxicity factors, the need to conduct animal-to-human extrapolation (e.g., dose, effect) is unnecessary, and the uncertainty is decreased. As mentioned previously, a consensus among the scientific community on how to evaluate and rate different types of epidemiology studies is needed. Money et al. (2013) proposed a systematic review process for evaluating and scoring human data that builds on previously published information, proposed by Klimisch et al. (1997) for animal studies. The authors adapted the reliability scores to human studies to provide a comparable categorization in addressing evidence integration. However, the authors note that the interpretation of human data is not as straightforward as animal data due to variability in study designs, human genetic variation, and the importance of accounting for confounding and bias. Therefore, assigning quality scores to human data is a challenge and professional judgment is a key factor in the process.

Table 8 is an example of how the TCEQ incorporates the assessment of study quality for human data. Table 8 is used in conjunction with Tables 5 (general criteria) and 6 (reproductive/developmental criteria) to identify additional study quality and ROB scoring criteria when evaluating human studies. These criteria may be revised as needed to better assess the available data. An example of how these criteria were used in the development of toxicity factors for ethylene glycol can be found in Table 24 in the Appendix.

Table 8. Study Quality and ROB Scoring Criteria for Human Studies

Score Criteria	1	0	-1
Study Strength	Experimental Study (Clinical Trial, Human Controlled Study)	Observational (Case-Control Study)	Observational (Cross-Sectional, Ecological, Case Report/Series)
Appropriate comparison groups	Similar baseline characteristics exist between comparison groups	Minor differences exist between groups, or differences are unclear	Significant differences exist between groups
Follow up of subjects (cohort)	Subject follow up was complete and sufficient to develop endpoint of interest.	Unable or unnecessary to complete follow up (mortality study)	Subject follow up was needed but not completed or documented
Selection and Response Bias	Low selection and response bias (e.g., Adequate response rate, data completeness reported, cross-sectional studies should have relatively high response rates that are similar across groups)	Response rate and selection criteria not reported.	High selection and response bias.
Temporal relation	Exposure of interest precedes the outcome	Unclear if the exposure of interest precedes the outcome	Outcome precedes the expected exposure period
Study results consistent with other available evidence	Study outcome is consistent with other available evidence	Study outcome is partially consistent or no other evidence is available for comparison	Overall study outcome is not consistent with other available evidence

4.1.2 Animal Studies:

Klimisch et al. (1997) proposed a systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. The authors identified three categories (Reliability, Relevance, and Adequacy) to evaluate data quality in animal studies; however, the authors focused only on the reliability category to determine the Klimisch score. Relevance and adequacy were not evaluated. By using Klimisch codes in evaluating study data, the information gathered is ordered so that the most reliable and relevant studies are assessed. The TCEQ uses a variation of the Klimisch scoring method to include relevance and adequacy in the final score criteria (Table 9). The TCEQ uses the study quality and ROB scoring criteria for general (Table 5), reproductive/development (Table 6), and animal studies (Table 9) as a mechanism to evaluate reliability, relevance, and adequacy as proposed by Klimisch et al. (1997).

Klimisch et al. (1997) state that the more details provided on procedures, methodology and analytics, the more reliable and thorough the evaluation will be. In addition, the authors recommend that data reported in compliance with the principles of good laboratory practices (GLP) should have the highest grade of reliability. For relevance, as mentioned in TCEQ 2015

guidance, studies that contribute most significantly to the evidence integration and that identify adverse effects relevant to humans are selected as key studies. For example, inhalation exposure studies usually take precedence over oral exposure studies for deriving inhalation toxicity factors and, conversely, oral exposure studies typically take precedence over inhalation studies for deriving oral toxicity factors. In addition, in the absence of adequate human data, animal studies and adverse effects that are known or likely to be relevant to humans are preferred as key studies. Section 3.3.3.4, Section 3.4, and Figure 3-1 of the TCEQ (2015) guidance depicts the main steps in evaluating the human relevance of an animal MOA to humans. Section 3.15 of TCEQ (2015) guidance provides considerations for chemicals that are limited in data.

Table 9 should be used in conjunction with Tables 5 (general criteria) and 6 (reproductive/developmental general criteria) to identify additional study quality and ROB scoring criteria when evaluating animal studies. An example of how these criteria were used in the development of toxicity factors for ethylene glycol can be found in Table 21 of the Appendix.

Table 9. Study Quality and ROB Scoring Criteria for Animal Studies

Score Criteria	1	0	-1
Multiple species	Study examined effects in multiple species	Study examined effects in a single species	Study did not clearly state the species
Both sexes	Study examined effects in both sexes	Study examined effects in a single sex	Study did not specify sex
Exposure regimes (repeated vs continuous)	Study examined effects following different exposure regimes	Study examined effects following a single exposure regime	Study did not state the exposure regime
Study design reporting	Study design was clearly defined and detailed in methods	Study design was not adequately defined and detailed information not provided	Study design contained an obvious flaw or problem
General Experimental Conditions	Study used identical experimental methods across study groups	Study used experimental methods with minor differences or use of identical experimental methods is unclear	Study used experimental methods with significant differences that could affect the outcome
Randomization	Explicitly stated whether animals were randomized into treatment or control groups	Unclear if animals were randomized into treatment or control groups	Animals not randomized or no discussion of randomization included.
Concentration relevant to human exposure	Study used a biologically and environmentally relevant exposure concentration	Unclear whether exposure concentration used was biologically and/or environmentally relevant	Study used an exposure concentration that was not biologically and/or environmentally relevant
Dose applicable to ReV development	Dose can be used directly to establish a POD for ReV development	Dose must be converted/calculated in order to establish a POD	Dose cannot be converted into an appropriate POD
Control Groups	Appropriate control group used.	Unclear if appropriate control group was used.	No control group used or inappropriate control group used.
Sample Size	Sufficient number of animals used (n = 5/sex/group, or power calculation showing sufficient size).	Sufficient sample size was used but no power calculation.	Insufficient number of animals used.
Exposure or Test Substance Characterization	Details regarding source, composition, purity, and stability of test substance reported.	One or more details regarding test substance missing.	Details regarding source, composition, purity, and stability of test substance were not reported.
QA/QC Protocols	Provided details on any biological sample collection, handling, and storage methods (e.g., temperature).	Provided some details on QA/QC protocols but not complete.	QA/QC protocol details missing.
Statistical Methods	Appropriate statistical methods used, given the type of exposure and outcome tested (e.g., mixed effects models for outcomes with repeated measures).	Study did not use statistical methods appropriate for study design.	Study did not include statistical methods.

4.1.3 Mechanistic Studies

Traditional risk assessments that rely primarily on *in vivo* testing have several limitations. For example, *in vivo* testing typically focuses on apical endpoint testing that makes the whole toxicity testing process very resource intensive and expensive. The time and expense needed for *in vivo* toxicity testing are often prohibitive in terms of testing the vast influx of chemicals in commerce. *In vitro* testing has gained popularity because *in vitro* assays, in theory, can generate molecular, biochemical, or histological data. *In vitro* testing, can also provide information on perturbations of critical pathways that supplement the toxicity information for a specific chemical. *In vitro* assays can also be easily scaled to high-throughput systems, and therefore can potentially be used to screen a large number of chemicals in a short period of time. However, although *in vitro* assays can provide useful mechanistic information, there is insufficient evidence regarding translation of pathway perturbations to quantifiable adverse effects. A critical challenge to using this type of mechanistic information is translating outcomes to relevant risk assessment and risk management objectives (i.e., protection of individuals or populations) in a toxicologically predictive manner. Computational systems biology toxicity pathway models must be further developed and validated to reliably distinguish non-adverse responses (or levels of responses) for in vitro endpoints (e.g., adaptive) from those that should be deemed adverse at the cellular level (e.g., produce progressive toxicity pathway perturbations sufficient to cause adverse effects in vivo) (TCEQ 2015). When available and appropriate, the TCEQ will use *in vitro – in vivo* extrapolation (IVIVE) tools to predict *in vivo* effects.

Mechanistic data may be used to evaluate toxicokinetics, metabolism, structure-activity relationships, susceptibility, carcinogenic mechanisms, and target-organ toxicity (IARC 2017). As stated in TCEQ (2015) Guidelines, once the POD for each key study is determined, adjustments must be made to account for differences between experimental and desired exposure durations and/or differences in anatomy and physiology in experimental animals and humans. A comprehensive biologically-based dose-response model links mechanistic determinants of chemical disposition, toxicant-target interactions, and tissue responses into an overall model of pathogenesis. The proposed stages between exposure and response include processes relating exposure to consequent tissue dose (i.e., toxicokinetics) and processes that determine response to the tissue dose (i.e., toxicodynamics). If empirical data are not available to construct a comprehensive biologically-based dose-response model for a chemical, then response can be related to exposure by incorporating and integrating as much mechanistic data as possible to allow a more accurate characterization of the pathogenic process (TCEQ 2015). When possible, the TCEQ uses verified physiologically-based pharmacokinetic (PBPK) compartmental models to characterize pharmacokinetic (a.k.a. toxicokinetic) behavior of a chemical and to perform dosimetric adjustments (TCEQ 2015).

Table 10 should be used in conjunction with Tables 5 (general criteria) and 6 (reproductive/developmental criteria) to identify additional study quality and ROB scoring

criteria when evaluating mechanistic studies. An example of how these criteria were used in the development of toxicity factors for ethylene glycol can be found in Table 22 in the Appendix.

Table 10. Guidelines for Study Quality and ROB for Mechanistic Studies

Score Criteria	1	0	-1
Study design reporting	Study design was clearly defined and detailed in methods	Study design was not adequately defined and detailed information not provided	Study design contained an obvious flaw or problem
Concentration is relevant to human exposure	Study used a biologically and environmentally relevant exposure concentration	Study was unclear about using biologically and/or environmentally relevant exposure concentrations	Study did not use a biologically and/or environmentally relevant exposure concentration
Dose is applicable to ReV development	Dose can be used directly to establish a POD for ReV development	Dose must be converted/calculated in order to establish a POD	Dose cannot be converted into an appropriate POD
Tissue of Interest	Study evaluated tissues/cells related to critical effect	Unclear whether tissue/cells are related to critical effect	Study did not evaluate appropriate tissue/cells related to critical effect.
Biologically significant effect	Study used PBPK model to predict the absorption, distribution, metabolism, and excretion of chemical.	Study used appropriate cross-species adjustment using toxicokinetic and toxicodynamics of the chemical.	Study uses default due to lack of useful information.
Control Groups	Appropriate control group used.	Unclear if appropriate control group was used.	No control group used or inappropriate control group used.
Sample Size	Replicates reported; sufficient number of replicates used given method/test kit specifications.	Unclear if sufficient number of replicates are used.	Insufficient number of replicates.
Test Substance Characterization	Details regarding source, composition, purity, and stability of test substance reported.	One or more details regarding test substance missing.	Details regarding source, composition, purity, and stability of test substance were not reported.
QA/QC Protocols	Details provided on precision of test system kits and any storage conditions for test materials.	Provided some details on QA/QC protocols but not complete.	QA/QC protocol details missing.
Statistical Methods	Appropriate statistical methods used, given the type of exposure and outcome tested (e.g., mixed effects models for outcomes with repeated measures).	Study did not use statistical methods appropriate for study design.	Study did not include statistical methods.

Step 5: Evidence Integration

The NRC recently released its evaluation of the USEPA's IRIS program (May 6, 2014) in which they suggested the term "evidence integration" instead of weight of evidence (WOE) (NRC 2014). The TCEQ agrees with this terminology and uses principles of evidence integration when conducting a WOE analysis. Evidence integration is a two-step process. In the first step, evidence from each stream of data (animal studies, human studies, and mechanistic) is identified. In the second step, the evidence from the individual streams is combined with the other streams of data.

Because chemicals differ in the amount and quality of each stream of data, prescribing universally applicable rules for evidence integration is difficult. Additionally, the different types of data also have different strengths and weakness. The challenge is to determine objectives *a priori* so that evidence integration can be conducted in a transparent and consistent manner. Properly conducted evidence integration of the available data from the different streams allows confidence in the body of evidence as a whole to be rated when making determinations. As a general guideline, the following steps should be considered during the evidence integration step (adapted from Goodman et al., 2013, 2015):

- Integrate data across all realms of evidence (*e.g.*, animal, epidemiology, and mechanistic;
- Assess all data;
- Assign less weight to the results of studies that are of lower quality;
- Incorporate peer and public comments;
- Formulate conclusions.

The TCEQ provides evidence integration tables to summarize the available data for toxicity factor derivation in its DSDs. Information on the type of POD (e.g., free-standing NOAEL, minimal LOAEL) or exposure method (e.g., single dose, data amenable to benchmark dose modeling) are provided as a means to measure a study's strength for toxicity factor development. Some additional considerations when developing evidence integration tables include strength and consistency of association, biological plausibility and dose-response, coherence across data streams, and biological and clinical relevance (Goodman et al., 2013, 2015). These tables are also indicative of the considerations behind designating studies as key, supporting, or informative (See section A.5). Examples of evidence integration tables used for the ethylene glycol DSD can be found in Tables 27-29 in the Appendix. Due to the variety of chemicals and toxicity factors that are developed, these tables may be altered by TCEQ as needed.

Step 6: Rate the Confidence in the Toxicity Assessment

In this step, the confidence in the whole body of evidence is evaluated. The confidence in the body of evidence is determined by evaluating all of the elements, including type of data, study design, study quality, sample size, human relevance, and ROB that are discussed in detail in the previous steps. For example, good quality studies and lower ROB can translate to higher ratings that, in turn, indicate greater confidence and lower uncertainty that the key study findings accurately depict a true association between exposure and effect. Section 7.13 of the TCEQ (2015) guidance briefly describes the importance of recognizing and characterizing uncertainties. Higher confidence ratings generally coincide with lower uncertainty factors. Appropriately applying uncertainty factors is critical because the evidence integration approach requires some scientific judgment, use of assumptions, and data extrapolations. In addition, toxicity assessments often differ amongst scientists and regulatory agencies, and documenting uncertainties of the final toxicity values provides a transparent approach to illuminating differences in derivations.

Beck et al. (2016) developed an assessment tool that deconstructs toxicity development into elements (database completeness, systematic review, key study quality, critical effect, relevance of critical effect, point of departure, human equivalent point of departure, sensitive populations, peer review, and toxicity value comparison), and recommends scoring confidence and uncertainty for each element separately. Evaluating the elements separately allows users of toxicity values to clearly understand the inherent uncertainty of each step of the process. The authors identified major elements for both non-cancer and cancer assessments. Because many of the aspects of the elements are interrelated, the TCEQ combined the evaluations for simplicity. However, adjustments to the assessment may be made on a case-by-case basis. Table 11 provides the name of the element and the magnitude of the confidence in the elements using a qualitative ranking system of low, medium, or high confidence. Table 31 in the Appendix provides an example of how Table 11 would be used in an actual assessment for displaying the overall confidence in a toxicity assessment (for ethylene glycol) using a single metric/table. The format portrays the relative picture of the overall uncertainty and provides a rapid visualization of the confidence scoring for the overall toxicity assessment (Beck et al. 2016). In addition to the confidence table, narrative discussion of the overall uncertainty may be added to strengthen the assessment, including details on study quality, existing data gaps, uncertainty, variability, and sensitivity analyses, and how animal, human, MoA, dose-response relationships, and all relevant data are integrated as part of the conclusion.

Table 11. Confidence Scoring for Reference Values

Element	Low	Medium	High
Database Completeness	A single acute and/or chronic study was available.	Several studies were available, but some important studies were missing.	Two studies in different species, one 2-generation reproductive study, and two developmental studies were available.
Systematic Review	A systematic approach was not used.	A systematic approach was considered and some criteria were applied, but a full review was not conducted.	A systematic approach was used in study evaluation and clear criteria were established for judgment.
Key Study Quality	Selected study has deficiencies, but is still considered useful.	Selected study was reasonably well done but limitations must be considered.	Selected study was well done and can be used without restriction.
Adverse effect	Adverse effect or dose-response curve was moderate to severe. MOA information was not available.	Adverse effect was moderate; other studies are deemed necessary to determine the adverse effect.	Adverse effect was minimal severity, or the confidence in the adverse effect was high; MOA information was available.
Relevance of Adverse Effect	Adverse effect identified in animal studies is only assumed to be relevant to humans; MOA is not known for the adverse effect.	Adverse effect appears to be relevant to humans; MOA is assumed for the adverse effect and possibly relevant to humans	Adverse effect was based on a human study or matches observed human experience; MOA is well understood so adverse effect is known or assumed relevant
Point of Departure (POD)	Many uncertainties exist in POD; only a free-standing NOAEL or LOAEL is identified; few dose groups were studied; BMD modeling not possible.	Some uncertainty exists in POD, NOAEL or LOAEL; few dose groups; and difference between BMD and BMDL is large.	Basis for POD well understood (NOAEL and LOAEL); multiple dose groups were studied, BMD modeling was conducted; and the difference between BMD and BMDL is less than 2-fold.
Human Equivalent POD (PODHEC)	Many uncertainties exist in the PODHEC; no dosimetric adjustment could be made from animal POD to PODHEC.	Some uncertainty exists in adjustment to a HEC; default adjustments were used and are considered conservative	Little uncertainty exists because human data are available; or the HED/HEC is known from PBPK or dosimetry model or CSAF.
Sensitive Populations	Many uncertainties on sensitive populations exist and are not addressed	Uncertainties on sensitive populations exist but default procedures are presumed to be conservative	Human data on sensitive populations are available and uncertainties are addressed
Peer Review	Limited or no peer review; unaddressed comments would significantly change risk value; no independent check	Adequate peer review; most substantive comments addressed; disregarded comments would not significantly change value	High quality panel peer review with appropriate experts; all substantive comments addressed as per independent check
Toxicity Value Comparison	Relevant risk values show a greater than 10-fold difference	Some relevant risk values agree within 3-fold of each other, and others disagree within 10-fold of each other	All relevant risk values agree within 3-fold of each other

* Criteria for scoring the individual elements adapted from Beck et al. (2016)

Limitations

The TCEQ recognizes that there are complex limitations in the proposed systematic review approach regarding the development of chemical-specific toxicity factors. In general, we understand that ultimately relying on "expert judgment" is a limitation of systematic reviews. However, the nature of regulatory risk assessment is inherently reliant on sound scientific judgment informed by the body of scientific data. The point of the systematic review process aims to document the scientific basis for those judgments and ultimately improve the decision-making process, increase transparency, minimize bias, and improve consistency between different risk assessments.

1. Problem Formulation and Protocol Development: In this framework, the problem formulation and the protocol development sections are presented in a general manner because they must be applicable to a wide array of chemicals, data sets, and endpoints. Due to the regulatory nature of the TCEQ toxicology program, this guidance must also be applicable for chemicals for which limited toxicity data are available. Therefore, the guidelines are designed to be versatile depending on the chemical under assessment. Documenting these changes will minimize bias and increase transparency in the scientific basis for regulatory decisions.

2. Literature Searches: The TCEQ conducts thorough literature searches of relevant databases and takes prudent steps to identify all relevant studies during the literature review process. While initial search terms may limit the identification of relevant literature, the review of additional sources (e.g., toxicological reviews) allows for identification of other relevant literature for inclusion and the refinement of search criteria.

3. Inclusion and Exclusion Criteria: Defining one set of inclusion and exclusion criteria for all chemicals is difficult since often the criteria will be chemical and/or purpose-specific. Therefore, inclusion and exclusion criteria may be modified based on scientific judgment as needed to allow for the identification of other relevant literature as long as the changes are documented.

4. Scoring criteria: Defining a distinct set of rules across the different types of studies and data streams can be difficult. These criteria may be revised as needed to better assess the available data depending on the chemical under assessment. In addition, the study quality criteria provided are general and subject to scientific judgement and interpretation. The TCEQ will consider reliability testing in the future to enhance the application of these criteria.

Conclusions

Systematic reviews and evidence integration are becoming increasingly important in chemical risk assessments (Rooney et al. 2014, NRC 2014, and Rhomberg et al. 2013). Each phase of the systematic review and evidence integration process plays an important role in improving confidence and transparency in the risk assessment process. In conducting systematic reviews, the TCEQ:

- Sets clear inclusion and exclusion criteria to promote transparency and limit subjective scientific judgment;
- Assesses data quality and conducts ROB analysis that result in higher confidence in the key studies and lessen uncertainty; and
- Weighs the evidence from different data streams prior to integrating the evidence, creating greater confidence in the final toxicity factor.

This guidance document may be revised based upon experience with its implementation or as additional tools and resources become available.

References

Beck, N., Wise,K., Becker, R., Dourson, M., Nancy, P., Erraguntla, N., Grant, R.L., Shirley, S., Gray, G., Farland, B., Lakind, J., Simon, T., Santos, S., Kirman, C., Lewis, R.J., Pottenger, L. (2016). Approaches for Communicating Overall Uncertainty in Hazard Assessment and Dose-Response Assessment: U.S. Environmental Protection Agency's Integrated Risk Information System (IRIS) as a Case Study. Environmental International 89-90:110-128.

Becker, R., Dellarco, V., Seed, J., Kronenber, J., Meek, B., Formean, C., Kirman, C., Linkov, I., Schoeny, R., Dourson, M., Pottenger, L., Manibusan, M. (2017). Quantitative weight of evidence to assess confidence in potential modes of action. Regulatory Toxicology and Pharmacology 86:205-220.

Cochrane Collaboration. (2011). Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. http://handbook.cochrane.org/front_page.htm

Grimes, D., Schulz, K. (2002). An overview of clinical research: the lay of the land. Lancet (London, England) 359(9300): 57-61.

IARC. (2017). International Agency for Research on Cancer. Instructions to Authors for the Preparation of Drafts for IARC Monographs.

Klimisch, H. Andreae, M., Tilmann, U. (1997). A Systematic Approach for Evaluating the Quality of Experimental Toxicological and Ecotoxicological Data. Regulatory Toxicology and Pharmacology 25(1): 1-5.

Künzli, N., Tager, B. (1997). The semi-individual study in air pollution epidemiology: A valid design as compared to ecologic studies. Environ Health Perspect 105(10): 1078-1083.

Money, C., Tomenson, J., Penman, M., Boogaard, P., Lewis, J. (2013). A systematic approach for evaluating and scoring human data. Regulatory Toxicology and Pharmacology 66(2): 241-247.

National Research Council (NRC). *Review of EPA's Integrated Risk Information System (IRIS) Process*. Washington, DC: The National Academies Press, 2014.

Phillips, C. and Goodman, K. (2004). The missed lessons of Sir Austin Bradford Hill. Epidemiologic Perspectives & Innovations 1(1): 1-5.

Rhomberg, L., Goodman, J., Bailey, L., Prueitt, R., Beck, N., Honeycutt, M., Kaminski, N., Paoli, G., Pottnege, L., Scherer, R., Wise, K., Becker, R. (2013). A survey of frameworks for best practices in weight-of-evidence analyses. Crit Rev Toxicology 43(9): 753-84.

Rooney, A., Boyles, A., Wolfe, M., Bucher, J., Thayer, K. (2014). Systematic review and evidence integration for literature-based environmental health science assessments. Environ Health Perspect 122(7): 711-718.

Rushton, L., Elliott, P. (2003). Evaluating evidence on environmental health risks. Br Med Bull. 68:113-28.

Szklo, M. and Nieto, F. (2007). Epidemiology : beyond the basics. Sudbury, Mass., Jones and Bartlett Publishers.

TCEQ. (2015). Texas Commission on Environmental Quality. Guidelines to develop effects screening levels, reference values, and unit risk factors (RG-442). Chief Engineer's Office.

Uman, L. (2011). Systematic Reviews and Meta-Analyses. Journal off The Canadian Academy of Child & Adolescent Psychiatry. 20(1): 57-59.

USEPA. (2005). United States Environmental Protection Agency. Science Issue Paper: Mode of Carcinogenic Action for Cacodylic Acid (Dimethylarsinic Acid, DMA) and Recommendations for Dose Response Extrapolation. Health Effects Division of Office of Pesticide Programs. Accessed 9/15/2017: https://archive.epa.gov/pesticides/reregistration/web/pdf/dma_moa-2.pdf

Appendix: Example of the Systematic Review and Evidence Integration Used in the Ethylene Glycol Development Support Document (DSD)

A.1 Problem Formulation and Protocol

Problem formulation identifies and describes the extent of the evaluation. These questions structured the systematic review for ethylene glycol (EG):

- What are the physical and chemical properties of EG?
- What is the critical effect following exposure to EG?
- Are there sensitive subpopulations?
- What is the mode of action (MOA)?
- Does route of exposure play a role?
- Is EG carcinogenic, and if so, is it carcinogenic by a specific route of exposure?
- Is EG a reproductive or developmental toxicant?

Protocol development is another important aspect in the initial process. A protocol is typically developed around a PECO statement: Populations, Exposure, Comparator/Control, and Outcomes. These identifiers are used to lay out the framework for the literature search and inclusion/exclusion criteria. The PECO statement for EG followed the criteria in Table 12:

Table 12. PECO statement used by the TCEQ to develop toxicity factors for Ethylene Glycol (EG)

Population	General human population and any relevant sensitive subpopulations, animals, and vegetation
Exposure	Exposure to EG, surrogates with demonstrated similar MOAs, and any identified metabolites
Comparator/ Control	Populations exposed to concentrations below the concentration that causes the most sensitive critical effect
Outcome(s)	The most sensitive critical effect directly related to EG exposure

The protocol used for the systematic review and the development of toxicity factors for EG is as follows:

1. Identify the chemical of interest and define the causal questions

2. Conduct a systematic review
    a. Conduct a systematic literature search
    b. Identify the inclusion/exclusion criteria
    c. Extract the relevant data from each data stream (human, animal, mechanistic)
    d. Assess the study quality and conduct a risk of bias analysis

   e. Weigh the evidence in each data stream and then integrate the evidence across the data streams
   f. Rate the confidence in the evidence
3. Derive toxicity factors (TCEQ 2015)
   a. Review the essential data, including chemical/physical properties and selected key studies from the systematic review
   b. Conduct MOA analysis
   c. Choose the most appropriate dose metric available (e.g., tissue dose, air concentration)
   d. Select critical effect, based on human equivalent exposure considering each key study
   e. Extrapolate from the adjusted POD to lower exposures based on MOA analysis

A.2 Systematic Literature Review and Study Selection

As a first step, publically available databases were searched using explicitly stated search criteria. Please see TCEQ (2015) for a list of available databases that were searched. The search terms used in the literature review for EG, along with the number of results from PubMed, are found in Table 13. Additional references were also identified using the reference sections from some of the selected studies. This literature review was conducted in June, 2015, and therefore studies published after this date were not available at the time of the review.

Table 13. Search strings used in the literature review of EG

Search Term/String	PubMed Results
ethylene glycol	20205
“ethylene glycol”	18895
“ethylene glycol” [mesh]	2093
“ethylene glycol” [mesh] NOT “ethylene oxide”	2077
“ethylene glycol” [mesh] NOT “ethylene oxide” AND (inhal* OR air OR carc* OR onco* OR oral)	168
“ethylene glycol” [mesh] NOT “ethylene oxide” AND (inhal* OR air OR carc* OR onco*)	106

An additional PubMed search was conducted using the search terms “ethylene glycol” AND inhalation, which resulted in 105 references. These references were compared to the list generated above, and added as needed. The selected studies were imported into the Health Assessment Workspace Collaborative (HAWC) systematic literature review tool. As a team, each title and abstract was reviewed for relevance and tagged for either inclusion (human, animal, or

mechanistic), or exclusion (not a relevant/applicable study). For EG, a number of studies involving cryopreservation and chemical synthesis were excluded due to the lack of relevance in a health-based risk assessment. Other reasons for initial exclusion included studies using chemicals other than EG (di- or triethylene glycol, ethylene glycol ethers, etc.), studies that did not look at toxic effects (bactericidal or solvent effects), and unrelated mechanistic studies.

Additionally, several governmental and private sector organizations were searched for published literature and toxicity values for EG, and the available documents are listed in Table 14. Relevant referenced articles from documents listed in Table 14 were then added to the pool of selected material.

Table 14. Available reviews and toxicity values for EG

Organization	Year	Toxicity Value
Agency for Toxic Substances and Disease Registry (ATSDR) Toxicological Profiles	2010	Acute MRL*
Integrated Risk Information System (IRIS) USEPA	1989	Oral RfD*
Office of Environmental Health Hazard Assessment (OEHHA) CalEPA	2000	Chronic REL*
Health Canada	2000	NA
International Programme on Chemical Safety (IPCS)	2002	NA

MRL – minimal risk level, RfD – reference dose, REL – reference exposure level

Following this initial review, which produced a pool of ~170 articles and documents, specific inclusion and exclusion criteria were used to narrow down the pool of available data. The criteria, along with examples of the kinds of studies that were excluded, can be found in Table 15.

Table 15. Inclusion/exclusion criteria used in the review of EG

Study Type	Inclusion Criteria	Exclusion Criteria
General	Complete study available for review	- Only abstract is available - Study in a language other than English - Unpublished report/unable to retrieve
	Exposure concentration is environmentally relevant	- Significantly high concentrations used - Study focused on overdose/poisoning or mortality - Exposure concentration unknown
	Study contains original data	- Study is a review article
	Study examines effects related to chemical exposure	- Study measures concentration in products, etc. - Study does not examine health effects
	Study focused on the chemical of concern or active metabolites	- Study examined mixture effects (i.e. antifreeze) - Study on treatment following EG exposure
Animal	Route of exposure is relevant to environmental exposure and to toxicity factor development	- Exposure through i.v., i.p., or subcutaneous injection - Study examining dermal exposure - Study examining oral exposure*
	Relevant animal model and endpoints examined	- Study used non-mammalian animal models - Endpoint studied not relevant to human health - Endpoint not applicable to toxicity factor development
Human/Epi	Route of exposure is relevant to toxicity factor development	- Study examining dermal exposure - Study examining oral exposure* - Multiple routes possible/unknown route of exposure
	Relevant endpoints examined	- Study focused on mortality/intentional ingestion

i.v. – intravenous, i.p. – intraperitoneal

* Studies using the oral route of exposure were initially excluded from the key study selection due to the inhalation route being more applicable to the development of a ReV/ESL. Oral data may be used to fill gaps in the inhalation data as needed.

Using these inclusion/exclusion criteria, the pool of available data was narrowed down to 18 included studies: 7 human studies, 6 animal studies, and 5 mechanistic/*in vitro* studies. These studies were collected and reviewed in detail by each of the authors.

A.3 Data Extraction

Each of the identified studies was reviewed in detail and the primary data was extracted for potential use in the Ethylene Glycol Development Support Document (DSD). Data from the studies can be found in Table 16 (human studies), Table 17 (animal studies), and Table 18 (*in vitro* studies). Data that were applicable to the development of the acute and chronic ReVs and ESLs are also in sections 3.1.2 and 4.1.2, respectively.

Table 16. Data extraction from human studies

Reference	Exposure Concentration	Exposure Duration	NOAEL	LOAEL	Notes
Bond et al. (1985)	Unknown	Varied	--	--	Case-control study of chemical plant workers
Carstens et al. (2003)	25, 28 mg/m^3 (vapor)	4 h	28 mg/m^3	--	Health effects not measured or reported
Gérin et al. (1997)	Varied	Sampled 42 working days over 2 months	<22 mg/m^3 (vapor), 190 mg/m^3 (aerosol)	--	No changes in measured biomarkers for kidney effects
Laitinen et al. (1995)	<1.9 ppm (vapor)	Varied	--	--	Changes in urinary markers, possible dermal exposure
Troisi et al. (1950)	Unknown	Varied	--	--	Noted symptoms in chemical plant workers
Upadhyay et al. (2008)	25, 30 mg/m^3 (vapor)	4 h	30 mg/m^3	--	Health effects not measured or reported
Wills et al. (1974)	0.8-75 mg/m^3, 188, 244, 308 mg/m^3 (aerosol)	Varied	34 mg/m^3 (mean 7 d), 75 mg/m^3 (high)	140 mg/m^3 (duration not reported)	Respiratory irritation occurred after 140 mg/m^3, no changes in urinary markers

Table 17. Data extraction from animal studies

Reference	Species	Exposure Concentration	Exposure Duration	NOAEL	LOAEL	Notes
Coon et al. (1970)	Rats, guinea pigs, rabbits, monkeys, dogs	10 and 57 mg/m^3 repeatedly or 12 mg/m^3 continuously (vapor)	8 h/d, 5 d/wk for 6 wk (repeated) or 90 d (continuously)	--	10 mg/m^3 (repeated) 12 mg/m^3 (continuous)	Moderate to severe eye irritation in rabbits and rats, nonspecific inflammatory changes in the lungs of all the species
Corley et al. (2005)	Various	Various	Various	--	--	PBPK model development using various studies
Corley et al. (2011)	Various	Various	Various	--	--	PBPK model development using various studies
Marshall and Cheng (1983)	Rats	32 mg/m^3 (vapor), 184 (aerosol) mg/m^3	30 min (vapor), 17 min (aerosol)	32 mg/m^3 (vapor), 184 mg/m^3 (aerosol)	--	Health effects not measured or reported
Tyl et al. (1995a)	Rats and mice	0, 150, 1000, and 2500 mg/m^3 (aerosol, whole body)	6 h/d on GD 6-15	1000 mg/m^3 (maternal) 150 mg/m^3 (fetal)	2500 mg/m^3 (maternal) 1000 mg/m^3 (fetal)	Increased resorptions, decreased fetal body weight, possible oral exposure
Tyl et al. (1995b)	Mice	0, 500, 1000, and 2500 mg/m^3 (aerosol, nose-only)	6 h/d on GD 6-15	500 mg/m^3 (maternal) 1000 mg/m^3 (fetal)	1000 mg/m^3 (maternal) 2500 mg/m^3 (fetal)	Increased maternal kidney weights, fetal skeletal variations

Table 18. Data extraction from mechanistic studies

Reference	Model	Exposure Concentration	Exposure Duration	NOAEL	LOAEL	Notes
Capo et al. (1993)	Rat embryonic nerve cells	0.01, 0.1, 1, 10, 100 µM	24 h	--	0.01 µM (IC50 0.26 µM)	Neuronal degeneration, decrease in cell number
Carney et al. (1996)	Rat whole embryo culture	0.5, 2.5, 12.5, 25, 50 mM EG or GA	48 h	50 mM EG, 2.5 mM GA	12.5 mM GA	Inhibition of embryo growth and development
Carney et al. (2008)	Rabbit whole embryo culture	2.5, 6, 12.5, 25, 50 mM GA	48 h	50 mM GA	--	No significant adverse effects on developing embryos
Guo et al. (2007)	Human proximal tubule cells	0-25 mM EG or metabolites	6 h	25 mM EG	2 mM oxalate	Cytotoxicity and decreased cell viability
Klug et al. (2001)	Rat whole embryo culture	0-200 mM EG or metabolites	48 h	200 mM EG	0.1 mM GAl, 3 mM GA	Embryotoxicity, morphological changes

GA – glycolate, GAl - glycoaldehyde

A.4 Study Quality and Risk of Bias (ROB)

Each of the selected studies was evaluated for study quality and ROB based on a number of attributes determined prior to this review. The attributes were scored on a scale of 1 to -1, with 1 meaning the study possessed the specific attribute, 0 meaning the study did not examine the attribute, and -1 meaning the study lacked the attribute. Each of these study quality attributes along with the criteria used in scoring them can be found in Table 19 (general studies), Table 20 (human studies), Table 21 (animal studies), Table 22 (*in vitro* studies), and Table 23 (reproductive and developmental studies).

Table 19. Study quality and ROB scoring criteria for general studies

Score Criteria	1	0	-1
Original data	Authors generated primary data	Authors used data from another source to draw their own conclusions	Review study, data from other sources mentioned but not further analyzed
Applicable route of exposure	Study looks at specific route of exposure relevant to ReV development	Unknown what the exact route of exposure was	Study states that a different route of exposure was studied
Single route of exposure	Study looks at a single route of exposure relevant to ReV development	Unknown if multiple routes were accounted for during exposure	Study states that multiple routes were examined
Single chemical exposure	Single chemical of interest or activate metabolite was used	Unknown whether additional chemicals may have been present	Study used multiple chemicals/mixture
Range of doses/ exposures	Study examines >2 exposure concentrations	Study examines one or two exposure concentrations	Exposure concentration unknown
Exposure concentration known/ measured	Study measures the exposure concentration (analytical)	Exposure concentration assumed but not measured/tested (nominal)	Exposure concentration unknown
Blinded study	Study specifically states that blind testing was used	Unclear whether blind testing was used	Study specifically states that blind testing was not used
Health effects relevant to ReV development	Measured health effects relevant to ReV development	Measured effects not relevant to ReV development (e.g. measured changes in protein expression, urinary excretion)	No health effects were measured (e.g. measured air or mixture concentrations)
Appropriate endpoints measured	Study examines target organ or relevance of adverse effects known or suspected in be involved in MOA	Study lacks information about certain relevant endpoints (e.g. measured urinary excretion but not irritation or other effects)	Appropriate endpoints not measured (study did not examine relevance of adverse effects or effects not part of MOA)
Measured outcomes reported	All measured outcomes were reported in a consistent manner	Some outcomes were reported, but not consistently	All measured outcomes were not reported
Study design sufficient/ clearly defined	Study designed clearly defined and detailed in methods	Study design not defined, detailed information not provided	Study design contains an obvious flaw or problem
Calculation of sample size	Study conducts calculation to determine appropriate sample size	Study does not calculate sample size but sample size appears to be appropriate	Study does not calculate sample size and size does not appear to be sufficient
Confounding factors	Study eliminates or controls for any possible confounding factors	Confounding factors not identified or addressed	Study has confounding factors (e.g. smoking, behavioral patterns)
Appropriate research practices	Study provides enough detail to assume quality, uniformity, consistency, and reproducibility	Study qualities not clearly or specifically stated	Study lacks a specific aspect of quality, uniformity, consistency, or reproducibility

Table 20. Study quality and ROB scoring criteria for human studies

Score Criteria	1	0	-1
Appropriate comparison groups	Comparison groups have similar baseline characteristics	Minor differences exist between groups, or are differences unclear	Significant differences exist between groups
Follow up of subjects	Subject follow up was complete and thorough	Unable or unnecessary to complete follow up (mortality study)	Subject follow up was needed but not completed
Temporal relation	Exposure of interest precedes the outcome	Unclear if the exposure of interest precedes the outcome	Outcome proceeds the expected exposure period
Study results consistent with other available evidence	Study outcome is consistent with other available evidence	Outcome is partially consistent or no other evidence is available for comparison	Overall study outcome is not consistent with other available evidence

Table 21. Study quality and ROB scoring criteria for animal studies

Score Criteria	1	0	-1
Multiple species	Studied examined effects in multiple species	Studied examined effects in a single species	Species not clearly stated
Both sexes	Studied examined effects in both sexes	Studied examined effects in a single sex	Sex not specified
Exposure regimes (repeated vs continuous)	Studied examined effects following different exposure regimes	Studied examined effects following a single exposure regime	Exposure regime not stated
Identical experimental conditions across study groups	Study used identical experimental methods across study groups	Minor differences exist, or use of identical experimental methods are unclear	Significant differences exist that could affect the outcome
Concentration relevant to human exposure	Study used a biologically and environmentally relevant exposure concentration	Unclear whether exposure concentration used was biologically and/or environmentally relevant	Exposure concentration was not biologically and/or environmentally relevant
Dose applicable to ReV development	Dose can be used directly to establish a POD for ReV development	Dose must be converted/calculated in order to establish a POD	Dose cannot be converted into an appropriate POD
Dose-response relationship	Critical effect showed a significant positive dose-response curve	Critical effect failed to show a significant dose-response curve	Critical effect showed a significant negative dose-response curve

Table 22. Study quality and ROB scoring criteria for mechanistic studies

Score Criteria	1	0	-1
Concentration is relevant to human exposure	Study used a biologically and environmentally relevant exposure concentration	Unclear whether exposure concentration used was biologically and/or environmentally relevant	Exposure concentration was not biologically and/or environmentally relevant
Dose is applicable to ReV development	Dose can be used directly to establish a POD for ReV development	Dose must be converted/calculated in order to establish a POD	Dose cannot be converted into an appropriate POD
Dose-response relationship	Critical effect showed a significant positive dose-response curve	Critical effect failed to show a significant dose-response curve	Critical effect showed a significant negative dose-response curve

Table 23. Study quality and ROB scoring criteria for reproductive/developmental studies

Score Criteria	1	0	-1
Critical window for effects	Exposure model based on appropriate critical window (e.g. GD 6-15 for rodents)	Study uses alternate exposure window than would be expected for the measured effect	Exposure window not described or detailed
Maternal and fetal toxicity	Study examines both maternal and fetal toxicity	Study examines either maternal or fetal toxicity	Study fails to appropriately measure maternal or fetal toxicity

Rankings for each of the identified studies can be found in Table 24 (human studies), Table 25 (animal studies), and Table 26 (in vitro studies). Note that total scores were added as a guide to compare within the study groups; however, because each study group has a different number of scoring criteria, totals should not be compared across groups.

Table 24. Study quality and ROB scoring for the selected EG human studies

Study criteria	Bond 1985	Carstens 2003	Gerin 1997	Laitinen 1995	Troisi 1950	Upadhyay 2008	Wills 1974
General							
Original data	1	1	1	1	1	1	1
Applicable route of exposure	0	1	1	1	1	1	1
Single route of exposure	0	1	-1	-1	0	1	0
Single chemical exposure	-1	1	-1	-1	-1	1	1
Range of doses/exposures	-1	0	1	0	-1	0	1
Exposure concentration known/ measured	-1	1	1	1	-1	1	1
Blinded study	0	0	0	0	0	0	0
Health effects relevant to ReV development	1	0	0	0	0	0	1
Appropriate endpoints measured	1	0	0	0	0	0	1
Measured outcomes reported	1	1	1	1	0	1	1
Study design sufficient/ clearly defined	0	1	1	1	-1	1	0
Calculation of sample size	0	-1	0	-1	0	-1	0
Confounding factors	-1	0	0	0	-1	0	-1
Appropriate research practices	1	1	1	1	0	1	-1
Human							
Appropriate comparison groups	0	0	-1	1	-1	-1	1
Follow up of subjects	0	0	0	0	1	0	0
Temporal relation	1	1	1	1	1	1	1
Study results consistent with other available evidence	0	1	1	1	0	1	1
Total Points	**2**	**9**	**6**	**6**	**-2**	**8**	**9**
Study Selection – Key, supporting, or informative	**I**	**S**	**I**	**I**	**I**	**S**	**K**
Acute or chronic	**C**	**A**	**C**	**C**	**C**	**A**	**A/C**

Table 25. Study quality and ROB scoring for the selected EG animal studies

Study criteria	Coon 1970	Corley 2005	Corley 2011	Marshall 1983	Tyl 1995a	Tyl 1995b
General						
Original data	1	0	0	1	1	1
Applicable route of exposure	1	1	1	1	1	1
Single route	1	-1	0	1	-1	0
Single chemical exposure	1	1	1	1	1	1
Range of doses/ exposures	1	0	0	0	1	1
Exposure concentration known/ measured	1	0	0	1	1	1
Blinded study	0	0	0	0	0	0
Health effects relevant to ReV development	1	0	0	0	1	1
Appropriate endpoints measured	1	0	0	0	1	1
Measured outcomes reported	1	0	0	1	1	1
Study design sufficient/ clearly defined	0	1	1	1	1	1
Calculation of sample size	0	0	0	0	0	0
Confounding factors	0	0	0	0	0	-1
Appropriate research practices	0	0	0	1	1	1
Animal						
Multiple species	1	1	1	0	1	0
Both sexes	1	1	1	1	1	1
Exposure regimes (repeated vs continuous)	1	0	0	0	0	0
Concentration relevant to human exposure	0	0	0	0	0	0
Dose applicable to ReV development	1	0	0	1	1	1
Dose-response relationship	0	0	0	0	1	1
Reproductive/developmental						
Critical window for effects	-	-	-	-	1	1
Maternal and fetal toxicity	-	-	-	-	1	1
Total Points	**13**	**4**	**5**	**10**	**15**	**14**
Study Selection – Key, supporting, or informative	**S/K**	**I**	**I**	**S**	**I**	**S**
Acute or chronic	**A/C**	**A**	**A**	**A**	**A**	**A**

Table 26. Study quality and ROB scoring for the selected EG mechanistic studies

Study criteria	Capo 1993	Carney 1996	Carney 2008	Guo 2007	Klug 2001
General					
Original data	1	1	1	1	1
Applicable route of exposure	-1	-1	-1	-1	-1
Single route	1	1	1	1	1
Single chemical exposure	1	1	1	1	1
Range of doses/ exposures	1	1	1	1	1
Exposure concentration known/ measured	1	1	1	1	1
Blinded study	0	1	0	0	0
Health effects relevant to ReV development	0	1	1	0	1
Appropriate endpoints measured	0	1	1	1	1
Measured outcomes reported	1	1	1	1	1
Study design sufficient/clearly defined	0	1	1	0	1
Calculation of sample size	0	0	0	0	0
Confounding factors	0	0	0	0	0
Appropriate research practices	1	1	1	1	1
Mechanistic					
Concentration is relevant to human exposure	0	1	1	0	0
Dose is applicable to ReV development	0	0	0	0	0
Dose-response relationship	1	1	0	1	1
Reproductive/developmental					
Critical window for effects	-	1	1	-	1
Maternal and fetal toxicity	-	0	0	-	0
Total Points	**7**	**13**	**11**	**8**	**11**
Study Selection – Key, supporting, or informative	**I**	**I**	**I**	**I**	**I**
Acute or chronic	**A**	**A**	**A**	**A**	**A**

A.5 Evidence Integration

After addressing the study quality and ROB for each of the selected studies, points were totaled to gain a better understanding of the quality of each paper and to compare the studies within each of the study groups. The information from each of the data streams (human, animal, and mechanistic) were then assessed for use as key, supporting, orinformative studies based on the study quality criteria. As mentioned previously, because each study group has a different number of scoring criteria, totals were not compared across groups. The reasoning behind identifying each study as key, supporting or informative was compiled into the evidence integration tables found in Tables 27-29. As seen in Table 24 and 27, Wills et al. (1974) was chosen as the key study for the derivation of the acute toxicity factors because a LOAEL of 140 mg/m^3 was identified for common complaints of respiratory irritation. As seen in Table 25 and 28, there were several animal studies of high quality; however, because human data are preferred when developing toxicity factors, the Wills et al. (1974) study was chosen as the key study. For derivation of the chronic toxicity factors, the Coon (1970) study was selected even though it did not score highest in the study quality scoring. Although overall both Tyl et al. (1995a &b) studies scored higher, these studies did not provide adequate POD values for use in toxicity factor development, and therefore were not selected as the key study. The human, animal and mechanistic studies that were not chosen as the key study were used either as supporting data or background information in the DSD.

Table 27. Evidence Integration Table for Human Studies

Study	Species	Type	Reasoning
Bond et al. (1985)	Human	Informative	- No exposure concentrations available - Health effects not associated with exposure
Carstens et al. (2003)	Human	Supporting	- Health effects not directly measured, but also not reported - Free-standing NOAEL
Gérin et al. (1997)	Human	Informative	- Measured air concentrations, but actual exposure unknown - No measured health effects
Laitinen et al. (1995)	Human	Informative	- Measured air concentrations, but actual exposure unknown - No measured health effects
Troisi et al. (1950)	Human	Informative	- No exposure concentrations available - Multiple chemical exposure
Upadhyay et al. (2008)	Human	Supporting	- Health effects not directly measured, but also not reported - Free-standing NOAEL
Wills et al. (1974)	Human	Key	- Acute respiratory irritation, free-standing LOAEL - Subacute free-standing NOAEL for kidney toxicity biomarkers - Exposure concentration suitable for toxicity factor derivation

Table 28. Evidence Integration Table for Selected Animal Studies

Study	Species	Type	Reasoning
Coon et al. (1970)	Rats, guinea pigs, rabbits, monkeys, dogs	Key	- Multiple species examined - Acute ocular irritation free-standing LOAEL - Chronic systemic free-standing LOAEL - Few dose groups, NOAEL not identified
Corley et al. (2005)	Various	Informative	- PBPK model based on previous studies - No exposure/dose response data available
Corley et al. (2011)	Various	Informative	- PBPK model based on previous studies - No exposure/dose response data available
Marshall and Cheng (1983)	Rats	Supporting	- Health effects not directly measured, but also not reported - Free-standing NOAEL
Tyl et al. (1995a)	Rats and mice	Informative	- NOAEL and LOAEL for maternal and fetal toxicity - Two species tested - Significant oral exposure from grooming behaviors
Tyl et al. (1995b)	Mice	Supporting	- NOAEL and LOAEL for maternal and fetal toxicity - Minimum oral exposure due to nose-only exposure - Skeletal malformations linked to restraining apparatus

Table 29. Evidence Integration Table for Selected Mechanistic Studies

Study	Species	Type	Reasoning
Capo et al. (1993)	Rat embryonic nerve cells	Informative	- Informative for EG MOA - Not clear if dose is relevant to human inhalation exposure
Carney et al. (1996)	Rat whole embryo culture	Informative	- Informative for MOA of EG and metabolites - Developmental study, fetal toxicity - Not clear if dose is relevant to human inhalation exposure
Carney et al. (2008)	Rabbit whole embryo culture	Informative	- Informative for MOA of EG and metabolites - Developmental study, fetal toxicity - Not clear if dose is relevant to human inhalation exposure
Guo et al. (2007)	Human proximal tubule cells	Informative	- Informative for MOA of EG and metabolites - Not clear if dose is relevant to human inhalation exposure
Klug et al. (2001)	Rat whole embryo culture	Informative	- Informative for MOA of EG and metabolites - Developmental study, fetal toxicity - Not clear if dose is relevant to human inhalation exposure

A.6 Confidence Rating

Table 30 provides scoring criteria to rate the confidence and uncertainty for each aspect or element of the toxicity assessment. The table provides the name of the element and the magnitude of the confidence in each element using a qualitative ranking system of low, medium, or high confidence. Table 31 displays the overall confidence in the ethylene glycol toxicity assessment. **As seen in Table 31, overall, the TCEQ has medium confidence in the development of the ethylene glycol reference values based on the scoring elements defined in table 30.**

Table 30. Confidence Scoring Criteria

Element	Low	Medium	High
Database Completeness	A single acute and/or chronic study was available	Several studies were available, but some important studies were missing.	Two studies in different species, one 2-generation reproductive study, two developmental studies
Systematic Review	A systematic approach was not used.	A systematic approach was considered and some criteria were applied, but a full review was not conducted	A systematic approach was used in study evaluation and clear criteria are established for judgment
Key Study Quality	Selected study has deficiencies, but is still considered useful	Selected study was reasonably well done but some restrictions must be considered	Selected study was well done and can be used without restriction
Critical effect	Critical effect or dose-response curve was moderate to severe. MOA information not available.	Critical effect was moderate; other studies are deemed necessary to determine the critical effect.	Critical effect was of minimal, or the confidence in the critical effect was high. MOA information available.
Relevance of Critical Effect	Critical effect identified in animal studies is only assumed to be relevant to humans; MOA is not known for the critical effect	Critical effect appears to be relevant to humans. MOA is known for the critical effect and possibly relevant to humans.	Critical effect based on a human study or matches observed human experience; MOA is well understood so critical effect is assumed relevant.
Point of Departure (POD)	Many uncertainties exist in POD; only a free-standing NOAEL or LOAEL identified; few dose groups; BMD modeling not possible	Some uncertainty exists in POD, NOAEL or LOAEL; few dose groups; difference between BMD and BMDL is large	Basis for POD well understood: NOAEL and LOAEL; multiple dose groups, BMD modeling conducted; difference between BMD and BMDL less than 2-fold
Human Equivalent POD (PODHEC)	Many uncertainties exist in the PODHEC; no dosimetric adjustment from animal POD to PODHEC	Default adjustments used and considered conservative; some uncertainty exists in adjustment to a HEC.	Human data available; HED/HEC is known from PBPK or dosimetry model or CSAF
Sensitive Populations	Many uncertainties on sensitive populations exist and are not addressed.	Information on sensitive population is not known but default procedures are presumed to be conservative.	Human data on sensitive populations are available and uncertainties are addressed.
Peer Review	Limited or no peer review; disregarded comments would significantly change risk value; no independent check	Adequate peer review. Most substantive comments addressed; disregarded comments would not significantly change value	High quality panel peer review with appropriate experts; all substantive comments addressed as per independent check
Toxicity Value Comparison	Relevant risk values show a greater than 10 fold difference.	Some relevant risk values agree within 3-fold of each other, and others disagree within 10-fold of each other	All relevant risk values agree within 3-fold of each other

Table 31. Confidence in the Toxicity Assessment

Element	Score	Basis
Database Completeness	Medium	- Several acute and chronic studies in multiple species - Two developmental studies in two species - Lacking a 2-generation reproductive study and additional chronic information
Systematic Review	High	- Systematic review conducted
Key Study Quality	Medium	- Acute study had confounding factors (smoking, varying chamber concentrations) - Chronic study lacked a NOAEL and detailed histopathology information
Critical effect	Medium	- Acute and chronic critical effects were mild - Both lacked NOAEL information
Relevance of Critical Effect	Medium	- Acute critical effect based on human study - Chronic critical effect is possibly relevant to humans
Point of Departure (POD)	Low	- Only free-standing LOAELs available - Few dose groups, BMD modeling not possible
Human Equivalent POD (POD_{HEC})	Medium	- Default adjustments used, considered conservative
Sensitive Populations	Medium	- No information on sensitive subpopulations - Default UF_H of 10 used and considered protective
Peer Review	-	- DSD will be proposed for public comment
Toxicity Value Comparison	-	- No other agencies have derived relevant inhalation toxicity factors

Confidence Scoring Summary			
Not Evaluated	**Low Confidence**	**Medium Confidence**	**High Confidence**
Peer Review Toxicity Value Comparison	Point of Departure	Database Completeness Key Study Quality Critical Effect Relevance of Critical Effect Human Equivalent POD Sensitive Populations	Systematic Review

* Criteria for scoring the individual elements adapted from Beck et al. (2016).