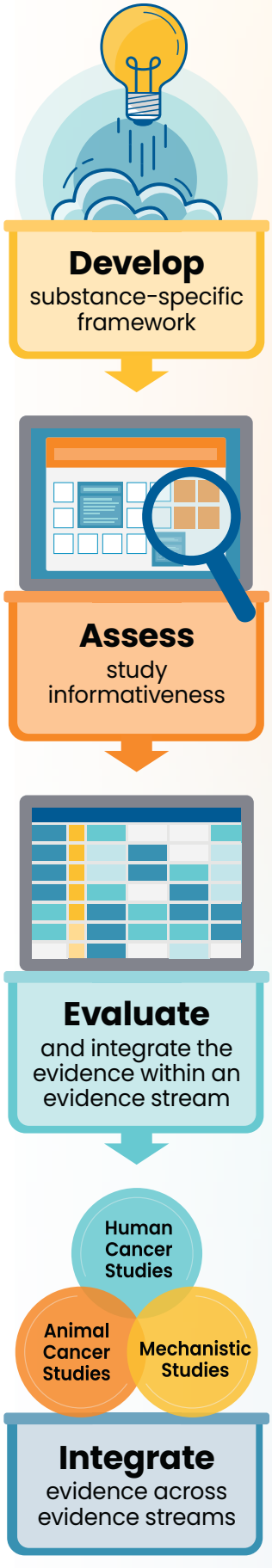
NTP | National Toxicology Program
U.S. Department of Health and Human Services



NTP Report on Carcinogens Handbook on Methods for Conducting Cancer Hazard Evaluations

June 2025

NTP
National Toxicology Program
U.S. Department of Health and Human Services

NTP Report on Carcinogens Handbook on Methods for Conducting Cancer Hazard Evaluations

June 2025

National Toxicology Program
Public Health Service
U.S. Department of Health and Human Services

Research Triangle Park, North Carolina, NC USA

Foreword

The National Toxicology Program (NTP), established in 1978, is an interagency program within the Public Health Service of the Department of Health and Human Services. Its activities are executed through a partnership of the National Institute for Occupational Safety and Health of the Centers for Disease Control and Prevention, the National Center for Toxicological Research of the Food and Drug Administration, and the National Institute of Environmental Health Sciences of the National Institutes of Health where the program is administratively located. The NTP offers unique opportunity for addressing contemporary toxicological issues through research conduct, collaboration, or coordination. In partnership with others, the NTP works to develop and apply new or improved methods and approaches that will advance toxicology and better assess health effects from environmental exposures and to generate knowledge that will strengthen the science base and inform decisions by health regulatory and research agencies to safeguard the public health.

The NTP identifies environmental substances that pose a cancer risk to humans through the conduct of literature-based cancer hazard evaluations to inform their potential listing in the Report on Carcinogens (RoC), a congressionally mandated document. This handbook presents guidance for those assessments. The Report on Carcinogens and the RoC handbook are available free of charge on the NTP website and cataloged in PubMed, a free resource developed and maintained by the National Library of Medicine (part of the National Institutes of Health).

Preface

The NTP has responsibility for preparing the congressionally mandated Report on Carcinogens (RoC). This cumulative report lists and discusses chemical, physical, and biological agents, mixtures, or exposure scenarios (collectively referred to as "substances") that are *known* or *reasonably anticipated to be human carcinogens*. A substance profile and the listing category are provided for each listed substance, along with a concise discussion of the scientific evidence supporting the listing, information on relevant human exposures, and existing regulations and guidelines to decrease exposure to the substance.

This handbook (RoC Handbook 2025) is the guidance used for preparing literature-based cancer hazard assessments on substances under consideration for listing in the RoC. Those assessments, which are published in NTP RoC monographs, identify whether a substance is a potential cancer hazard for humans. They do not estimate the actual cancer risk to individuals associated with exposure to a substance in their daily lives, which would depend on many additional factors such as the amount and duration of exposure and each individual's susceptibility. Importantly, identifying cancer hazards is a first step in cancer prevention.

The conduct of high-quality cancer hazard assessments on selected substances for RoC listing consideration requires the use of scientifically rigorous evaluation methods. Those methods evolve and improve over time, hence the need to update the RoC Handbook. A transparent, multi-step process is used for the review and evaluation of selected substances that applies established criteria to determine whether a substance should be listed in the RoC. As described in this updated handbook, the current four-step process includes: (1) identifying a substance to review, (2) conducting a cancer hazard assessment on the substance and applying the RoC listing criteria to the scientific evidence, (3) conducting external peer review of the draft assessment presented in an NTP RoC monograph, and (4) finalizing the assessment and publishing the monograph.

The RoC listing criteria date back to 1996 and were developed by a multi-step process with opportunities for public comment and scientific input. Recognizing the need to incorporate emerging best practices for more structured and transparent assessments, NTP developed and applied systematic review methods to the evaluation of scientific evidence from published animal and human cancer studies. These systematic methods, including guidance for identifying, selecting, and critically assessing the evidence from published cancer studies, were published in the RoC Handbook 2015.

Since release of that handbook, advances in systematic review methods and other tools have prompted the need for updated guidance. The updated guidance in this current handbook (RoC Handbook 2025) builds on those robust methods and best practices for conducting literature-based cancer hazard assessments and provides an updated suite of tools, approaches, and resources to assure scientific rigor, transparency, and confidence in the conclusions of those evaluations. The most significant updates are the development of systematic approaches for assessing the evidence from mechanistic studies followed by enhanced clarity to approaches for integrating evidence across data streams (human cancer epidemiology data, experimental animal cancer data, and cancer mechanistic data) to reach cancer hazard conclusions. Other updates to the handbook include methods to create systematic evidence maps for exploring a collection of studies, improved tools for assessing the informativeness of human cancer and experimental animal studies, greater emphasis on evaluating the impact of bias on findings from individual

studies and across studies, and enhanced reporting methods to increase transparency of the assessments.

Importantly, the guidance and approaches detailed in this handbook can also serve as a resource for the scientific community. It is anticipated that this handbook will continue to be updated in the future as systematic review tools and evidence integration approaches evolve and new tools applicable to literature-based cancer hazard assessments become available.

Table of Contents

Tables

Figures

About This Handbook

National Toxicology Program[1]

[1]Division of Translational Toxicology, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

Collaborators

Ruth M. Lunn, Whitney D. Arroyave, Stanley T. Atwood, William Bisson, Andrew D. Ewens, Sanford C. Garner, Neela Guha, M. Elizabeth Hodgson, Gloria D. Jahnke, M. Elizabeth Marder, Alton F. Peters, Pamela J. Schwingl, Amy Wang, Suril S. Mehta

Division of Translational Toxicology, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA
Responsible for oversight and coordination, developing the overall framework, drafting, and critically reviewing the handbook
Ruth M. Lunn, Dr.P.H., Project Leader

Contributed to drafting and/or critically reviewing the handbook
Gloria D. Jahnke, D.V.M. (Retired)
Suril S. Mehta, Dr.P.H.
Amy Wang, Ph.D.

Integrated Laboratory Systems, LLC, an Inotiv Company, Research Triangle Park, North Carolina, USA
Contributed to the coordinating activities, drafting, and critically reviewing the handbook
Whitney D. Arroyave, Ph.D., Project Manager

Contributed to drafting or critically reviewing the handbook
Stanley T. Atwood, M.S. (Retired)
William Bisson, Ph.D.
Andrew D. Ewens, Ph.D.
Sanford C. Garner, Ph.D. (Retired)
M. Elizabeth Hodgson, Ph.D.
Alton F. Peters, M.S.
Pamela J. Schwingl, Ph.D. (Retired)

Office of Environmental Health Hazard Assessment, California Environmental Protection Agency, Sacramento, California, USA
Contributed to drafting and reviewing the human cancer section
Neela Guha, Ph.D.
M. Elizabeth Marder, Ph.D.

Contributors

National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA
Internal peer reviewers: provided critical internal scientific review of specific sections of the draft handbook
Brandy Beverly, Ph.D.
John Bucher, Ph.D. (retired)
Stephen Ferguson, Ph.D.
William Gwinn, Ph.D.
Kamel Mansouri, Ph.D.
B. Alex Merrick, Ph.D. (Retired)
Katie O'Brien, Ph.D.
Kristen Ryan, Ph.D.
Stephanie Smith-Roe, Ph.D.
Kyla Taylor, Ph.D.
Vickie Walker, B.S.

Provided oversight of external peer review
Milene L. Brownlow, Ph.D.
Mary S. Wolfe, Ph.D.
Provided contract oversight
Kelly A. Shipkowski, Ph.D.

U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, USA
Internal peer reviewers: provided critical internal scientific review of specific sections of the draft handbook
Catherine Gibbons, Ph.D.
Lucina Lizarraga, Ph.D.
Esra Mutlu, Ph.D.

Integrated Laboratory Systems, LLC, an Inotiv Company, Research Triangle Park, North Carolina, USA
Provided technical input or reviewed specific subsections or tables
Alexandre Borrel, Ph.D.
Danila Cuomo, Ph.D.
Mona Sethi, Ph.D.

Provided administrative assistance and document preparation
Ella Darden, B.S.
Tracy L. Saunders, B.S.

Provided support for literature identification
Rachel Kalsch, M.L.I.S.

ICF, Reston, Virginia, USA
Provided contract oversight
David F. Burch, M.E.M., Principal Investigator

Jessica A. Wignall, M.S.P.H.

Provided editorial assistance
Kevin O'Donovan, B.A.

Supported external peer review
Leah W. Hennelly, B.S.
Aishwarya Javali, M.S.

Independent Consultant
Provided editorial assistance
Susan Dakin, Ph.D.

Peer Review

The National Toxicology Program (NTP) conducted an external peer review of the draft *Handbook for Preparing Report on Carcinogens Monographs, 2nd Edition* (retitled following peer review) by letter in January 2024 by the experts listed below. Reviewer selection and document review followed established NTP practices. The reviewers were charged to:

(1) Peer review the draft handbook.

(2) Comment on the completeness of each section.

NTP carefully considered the reviewers' comments in finalizing this report**.**

Peer Reviewers

Laura Beane-Freeman, Ph.D.
Senior Investigator, Division of Cancer Epidemiology and Genetics
National Cancer Institute
Bethesda, Maryland, USA

Lisa Bero, Ph.D.
Professor, School of Medicine
Chief Scientist, Center for Bioethics and Humanities
University of Colorado CU Anschutz Medical Center
Denver, Colorado, USA

Terry Gordon, Ph.D.
Professor, Department of Environmental Medicine
New York University School of Medicine
New York City, New York, USA

Thomas Hartung, Ph.D.
Professor, Department of Environmental Health and Engineering
Johns Hopkins Bloomberg School of Public Health
Baltimore, Maryland, USA

David Kriebel, Ph.D.
Director, Lowell Center for Sustainable Production
Professor Emeritus, Department of Public Health
University of Massachusetts Lowell
Lowell, Massachusetts, USA

Lisa Peterson, Ph.D.
Professor, Division of Environmental Health Sciences
Co-Program Leader, Masonic Cancer Center
University of Minnesota
Minneapolis, Minnesota, USA

David Philips, Ph.D.
Professor, Department of Analytical, Environmental and Forensic Sciences
King's College
London, UK

Leslie Stayner, Ph.D.
Professor Emeritus
University of Illinois at Chicago
Chicago, Illinois, USA

Emily Watkins, Ph.D.
Senior Lecturer, Environmental and Exercise Physiology
University of Roehampton
London, UK

Publication Details

Publisher: National Toxicology Program

Publishing Location: Research Triangle Park, NC

DOI: https://doi.org/10.22427/NTP-OTHER-1008

Official citation: National Toxicology Program (NTP). 2025. NTP report on carcinogens handbook on methods for conducting cancer hazard evaluations. Research Triangle Park, NC: National Toxicology Program.

Acknowledgments

This work was supported by the Intramural Research Program (ES103380) at the National Institute of Environmental Health Sciences, National Institutes of Health and performed for the National Toxicology Program, Public Health Service, U.S. Department of Health and Human Services under contract GS00Q14OADU417 (Order No. HHSN273201600015U).

Introduction

Overview

The NTP Report on Carcinogens Handbook on Methods for Conducting Cancer Hazard Evaluations (herein "RoC Handbook 2025") is an update to the handbook published in 2015 (NTP 2015). This updated handbook details systematic review and evidence-integration methods for conducting the cancer hazard evaluation of an agent, substance, mixture, or exposure circumstance (collectively referred to as "substance") under consideration for listing in the Report on Carcinogens (RoC). Completed evaluations are published as RoC monographs.

The cancer hazard evaluation of most substances will follow the methods described in this handbook. However, in a few instances, the NTP may instead implement a streamlined approach that directly applies the RoC listing criteria to information from peer reviewed cancer hazard assessments conducted by authoritative groups such as the International Agency for Research on Cancer or to the outcomes of peer reviewed NTP toxicology and carcinogenesis studies.

Table 1 lists the sections in the handbook and the extent of revision (systematic or narrative) to each section relative to the RoC Handbook 2015. Changes to the systematic review and evidence integration methods in the RoC Handbook 2025 range from minor revisions (e.g., Section 4. Evaluation of Cancer Studies in Experimental Animals) to more substantial revisions (e.g., Section 6. Evaluation of Mechanistic Information). Noteworthy additions are: (1) systematic review methods for assessing scientific evidence from mechanistic studies and (2) approaches for providing greater clarity on evidence integration across data streams (human cancer epidemiology data, experimental animal cancer data, and cancer mechanistic data) to reach cancer hazard conclusions. The development of new systematic review approaches for assessing carcinogenicity mechanistic studies (Section 6. Evaluation of Mechanistic Information) responds to the need for well-developed methods given the increasing role of mechanistic data in cancer hazard identification (Lunn et al. 2022). Other updates to the handbook include methods to create systematic evidence maps for exploring a collection of studies, improved tools for assessing the informativeness of human cancer and experimental animal studies, greater emphasis on evaluating the impact of bias on findings from individual studies and across studies, and enhanced reporting methods to increase transparency of the assessments.

Lessons learned from applying the systematic review and evidence integration tools and approaches in this handbook and scientific advancements (such as New Approach Methodologies, machine learning, see Section 6.7, New Directions) will inform future updates on methods for conducting cancer hazard evaluations.

Table 1. Overview of the Handbook Sections

Section	Description[a]
Introduction	Overview of the revised handbook
1. Initial Planning Phase	Scoping and problem formulation activities are included in each section Text from RoC Handbook 2015 rewritten Additional information included in Appendix B and Appendix C
2. Evaluation of Human Exposure Data	Not a systematic review or a quantitative exposure assessment; high-level guidance Text from RoC Handbook 2015 rewritten Additional information included in Appendix B
3. Evaluation of Human Cancer Epidemiological Studies	Systematic review methods applied in this section Moderate to major revisions to text in RoC Handbook 2015
4. Evaluation of Cancer Studies in Experimental Animals	Systematic review Minor to moderate revisions to text in RoC Handbook 2015
5. Evaluation of Disposition and Toxicokinetic Data	Not a systematic review; structured guidance on approach Text from RoC Handbook 2015 rewritten Information about ADME also included in Sections 6 and 7
6. Evaluation of Mechanistic Information	Systematic review methods applied in this section Text from RoC Handbook 2015 rewritten Additional information included in Appendix D
7. Evidence Integration	Text from RoC Handbook 2015 rewritten
Appendix A: Glossary	General systematic review terms and definitions of terms used in all sections
Appendix B: General and Exposure-specific Authoritative Sources	Provides information applicable to all sections
Appendix C: Systematic Review-related Tools Used by the Report on Carcinogens Monographs	Provides information applicable to all sections
Appendix D: Background Information on Key Characteristics of Carcinogens Biomarkers and Indicators	Provides information applicable to Section 6

[a]Description provides information about revisions to the sections relative to RoC Handbook 2015, the relationship of a specific section to other sections, and the type of review.

Identifying Carcinogens: Report on Carcinogens

The Report on Carcinogens is a congressionally mandated (see below) science-based document that identifies potential cancer hazards for people living in the United States (Lunn et al. 2022). Substances are listed in two categories: *known to be a human carcinogen* and *reasonably anticipated to be a human carcinogen*. The National Toxicology Program (NTP) prepares the report for the secretary of the Department of Health and Human Services (HHS) using a four-part process and established criteria (NTP 2023b). For each listed substance, the RoC includes a substance profile with information from cancer studies that is key to the listing, as well as information about use and production, potential sources of exposure, and current federal regulations to limit exposure. Each edition of the RoC is cumulative and consists of substances newly reviewed in addition to those listed in previous editions. The RoC can be used by policymakers and the public to make informed decisions about human health. A substance not meeting the RoC listing criteria is not listed in the RoC and its review is captured in an RoC appendix.

Box 1. Congressional Mandate

Section 301(b)(4) of the Public Health Service Act, as amended, requires that the Secretary, HHS, publish an annual report that contains a list of all substances (NTP 2023a)

- Which either are known to be human carcinogens or may reasonably be anticipated to be human carcinogens, and
- To which a significant number of persons residing in the United States are exposed.

Evaluating Cancer Hazards: Report on Carcinogens Monographs

A systematic evaluation of relevant cancer studies is conducted by the RoC evaluation team to determine whether a substance should be listed in the NTP's Report on Carcinogens. The RoC monograph captures the cancer hazard evaluation, which discusses, evaluates, and integrates evidence on:

- Human exposure
- Disposition and toxicokinetics
- Human epidemiological cancer studies
- Experimental animal carcinogenicity studies
- Mechanisms of carcinogenicity

The cancer hazard evaluation process (see Figure 1) starts with a planning phase, which includes scoping and problem formulation activities to identify the substance for review and develop the framework and methods to evaluate the evidence (see Section 1). Next, the studies are assessed for informativeness (bias assessment and study sensitivity, which is the study's ability to inform the cancer hazard evaluation), and the evidence is integrated across studies to reach a level-of-evidence conclusion (see Sections 2 to 6 for evidence-specific methods). The last step integrates the evidence across evidence streams (Section 7) to determine whether it fulfills the RoC listing criteria. Monographs are peer reviewed by substance-specific and discipline (e.g., different evidence streams) experts.

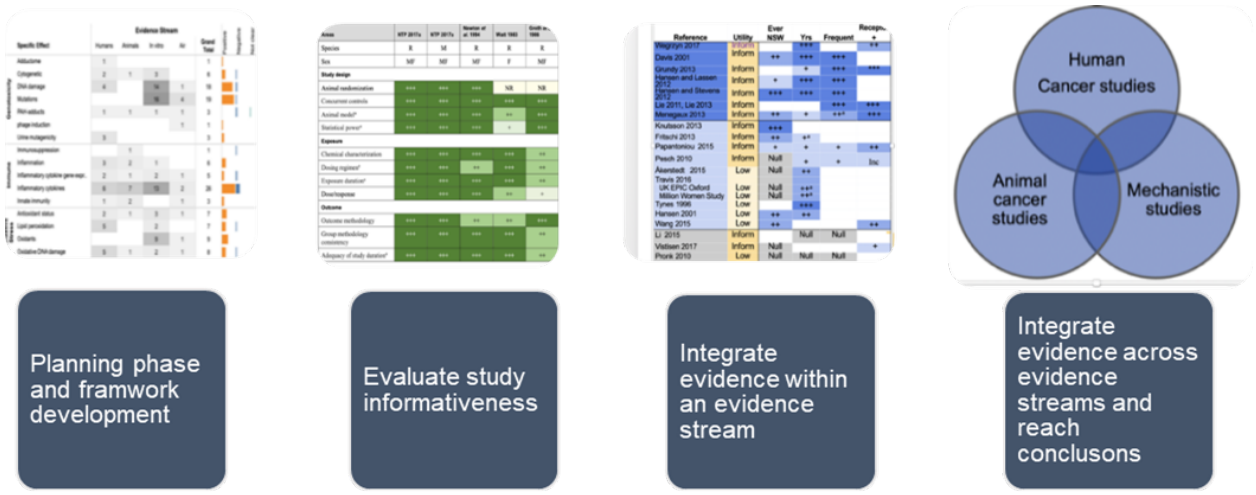


Figure 1. Cancer Hazard Evaluation Process

The schematic shows the major steps in the cancer hazard evaluation review process.

1. Initial Planning Phase

This section primarily describes scoping and problem formulation related to identifying a substance for review and general scoping methods that are relevant across all evidence streams (e.g., human cancer, animal cancer, and mechanistic studies). Scoping methods specific for an evidence stream are described in the relevant handbook sections (especially for evidence evaluated by systematic review methods, see Introduction). The planning process is iterative and includes scoping activities to identify the nominations and develop its review. A glossary of general systematic review terms as well as terms for all sections of the handbook is available in Appendix A.

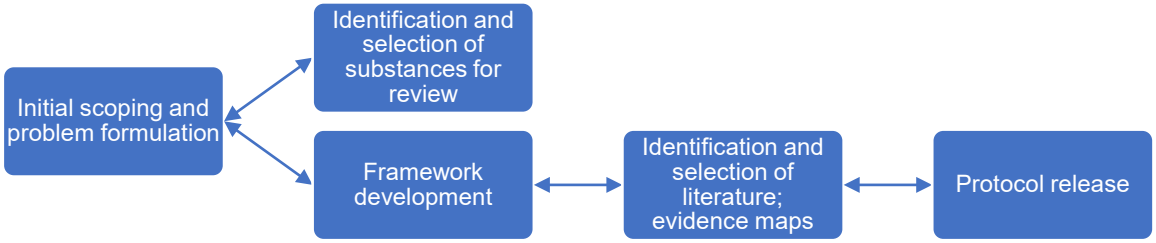


Figure 1-1. Initial Planning and Scoping of a Cancer Hazard Assessment

This process is iterative, starting with scoping and problem formulation to identify the substances for review (Section 1.1) and development the framework (Section 1.2). The framework is captured in the substance-specific protocol for the evaluation and is informed by the scoping activities, including identification, selection, and mapping of the literature. The protocol is made publicly available after peer review.

1.1. Identify and Select Substances for Review

Substances are selected for review according to the Report on Carcinogens (RoC) process for preparing the RoC using a variety of sources (e.g., public or government nominations, authoritative evaluations, and searches of peer-reviewed literature, government, and public health-related websites) to identify substances for review for the RoC (NTP 2023b).

We determine the scope of the literature review to determine whether there is a potential public health concern to warrant a cancer hazard evaluation, e.g., evidence of U.S. exposure and an adequate database (e.g., number of studies) for evaluation. The multistep process begins with searching authoritative sources (starting with EPA CompTox) and websites (such as Agency for Toxic Substances and Disease Registry, National Institute of Occupational Safety and Health, and EPA Integrated Risk Information System, see Appendix B) for exposure and carcinogenicity information and then developing more specific literature searches of citation databases. These initial searches inform the development of the research question (exposure and cancer outcome) and the framework (e.g., included literature) for answering the question. Systematic review methodology defines the framework for studies included in the review using the terminology "PECO" (population, exposure, comparison group, and outcome). Because our cancer hazard evaluations include multiple types of evidence streams, (animal models, human populations, and cells), we have modified the PECO to EECO and replaced "population" with "evidence stream"

for the overall monograph; we retained the term PECO for the human epidemiology studies and use MECO for animal studies in which M stands for model.

1.2. Develop the Framework

This section describes the first step in framework development, (e.g., identifying and mapping the literature). Next, the evaluation team reviews the literature and characterizes scientific and substance-specific issues for conducting the assessment, including methods for evaluating study informativeness and integrating the evidence within and across evidence streams (see the relevant sections). The monograph protocol captures the frameworks for each evidence stream.

1.2.1. Identify the Literature

The initial EECO informs the literature search strategy and associated inclusion/exclusion criteria. General approaches for identifying the literature are listed in Box 1-1 (see Sections 2 to 6 for guidelines for specific types of evidence).

In consultation with an information specialist, we develop a literature search strategy based on the EECO and typically search three citation databases (PubMed, Web of Science, Scopus); however, the databases and dates searched may vary on a case-by-case basis, depending on the nature of the substance and the topic.

Box 1-1. Types of Informational Searches

Database searches: The major source of identifying relevant peer-review publication papers on the relevant topics (see text for more details).

General sources: Examples include authoritative reviews, government reports, and web-based databases (see Appendix B).

Focused searches for specific scientific issues: Typically, issues that are identified at the beginning and during the literature-based review.

Secondary citations: Citations identified from the literature cited in authoritative review or primary references located by the literature search.

Updated searches: Literature searches are updated by either saving the search strategy and rerunning them or by creating monthly alerts in the appropriate databases (e.g., PubMed, Scopus, Web of Science).

Literature searches of the databases generally combine search terms for the substance with search terms for the outcome (e.g., cancer or biological effects) and, for human and animal cancer studies, search terms for the relevant evidence stream (e.g., epidemiology terms) (a list of search strings is available on the RoC web page). For chemical substances, search terms usually include the substance, its synonyms, trade name(s) when relevant, the metabolites of the substance, and the chemical class to which the substance belongs. Terms for exposure scenarios or settings are often used to identify human epidemiology studies.

Before screening the literature, we evaluate the adequacy of the literature search using "seed" studies (e.g., known relevant studies, often identified from reviews).

1.2.2. Select and Map the Literature

Citations retrieved from literature searches (and other sources) are uploaded to web-based systematic review software for multilevel screening using inclusion/exclusion criteria based on the initial EECO. Level 1 screening is based on title and abstract and Level 2 is based on full text. Level 1 screening is typically done by one reviewer, and a second reviewer may screen a certain percentage (e.g., 10%) of the excluded literature. Included studies are manually tagged

(but could be replaced by machine learning as those technologies evolved) or mapped according to evidence stream and other characteristics, such as cancer type, study design, biological effect, and publication type. These tags will vary by project, and in general, more detailed mapping occurs at Level 2 screening when the full PDF is available. Pilot screening tests are conducted to ensure that understanding of the inclusion/exclusion criteria by the reviewers. Level 2 screening and mapping is checked by a second reviewer and differences are discussed, and if needed a third reviewer is consulted.

The software and screening procedures depend on the size and scope of the retrieved literature. Screening and mapping may be conducted manually by staff or aided by machine learning.

Table 1-1 provides examples of commonly used tools for literature screening (see Appendix C for a list and description of systematic review tools). We will assess the utility of our tools and availability of new tools on a regular basis.

To assure the accuracy of the literature, we perform quarterly retraction checks for all citations for a given substance using Retraction Watch (https://retractionwatch.com/retraction-watch-database-user-guide). Additionally, we perform a retraction check prior to conducting study informativeness evaluations for human cancer, animal cancer, and mechanistic studies.

Table 1-1. Examples of Tools Used to Screen Literature

Tool[a]	URL	Description
Health Assessment Workspace Collaborative (HAWC)	https://hawcproject.org/	Content management system allows users to document the assessments including screening literature
Sciome Workbench for Interactive Computer-Facilitated Text-mining (SWIFT) Review	https://www.sciome.com/swift-review/	Uses machine learning to prioritize and sort literature
SWIFT Active screener	https://www.sciome.com/ufaq-category/introduction	Uses active learning to prioritize literature

[a]Tools used to screen literature may change or expand over time.

In some cases, limited data extraction of selected characteristics or findings may be conducted on subsets of studies to create more detailed evidence maps. These interactive maps can be visualized and explored using Tableau or similar software.

The evidence maps and visualizations of the literature can be useful to determine which elements of the literature have an adequate database (the final EECO) for the cancer hazard evaluations (e.g., human cancer types, exposure scenarios, specific mechanistic data). More detail is available in the specific evidence stream sections. In many cases, these visualizations will be made publicly available as part of the protocol or monograph for a substance.

1.2.3. Develop the Protocol

The monograph evaluation team develops a protocol that adapts the handbook methods to issues specific to the substance. It consists of multiple parts, one for each relevant evidence stream; the protocols for human, animal, and mechanistic studies incorporate systematic review methods and are made publicly available on the RoC website following peer review. Protocol sections that are

part of the systematic review may be written and reviewed separately. A protocol includes the following sections: (1) activities related to framework development including a summary of the scoping activities, e.g., literature search strategy, evidence maps, the initial and final EECOs, and scientific issues, (2) methods for evaluating study informativeness, and (3) methods for evidence integration. Scientific issues are substance-specific issues that are important for the assessment, including study informativeness and evidence integration, and can be identified by a variety of sources, such as reviews, primary articles, and experts. For example, a scientific issue for the night shift work evaluation was the timing that women started work. For all monograph sections, data extraction is checked for accuracy by a second reviewer. Study informativeness (bias assessment and sensitivity) for relevant evidence streams is independently completed by two reviewers using substance-specific guidance adapted from the handbook for each evidence stream. (See specific sections of the handbook for developing protocols for each evidence stream).

References

Lunn RM, Mehta SS, Jahnke GD, Wang A, Wolfe MS, Berridge BR. 2022. Cancer hazard evaluations for contemporary needs: Highlights from new National Toxicology Program evaluations and methodological advancements. J Natl Cancer Inst. 114(11):1441-1448. https://doi.org/10.1093/jnci/djac164

National Toxicology Program (NTP). 2023a. History of the Report on Carcinogens. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. [Last updated: March 7, 2023]. https://ntp.niehs.nih.gov/whatwestudy/assessments/cancer/history/index.html

National Toxicology Program (NTP). 2023b. Report on Carcinogens process & listing criteria. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. [Last updated: March 7, 2023]. https://ntp.niehs.nih.gov/go/rocprocess

2. Evaluation of Human Exposure Data

Summary

The aim of the human exposure section is to provide information on evaluating the extent of exposure and how people are exposed to a specific substance. This task is accomplished by: (1) establishing a human exposure data evaluation framework, (2) extracting and organizing the data, and (3) evaluating, integrating, and summarizing the evidence.

Establishing the evaluation framework (Section 2.1) includes defining and characterizing the substance to represent real-world human exposure, developing a literature search strategy and mapping studies by type of information and type of publication, and developing a short protocol. The human exposure protocol identifies any substance-specific exposure issues and provides transparency regarding the literature searches and approach for reaching decisions on key questions guiding the search for human exposure data.

Extracting and organizing the data (Section 2.2) involves taking relevant exposure information from sources identified from literature searches and arranging the information in tables based on the type of information (e.g., levels in the environment or biomonitoring data) and exposure sources. General study quality considerations are evaluated to provide context for understanding how the data affect answers to key questions.

Evaluating, integrating, and summarizing the evidence (Section 2.3) involves assessing extracted data to answer key questions. Data types are integrated to support conclusions about level and extent of human exposure.

Introduction and Objective

This section describes the methods used to identify, summarize, and interpret human exposure data for substances under evaluation for carcinogenicity for either Report on Carcinogens (RoC) monographs or other cancer hazard evaluation documents. The approach has been updated since the 2015 RoC handbook. The objective of this human exposure section is to provide information on evaluating extent of exposure and how people are exposed to a specific substance. It does not include a formal exposure assessment or evaluate health risks associated with various levels of exposure or provide a comprehensive review of all exposure information.

The specific objective of this section is to provide the information necessary to determine whether a significant number of people residing in the United States are exposed to the substance under review, as required by the congressional mandate (see Introduction). Past exposure can fulfill this criterion due to, in part, the long-latency period for many types of cancer. The congressional mandate does not provide guidance to interpret “significant” exposure, and information on numbers of exposed individuals is rarely available. However, significant exposure can be inferred from other types of information, such as use and production; occurrence in the environment, workplace, food, consumer, or medical products; and exposure levels in people.

Human exposure data typically are summarized using authoritative sources supplemented with contemporary reviews and more recent primary studies if they are needed to inform the

objectives and answer key questions. The amount of available exposure information varies for different substances and scientific judgment is required to determine when the level of evidence meets the criterion. Data from all geographical locations are included as non-U.S. studies can provide information on the extent of exposure and how people are exposed to a specific substance that is relevant to U.S. residents and globally.

Key Questions

The questions below guide the search for information summarized in the exposure section.

Primary Question

- Are a significant number of people living in the United States exposed to the candidate substance and, if so, what is the evidence supporting this conclusion?

Secondary Questions

- How should the substance be defined and characterized? Ideally, the definition should represent real-world human exposure.
- How and where are (were) people exposed to the substance (sources, settings, levels, frequency, trends)?
- What are the most important sources of exposure?
- What federal regulations and guidelines limit (or potentially limit) exposure?
- What are other ways to decrease or prevent exposure?

Components of the Evaluation of Human Exposure Data

- Develop the framework (Section 2.1).
  - Develop a literature search strategy and search, select, and map the literature.
  - Develop approach and protocol.
- Extract and organize the data (Section 2.2).
- Evaluate, integrate, and summarize the evidence (Section 2.3).

2.1. Evaluation Framework Development

Scoping and problem-formulation activities are conducted to identify and determine the key issues and relevant literature.

2.1.1. Substance Characterization

The candidate substance should be defined and characterized to represent real-world human exposure, which is especially important when evaluating complex exposure scenarios or mixtures. For example, pentachlorophenol was redefined as a mixture based on exposure data showing that people who are exposed to pentachlorophenol are also exposed to by-products of its synthesis. Another example is night shift work, which is defined as working during the biological night (e.g., 12 a.m. to 5 a.m.) and as a complex exposure scenario including exposure to light at night, sleep disruption, altered meal timing, and stress and behavior factors.

2.1.2. Literature Search Strategy and Evidence Mapping

Before selecting the substance for evaluation, the initial step is to estimate the extent of U.S. exposure to the substance using authoritative and online sources, starting with EPA's CompTox toxicological profiles, which include sources such as IRIS (Integrated Risk Information System) and ATSDR (Agency for Toxic Substances and Disease Registry), and supplementing with other sources (e.g., USGS Commodity Reports and NIOSH Health Hazard Evaluations) or journal review articles. A full list of the general sources can be found in Appendix B. Reviews and other exposure studies are typically retrieved from citation database searches for cancer and mechanistic data that are conducted in the scoping and problem-formulation activities. Targeted literature searches for exposure information may be conducted to address gaps.

Once the substance has been selected for review, the extent of the literature search is determined on a case-by-case basis, largely depending on the availability of data from the authoritative sources and information retrieved during the scoping and problem-formulation searches. Focused citation database searches using selected standard exposure search terms (see Table 2-1) and possible searches of grey literature may be used. This process is iterative and answers key questions, updates literature from authoritative sources or reviews, and usually focuses on the most informative type of exposure information, such as biomonitoring studies.

The literature search strategy, substance-specific search terms, and associated inclusion/exclusion criteria used to identify the relevant literature are constructed in consultation with an information specialist. Table 2-1 lists examples of general exposure-related search terms, including text words and Medical Subject Heading [MeSH] terms used in PubMed. These terms are used in conjunction (using the Boolean operator "AND") with the substance-specific terms or metabolites of the substances. Additional exposure or substance-related search terms may be added to fully capture all exposure scenarios.

Search terms for the substance may include chemical synonyms or exposure scenarios associated with exposure to the specific substance. The former is usually identified from EPA's CompTox Dashboard or National Library of Medicine databases (e.g., ChemIDplus, HSDB), and exposure scenarios are identified from secondary sources. Standard RoC exposure-related search strings can be found in the search string document on the RoC website.

Table 2-1. Examples of Concepts for Searches for Exposure Information

PubMed, Scopus, and Web of Science	MeSH Terms Used in PubMed
Occurrence	Environmental pollutants
Environmental pollution	Environmental pollution
Environmental exposure/monitoring	Occupational exposure
Occupational exposure/monitoring	Biomarkers
Biomonitoring	Prevention and control [subheading]
Indoor air or air pollution	Climate change
Prevention/intervention	

The literature search is followed by selection and mapping of studies by the type of information (e.g., uses or sources of exposure) and type of publication (e.g., review or primary research publication). Citations retrieved from literature searches are uploaded to a web-based systematic

review software such as the Health Assessment Workspace Collaborative (HAWC) and screened by a reviewer using predefined inclusion/exclusion criteria based on the research question. Exposure information should be specific for the substance. Data from authoritative sources or published articles are initially included in the review if they meet the inclusion criteria for providing relevant information (see Table 2-2 for more details):

- Information on production, export, import, and uses
- Occurrence data (such as levels in the environment or workplace, food, consumer, or medical products; toxics release data)
- Exposure data, such as intake and personal and biomonitoring data (e.g., urine, blood, or tissue data)

2.1.3. Framework and Protocol Development

Typically, a short protocol for evaluating human exposure data is included with the protocol for other parts of the evaluation (such as human cancer or mechanistic data). The protocol identifies any substance-specific exposure issues and provides transparency for the literature searches (see Section 1.2.2) and approach for reaching decisions on the key questions. This information may include data on uses; production; occurrence; and environmental, occupational, and consumer product exposures (see Table 2-2). In cases for which highly informative data (e.g., estimates of numbers of exposed people, modeled exposure levels from specific sources, biomonitoring data) are not available from the standard sources but are available from published articles (i.e., primary data), these data can be used to address primary and secondary questions. For example, a study reporting biological measurement data (such as urinary metabolite levels) or air concentrations for a group of workers employed by a company in an emerging industry or a company using a new manufacturing process could provide key occupational exposure information for the review.

2.2. Organization and Extraction of Exposure Data

Once studies are identified that meet the inclusion criteria for providing relevant exposure information, the data are extracted into tables and organized by the type of information and exposure sources. General quality considerations are evaluated; however, there is no formal risk-of-bias assessment.

2.2.1. Types of Information

In general, the types of data included in exposure sections of either RoC monographs or other cancer hazard evaluation documents include: (1) use and production-related information, (2) occurrence (e.g., levels in the environment), (3) measured or modeled human exposure (e.g., biomonitoring data such as urine levels), and (4) prevention related information (e.g., regulations and interventions) Information presented in the human exposure section can vary considerably depending on the substance. Table 2-2 lists general categories of exposure information, types of data in each category, and examples of literature sources for each type of data, which is organized by the type of data and source of exposure. Each source of exposure may include different types of data, e.g., environmental exposure may include releases and occurrence, as well as modeled exposure.

Table 2-2. Exposure-related Data

Data Category	Types of Data	Literature Sources
Production	Present and past production, import, export, or consumption	United States International Trade Commission DataWeb (USITC 2023)
	Trends in use or production over time	USEPA Chemical Data Reporting rule (USEPA 2023b)
Uses	Present and past uses	USEPA CompTox Chemicals Dashboard (USEPA 2023c)
	Identification of most widespread or important uses	Ullmann's Encyclopedia of Industrial Chemistry (Ley 2002)
		Trade association publications
		Authoritative sources, reviews, and primary literature
Occurrence and Exposure Data		
Occupational Exposure	Number of exposed workers, types of industries, exposure levels (personal, ambient, and biomonitoring data), and exposure trends; modeling data if available	NIOSH Health Hazard Evaluations (NIOSH 2019)
		OSHA Chemical Exposure Health Data (OSHA 2023)
		Authoritative sources, reviews, and primary literature
Environmental exposure/ occurrence (group not individual)	Release data, occurrence levels in ambient and indoor air, water, soil, and dust, modeling data, if available, of numbers of exposed people	USEPA Toxics Release Inventory (USEPA 2023e)
		Authoritative sources, reviews, and primary literature
Consumer products including smoking	Percentage of products containing chemical, percentage of total volatile organic compounds (VOCs) emitted from product groups, emissions from various products (e.g., emissions from wood and wood-based materials used in furniture and building products), or occurrence (levels)	Chemical and Products Database (CPDat)[a] (USEPA 2023a)
		Consumer Product Information Database (CPID) (DeLima Associates 2023)
		Authoritative sources, reviews, and primary literature
Medical use	Prevalance and dose	Medical databases such as the Medical Expenditure Panel Survey (AHRQ 2024) and FDA Drug Databases, such as Drugs at FDA (FDA 2024a)
Food and drinking water	Occurrence (levels) and intake, exposure (modeled intake)	FDA Total Diet Study (FDA 2024b)
		USEPA Safe Drinking Water Act (USEPA 2023d) national compliance data
		Authoritative sources, reviews, reviews, and primary literature
Prevalence and transmission of biological agents	Exposure in different populations	CDC, NIAID (NIAID 2023)

Data Category	Types of Data	Literature Sources
		Authoritative sources, reviews, and primary literature
Personal and biological monitoring	Personal monitoring data Information related to interpreting biological indices used in exposure studies Data on levels and trends of the substance (or metabolite when relevant) in human tissues or samples measured in studies Modeled intake levels from various environmental, occupational, or other sources	National Health and Nutrition Examination Survey (NHANES) (CDC 2023) Agency for Toxic Substances and Disease Registry Toxicological Profiles (ATSDR 2023) Authoritative sources, reviews, and primary literature
Prevention and Intervention		
Prevention and intervention	Occupational: engineering controls, modified work practices, PPE Environmental: engineering controls and PPE Personal behaviors including consumer products, diet, etc.; behavior risk factor modification Biologicals: behavioral risk factor modification, blood supply screening, and vaccines and drug treatments	Substance-specific authoritative sources, reviews, and primary literature
Guidance and regulations	The congressional madidate requires providing information regarding federal regulations and guidelines to limit exposure	Websites for U.S. government agencies (e.g., USEPA, NIOSH) Websites for authoritative nongovernmental sources (e.g., ACGIH) (ACGIH 2023) The CFR (GPO 2023), using a keyword search across all 50 titles of the CFR simultaneously to identify regulations beyond those listed on government websites

ACGIH = American Conference of Governmental Industrial Hygienist; CDC = Centers for Disease Control and Prevention; CFR = Code of Federal Regulations; FDA = Food and Drug Administration; NHANES = National Health and Nutrition Examination Survey; NIAID = National Institute of Allergy and Infectious Disease; NIOSH = National Institute of Occupational Safety and Health; OSHA = The Occupational Safety and Health Administration; PPE = personal protective equipment; USEPA = U.S. Environmental Protection Agency; VOCs = volatile organic compounds.
[a]The Chemical and Products database (CPDat) is part of EPA's Computational Toxicology (CompTox) Dashboard.

2.2.2. Data Extraction Methods

Data are extracted in a systematic manner using web-based tools such as Table Builder (Shapiro et al. 2018) to create Word tables, manually created Word tables, or text. Examples of the type of exposure information extracted include sample matrix and method, number of measurements,

measurement duration, exposure level and range, chemical composition data for consumer products, and description of exposed group.

Quality assurance of data extraction and data entry into tables (such as those created by Table Builder) are accomplished by review of data entry by an independent reviewer, and discrepancies are resolved via follow-up discussion of the source data.

2.2.3. Study Informativeness Considerations

Exposure data are evaluated to identify any general limitations such as reliability of the analytical methods, the choice of biological media as they relate to the likely route of exposure, and the final or preliminary nature of a data set (e.g., a draft version of an emissions report). For example, these evaluations could be accomplished by review of analytical method development and demonstration documentation or case-by-case consideration of whether to include draft report data depending on how informative the data set is for answering primary and secondary questions. No formal risk-of-bias review is conducted for exposure studies; however, consideration of general strengths and limitations of exposure data and studies provides context for understanding how the data affect conclusions about key questions.

2.3. Evidence Summarization

The extracted data are reviewed and evaluated to answer the key questions (e.g., whether a significant number of people in the United States are exposed to the substance and the nature of exposure). All evidence streams are integrated to support the conclusions.

2.3.1. Informativeness of Data Types

The estimated number of exposed people directly addresses the congressional madidate. Other types of data provide information on the level of exposure, such as biomonitoring data—substances or metabolites measured in blood, tissue, or urine—especially for samples for large numbers of people (e.g., NHANES) or of workers, personal monitoring data, or modeled biological exposure estimates (e.g., intake from food). The next tiers of informative data relate to occurrence: measured levels in the environment (e.g., ambient air), in the workplace, or in consumer products, food, and medicines, etc., followed by estimates of releases to the environment (e.g., industrial releases to air from TRI) and data on prevalence of chemicals in consumer product compositions or market use. Ideally, data should be available for higher exposure scenarios (e.g., environmental measurements taken near a manufacturing plant versus general ambient air) or more directly related to how people are exposed (e.g., drinking water versus surface water). Production, import, and export levels are the least informative data. Integration across these types of data may compensate for limitations for specific types of data and all data are useful for determining how and where people are exposed. Table 2-3 lists examples of exposure data types in their general order of prioritization. Appendix B contains examples of sources of exposure information for the U.S. and other countries.

Significant U.S. exposure is not always defined only by the extent of exposure (numbers of exposed people) and can include consideration of the level of exposure. In some cases, available occupational exposure data might indicate that exposure levels are high but germane to few people in that setting. Similar considerations include sensitive subpopulations (e.g., infants,

elderly, immunologically compromised) and environmental justice concerns, resulting in potential disparities to disproportionally affected populations (e.g., race and ethnicity, gender, socioeconomic status) that may be more sensitive to exposure, have higher exposure levels, or both.

Because the overall objective is to determine whether the exposure information meets the requirement for listing a substance in the RoC (i.e., significant number of U.S. residents are exposed to the substance) generally, U.S. data are most informative; however, these data can be supplemented with data from other countries to provide context for answering specific key questions. For example, data from countries in Europe, Asia, or other region on exposure associated with manufacturing processes, industrial uses, or the general population's use of products containing the substance can also be useful if the processes and products are comparable with those in the United States.

Table 2-3. Examples of Exposure Data Types

Data Type/Category	Level or Extent of Exposure	Example
Estimated Number of Exposed People	Extent	Workers: Company-reported or government agency estimates of number of exposed workers
		General population: Census or other government records (USCB 2023), number of people living near sources of pollution, utility users (e.g., public water)
Exposure Models	Level	Estimated exposure from specific sources [e.g., intake from food from the FDA total diet study (FDA 2024b)] or from multiple sources [e.g., EPA's human exposure modeling databases (USEPA 2024)]
		Reporting of typical and worst-case occupational exposure estimates via use of analogous or surrogate data (e.g., extrapolation from data on exposure from production to estimate exposure in other exposure scenarios)
Biomonitoring	Level	Blood, tissue, or urine samples in a discrete group with a known source of exposure (e.g., workers or residents exposed to pollution from a chemical plant or spill) or the general population with unknown sources of exposure (e.g., NHANES)
Personal Monitoring	Level	Air concentrations measured by workers wearing badge or other samplers, or study participants in the general public wearing exposure monitoring bracelets
Occurrence	Level	Levels in the environment (outdoor, indoor air, water, soil), workplace, medicine or consumer products, food; the source of exposure may be known (e.g., workers, environmental spill, downstream from factories) or unknown (e.g., ambient air in cities)
Releases to the Environment	Extent and levels	EPA's TRI data (USEPA 2023e) (industrial manufacturing and processing emissions)

Data Type/Category	Level or Extent of Exposure	Example
Prevalence in Industrial, Medical, Consumer Products, Food	Extent	Chemical use data reported in EPA's CompTox Chemicals Dashboard (USEPA 2023c)
		Chemical composition data reported in the Consumer Product Information Database (DeLima Associates 2023)
		Trade association consumption pattern reports
Production or Import/Export Information for the Substance	Level	Import/export data reported in United States International Trade Commission DataWeb database (USITC 2023)

2.3.2. Conclusions

Exposure data are synthesized by concisely summarizing data from the different data types to answer the key questions and draw conclusions about whether a significant number of people in the United States are (or were) exposed; how and where people are exposed; what the major sources of exposure are; trends in sources, settings, levels, and frequency of exposure over time; and ways to decrease or prevent exposure.

The importance of the available exposure data can vary by substance and is outlined in the protocol. For example, the conclusions could be presented by the most important source of exposure on the basis of exposure levels (e.g., occupational exposure; general population exposure from air, water, and food; and environmental exposure); the most important exposure route (e.g., inhalation, ingestion, dermal); or most recently available updated data (e.g., updated general population emissions data).

References

Agency for Healthcare Research and Quality (AHRQ). 2024. Medical Expenditure Panel Survey (MEPS). Rockville, MD: Agency for Healthcare Research and Quality. https://meps.ahrq.gov/mepsweb/

Agency for Toxic Substances and Disease Registry (ATSDR). 2023. [Homepage]. Atlanta, GA: Centers for Disease Control and Prevention, Agency for Toxic Substances and Disease Registry. https://www.atsdr.cdc.gov/

American Conference of Governmental Industrial Hygienists (ACGIH). 2023. [Homepage]. Cincinnati, OH: ACGIH. https://www.acgih.org/

Centers for Disease Control and Prevention (CDC). 2023. NHANES: National Health and Nutrition Examination Survey. Hyattsville, MD: Centers for Disease Control and Prevention, National Center for Health Statistics. [Accessed: May 31, 2023]. https://www.cdc.gov/nchs/nhanes/index.htm

DeLima Associates. 2023. CPID: Consumer Product Information Database. McLean, VA: DeLima Associates. https://www.whatsinproducts.com/

Ley C. 2002. Ullmann's encyclopedia of industrial chemistry. 6th ed. Weinheim, Germany: Wiley-VCH. https://doi.org/10.1002/14356007

National Institute for Occupational Safety and Health (NIOSH). 2019. Health Hazard Evaluations (HHEs). Cincinnati, OH: Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health. [Last reviewed: November 30, 2019]. https://www.cdc.gov/niosh/hhe/default.html

National Institute of Allergy and Infectious Diseases (NIAID). 2023. [Homepage]. Bethesda, MD: National Institutes of Health, National Institute of Allergy and Infectious Diseases. https://www.niaid.nih.gov/

Occupational Safety and Health Administration (OSHA). 2023. Chemical Exposure Health Data. Washington, DC: U.S. Department of Labor, Occupational Safety and Health Administration. https://www.osha.gov/opengov/health-samples

Shapiro AJ, Antoni S, Guyton KZ, Lunn RM, Loomis D, Rusyn I, Jahnke GD, Schwingl PJ, Mehta SS, Addington J, et al. 2018. Software tools to facilitate systematic review used for cancer hazard identification. Environ Health Perspect. 126(10):104501. https://doi.org/10.1289/EHP4224

U.S. Census Bureau (USCB). 2023. [Homepage]. Suitland, MD: U.S. Department of Commerce, U.S. Census Bureau. https://www.census.gov/

U.S. Environmental Protection Agency (USEPA). 2023a. Chemical and Products Database (CPDat). Washington, DC: U.S. Environmental Protection Agency. [Last updated: March 6, 2023]. https://www.epa.gov/chemical-research/chemical-and-products-database-cpdat

U.S. Environmental Protection Agency (USEPA). 2023b. Chemical Data Reporting under the Toxic Substances Control Act. Washington, DC: U.S. Environmental Protection Agency. [Accessed: January 26, 2023]. https://www.epa.gov/chemical-data-reporting

U.S. Environmental Protection Agency (USEPA). 2023c. CompTox Chemicals Dashboard. Washington, DC: U.S. Environmental Protection Agency. [Last updated: November 29, 2023]. https://www.epa.gov/chemical-research/comptox-chemicals-dashboard

U.S. Environmental Protection Agency (USEPA). 2023d. Safe Drinking Water Act (SDWA). Washington, DC: U.S. Environmental Protection Agency. [Last updated: March 30, 2023]. https://www.epa.gov/sdwa

U.S. Environmental Protection Agency (USEPA). 2023e. Toxic Release Inventory (TRI) Program. Washington, DC: U.S. Environmental Protection Agency. [Last updated: July 24, 2023]. https://www.epa.gov/toxics-release-inventory-tri-program

U.S. Environmental Protection Agency (USEPA). 2024. Air: Fate, Exposure, and Risk Analysis (FERA): Human exposure modeling - Databases to support exposure modeling. Washington, DC: U.S. Environmental Protection Agency. [Accessed: May 9, 2024]. https://www.epa.gov/fera/human-exposure-modeling-databases-support-exposure-modeling

U.S. Food and Drug Administration (FDA). 2024a. About Drugs@FDA. Silver Spring, MD: U.S. Food and Drug Administration. https://www.fda.gov/drugs/drug-approvals-and-databases/about-drugsfda

U.S. Food and Drug Administration (FDA). 2024b. FDA Total Diet Study (TDS). Silver Spring, MD: U.S. Food and Drug Administration. https://www.fda.gov/food/science-research-food/fda-total-diet-study-tds

U.S. Government Publishing Office (GPO). 2023. Code of Federal Regulations (annual edition). Washington, DC: U.S. Government Publishing Office. https://www.govinfo.gov/app/collection/cfr

United States International Trade Commission (USITC). 2023. DataWeb. Washington, DC: United States International Trade Commission. [Accessed: May 27, 2023]. https://dataweb.usitc.gov/

3. Evaluation of Human Epidemiology Studies

Summary

The aim of this part of the cancer hazard evaluation is to determine the level of evidence for carcinogenicity (sufficient or limited) from human cancer epidemiology studies. Systematic review and evidence-integration methods are used to conduct the evaluation.

The first step is to develop a human health hazard framework for epidemiology studies (Section 3.1). This involves developing a **scoping review** (a structured literature search to determine the extent of the body of literature on a particular topic, as well as identify key issues and potential data gaps) and **evidence map** (a visual representation of literature scoping); creating and refining a Population, Exposure, Comparison group, and Outcome **(PECO) statement**, to define the exposure(s) and outcome(s) under review; and developing a **study protocol** (a detailed description of reasoning, methods, and considerations for evaluating study informativeness).

The second step (Section 3.2) is to evaluate each study (irrespective of its findings) and determine its **informativeness** (the study's ability to inform the cancer hazard evaluation, somewhat analogous to study utility). This involves a domain-based approach to assessing **biases** (an evaluation of internal validity that assesses the potential that a specific bias is present and, if so, the direction and magnitude of the bias and its impact on the study findings) and **study sensitivity** (the study's ability to detect a true effect) and making an overall judgment about the study's informativeness. It is important to note the direction and magnitude (or distortion) of a potential bias are considered in the judgment calls when evaluating the potential for bias for each study.

The third step is to integrate the evidence to reach health hazard conclusions for each cancer type based on the human epidemiology studies (Section 3.3). This includes an evaluation of confidence (e.g., moderate or strong, some, null, inconclusive) in the **study's findings** (evidence for or against an association between the exposure and the outcome). To aid this step, a meta-analysis may be prepared, if feasible. The evaluation of the level of confidence integrates the study findings with **bias impact analyses** (quantification or qualitative assessment of the overall impact of one or more biases on the identified association, informed by the bias analysis), followed by **evidence integration** (evaluating the evidence across studies) to determine whether a credible association exists between exposure and cancer and, if so, whether alternative explanations (chance, bias, or confounding) can be ruled out with reasonable confidence.

The Report on Carcinogens (RoC) uses several approaches to integrate evidence across all epidemiology studies for a particular cancer type. These approaches use **triangulation** (integration of data from different methods, designs, theoretical approaches, and unrelated sources of bias to see whether the evidence converges on one conclusion) methods and **guidelines for causality such as Bradford Hill** (e.g., the strength of the association, consistency across studies, evidence of an exposure-response gradient, and temporality of exposure). **Meta-analysis** (a statistical method for combining the results of several studies) may be used to explore heterogenicity and evaluate consistency when appropriate. Lastly, the **level of evidence** (sufficient, limited, or inadequate) for the carcinogenicity of the substance from studies in human populations is determined by applying the RoC criteria to the body of evidence. The assessment

is made for each cancer outcome, and the overall conclusion is based on the highest level of evidence.

Introduction

This section describes the methods for systematic review and integration of evidence from human epidemiology studies to assess the level of evidence for the carcinogenicity (primarily) of a substance, agent, mixture, or exposure circumstance (collectively called "substances"). The approach to the cancer hazard evaluation of human studies has been updated from the 2015 RoC Handbook and is informed by methods of systematic review developed by other organizations and groups, by standard epidemiological principles (e.g., IARC 2021; Rothman et al. 2008; Sterne et al. 2014; USEPA 2022), and with input from epidemiologists and other scientists developing systematic review procedures. Although this section focuses on cancer epidemiology studies, the study informativeness evaluation questions and guidelines also apply to cancer mechanism studies with human participants. Section 6 describes the methods for evaluating mechanistic endpoints.

The key scientific questions and major components in the cancer hazard evaluation are described below.

Key Questions for Cancer Hazard Evaluations

Primary Questions

- Is there a credible association between exposure to the substance and cancer (for specific cancer types or all cancers combined)?
- If so, can the association between exposure to the substance and cancer types be reasonably explained by chance, bias, or confounding?
- What are key scientific issues for evaluation of the studies?

Secondary Questions (Scoping Phase)

- Which epidemiologic studies should be included in the review?
- Which cancer types should be the focus of the review?
- What are the potential confounders (including co-exposures) important for assessing the association between the substance and specific cancer types?
- What are the key issues for evaluating the quality of the exposure assessment for the substance and cancer type under consideration, such as the methods applied, the time windows of exposure considered, the most relevant exposure metrics, and the timing of the assessment? Is there biological or mechanistic information to inform these metrics?
- What are the methodological strengths and limitations of these studies?

Process and Components of the Cancer Hazard Assessment

In general, the cancer hazard assessment for epidemiology studies consists of three major components, as shown in Figure 3-1.

Components of the Cancer Hazard Assessment of Epidemiology Studies for the RoC

- Develop the cancer hazard evaluation framework (Section 3.1):
  a. Develop and execute a literature search strategy to search, identify, and map the literature.
  b. Develop the protocol that captures the approach to conduct the assessment.
- Evaluate the informativeness of the epidemiology studies (Section 3.2).
- Evaluate and integrate the evidence to reach cancer hazard conclusions from the human studies (Section 3.3).

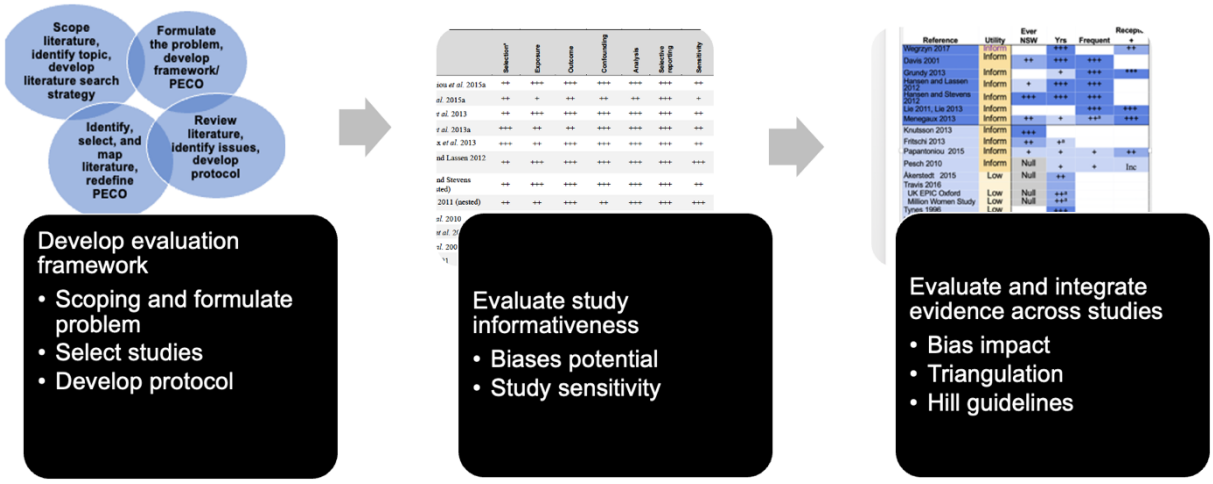


Figure 3-1. Components of the Assessment of Epidemiology Studies for the Cancer Hazard Evaluation

The first step is to develop a framework for the review. This involves scoping and mapping the literature, refining the criteria or the database for inclusion (e.g., the PECO statement), identifying key issues, and developing the methods. Next, through a structured approach using questions and guidelines, the included studies are evaluated for their ability to inform the cancer evaluation based on the bias assessment (internal validity) and study sensitivity (the ability to detect a true effect, see Section 3.2.2 and Table 3-9). The last step includes evaluating the individual studies for the level of confidence in their findings and the impact of bias (see Section 3.3.1) and integrating the evidence across studies using formal approaches and guidelines to reach a hazard conclusion judgment (see Section 3.3.2).

3.1. Cancer Hazard Framework Development

Conducting a cancer hazard assessment for a substance begins with scoping and problem formulation activities to develop the research question, the initial PECO framework, literature search strategies, and creation of a protocol for the cancer hazard evaluation of a specific substance (Figure 3-2). Section 1 discusses scoping and problem formulation for the overall cancer hazard evaluation (including selecting the substance and developing the initial PECO), whereas this section focuses on methods related specifically to the review of epidemiology studies (e.g., identifying, selecting, mapping the literature, redefining the PECO, identifying issues of concern, and developing the protocol). The scoping activities and systematic evidence maps are usually not reported as an independent review.

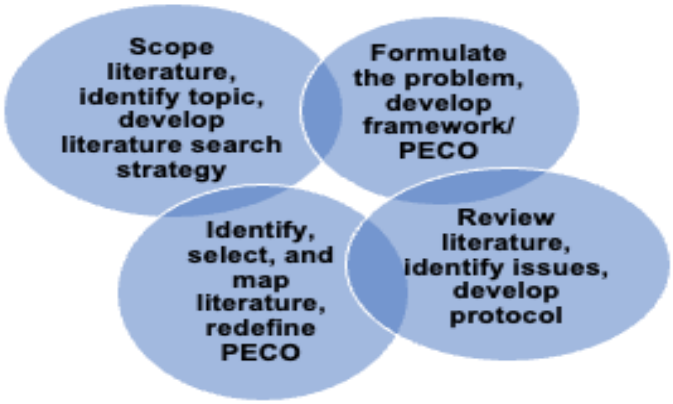


Figure 3-2. Developing the Cancer Hazard Evaluation Framework

Early steps in the process are to identify technical advisors with subject-specific expertise who can provide input throughout the systemic review, and to obtain scientific and public input via webinars, information groups, or other relevant mechanisms. Examples include past webinars on pentachlorophenol and trichloroethylene and a workshop on night shift work and light at night. The process is necessarily iterative (i.e., there may be several cycles of literature searches and evidence mapping).

3.1.1. Literature Search Strategy, Study Selection, and Systematic Evidence Mapping

The initial PECO statement (which was informed by scoping activities) indicates the first step is to develop and conduct a targeted literature search strategy (in consultation with an information specialist) and associated inclusion/exclusion criteria. The second step is to select and map the primary epidemiology studies from this database of identified studies. As per the RoC process, studies must be peer reviewed and publicly available. In addition to primary epidemiology studies, the review will include supporting literature (e.g., independent exposure assessment studies) that may be relevant for interpretation of the epidemiology studies. Toxicological and mechanistic data may inform the evaluation of human studies, e.g., biologically relevant exposure metrics, latency, tumor subtypes, and effect modifiers (see Sections 4, 5, and 6). Recent high-quality meta-analyses may also be included in the evaluation.

Searches are conducted in PubMed and at least one other database (such as Scopus or Web of Science) using search terms for the substance and exposure scenarios related to the substance combined (using the Boolean operator "AND") with search terms for epidemiology studies and with search terms for the outcome (cancer). The RoC staff maintains standard lists of search terms for epidemiological concepts and cancer types found on the RoC website. Table 3-1 lists some general search concepts used to identify epidemiology and cancer studies in most evaluations. Search terms may be modified (usually to a minor extent) on a case-by-case basis, depending on the specific substance. Section 6 describes search strategies specific to mechanistic studies in humans.

Table 3-1. Examples of Concepts Used in Searches for Cancer Epidemiology Studies in Humans

PubMed, Scopus, and Web of Science		MeSH Terms Used in PubMed	
Epidemiology Terms	**Cancer Terms**	**Epidemiology Terms**	**Cancer Terms**
Case-control Cohort Case-referent/cohort Case-report/series Cross-sectional[a] epidemiology Meta-analysis [publication type] Systematic review [publication type] Workers Workmen Ecological study Randomized controlled trial	Cancer Leukemia Lymphoma "lymphohematopoietic cancer"[b] "multiple myeloma" Neoplasm Tumor	Epidemiological studies Epidemiological methods Occupational exposure/ Adverse effects Epidemiology[subheading] Etiology[subheading]	Neoplasms

Note that these are examples of search terms, and not the detailed or fully developed search strings used in the actual literature search found in the search string document on the RoC website.
[a]May be useful for specific candidate substances and tumor types when a temporal sequence can be established (e.g., the presence of a virus before detection of cancer), and reverse causation can be ruled out with reasonable confidence.
[b]More specific search terms for lymphohematopoietic cancer may be developed for specific candidate substances.

Relevant literature may also be identified from sources such as authoritative reviews and citations from identified publications, and searches may also be conducted on specific topics. In addition, to supplement our standard searches and searches of authoritative reviews, we have created a custom library of PDFs of occupational case-control studies identified through searches of the three literature databases using terms for occupational exposure, case-control studies, and cancer. This library was created to identify studies that report on a large number of substances, which are not identified in the title or abstract, which is common in occupational case-control studies. Full-text searches of this library are conducted (using Adobe Acrobat full-text search tools) to identify cancer studies of substances or chemicals that may not have been identified through the database searches alone.

Citations retrieved from literature searches are uploaded to a web-based systematic review software application (such as Health Assessment Workspace Collaborative [HAWC], SWIFT Active Screener, and SWIFT Review) and screened by reviewers using predefined inclusion/exclusion criteria based on the initial PECO statement. For reviews with large databases, machine-learning software, such as SWIFT Active Screener, may be used. SWIFT Active Screener has been shown to have a high rate of recall (sensitivity) while minimizing the number of articles screener (Howard et al. 2020). In general, primary studies may be excluded if they (1) do not adequately evaluate exposure specifically to the substance, and (2) do not evaluate health effects related to carcinogenicity. Inclusion of studies such as case reports, case series, cross-sectional, or ecological studies is decided on a substance-by-substance basis and will be delineated in the protocol (see below).

For planning the hazard assessment, it may be useful to visualize the literature retrieved. Studies may be tagged by keyword in review software programs (e.g., HAWC). The evidence for included studies may be mapped according to cancer type, exposure source (e.g., occupational or environmental), exposure assessment method, study design, or other relevant issues. Systematic

evidence maps present a broad scope of the literature in a visual format and can also be visualized in an interactive format, such as on platforms created using Tableau software. During the problem formulation stage, these maps can be key for deciding which elements (e.g., study population characteristics, exposure metrics) to carry forward to the systematic review, how many studies are available for a specific cancer type and can be used to inform or refine the PECO statement and protocol. For example, evidence maps may help identify on which study characteristic to conduct stratified analyses.

Final selection of studies for inclusion in the systematic review is based on the final PECO statement. For example, the systematic review and hazard assessment may be restricted to cancer types with an adequate database that suggests a cancer hazard; findings for cancer types with sparse or inconsistent data may be briefly summarized. Note that there is no formal guidance on what constitutes an adequate number of studies for a cancer hazard evaluation. For example, a single study reviewed during the scoping process may be influential for reaching a hazard conclusion at the evaluation stage if it is a large multicenter study (i.e., replicated in different study populations), conducted in populations with high exposure levels and large exposure contrasts, adequately controlled for confounding, and presents multiple sensitivity analyses to demonstrate with reasonable confidence that any associations detected are unlikely to be due to chance, bias, or confounding.

There is also no formal guidance on which study design is considered the most informative for causal inference; this must be judged on a substance-by-substance basis (Arroyave et al. 2021; Steenland et al. 2020). For example, ecological studies with large exposure contrasts have been influential in evaluating the carcinogenicity of arsenic in drinking water (IARC 2004; 2012). Cross-sectional studies and case series/case reports are generally not informative for cancer hazard assessment and will be decided on a case-by-case basis. By design, cross-sectional studies assess exposure and outcome simultaneously; therefore, it is challenging to establish whether the estimated exposure precedes disease occurrence (e.g., a temporal association), particularly for studying short-lived biomarkers and long-latency diseases. Cross-sectional studies are carefully assessed for their accuracy in assigning individuals to exposure levels or categories (considering the potential for reverse causation or recall bias); and for whether prevalent disease is an appropriate proxy for disease incidence (Savitz and Wellenius 2023). Typically, cohort studies (and their variants) and case-control studies have provided the bulk of the evidence (IARC 2019).

3.1.2. Protocol Development

The protocol describes the systematic review methods for evaluating the human cancer studies and may be part of the monograph protocol and may be published on the RoC website as a separate document after peer review. It consists of the following sections: (1) developing the framework, which provides information on the objectives, identifying and mapping the evidence, the PECO statement, and substance-specific scientific issues, (2) detailed methods for evaluating study informativeness, such as the potential for exposure and outcome misclassification, confounding, and other potential biases that may be important in evaluating the findings for the hazard evaluation and (3) methods for evaluating and integrating the evidence across studies. Information on data extraction and the roles of the evaluation team may also be included. Developing the protocol requires understanding the types of studies and scientific issues that are available to inform the hazard assessment and requires background research on the substance, the

cancer type(s), and any co-exposures and their measurement, taking into consideration input from subject-matter and methodologic experts.

Identification of Potential Confounders, including Co-exposures

A key question for reaching a level-of-evidence conclusion from observational studies is whether any association between exposure and cancer is likely to be explained by confounding. Potential confounders are factors that are moderately to strongly associated with both exposure and the disease outcome(s) of interest. A confounder is not an intermediate in the disease pathway (Figure 3-3). For example, in cohort study designs, confounding occurs when the comparison groups under study (e.g., the exposed vs. the unexposed groups) have different background risks of disease (Pearce et al. 2007), thereby mixing the association of interest with the effects of other factors.

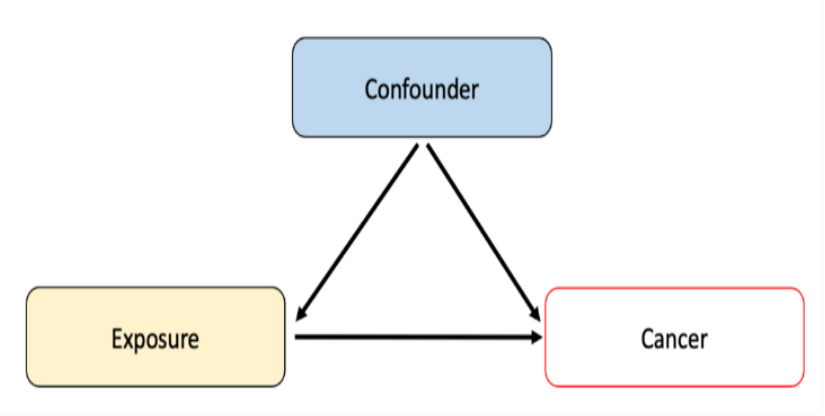


Figure 3-3. Diagram to Explain Confounding in the Exposure-Outcome Relationship

Adapted from Jager et al. (2008), this diagram visualizes how a factor can confound an exposure-outcome relationship of interest.

Potential confounders, including co-exposures, may be quantified or noted by the study authors or identified from authoritative sources, literature review, or consultation with experts during the planning phase. Authoritative sources of risk factors include substances listed in the RoC, such as the 15th RoC Dashboard (NTP 2023), the International Agency for Research on Cancer (IARC), the National Cancer Institute, the American Cancer Society, and the World Cancer Research Fund. Peer-reviewed and published evidence that a risk factor is also associated with the exposure of interest is critical for its being considered a potential confounder.

Co-exposures are environmental and/or occupational exposures that study participants may be simultaneously exposed to and are correlated with the exposure of interest. Whether a given co-exposure should be considered as a potential confounder depends on the availability of evidence that the co-exposure is potentially associated with a specific cancer(s) of concern. Occupational co-exposures, as assessed in the general population (e.g., population-based, or hospital-based case-control studies), may be of less concern. The broad diversity of jobs among study participants and the low prevalence of specific jobs and exposures (Cocco et al. 2013) reduces the sensitivity to detect an association. Visual diagrams (e.g., directed acyclic graphs) may be used in some cases to help identify potential confounders and to identify whether confounders were controlled for correctly in the analyses (e.g., control for factors on the causal pathway is inappropriate).

The estimated impact of potential confounding bias on study results may vary by confounder (Steenland et al. 2020). Importantly, reviewers should distinguish major or key confounders from confounders with minimal impact. Major or key confounders are factors that are expected to substantially impact the magnitude and/or direction of the study results. Note that a factor

identified as a confounder can also be an effect modifier (i.e., a factor that differentially impacts the magnitude of effect of an exposure-outcome relationship). Within the protocol, key confounders are identified apart from factors expected to weakly impact the magnitude or direction of the study results. Use of causal diagram tools (e.g., www.dagitty.net) may aid in visualizing and identifying key confounders across studies having similar populations.

Background Research on Exposure Metrics

Characterization of exposure in observational studies involves various tools and methods for assigning participants to exposure groups, levels, or categories. Several examples relevant to the evaluation of cancer epidemiology studies by exposure type are presented in Table 3-4. Relevant types of exposures can be broadly classified as follows:

- ***Occupational exposures***: Assessment is often based on job-exposure or job-task-exposure matrices or expert assessments that link the subject's occupational history (e.g., job or department titles, task descriptions and frequency, duration of employment) with workplace exposure data (e.g., monitoring data, production methods or applications, or protection procedures). Biological monitoring data and environmental monitoring data (e.g., levels of a chemical in facility air, personal monitors such as radiation badges) may also be used to assess occupational exposure.
- ***Environmental exposures***: Assessment is often based on measurements of the likely source of individual exposure (e.g., the concentration of an air pollutant) and/or biological monitoring data. In these studies, a range of relevant surrogate measures may be included and can be supplemented through the use of questionnaires to establish individual patterns of exposure (e.g., consumption of drinking water or duration of residence near a pollution source). Assessment can also be conducted at the individual level through the use of personal monitors or by measuring biomarkers.
- ***Personal behaviors and related exposures***: Assessment of other types of nonoccupational exposures (e.g., biological agents, consumer products, or personal behaviors such as smoking or diet) typically relies on a combination of one or more methods, including medical and clinical data or records, biological monitoring (e.g., for cotinine in urine), or participant questionnaires collected at the individual level.
- ***Social determinants of health***: Examples include race, gender, socioeconomic status (SES), healthcare access, and community-level factors such as community cohesion and built environment (e.g., transportation access, green space). Race and/or ethnicity are typically used as proxies for racism; biological or self-reported sex for gender or gender roles; and education and/or income for SES; however, these proxies may not capture the determinants of interest. Many individual-level factors are usually self-reported or obtained from records (e.g., medical, occupational); indicators of community-level factors may be obtained from a range of administrative and governmental databases, including census data, although an increasing number of tools are also available (see, for example, https://www.cdc.gov/socialdeterminants/data/index.htm)

The ability of epidemiology study findings to inform the hazard assessment depends on the type and quality of assessment used to characterize exposure, a detailed understanding of relevant methods used to collect and process exposure data, and the exposure metrics and timing of

assessment. A guide to identifying potential biases in the exposure characterization of a specific substance is provided in Table 3-4 and Table 3-5, found in Section 3.2.

In addition to the considerations above, when at all possible, selection of the most appropriate exposure metrics should be done with consideration of the underlying biologic mechanism. Concepts such as latency, susceptibility windows, the reversibility of the exposure or effect, and peak exposure are important to consider when selecting the exposure metric (Checkoway et al. 2019; Smith and Kriebel 2010). Although these considerations and an understanding of the underlying biology are important, in epidemiology studies; however, some or all of these factors may not be known.

Therefore, searches for information regarding the most appropriate exposure metrics and methods used to characterize exposure should be conducted in the context of both the substance under evaluation and the study design, along with an understanding of the underlying mechanisms when available. This includes method validation studies, as well as environmental scenarios related to exposure, consumer products and uses, production methods, anticipated levels of exposure to the substance, and interpretation of various exposure metrics, such as intensity, duration, or calendar years employed. For example, for studies assessing exposure using biological markers, it is important to know the specificity of the exposure biomarker of interest and how the magnitude, frequency, and timing of exposure are relevant to the etiologic time (Smith and Kriebel 2010) interval. Similarly, it is important to identify relevant time windows of exposure in relation to the cancer type(s) evaluated.

Background Research on the Outcome Assessment

Cancer Types

Prior to the outcome assessment, it is important to consider the methods used to obtain data on cancer incidence, vital status, and cause of death; the expected rates of cancer incidence, mortality, and survival for the cancer types of interest; and the implications of survival rates for interpreting mortality or incidence rates. Because the completeness and reliability of cancer registry incidence data can vary (e.g., by collection methods, country or region, and calendar period), relevant registries should be researched prior to study evaluation. For instance, the United States has no central national cancer registry; therefore, it can be difficult to obtain complete follow-up information for a cohort, especially for nationwide studies and for individuals who migrate from one area of the country to another. Currently, the U.S. Surveillance, Epidemiology, and End Results (SEER) (NCI 2022) database and other relevant data [e.g., Globocan (IARC 2022)] are used to understand expected rates of incidence, mortality, and survival for specific types of cancer in a given country or region. However, future RoC reviews may make use of an effort to create a pooled virtual registry that is currently underway. Diagnostic methods, criteria, and coding systems for cancer, such as the International Classification of Diseases cancer codes, change over time. These changes may have implications for particular cancer types and subtypes (especially some of the lymphohematopoietic cancers) and should also be researched prior to outcome assessment. Finally, the latency period between etiologically relevant periods of exposure and cancer diagnosis can differ among various types of cancer and patterns of exposure, and each should be researched prior to the evaluation. This will provide information pertinent to understanding a study's sensitivity to detect an effect.

Mechanistic Evidence from Human Molecular Epidemiology Studies

In general, mechanistic outcomes will be biological effects identified as the key characteristics of carcinogens (Smith et al. 2016). Studies in exposed humans that report outcomes relevant to the key characteristics of carcinogens are emphasized when available. Research on issues related to evaluating these outcomes will be part of the mechanistic section of the protocol, noting any issues that may be pertinent to the evaluation of human epidemiology studies.

Meta-analysis

If relevant and feasible, a meta-analysis may be conducted as a part of the cancer hazard evaluation. A meta-analysis may be indicated when there are a sufficient number of studies with similar exposure and outcome measures than can be combined. A meta-analysis may be useful as a quantitative means of exploring heterogeneity and can complement the qualitative cancer hazard assessment. If a meta-analysis is planned, the protocol will include the methods for conducting the meta-analysis, including the PECO statement (which may differ from the overall PECO for hazard evaluation), the statistical methods, and methods for exploring heterogeneity (including direction of bias, key issues, and effect modifiers, which may be informed by evidence mapping). More information on key issues to consider in a meta-analysis protocol can be found in Section 3.3.2, under the meta-analysis and meta-regression subsection.

In some circumstances, we may rely on published meta-analyses provided they are up to date, free of biases and well-conducted by authoritative research groups that are free of conflicts of interest. Meta-analyses will be evaluated individually and may not be included in the assessment if they are not deemed relevant.

Table 3-2. Summary of Exposure Data Collection Tools and Assessment Methods Commonly Used in Epidemiology Studies

Exposure Measurement		Definition	Occupational Exposures	Environmental Exposures	Personal Behaviors and Social Determinants of Health (SDH)
Data Collection Tools					
Self or Proxy-report	Interview (mail, web, in-person, or telephone) Questionnaire	Series of questions on specific exposure or circumstances that can be used to infer exposure	Specific chemicals or agents Job title, tasks, or history	Specific environmental contaminants Geographical location	Use of specific consumer products or medications, personal behaviors, and proxy for SDH Questionnaires are available for assessing some factors, such as perceived racism or discrimination
Records	Routine or population data	Data routinely collected for other purposes than the study of interest, such as a census	Usually used in population-based studies Information on workers (usually just job title) but not on workplace conditions	Aggregated data on substances in the environment (e.g., chemicals, pesticides)	Aggregated data from population surveys on personal behaviors (e.g., NHANES)
	Administrative or specific	Data collected for a specific purpose, such as employment, health, or school records; often used for specific cohorts, but also in population-based studies	Detailed information on specific workplace exposures and detailed job history for that job	Data collected to describe statewide, regional, or national land, water, or groundwater characteristics	Medical (or occupational) records, which in some cases may have information on demographics, medication or personal behaviors related to disease
Environmental Monitoring	Area or spatial	Involves the collection of one or more measurements used to assess the status of an environment	Air monitoring over a defined time period in a specific job or work areas in plants	Measurements of water, soil, or air in various environments over time and space and in the vicinity of study participants	NA

Exposure Measurement		Definition	Occupational Exposures	Environmental Exposures	Personal Behaviors and Social Determinants of Health (SDH)
	Personal	Devices designed to collect data on personal environmental exposures occurring very close but external to the body	Devices used to collect data on substances in the personal airspace of workers	Personal monitoring devices measuring individual exposures, such as silicone bracelets	Devices used to collect data about diet, medications, and other personal behaviors
Biomonitoring	Biological sampling	Examples include urine, serum, plasma, saliva, breast milk, hair, nails, cells, or other tissues collected for the purpose of measuring a specific substance or exposure	Biological samples (often urine or serum) collected from workers	Biological samples collected from individuals living in study areas or in controlled indoor experiments	Biological samples collected from individuals in a study
Methods Used to Assess Collected Data					
Expert Assessment	Job-exposure matrix (JEM)	Standardized method used to translate job information obtained from administrative records into specific exposures for characterizing individuals as exposed or nonexposed, allocating the same exposure estimate to all workers within a job category and time period	Population-specific JEMs or general population JEMs can be developed based on questionnaire data; usually conducted for assessing exposures in occupational studies	NA	NA
	Expert assessment	Conducted by industrial hygienists or occupational experts based on case-control questionnaire data and historical information about past exposures for various locations and time periods	Usually conducted for occupational studies to assess exposure on the job; can be based on administrative records or questionnaire data	NA	NA

Exposure Measurement		Definition	Occupational Exposures	Environmental Exposures	Personal Behaviors and Social Determinants of Health (SDH)
Modeling	Dosimetric model	A conceptual or mathematical representation of the exposure process; similar to using appropriate weights for each subject's exposure history, taking into account biological variability (Smith and Kriebel 2010) Outputs can be an exposure concentration or the amount of a chemical that is absorbed into the body, including a dose at a target tissue or organ	May include biomarkers, toxicokinetic, and physiological data to estimate individual dose	May incorporate elements of environmental and biologic fate, population activity patterns, and biomonitoring	May incorporate elements such as population activity patterns and biomonitoring

Exposure Measurement		Definition	Occupational Exposures	Environmental Exposures	Personal Behaviors and Social Determinants of Health (SDH)
–	Other exposure models	Other mathematical models, such as cumulative, aggregated, probabilistic, deterministic, or stochastic models, used to best represent the exposure process and grounded in one or more exposure measures and/or exposure-response relationships Used to more accurately reflect more complex exposure scenarios, when other exposure methods are not feasible or require further characterization	May incorporate elements of biomonitoring (e.g., biomarkers) May be used to model exposure metrics such as e.g., cumulative, peak, frequency, or average exposure	May incorporate elements of environmental and biologic fate, population activity patterns, and biomonitoring	May use algorithms and software to automate the calculation of exposures, such as estimated daily intakes of nutrients or food groups based on food frequency questionnaires

NA = not applicable or not usually used, though not excluded.

3.2. Study Informativeness Assessment

Each study will be evaluated for its ability to inform the cancer hazard evaluation (informativeness) through the use of a domain-based approach (Section 3.2.1) and structured guidelines for assessing biases (that is, an evaluation of internal validity that identifies the potential for biases especially the most influential biases) and study sensitivity (the ability to detect a true effect, see Section 3.2.2 and Table 3-9). Biases in observational studies often are classified into three major categories: (1) selection bias; (2) information bias; and (3) confounding (Rothman et al. 2008). Studies should have adequate reporting methods (von Elm et al. 2007) and apply appropriate analytical methods to calculate effect estimates. Finally, studies with greater sensitivity to detect an effect (i.e., having adequate numbers of exposed cases and sufficient exposure contrasts, durations, ranges, windows of exposure, and lengths of follow-up) are considered to be more informative for the evaluation, although studies with less sensitivity may not suffer from bias per se (Cooper et al. 2016). Domain judgments (bias and study sensitivity), overall judgments of study informativeness, are made for each study (Section 3.2.3).

3.2.1. Domain-based Approach

Box 3-1. Bias Concern Judgments

***No or minimal concern*:** The study characteristics being evaluated for the domain closely resemble the ideal study characteristics. The potential for bias is considered minimal, recognizing the general limitations of observational studies.

***Some concern*:** The study design or methodologies are less than ideal for this domain. However, although the potential for bias exists, these studies are generally considered informative for the cancer hazard evaluation.

***Moderate or major concern*:** The study design or methodologies suggest a high potential for a specific type of bias. Depending on the direction and distortion of the potential bias, the study may still be informative for cancer hazard evaluation but should be viewed with caution.

***Critical concern*:** The distortion resulting from bias likely makes the study findings unreliable for cancer hazard identification. This category is rare.

***No information*:** The information in the study is inadequate to evaluate the level of concern for the domain.

Direction of bias:

- ↑Away from the null, or overestimation of the effect.
- ↓Toward the null, or underestimation of the effect.
- Not known (unable to determine).

Magnitude of bias: Minimal, moderate or major, or unknown. In most cases, the magnitude will be unknown or subjective, but, where available, it will be based on the sensitivity analysis.

The potential for bias in each domain is captured by a core question for that domain. Core questions, together with a series of signaling and follow-up questions addressing specific issues related to the core question, are based on standard epidemiological principles (e.g., Rothman et al. 2008) and have been developed to reflect concerns that epidemiologists usually consider when evaluating studies. These concerns are also reflected in a peer-reviewed article presenting a "toolkit" that can be used to identify methodological concerns in epidemiology studies (e.g., validity, sensitivity, transparency) (Soskolne et al. 2021). The signaling questions are used to provide transparency in answering the core questions (i.e., the domain-level judgments), and a separate response is not required for each signaling question. The questions are intended to guide bias and sensitivity assessments; they are not meant to be a checklist. Although some of these concerns (such as the healthy worker effect) could be considered in more than one domain, they are

evaluated in only one domain in this handbook.

To determine the potential for bias within a study, each characteristic of the actual study is compared with that of an "ideal" observational study for the study design and a specific end point and exposure, as defined in the protocol (see Section 3.1.2). However, the potential for a given bias in a study does not necessarily mean that the findings of the study should be disregarded. When there is adequate information, a judgment is made on the direction of the potential bias (whether it over- or under-estimates the effect or its direction is unknown) and the potential magnitude of the distortion caused by the bias (see Box 3-1 for guidance for bias concern judgments). (The impact of the bias on the effect estimate is discussed in Section 3.3). The overall evaluation of study informativeness is derived by integrating the domain-level judgments. In some cases, especially for exposure assessment, a study-level judgment may not be possible because of the complexity of the issues, and the evaluation will be captured by narrative text; or the categories could be expanded. An example of this was the evaluation of exposure assessment in the RoC Monograph on Trichloroethylene (NTP 2015).

Differences in reviewer rankings are resolved through mutual discussion with reference to the original data source. A small subset of studies may be used in a "pilot" phase, so that any ambiguity can be discussed and resolved before evaluation of the full set of studies. If the information to evaluate a signaling question is inadequate, the study authors may be contacted. The bias and sensitivity analyses are captured in a web-based content management system such as Table Builder. Terms used in the evaluation are defined below, and the evaluation of the specific domains follows the scheme shown in Figure 3-4.

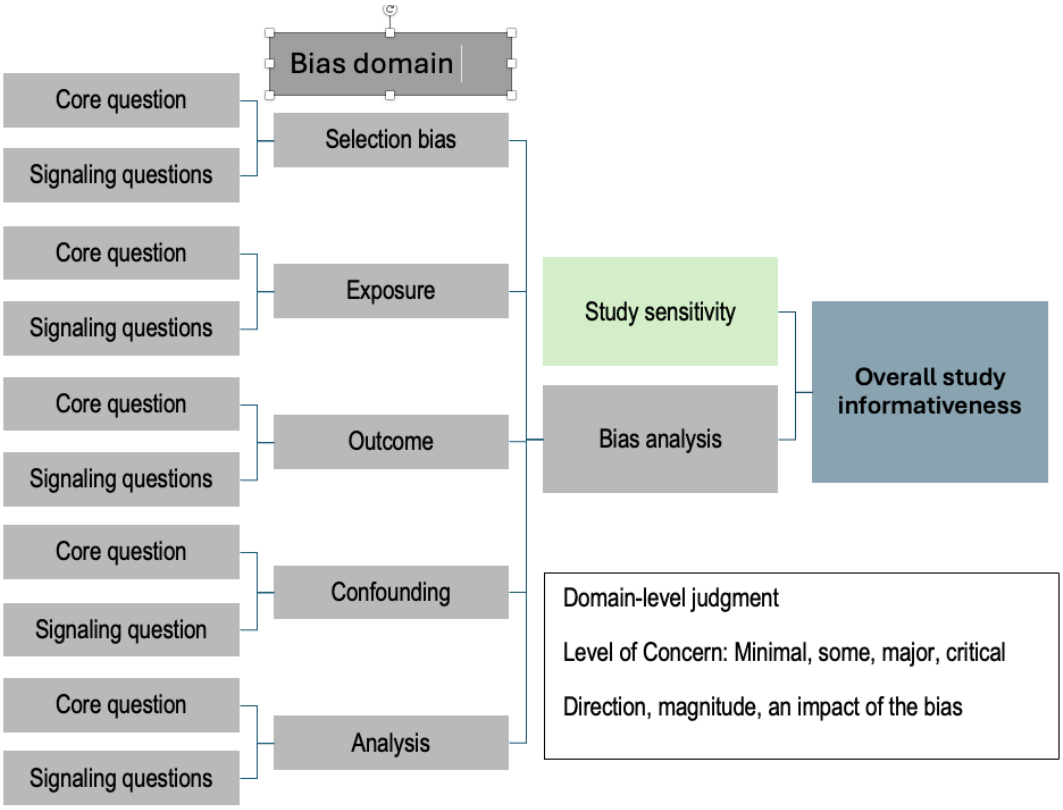


Figure 3-4. Schematic of the Approach to Systematic Review of Study Informativeness

The bias evaluation (grey) for each domain or sensitivity (green) is captured by a core question for that domain. A series of signaling questions is used to inform the core question and domain-level judgment, including the direction, magnitude, and impact of the bias (see Section 3.3.2). To evaluate bias within a study, characteristics of the study are compared with that of an "ideal" study conditions for a specific study design, end point, and exposure. The overall informativeness assessment is based on scientific judgment considering all the domains.

3.2.2. Study Informativeness Evaluation Questions and Guidelines

Selection and Attrition Bias

Selection bias occurs when selection into the study is related to both exposure and outcome; the relationship between the exposure and outcome is different for those who participated in the study (study population) than for the source population (all those who were eligible for the study, including those who did not participate) (Rothman et al. 2008). Generalizability is determined by how representative the sample (study population) is of the target population. This is known as external validity.

Concern about selection bias is greater in case-control studies, owing to different probabilities of selection for cases and controls (Pearce et al. 2007); however, selection bias in most instances is less likely if participation was high among both cases and controls, given that the initial selection strategy was not biased.

When there is complete recruitment and follow-up, selection bias is usually less of a concern in cohort studies than in case-control studies and cross-sectional studies, because the cohort itself acts as the source population (Pearce et al. 2007); however, attrition bias or selection out of the

cohort is a concern. Note that some specific biases could be considered in multiple domains. The RoC considers the healthy worker effect (HWE)—both the healthy worker-hire effect (HWHE) and healthy worker survivor effect (HWSE)—as a type of selection bias, but it can also be considered as confounding (Pearce et al. 2007). Additionally, in cancer studies, evaluation of the completeness of follow-up often overlaps with outcome determination.

Directed acyclic graphs can be used to visualize most types of selection bias, which can be described as a "collider bias," that is, bias that results from conditioning on a common effect (the collider) of both the exposure and disease under study. For example, conditioning on the presence of cardiovascular disease (CVD) when evaluating the relationship between a genetic polymorphism that causes CVD and smoking (also a cause of CVD) would yield a distorted estimate of the exposure-outcome relationship (relative risk, odds ratio, or standardized mortality ratio) (Hernán et al. 2004). The bias can be toward or away from any true association or can cause a spurious association to appear when none exists (Hernán et al. 2004). For human mechanistic studies, other study designs may be relevant, such as cross-sectional studies, intervention studies, and randomized trials. However, the guidelines for case-control studies generally apply to those studies. Specific details will be outlined in the evaluation protocol.

For further details, see Table 3-3.

Core question: Is there concern that selection into the study (or out of the study) was related to both exposure and outcome?

Table 3-3. Selection and Attrition Bias: Questions and Guidelines

Signaling Questions	Guidance	Response Options
Selection into the study: All study designs Is there concern that the selection methods were not adequate? Did the eligibility criteria (inclusion or exclusion) or recruitment strategies differ among study participants, such as between cases and controls (case-control studies) or exposed and unexposed subjects (cohort studies)?	Participation should not be an effect of both outcome and exposure status (e.g., conditioning on a collider). Creating a directed acyclic graph may help identify colliders and whether they have been conditioned on for selection. This usually biases an estimate away from the null, but in some cases can bias it toward the null (Hernán et al. 2004).	**Minimal concern** There is no evidence that selection of the subjects was related to both exposure and outcome. Cases and controls or exposed and nonexposed were selected from the same source population by similar methods and criteria. For cohort studies, the cohort is clearly defined (e.g., includes groups of those exposed and unexposed for a specific time period/location, with no evidence that follow-up differs between the exposed and nonexposed). **Some or major concern** These are evaluation-specific and will be defined in the protocol. **Critical concern** There is strong evidence that selection or attrition of subjects was clearly related to both exposure and outcome.
Is there concern that study participants were not selected from the same underlying (source) population during a similar time period?	Ideally, the study participants should be similar in all respects except for exposure status (cohort) or disease status (case-control) and be drawn from the same underlying population. Note: this does not apply to population-based cohorts where the cohort is not sampled based on exposure status. For **case-control studies**, controls should be from the same source population as the cases and be at risk of the outcome (e.g., participants with hysterectomies should not be controls in a study of endometrial cancer). Inappropriate selection of control groups that do not represent the underlying population from which the cases are selected can be a major concern. Selection bias is less likely if there is high participation among both cases and controls given the initial strategy was not biased (Pearce and Richiardi 2014; Snoep et al. 2014). The following scenarios indicate potential bias for **cohort and/or case-control studies**: • Differential participation and/or response related to outcome and exposure status (cohort, case-control). • Self-selection into the study (cohort, case-control). This can be assessed by comparing participants' characteristics	

Signaling Questions	Guidance	Response Options
	with those of the general population (cohort and case-control). • Berkson's bias (case-control). This can be largely attenuated by limiting enrollment to incident cases and further attenuated by excluding cases and controls (if controls are from a diseased group) with a co-occurring condition that is not the reason for hospitalization (Pearce and Richiardi 2014; Snoep et al. 2014). • Cohort studies: Bias can occur if both exposure and outcome change over time (dos Santos Silva 1999; Soskolne et al. 2021; Vandenbroucke et al. 2007). In **cross-sectional studies**, there may be concern that those with the outcome who survive long enough to participate in the study differ from those who do not survive as long or are not healthy enough to participate.	
• In cohort studies, is there concern that the health of a participant may have affected selection into the study (e.g., the HWHE or healthy volunteer effect), leading to underlying differences in disease risk between exposed and unexposed individuals in the target population? • In cohort studies, is there concern that the cohort was initiated as the result of a cluster?	In occupational cohorts, standardized mortality or incidence ratios below 1 for mortality from all causes, all cancers, cardiovascular disease, or non-malignant respiratory diseases relative to the general population may be an indication of the HWE. The HWHE[a] generally biases the findings toward the null.	

Signaling Questions	Guidance	Response Options
Selection out of the study: Cohort study Is there concern about attrition bias or incomplete follow-up? • Is there concern that selection out of the study was related to both exposure and outcome status? • Is there concern that an analysis conditioned on censoring is related to both exposure and outcome?	Ideally, rigorous methods should be used to ascertain case status and should not differ by exposure and outcome status. Incomplete follow-up unrelated to either exposure or outcome (and therefore nondifferential) can reduce the statistical power of the study, but does not result in a biased effect estimate (Kristman et al. 2004). Loss to follow-up can occur more often for certain subgroups of the population, such as women, non-white people, and low-socioeconomic-status participants (James et al. 1997; Jewkes and Wood 1998; NCHS 2013). If this bias in vital status ascertainment is associated with exposure, this can potentially bias the study results (Linet et al. 2020). Ideally, proxy-reported deaths should be supplemented or confirmed with data from a national mortality registry to reduce loss to follow-up whenever possible (James et al. 1997). The proportion of cases using or confirmed by supplemental data should be reported.	
Is there concern about the timing of follow-up (e.g., follow-up time did not coincide with the start of exposure, especially for outcomes with shorter latencies)? This may overlap with these concerns: • Prevalent hires • Left truncation • HWSE	Ideally, studies enroll participants at the start of exposure and follow them for an adequate period, which is determined by the relative latency periods for the specific type of cancer. However, start of exposure is not usually known (sometimes it is available for occupational studies), and consideration of latency period is not typical. Cohorts that consist entirely of workers identified at one point in time (i.e., including both prevalent and incident hires) have been found to over-represent long-term, healthy workers (possible HWSE) for outcomes not observable at the time of hiring (left truncation) (Applebaum et al. 2011; Picciotto et al. 2013). Although the HWSE typically biases results toward the null (Stayner et al. 2003), its magnitude can be estimated, given sufficient information on the proportions of prevalent participants or workers and the length of the follow-up period.	

Signaling Questions	Guidance	Response Options
All studies		
Were analyses conducted to control for any selection bias (in or out) or sensitivity analyses conducted to address the extent of any bias?	Ideally, studies should conduct sensitivity analysis to estimate the extent of selection bias or to control for it, if methods such as these are available: • Control for (time-related) employment status (Pearce et al. 2007), such as employment duration (Garshick et al. 2012), in the analysis. • Use of statistical methods to correct for left truncation and the HWSE, such as restricting the analysis to incident hires, or use of G-estimation, inverse-probability-of-treatment methods, or censoring weights (Sterne and Tilling 2002). However, the variables needed to correct the bias are unmeasured or unavailable in most longitudinal studies. • Minimization of the HWE by conducting an internal analysis that compares the exposed workers with the unexposed workers instead of with the general population. • Calculation of the relative odds ratio (ROR) to describe the effect on nonparticipation is sometimes used (Kleinbaum et al. 1982; Nohr and Liew 2018).	
If there is concern about the potential for selection or attrition bias, what is the predicted direction or distortion of the effect estimate (if there is enough information)?	When there is enough information to assess the predicted direction and/or distortion of selection or attrition bias, this assessment should be used in making the domain-level judgment for potential bias.	

Hover over underlined terms for pop-up definitions.

[a]The HWHE and HWSE are part of the healthy worker effect (HWE), which can be considered as both a type of selection bias and a type of confounding: however, it is addressed here under selection bias (Pearce et al. 2007).

Exposure Measurement Error and Misclassification

One of the most important aspects of an epidemiology study is its ability to correctly classify the study subjects at the individual level with respect to exposure status and level of exposure. This involves several dimensions: carefully defining the exposure used in the study, knowing sufficient information about the exposure setting (e.g., occupational, residential), selecting appropriate data collection tools and methods for using or modeling the exposure data, evaluating the quality of the exposure assessment methods, determining whether individuals can be adequately separated with respect to their exposure levels, and assessing whether knowledge of the outcome may have affected the reporting of exposure. Typically, exposure mismeasurement applies to continuous exposure data while misclassification applies to categorical (including discrete) variables.

As summarized in Section 3.1.2, a wide range of tools may be used in collecting exposure data, as well as a wide range of methods in analyzing these data. Table 3-2 shows that, depending on the type of study, certain summary measure(s) of exposure and a given method may have advantages over others with respect to potential to avoid bias. Assessing the potential for bias from measurement error (or exposure misclassification) includes consideration of: (1) how well the exposure proxy approximates the exposure of interest; (2) how accurately and precisely the exposure (or exposure proxy) is assessed (e.g., measured); and (3) differential recall bias or observational bias. Table 3-4 presents signaling questions and general considerations for assessing the risk of bias in measurement of exposure. Table 3-5 links considerations for risk of bias from Table 3-4 to the types of exposure assessments presented in Table 3-2 and the types of studies in which they are commonly used. Additional considerations are discussed below.

Any evaluation of exposure assessment includes complex consideration of the manner by which exposure data are collected, the way data are used to classify participants into exposure groups, and how missing data and datasets containing values too low to report as reliable numerical values (i.e., not detected or below the limit of quantitation) are addressed.

Biological information may also inform the evaluation of the exposure assessment. Such information, as available, can be relevant to multiple aspects of exposure assessment, including determination of whether an internal or excreted measure of exposure is most appropriate and identification of the relevant time window of exposure for a specific type of cancer (considering the relevant latency period).

Misclassification of exposure in cancer epidemiology studies is often nondifferential, which usually results in a bias toward or beyond the null (i.e., underestimation of the true risk). However, there are exceptions when non-misclassification may result in bias away from the null, such as when there is: (1) nondifferential but dependent misclassification of both exposure and outcome (e.g., self-reported exposure and outcome from the same questionnaire; however, for most cancer studies, information of cancer comes from an independent source, and this is usually not a problem) (Kristensen 1992); (2) nondifferential of multiexposure categories (however, this can attenuate or inverse the exposure-response relationship, which may bias the conclusion of the study away from the null); (3) Berkson bias results in imprecision and does not affect the magnitude of the risk estimate for linear regression models, but may bias the effect measure in log-linear models (Yland et al. 2022).

Differential bias can modify the effect estimate in any direction; in case-control studies, differential recall bias and reverse causation usually result in a bias away from the null.

Core question: Is there concern that the exposure assessment did not distinguish between exposed and nonexposed people or among exposure categories? If known, exposure assessment should be based on an etiologically relevant time window of susceptibility and exposure proxies for the cancer of interest.

Table 3-4. Exposure Measurement Bias: Questions and Responses

Signaling Questions	Guidance[a]	Response
Exposure proxy Is there concern that the exposure proxy did not adequately represent the exposure of interest at an appropriate time window of exposure for the outcome of interest? • If yes, was this true for all exposure metrics, or a particular metric?	In observational studies, the exposure of interest is often determined through the use of an exposure proxy (e.g., measurement of a chemical metabolite, job title) and not measured directly (exposure to the substance).	**Minimal concern** The exposure proxy closely approximates the exposure of interest at a relevant time window for the outcome of interest. The exposure is consistently assessed through the use of established or validated methods, with minimal missing exposure data, and any measurement error is small in relation to between-individual variation compared with differences between groups. Exposure groups are adequately separated. Any misclassification is nondifferential. **Some or major concern** These evaluation-specific concerns will be defined in the protocol. **Critical concern** Exposure assessment is not at the individual level, and/or the exposure proxy does not approximate the exposure of interest well, is not within a relevant time window, or lacks other relevant metrics, resulting in poor discrimination between exposed and unexposed and among exposure groups. The number of participants with missing exposure data is large enough that the estimated effect of exposure on outcome is likely to be substantially different from an estimate generated from a complete dataset, or there are indications that the data are not missing at random.
Exposure measurement Is there concern about error in measurement of the exposure of interest or exposure proxy? Is there concern about the use of the collected data to classify exposure groups?	Any evaluation of exposure assessment methods includes consideration of how well the data are collected (see Table 3-5) and how the data are used to classify participants into exposure groups. This is particularly important for studies in which the exposure contrasts in the population are small.	
If yes to either, is there concern that measurement error resulted in inadequate separation of groups with respect to exposure? • Did any misclassification vary by exposure category? Is there concern that the exposure classification did not capture the variability of exposure?	Ideally, the exposure would be consistently assessed across exposure groups through the use of established or validated methods, with any measurement error being small in relation to between-individual variation compared with differences between groups (e.g., unexposed vs. exposed, high vs. low exposure). Consideration should be given to how a study treats samples without detectable levels of analyte. • Exposure measurement may vary depending on the timing of exposure. In studies with large proportions of exposed individuals, attention to the definition of "unexposed" is important, as these individuals may not be completely unexposed, potentially introducing bias toward the null.	

Signaling Questions	Guidance[a]	Response
• If yes to either, is the measurement error classical, Berkson, or a mixture of both?	Random misclassification of exposure may introduce bias, depending on type of error. Berkson error results in lack of precision and does not usually bias the magnitude of the effect estimate much, if at all. Classical error tends to attenuate the risk estimates (i.e., bias toward the null).	
Observation and differential recall bias Is there concern that knowledge of the outcome (e.g., resulting in observation or recall bias) may potentially bias the exposure assessment (away from the null)?	Recall bias (potentially biasing toward overestimation of the effect) is not necessarily introduced when exposure information is collected after the outcome occurred. Authors may include additional data that help determine whether cases are overreporting exposures; this can help determine the likelihood of recall bias (e.g., by including an additional case group whose case status is unlikely to be related to exposure level [case-control designs] or including measures of symptoms or health outcomes unlikely to be related to exposure [cohort studies]). Differential recall bias is less likely to be a concern for self-reporting in case-control studies in which occupational exposure is assigned using participant recall of more objective data (such as job titles, occupations, work history) as a proxy for exposure than in studies using self-assessment of chemical-specific exposures or questionnaires with exposure checklists.	
Has a temporal association been established—i.e., does the exposure (or proxy) measure approximate the exposure that was relevant in the time period of susceptibility? Is there concern that presence of the outcome may potentially bias the exposure assessment (e.g., reverse causality)?	Exposure or proxy measures "Reverse causality" may be a concern in retrospective or cross-sectional designs (e.g., cross-sectional studies, some case-control studies, or cohort studies with cross-sectional analysis) that measure exposure at the time of or after disease diagnosis. This can be of particular concern in studies that use biomarkers for exposure assessment.	
If there is any misclassification, is it differential or nondifferential, and what is the predicted direction or distortion of the effect estimate (if there is adequate information)?	In general, nondifferential misclassification biasing toward the null occurs if there is equal exposure misclassification of cases and controls or of exposed and unexposed subjects. With more than two exposure categories, the direction of the bias is not always clear, but it may result in attenuation of the exposure-	

Signaling Questions	Guidance[a]	Response
	response relationship. When misclassification of exposure is not equal between cases vs. controls or exposed vs. unexposed, it is differential, and it causes a bias either toward or away from the null, depending on the proportions of subjects misclassified. When there is enough information to assess the predicted direction and/or distortion of exposure misclassification, this assessment should be used in making the domain-level judgment for potential bias.	

Hover over underlined terms for pop-up definitions.
[a]See Table 3-5 for guidance specific to different types of exposure assessments.

Table 3-5. Guidance on Exposure Assessment Methods that Reduce Various Types of Bias

Exposure Measure	Exposure Proxy	Exposure Measurements (Usually Nondifferential Misclassification)	Observational and Differential Recall Bias
Data Collection Methods			
Self-report	Use an exposure proxy that is representative of the exposure of interest in the appropriate time window and one that can be reported with minimal bias.	Link exposure questions to events or types of use to provide context for respondents and improve memory; exposure details for distant past events are more difficult to remember. Ask for detailed information on calendar time, location, duration, and frequency, with multiple prompts and visual aids (e.g., portion sizes, timelines) to help subjects remember; this provides more useful information for reliably classifying participants into different exposure groups. Use structured in-person or telephone interviews. Mail or web questionnaires can be preferable when asking about sensitive topics. Query study participants (actual cases and controls), rather than proxy respondents.	Differential recall bias: • Generating exposure information prior to the outcome (i.e., in a prospective study) eliminates the potential for differential recall bias, as exposure recall may be affected by knowledge of the outcome. • Differential recall bias is less of a concern in studies asking for more objective information, such as job title, than for personal estimates of specific exposures. Observer bias: • Reduced when participants are blinded to study objectives and asked information on other exposures in addition to the exposure of interest. • Reduced when interviewers/assessors are blinded to disease (case-control) or

Exposure Measure	Exposure Proxy	Exposure Measurements (Usually Nondifferential Misclassification)	Observational and Differential Recall Bias
			exposure (cohort) status; this is often difficult to do in case-control studies.
Records	Use record data that have additional exposure information, as job titles alone may or may not be representative of actual exposures across locations or industries.	Use detailed records; for example, information on tasks is more informative than job titles. Use records specific to the study population (e.g., company records) rather than routinely collected data (e.g., census).	Differential recall bias: • Not relevant. Observer bias: • Most studies blind exposure assessors to case status, so this is not usually a major concern.
Environmental Measurements	Use an indicator of exposure directly measured in environmental media (e.g., air, soil, dust, or water) and specific to the timing of exposure and the population.	Individual vs. area or aggregate measurements: • Use individual-level environmental measurements, if possible, instead of area measurements for ambient levels of pollutant in outdoor air, indoor air, workplace, water, soil, or household dust. • Avoid using aggregated measures of exposure, as they are subject to considerable error for individuals, for both ever-exposure and the exposure categories used for evaluating exposure-response relationships. Sampling: • Use appropriate sampling methods and strategies that are consistently applied, following guidelines and standards. • Select sampling location(s) at random and define the number of samples, with a maximum accepted level of error.	Differential recall bias: • Not relevant. Observer bias: • Most studies blind exposure assessors to case status, so this is not usually a major concern.
Biomonitoring	Use an indicator of exposure directly measured in biological media (e.g., urine, plasma, fat, or soft tissue) that is a valid indicator of the exposure of interest, and specific to the timing of exposure population, and current biological understanding.	Sampling: • Use appropriate sampling methods and strategies that are consistently applied, following guidelines and standards. Laboratory test methods: • Use high-sensitivity and -specificity methods to detect biomarkers in appropriate biological media (e.g., urine, plasma, fat, or soft tissue).	Differential recall bias: • Not relevant. Observer bias: • Most studies blind exposure assessors to case status, so this is not usually a major concern Reverse causation may be a problem in case-control or cross-sectional studies if the levels

Exposure Measure	Exposure Proxy	Exposure Measurements (Usually Nondifferential Misclassification)	Observational and Differential Recall Bias
	Caution the use of biomarkers with short half-lives in studies of disease with long disease latency unless there is evidence of chronic, consistent exposure (e.g., longitudinal sampling) or relevant toxicological information. Avoid the use of nonselective markers that can be markers for other compounds (e.g., with shared metabolites).	• Use assays with adequate limits of detection and quantification. • Consider factors that can alter concentrations of a metabolite as a result of individual variation, and, if necessary, adjust for them or consider them as effect modifiers.	of the biomarkers are affected by the disease process.
Methods Used to Assess Collected Data			
Expert Assessment or Job-exposure Matrix	Use exposure inputs for the JEM (job records and questionnaire data linking occupational exposures) or expert assessments that are representative of actual exposures either within specific locations or industries or across locations and industries.	• In general, JEMs or expert assessments applied to industrial cohorts are preferred over methods based on data from population-based or case-control studies. Assessment by experts (such as industrial hygienists or occupational physicians) is usually the most credible method for assessing occupational exposures in population-based case-control studies, as experts consider local differences in material usage, production processes, and control measures. JEMs: • Use data specific to the task, calendar year, location, and population to reflect variability in exposures within jobs from factory to factory and worker to worker. • Use multiple exposure metrics (e.g., cumulative, peak, frequency, or average intensity of exposure per unit time) that capture various dimensions of exposure, to increase the quality of the exposure assessment and increase the ability of the study to distinguish exposure groups from one another.	Differential recall bias: • Not relevant Observer bias: • Assessors are blinded to case status.

Exposure Measure	Exposure Proxy	Exposure Measurements (Usually Nondifferential Misclassification)	Observational and Differential Recall Bias
		• General population JEMs: Base algorithms on expert understanding of, and experience with, exposures in the population overall, independent of any specific subgroups. • Population-specific JEMs should include information on local conditions. Expert assessment: • Standardize assessments by independent experts across different settings in multicenter studies. In single-center studies, uniform assignment of exposure is usually not problematic, as assessment is done by only one expert or a group of closely associated experts. Misclassification of exposure levels by experts can differ greatly between the assessed exposures of interest; interrater agreement does not necessarily imply that assigned exposure levels are more accurate.	
Modeling	Use an appropriate conceptual model in which the (causal) relationships between the model parameters and the outcome are real, and the natures (or shapes) of these relationships are known. The study documents the rationale for the model.	Use the highest-quality input data, and parameters that are accurate and appropriate for the problem. Use exposure modeling grounded in measured data and biological understanding. Use correct assumptions and parameterization. Conduct sensitivity analyses given the uncertainty associated with modeled exposure data. Demonstrate that in a specific application, the model output agrees with measured data. When possible, use continuous measures of exposure rather than categories of exposure. Categorization of a continuous variable reduces the statistical power and may negatively bias the exposure-response relationship.	Less susceptible to observer and differential recall bias.

Outcome Misclassification

The outcome of interest is incidence of a specific cancer type or subtype. Assessment of the potential for bias due to measurement error or outcome misclassification considers: (1) how well the study outcome represents the outcome of interest; (2) the accuracy of the outcome measurement methods; and (3) the potential for observation bias. The adequacy of follow-up length is usually evaluated in the assessment of study sensitivity. Cancer incidence data are considered more informative than mortality data, because sources of incidence data (e.g., cancer registries, hospital records) are more accurate than sources of mortality data (e.g., death certificates) (Jewkes and Wood 1998; Linet et al. 2020; NCHS 2013; Patel et al. 2004) and mortality may not reflect incidence for cancers with high survival rates.

Table 3-6 provides signaling and follow-up questions, as well as general considerations, for assessing the potential for bias due to outcome misclassification.

Core question: Is there concern that the outcome measure did not reliably distinguish between the presence or absence (or degrees of severity) of the outcome?

Table 3-6. Outcome Misclassification: Questions and Responses

Signaling Question	Guidance	Response Options
Is there concern that the method of measuring the outcome did not represent the outcome of interest? • If mortality data were used, did they adequately reflect incidence?	Ideally, each study will measure cancer incidence. Mortality data selectively miss cases with longer survival. Mortality data are adequate for types of cancer with low survival rates and may be preferable for rapidly fatal diseases, when date of death approximates date of diagnosis (i.e., incidence), or in low-to-middle-income countries that lack cancer registries or have inadequate reporting systems (so that reporting rates do not reflect the true incidence of disease) (Siddiqui and Zafar 2018; Torre et al. 2016).	**Minimal or some concern** Outcome measurement methods clearly distinguish between diseased and nondiseased subjects. Follow-up and diagnoses are conducted independent of exposure status. **Some or major concern** These are evaluation-specific and will be defined in the protocol. **Critical concern** There is strong evidence that the outcome measurement methods do not discriminate between diseased and nondiseased subjects and/or that follow-up and diagnoses are likely related to exposure status.
Is there concern that the disease was not accurately diagnosed? • Does misclassification of outcome vary across exposure groups or levels of exposure? • If so, were any methods used to adjust for potential bias? • Is there concern that the non-diseased group may have had the disease of interest or that the diseased group may not have?	Ideally, cases of cancer should be histologically confirmed and obtained from population-based cancer registries (PBCR). Cancer incidence data obtained from PBCR are considered the gold standard, as they consolidate data from many sources and strive to provide a complete, unbiased estimate of cancer incidence in the population. In the absence of PBCR data, incidence data can be obtained from medical records and hospital pathology data. In the United States, incidence is obtained by linking to SEER databases and may be missed if cases occur in a region not covered by SEER (which covers ~48% of the population). Additionally, the North American Association of Central Cancer Registries (NAACR) combines data from different cancer registries in North America and can include populations not covered by SEER. If some cases are confirmed clinically or with imaging alone, the proportion of such cases should be noted. In some circumstances (e.g., difficult-to-	

Signaling Question	Guidance	Response Options
	diagnose cancer types or multicenter studies), it is preferable that at least a subset of cases undergo independent pathology review by a study investigator. There may also be inaccuracies in reporting cause of death (Flanders 1992). For cancer mortality, cause of death data from databases/registries that use algorithmic processing to standardize outcomes, such as the U.S. National Death Index, may be more reliable than death certificate data. Self-reported and proxy-reported incidence and cause of death are the least accurate sources, with proxy-reported being less accurate than self-reported. Any changes over time in subtype classification (e.g., of non-Hodgkin lymphoma subtypes and other lymphohematopoietic cancer) or deviation from specificity of the subtype of interest (e.g., leukemia vs. myeloid leukemia) should be noted, as this misclassification may dilute (reduce) effect estimates. Generally, if the number of cases is sufficient, subtypes with differing etiologies should be evaluated separately. However, in some contexts, a combined cancer type may be preferable.	
Is there concern about detection bias?	Ideally, observation for cancer incidence (e.g., screening or better health follow-up) should be similar for all subgroups in the study population, but particularly for exposed and unexposed participants. Mortality data can be used to examine detection bias for some, but not all, cancers. Cancers of concern for detection bias may be female breast, thyroid, or prostate cancers (Marjerrison et al. 2022).	

Signaling Question	Guidance	Response Options
Is there concern about observer bias?	Ideally, the outcome assessors do not have knowledge of the subjects' exposure status and are not influenced by exposure status. In general, cancer diagnosis is made objectively (e.g., by histopathology) and without knowledge of specific exposures (with the possible exception of smoking), so observer bias is relatively unlikely to be a concern for most cancer studies. Observer bias may be more likely for outcomes that are self-reported or reported by next of kin.	
If there is misclassification, is it differential or nondifferential, and what is the predicted direction or distortion of the effect estimate (if there is adequate information)?	Nondifferential misclassification of cancer (not related to exposure status) would most likely bias the effect estimate toward the null; however, there are some exceptions (see text above). Misclassification that is differential with respect to sociodemographic characteristics (income, race/ethnicity, gender, or age) may occur for self-reported cancers (D'Aloisio et al. 2017), and the possibility that these characteristics also are associated with exposure should be considered. When there is enough information to assess the predicted direction and/or distortion of outcome misclassification, this assessment should be used in making the domain-level judgment for potential bias.	

Hover over underlined terms for pop-up definitions.

Potential for Confounding

Confounders are factors that are moderately to strongly associated with both exposure and the disease outcome(s) of interest, as described in Protocol Development (Section 3.1.2). In lieu of randomization, the potential for confounding in observational studies can be controlled in the design phase or in the analysis phase. One option in the design phase is restriction—limiting the study to only those subjects for whom potential confounders fall within a narrow range of values (e.g., enrolling only males into a study). Another method of confounder control in the design phase is to match cases and controls by similar characteristics. In the analysis phase, confounding can be controlled through statistical techniques such as stratification and multivariable methods.

The ability to control for any confounding factor is predicated on that factor being accurately measured and quantified in the study. Assessment of the quality of measurement of exposure to the confounding factor is similar to that for measurement of the exposure of interest.

It is important to characterize the extent and impact of residual confounding (uncontrolled potential confounders), which, in turn, affects the magnitude and direction of the effect estimates. In some cases, adjusting for a factor that is not a confounder can also bias the risk estimate (i.e., if the variable is in the causal pathway). In other cases, overadjustment in statistical models for additional factors not considered to be true confounders may lower the precision of the estimates (widening the confidence intervals), but this is not likely to bias the magnitude (as discussed below, under the Analysis section).

Identifying key factors that could potentially confound the exposure-outcome relationship is a critical step in evaluating confounding bias. Our approach to this domain is twofold: first, we assess whether the studies adequately addressed the potential for confounding bias in a study, as seen in the questions below (Table 3-7). Subsequently, we assess the impact of potential confounding on the effect estimate within a study (detailed in Section 3.3.1.). In the process of considering potential confounders, particular emphasis is given to the “confounder matrix” [adapted from (Shapiro et al. 2018), as illustrated in Section 3.3.1]. The confounder matrix allows for each potential confounder to be identified, based on the protocol, and summarizes whether the confounder was addressed in the analysis, whether it was a co-exposure associated with the exposure of interest, the magnitude and direction of the association, and additional pertinent information. As detailed in Section 3.3.1, we then use the “confounder matrix” to determine whether confounding within a study can be ruled out based on scientific judgment. Potential confounders can be compared both within and across studies.

Note that the HWE can be considered as both a type of selection bias and a type of confounding (Pearce et al. 2007). For our purposes, it is addressed only under Selection and Attrition Bias (above).

Table 3-7 provides signaling and follow-up questions, as well as general considerations, for assessing the potential for bias due to confounding.

Core question: Is there concern that the effect estimate may be confounded, e.g., potential confounding is either not adequately addressed by the methods or there is inadequate information to allow for its evaluation?[a]

Table 3-7. Potential Confounding: Questions and Responses

Signaling Questions	Guidance	Response Options
Is there concern about the measurement of co-exposures or personal behaviors in the study? If no data are provided about confounders, are surrogate data on potential confounders available?	Ideally, quantitative information on personal behaviors and other likely confounders should be assessed by in-person interview conducted by interviewers blinded to the case status of the respondent, rather than via proxy respondents. Residual confounding is more likely when only limited qualitative information on a given risk factor (e.g., only "yes/no") is available. Studies should provide data on the distribution of potential confounders by exposure and disease status. Data may be available on potential confounders in subsamples, which can help provide interpretation of the prevalence of the potential confounders in the exposed and unexposed or cases and controls. In addition, data on diseases associated with exposure may provide indirect information about risk factors for specific cancer end points of concern.	**Minimal or some concern** The study measured key potential confounders and/or used appropriate analyses or designs to address them. **Some or major concern** These are evaluation-specific and will be defined in the protocol. **Critical concern** There is strong evidence that the effects of the exposure cannot be distinguished from the effects of potential confounders.
Is there concern that the design or analysis did not adequately address important confounding through matching, stratification, multivariable analysis, or other approaches? • Is there additional information available with which to evaluate potential confounding or conduct sensitivity analyses (indirect adjustment)? • Is there concern that controlling for one or more variables (such as those in the causal pathway) could cause bias? • Is there concern that not adjusting for one or more confounders would be expected to	Ideally, confounders should be controlled for in the design or analysis phase. Not controlling for unaccounted potential confounders is likely to bias the results. If there is no information on confounders, determine whether external information (e.g., smoking rates in population, strength of the risk factor for the outcome) can inform whether confounding is an issue. Comparing minimally adjusted (e.g., by age) to fully adjusted models can help inform whether a factor is a confounder. Some potential confounders could be effect modifiers. Ideally, studies should stratify by these	

Signaling Questions	Guidance	Response Options
differentially favor outcomes in those with higher or lower levels of exposure?	variables to indicate whether a variable is an effect modifier. Considering a variable only as a confounder may obscure possible effects in high-risk or vulnerable subgroup(s) within the study population. Statistical models over-adjusted for factors that do not meet the definition of a confounder may introduce bias. If collinearity from highly correlated co-exposures is likely to introduce confounding, determine whether additional analyses (e.g., correlation matrices, stratified models, multipollutant models, mixtures methods) can identify the individual effects while controlling for co-exposures.	
What is the predicted direction or distortion of the effect estimate (if there is adequate information)?	Confounding can lead to an over- or underestimation of the risk estimate. When there is enough information to assess the predicted direction and/or distortion of confounding, this assessment should be used in making the domain-level judgment for potential bias.	

[a]The healthy worker effect is both a special type of confounding and a type of selection bias (see Selection and Attrition Bias section).

Analysis

The assessment of analysis bias considers the appropriateness of data assumptions, statistical models and methods used in the statistical analysis to evaluate the overall findings. Bias can occur when adjusting for variables in the causal pathway between exposure and disease. Overadjustment refers to adjusting for additional variables in statistical models that are not considered to be true confounders (see Confounding section). This would decrease precision but does not bias the results.

When adequate data are available, studies should evaluate exposure-response relationships, periods of susceptibility, and latency, or conduct subgroup analyses (especially for subgroups exposed at higher levels for longer durations). This assessment overlaps somewhat with study sensitivity. In some studies, such as case-control studies evaluating exposure to numerous substances without clear hypotheses, appropriate consideration should be given when interpreting multiple comparisons. Note: The signaling questions and guidance in the Table below are common considerations (not an exhaustive list) of the analytical considerations that may arise in epidemiological studies. Working with a biostatistician may help elucidate any biases arising from data analysis and should be further detailed in the development of a protocol.

Core question: Is there concern that the data assumptions and analysis were not adequate or that the study did not conduct relevant analysis of the available data?

Table 3-8. Analysis: Questions and Guidelines

Signaling Questions	Guidance	Response Options
Data assumptions Is there concern about whether the data assumptions used in the statistical analysis were adequate?	For example, are data transformation methods (e.g., log transformation) appropriate Or is an assumption of linearity appropriate? Were outliers removed? (Ideally, outliers should not be removed without strong justification, as that may be where the effect is strongest.)	**Minimal or some concern** The study used relevant data and appropriate assumptions and analysis methods. **Some or major concern** These are evaluation-specific and will be defined in the protocol **Major or critical concern** There is strong evidence that the study's analytical methods were so limited that the findings were uninterpretable or were distorted such that no conclusion can be made.
Statistical model and methods Is there concern about the appropriateness of the statistical model for the study design or about the adequacy of the conduct of the analysis	Examples of models include Cox proportional hazards regression (hazard ratio), Poisson regression, multivariable logistic regression (odds ratio), and conditional logistic regression. Matching should be adequately described and accounted for in the analysis (e.g., for controls described as individually matched to cases, was the ratio of cases to controls included, and were conditional regression techniques applied?).	

Signaling Questions	Guidance	Response Options
If applicable, did the study use appropriate methods to adequately evaluate exposure-response and latency, or to conduct subgroup analyses?	Analyses of subgroups exposed at higher levels or for longer durations should use appropriate methods to delineate groupings. Questions about the adequacy of methods include whether the data were modeled continuously or divided into categories, and whether the analysis used linear tests for trend and appropriately stratified groups.	
Is there concern about "over-controlling" (i.e., controlling for variables unnecessarily)?	Controlling for variables that are not related to exposure or disease will most likely reduce the precision of the risk estimate.	
Missing data Is there concern that missing data may have biased the findings? • Is there concern that missing data on the outcome or on any potential confounders is substantial? • Is there concern that missing data were not handled by an analytically appropriate method (e.g., sensitivity analysis or imputation)?	Complete subject analysis (i.e., including only those participants with complete data for all variables modeled) is considered less than ideal. Ideally, there should be little or no concern that the data are missing for reasons related to both exposure and disease.	
What is the direction, magnitude, and impact of this bias on the effect estimate?	When there is enough information to assess the predicted direction and/or distortion of analysis bias, this assessment should be used in making the domain-level judgment for potential bias. This may be difficult to ascertain for most analyses.	

Study Sensitivity

Study sensitivity is the ability of a study to detect a true effect (Cooper et al. 2016) and is analogous to the term "informativeness" as used in the preamble to the IARC Monographs (IARC 2019; Samet et al. 2020). (Both IARC and RoC evaluate bias and sensitivity but used different terminology to describe study sensitivity). Studies that have a low risk of bias but are insensitive may not be informative for reaching public health decisions about a potential causal relationship between exposure and outcome, as they are less likely to be able to detect a true effect even if one exists. Both sensitivity and the potential for bias must be considered in order to identify the most informative studies and to identify those study elements that may help to explain heterogeneity across the body of literature. Failure to consider sensitivity may result in overweighting the results from insensitive studies or erroneously interpreting evidence as being conflicting (Cooper et al. 2016). Study sensitivity should be evaluated with the same rigor as bias assessment.

Assessment of study sensitivity includes consideration of: (1) study size or the numbers of exposed and unexposed participants or cases and controls; (2) exposure contrast and window; and (3) follow-up times based on estimated minimum latency period for the exposure and outcome of interest. The overall sensitivity evaluation requires integration of these factors. For example, a study evaluating effects from low levels of exposure most likely will need larger numbers of exposed subjects than studies of subjects exposed at higher levels. Table 3-9 provides signaling and follow-up questions pertinent to these issues and to general considerations for assessing study sensitivity.

In some cases, the line between biases and study sensitivity is not clear. Some systematic review methods consider questions in the study sensitivity domain that RoC considers in the bias assessment domain, such as: (1) the relationship between the participant's entry into the cohort and onset of exposure in the selection-bias domain;(2) the appropriateness of the exposure or outcome measure to the exposure or outcome of interest in the information-bias domain; and (3) the use of models inappropriately including factors that are not confounders in the analysis domain. Care should be taken to avoid considering the same issue in more than one domain.

Core question: Does the study have adequate sensitivity to detect an effect from exposure (if present)?

Table 3-9. Study Sensitivity: Questions and Guidelines

Signaling Questions	Guidance	Response Options
Statistical power Is there concern that the numbers of exposed cases were not adequate for detection of an effect in the exposed population or subgroups of the exposed population?	When both exposure and disease are rare, statistical power is determined largely by the number of exposed cases and exposed controls.	**No or minimal concern** The study had an adequate number of exposed subjects, with substantial exposure (level, duration, or range) and with adequate duration of follow-up for latency status. **Some or major concern** These are evaluation-specific and will be defined in the protocol **Critical or major concern** The study was modest or small, with few exposed subjects, and/or the exposure range was minimal.
Exposure contrast Is there concern that the levels, duration, or range of exposure of the population at risk in cohort and case-control studies was not sufficient or adequate for detection of an effect of exposure? • Did the exposed group include individuals with a low or unknown probability of exposure?	Dilution of risk estimates comparing exposed and referent groups can occur when exposure varies widely within the group(s) defined as exposed. Alternatively, in communities where a large proportion of subjects are exposed, those who are defined as "unexposed" may in fact be exposed, making it difficult to detect a signal. The ability to evaluate exposure-response relationships depends on an adequate range of exposure among the study participants and adequate numbers of participants in each exposure category.	
Is there concern that the exposure assessment did not capture the relevant exposure metric, ideally to the outcome of interest (if biological understanding is known)?	The relevance of the exposure metric could be considered in exposure misclassification instead of study sensitivity, which would be delineated in the substance evaluation protocol.	

Signaling Questions	Guidance	Response Options
Latency Was the elapsed time between exposure and the outcome measurement sufficient to allow for cancer induction?	The follow-up period should be relatively long, as cancers typically have latency periods ranging from years to decades. In addition, the age of the cohort should be considered. Assessment of relevant time windows of exposure, or use of analytic models that are lagged, consistent with knowledge of the latency of a specific type of cancer or other experimental data, are recommended, as the strength of the association between exposure and cancer risk may be stronger when capturing relevant exposure windows (Checkoway et al. 2019). Ideally, estimates of the minimum estimated time from initiation to cancer (i.e., minimum latency) from the exposure are based on direct observation of latencies.	

3.2.3. Overall Assessment of the Study

Box 3-2. Study Informativeness-Level Judgement

High: no or minimal concerns about most potential biases, high or moderate sensitivity.

Moderate: low, minimal, or some concerns about most potential biases.

Low: major concerns about several biases, sensitivity rating varies. Depending on the direction and distortion of the potential biases, the study may still be informative for the cancer hazard evaluation and can help explain heterogeneity of the findings.

Inadequate (very rare): critical concern about any bias; sensitivity rating varies.

The overall informativeness of a study is based on consideration of both the potential for bias (i.e., internal validity) and study sensitivity. Serious concerns about study quality will result in a lower informativeness judgment. However, a well-designed study with low sensitivity (e.g., having few exposed or expected cases for a specific end point) could be considered as having low informativeness for the cancer hazard assessment. In very rare cases, a study that meets inclusion criteria may ultimately be excluded from evidence integration if they are totally and obviously uninformative (such as documented evidence of severe exposure misclassification). For transparency, the study informativeness assessments are included in the monograph, and the findings may be briefly summarized, or evaluated in sensitivity analysis. Past example from the night shift work evaluation include a study using a JEM that predicted that only 0.06 of the participants were night shift workers in a country with a predicted estimate of 10% to 20% female night shift workers (Schwartzbaum et al. 2007). In the IARC Monograph evaluation of coffee and bladder cancer (IARC 2018), the working group focused its review on studies that adjusted for tobacco smoking (a strong risk factor and confounder), and studies that did not adjust for smoking were not carried forward for full review, because of the resource constraints associated with reviewing

a large number of studies with limited informativeness. When adequate information is available for a study, a judgment is made on the direction and distortion of its overall biases or whether it has low sensitivity to detect an effect (see Box 3-2).

The goal of the evaluation is to consider all the evidence and triangulate across the body of evidence, rather than exclude studies. Studies raising critical concern about bias (which is very rare) in at least one domain may be evaluated in the sensitivity analysis. The overall judgment of study informativeness is not meant to be an algorithm that sums up the ratings across domains. Different domains may be given greater weight depending on issues important for the specific candidate substance. For databases in which the quality of the studies varies considerably, informativeness-level categories may be combined, such as "moderate or low" (e.g., as in the case of trichloroethylene).

3.3. Evidence Evaluation and Integration

Following the assessment of study informativeness, evidence from individual studies is evaluated (Section 3.3.1) and integrated across studies (Section 3.3.2) to reach a level-of-evidence conclusion (sufficient, limited, or inadequate) about the carcinogenicity of the substance from studies in humans by applying the RoC criteria (see Box 3-3) to the assessment. The assessment is made for each cancer outcome, and the overall conclusion is based on the highest level of evidence (i.e., if the level of evidence for one cancer type is sufficient, the overall level of evidence is considered sufficient; levels of evidence for the other cancer types are noted). The evidence from the human cancer studies is integrated with the evidence from animal cancer experimental studies and mechanistic studies to reach the listing recommendation. Applying the RoC listing criteria to the body of studies on a specific substance involves evaluating: (1) whether there is credible evidence for an association between exposure to the substance and cancer and (2) whether such an observed association can be reasonably explained by chance, bias, or confounding.

Box 3-3. Report on Carcinogens Listing Criteria for Evaluating Carcinogenicity from Studies in Humans

Sufficient evidence of carcinogenicity from studies in humans: a causal relationship between exposure to the agent, substance, or mixture and human cancer.

Limited evidence of carcinogenicity from studies in humans: a causal interpretation is credible, but alternative explanations, such as chance, bias, or confounding factors, could not adequately be excluded.

The cancer hazard evaluation builds upon the assessment of study informativeness and assesses confidence in the findings from individual studies, which includes evaluating the impact of bias on the studies' findings (considering the magnitude and direction of the bias and the strength of the findings) (Section 3.3.1). The bias judgments (overall study judgment, domain judgment, and specific biases), effect modifiers, exposure metric, and other scientific issues are systematically explored across studies to evaluate consistency and potential sources of heterogeneity (Section 3.3.2). Finally, triangulation approaches and consideration of other causality factors (e.g., Bradford Hill considerations, causal inference) also guide the assessment, giving weight to the most informative studies and considering all the evidence (Section 3.3.2).

3.3.1. Evidence Evaluation

Conclusions regarding confidence in a given study are reached by evaluating the impact of specific potential biases and considering the strength and consistency of the findings. Information from other studies may also inform the evaluation of individual studies.

Evaluating the Impact of Bias on the Study Findings

The presence of potential bias (such as selection bias or information bias due to misclassification of exposure or outcome) or confounding does not necessarily mean that a study should be excluded from the hazard assessment. Conclusions about the evidence from each study should consider the strengths and weaknesses of that study, the direction and distortion of the biases, and the strength of the association between exposure to the substance and the specific type of cancer.

Lash et al. (2014) and others have proposed practices for quantitative bias analysis to estimate the direction, magnitude, and uncertainty of systematic bias. However, many but not all methods require the individual study data, which rarely are available for published studies. Nonetheless, there is growing interest in evaluating the impact of bias in published studies, and new methods may therefore be incorporated as these methods are advanced (IARC 2024). Most methods and approaches that have been developed focus on confounding, but some concepts could apply to other types of bias (e.g., from exposure misclassification). In addition, evaluations of recall bias and selection bias have been reported. However, the information needed to evaluate these types of bias (such as on nonparticipants) often is not available.

Confounding

The strength of the association can be important in evaluating whether specific confounders or biases can explain the observed association between exposure and cancer. When the magnitude is large, the effects of potential confounding (known, residual, or unknown) or bias typically are minor (Blair et al. 2007). However, in judging the magnitude of the risk estimate, the direction and distortion of the bias or confounding must be considered. For example, there may be data (e.g., sampling) to suggest that potential confounding from smoking could explain only 10% of an increase in cancer; therefore, one could have confidence in a study reporting an effect estimate of a relatively low but larger than 10% magnitude. Evidence that risk increases with increasing levels of exposure can help rule out bias, confounding, and chance with reasonable confidence and can provide convincing evidence of a credible association between exposure and disease. This is important for both identified confounders and unknown confounders.

RoC considers several qualitative and quantitative approaches for evaluating residual and unmeasured confounding at both the individual study level and across studies. It may also be possible to conduct sensitivity analyses (indirect adjustment) to evaluate the direction and extent of the potential confounding. This usually requires that the magnitude of the effect estimate for the confounder and disease be known and that information is available to estimate the prevalence of the confounder among the exposed and comparison groups (Pearce et al. 2007).

Axelson and Steenland (1988) proposed indirect methods (e.g., the following formula) for evaluating potential confounding from tobacco use when little or no smoking data are available. This method can be applied to other confounders. Using the value obtained for I0, the risk ratio due to the confounder can be estimated for various hypothetical populations with different proportions of the confounders (Box 3-4).

Box 3-4. Axelson and Steenland Formula for Bias Adjustment

$I = RI_0\ PCF + I_0(1- P_{CF})$

I = incidence of the disease in the population

I0 = rate of disease in non-smokers (i.e., without the confounder)

R = risk ratio of the confounder

PCF = proportion of the population with the confounder

Another approach is to calculate the E-value, which VanderWeele and Ding (2017) define as the minimum strength of association that an unmeasured confounder would need to have with both the exposure and the outcome to fully explain the exposure-outcome association. The larger the E-value, the greater the unmeasured confounding would need to be in order to explain the risk estimate. A major limitation of this approach is that it assumes that the prevalence of the uncontrolled confounder among the exposed is 100%, which is highly unlikely and limits the utility of this approach (MacLehose et al. 2021). An online calculator is available for calculating e-values.

The RoC process uses a confounder matrix [as shown in Figure 3-5, below, adapted from Shapiro et al. (2018)], which can incorporate the quantitative approaches discussed above or qualitative considerations for considering potential confounders. The confounder matrix allows for each potential key confounder to be identified based on the protocol and indicates whether the confounder was addressed in the analysis, whether the confounder was a co-exposure associated with the exposure of interest and the strength of the association is known, the magnitude and direction of the association, additional pertinent information, and whether confounding within a study can be ruled out based on scientific judgment. Potential confounders can be compared both within and across studies. Importantly, the confounder matrix is not a checklist but a tool to assist with determining the impact of confounding bias on an individual study's findings and across studies.

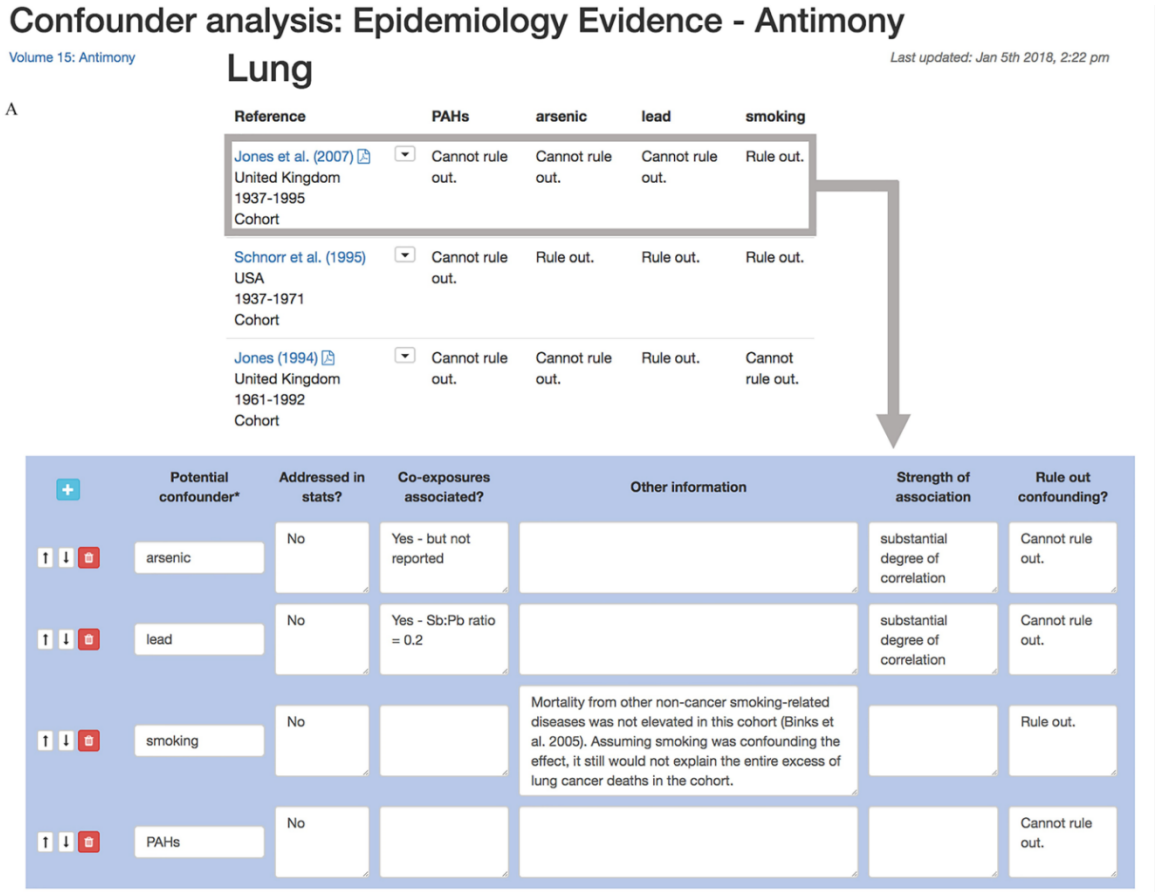


Figure 3-5. Confounder Analysis Matrix (Shapiro et al. 2018).

The summary table allows users to identify and view potential confounders in a matrix format and evaluate whether handling of these potential confounders can explain the findings within and across studies. This example displays studies examining the association between exposure to antimony and lung cancer mortality. Below the summary table, pertinent information on potential confounders from each study is extracted and evaluated.

Other Biases

The magnitude of the impact varies by the degree of misclassification and prevalence of the exposure. Sensitivity analyses can be performed to estimate the theoretical impact of exposure misclassification on risk estimates with data on the validity of exposure measurements (e.g., sensitivity and specificity) and predicted relative risks (Checkoway et al. 2004; Columbia Mailman School of Public Health 2024; Rothman et al. 2008). This information is rare, but assumptions can be made. Publications describing this theoretical exercise have demonstrated that relatively small errors (10% to 20%) can have large impacts on risk estimates (Copeland et al. 1977; Flegal et al. 1986). The potential for differential exposure misclassification can be inherent to case-control studies (e.g., due to recall bias or reverse causation), but the impact is generally thought to be minimal (Blair et al. 2007).

Methods have been developed to examine the effects of both selection and recall bias in case-control studies. For example, two large case-control studies on mobile phone use and cancer, the INTERPHONE study (Vrijheid et al. 2006), and the CEDALO study (Aydin et al. 2011) used Monte-Carlo simulation methods to determine the impacts of these biases on the risk estimates. These analyses are rare, as they rely on additional data (e.g., information on nonrespondents or

validation data), and it is uncommon for studies to have the information needed to conduct these analyses (Greenland 1996).

Evaluating Confidence in Individual Study Findings

Confidence in a study's findings (i.e., evidence for or against an association) involves considering the strength of the association, the potential for specific biases or confounding, the direction and distortion of those biases or confounding (i.e., their impact on the findings), and the sensitivity of the study to detect an effect (see Box 3-5 for guidance). Study confidence is based on scientific judgment and not criteria or checklists, which is especially important for studies for which there is a major concern about a potential bias. For example, if a study finds an association between exposure and disease despite concern about bias toward the null, the findings could be considered as supporting evidence. However, if the direction of the bias is unknown or away from the null, that study would probably not be considered in the integration of the evidence across studies.

Box 3-5. Guidance for Study Confidence Judgments

Moderate or strong evidence of an association (increased or decreased): Integration of the following considerations including (not all required): patterns showing internal consistency, evidence of an exposure-response relationship, **or** presence of a statistically significant risk from a well-designed study. These studies have a limited potential for (or small distortion from) bias, or any bias that may be operating tends mainly toward the null, underestimating the risk estimate (for a positive association). Methods used to assess confounding or information available on potential confounders indicate that potential confounding is unlikely to account for all of the excess or reduced risk.

Some evidence of an association: Evidence of an association, but the strength of the association is not likely to be fully accounted for by potential confounding or bias.

Null: Effect estimates are close to 1.0, but most potential bias is toward the null, or the study has low sensitivity to detect an effect.

Inconclusive: Study findings vary, but it is unclear whether all the excess or decreased risk can be explained by potential bias or confounding, or the direction of bias is unknown.

In addition to considering the potential for biases in the context of the strength of the findings (magnitude and exposure-response relationships), confidence in a study also is based on consideration of factors such as internal consistency. Examples of internal consistency include findings that are similar in both external and internal analyses or across different metrics of exposure. However, inconsistency may be attributed to design features (e.g., such as the HWE) or biological reasons (e.g., a specific metric may be related to the mode of action of a specific substance).

Statistical significance depends on power (sample size) and is not needed to assess the role of chance in a study's findings (Rothman et al. 2008). Rather, other factors less dependent on sample size, such as the width of the confidence interval (a narrower confidence interval indicating a more precise estimate) and consistency of patterns of findings within a study, can better be used to address the role of chance. Additionally, the conventional cutoff value for significance of $p < 0.05$ is somewhat arbitrary and does not have the same meaning in observational studies as it does in experimental studies (Brownstein et al. 2019; Wasserstein et al. 2019). Study power (usually gauged by the reported number of exposed cases and nondiseased) is assessed as one factor that contributes to study sensitivity, and the full range of considerations outlined above (strength of association, potential for and impact of specific biases such as confounding or selection bias, consistency across study metrics, and additional factors

related to sensitivity, such as the timing of exposure measurement) determine the level of confidence in a study's findings.

3.3.2. Evidence Integration Across Cancer Epidemiology Studies

The finding of consistent positive associations that are replicated across studies in different populations, with different study designs, and in different occupational settings reduces the likelihood that specific biases or potential confounders in individual studies explain the positive associations.

Evaluating Bias Across Studies

Domain or Specific Bias Analyses

In addition to analyzing biases in individual studies, evaluation of the most influential biases (e.g., nondifferential exposure misclassification) across studies is a useful step to understanding the overall impact of a specific bias on a body of literature. This includes a broader qualitative and/or quantitative understanding of the magnitude, direction, and impact of each common bias, when possible.

Each common bias may have unique aspects that should be considered. For example, if evaluation of the impact of confounding across studies found no appreciable differences between unadjusted and adjusted effect estimates, this would increase confidence in ruling out confounding from measured confounders. Relatedly, confounding by smoking may be minimized if no differences are seen between smoking-specific and nonsmoking populations.

Triangulation

Triangulation is an approach to evidence synthesis that requires a more inclusive view, integrating data from different methods, designs, theoretical approaches, and unrelated sources of bias to see whether the evidence suggests a single conclusion (Lawlor et al. 2016; Vandenbroucke et al. 2016). In this approach, a study would not automatically be downgraded because of bias; rather, the potential for influential biases for each study is recognized, and, if possible, the effect of that bias on the direction or magnitude of the effect estimate is identified. In a review, if different sources of bias exist across studies, but the results are consistent, given these potential biases, then this triangulation of the data can help increase the certainty of the reviewer's conclusion (Arroyave et al. 2021; Lawlor et al. 2016). Triangulation approaches can be used in integrating the evidence across studies of a similar discipline (human cancer epidemiology studies) or across disciplines. The latter approach can help inform an evaluation of cohesiveness or biological plausibility.

Bradford Hill Guidance

Several additional considerations—strength of the association, consistency across studies, evidence of an exposure-response gradient, and temporality of exposure (Bradford Hill 1965)—can assist in determining whether the findings can be better accounted for by alternative explanations. However, it should be noted that that these are not criteria; with the exception of temporality, not every element is required in order to demonstrate causality (Rothman and Greenland 2005). Biological plausibility is addressed in the assessment of mechanistic data (see Section 6).

Consistency and accounting for sources of heterogeneity across studies

Evaluation of the consistency and sources of heterogeneity across studies are key considerations in determining whether there is a credible association between the substance and cancer incidence. The evidence from studies should be systematically evaluated in the context of study informativeness (overall and specific biases, such as type of exposure assessment or potential confounders) and other key scientific issues or factors identified during the scoping activities—e.g., study design, population characteristics (including whether disproportionally affected populations have been included), exposure metrics, latency, cancer subtypes, and effect modifiers. This analysis may help explain the heterogeneity of findings (e.g., the associations between exposure to trichloroethylene and kidney cancer were stronger in the most informative studies or in the studies that estimated the highest exposure) or help contextualize the exposure (e.g., the strongest associations of night shift work and breast cancer were for long-term and frequent night shift work). Risk estimates can be visualized by stratification (for each factor), and a heatmap can be used to visualize the analysis across forest plots [see Table 3.5 in the Night Shift Work Monograph (NTP 2021a) for an example of a heatmap for night shift work and breast cancer]. In many cases, there are limited numbers of studies to stratify by multiple factors and it may be challenging to determine whether lack of consistency (if any) is due to biases or unmeasured but real effects (such as effect modifiers). Meta-analyses and meta-regression can also be used to evaluate consistency and sources of heterogeneity.

For a level-of-evidence conclusion to be reached, a positive (or negative) association needs to be replicated in more than one study or provide internal consistency within a multicenter or multicohort study; however, the precise considerations (e.g., number of studies) cannot be established, because the degree of replication may depend on the nature of the studies and the strength of the association observed. For example, findings from multicenter or multicohort studies of different populations would have greater weight than findings from a single factory or small case-control studies. In addition, weak associations may need to be replicated in more studies than strong associations (e.g., the evidence from studies on environmental tobacco smoke comes from many studies reporting relatively modest risk estimates (~20% in the meta-analysis), whereas the evidence for ortho-toluidine comes from a few studies reporting higher risk estimates (greater than fivefold) (NTP 2021b; 2021c).

Consistency of findings across studies is a method of ruling out chance as a driver of association. Although determining whether or not the statistical significance of the findings is replicated across studies may be involved, it is important to note that statistical significance is not the primary or sole indicator of consistency across studies, and in fact is not required to show consistency of findings or evidence of an association (Brownstein et al. 2019; Wasserstein et al. 2019). A full evaluation of consistency takes into consideration the factors outlined above (e.g., replication, heterogeneity, strength of the association) to determine whether or not the various findings are compatible with a consistent underlying effect (Amrhein et al. 2019).

Temporality

Exposure must occur before the disease outcome. In some studies (such as case-control studies with a cross-sectional exposure assessment), mechanistic and other relevant data may be used to inform temporality (see Viruses, NTP 2021d).

Strength of observed associations

The strength of the association, as measured by the magnitude of the effect estimate, may be difficult to evaluate across studies (in the absence of a meta-analysis), since effect estimates are likely to vary across studies for several reasons (e.g., differences in exposure conditions, outcome measurements, or populations). Although a higher magnitude may provide greater confidence that an association is not likely due to chance, bias, or confounding, this is not required in order to demonstrate causality. There are many examples of weak associations between exposure to a substance and an end point that are nevertheless considered to be causal (e.g., environmental tobacco smoke and lung cancer).

Evidence for an exposure-response gradient

As with the magnitude of an association, a positive exposure-response relationship can help rule out bias, confounding, and chance with reasonable confidence, as well as provide convincing evidence of a credible association between exposure and disease, which is important for both identified confounders and unknown confounders. Dose-response curves for established carcinogens include direct monotonic, inverse monotonic, J- or U- shaped, or plateau-shaped relationships. For example, radiation has a dose-response curve that plateaus, because it kills cells at high doses. For many occupational exposures, risks are attenuated at high doses, for a variety of reasons (Stayner et al. 2003). There may be biological or methodological reasons for not observing a gradient, and the absence of evidence for an exposure-response relationship is not strong evidence per se for the absence of a causal association. If adequate information on exposure levels (or duration) is available, exposure-response relationships can be evaluated across studies, in addition to within individual studies.

Meta-analysis and Meta-regression

A meta-analysis is a valuable tool and may be conducted in parallel with a qualitative hazard assessment to explore heterogeneity and inform the cancer hazard evaluation. However, it may be difficult, and sometimes inappropriate, to combine study data from observational studies for meta-analysis, particularly studies in environmental and occupational health, because of differences in exposure definitions, exposure levels, and outcome. For example, in the NTP cancer assessment report on night shift work, a reviewed meta-analysis of night shift work and breast cancer was not considered informative for evaluating potential causality because of heterogeneity in the definition of night shift work (the exposure) across study populations.

Meta-analytic techniques and visualization tools showing point estimates and confidence intervals (such as forest plots) can be used to assess factors contributing to heterogeneity between studies and the potential for publication bias. It is important to define a priori the sources of heterogeneity that will be explored in a meta-analysis. Heterogeneity across studies is expected, as studies are conducted in populations that differ in geographical location, socioeconomic conditions, and exposure patterns (Higgins et al. 2009). The I^2 statistic is the percentage of variation across studies due to heterogeneity rather than chance. However, it is not a measure of absolute heterogeneity (i.e., does not provide the predicted range of effect sizes due to heterogeneity) (Higgins and Thompson 2002; Higgins et al. 2003). The I^2 statistic should be interpreted with expert scientific judgment. For example, hypothetically, it is possible that combining studies that report different exposure matrices but with risk estimates of similar magnitude could result in a low value for the I^2 statistic. Thus. the studies would appear to be

homogeneous but the meta-analytic estimate would be impossible to interpret as there is inherent inconsistency when combining results across different exposure matrices.

Publication bias occurs when the findings of published studies differ from those of unpublished studies; in particular, null findings may be more likely to be unpublished, whereas published studies may be more likely to report an effect. However, publication bias may be less of a concern for qualitative evaluations that rely on more informative studies. Most of the available methods for evaluating publication bias are meta-analytical techniques (such as funnel plots and "trim and fit") that may be subject to error (Macaskill et al. 2001). Interpretation of funnel plot asymmetry (visual inspection and related analysis) can be affected by several factors including methodologic quality, statistical significance, and heterogeneity.

Some examples of quantitative meta-analyses that informed the qualitative evaluation of the evidence for a hazard assessment were analyses conducted by the U.S. Environmental Protection Agency (Scott and Jinot 2011) to assess the association between trichlorethylene and kidney cancer and the IARC Working Group's analysis of welding fumes and lung cancer (Honaryar et al. 2019). Both meta-analyses were rigorously conducted by a panel of subject-matter experts, with an in-depth exploration of factors that contributed to heterogeneity. These meta-analyses transparently illustrate the qualitative hazard evaluations and led to evidence-based policy changes shortly following publication of the hazard assessment (Cherrie and Levy 2020; HSE 2019).

When properly conducted, meta-analysis is a valuable tool to explore heterogeneity and quantify an exposure-outcome relationship explored in a systematic review. (Quantifying biases is not an objective for hazard identification.) However, the potential for bias must be investigated with the same rigor in a meta-analysis as in the individually published studies. The proliferation of published meta-analyses in a very short time has resulted in the use of incorrectly extracted data, effect estimates from multiple studies with overlapping participant populations, combining studies using entirely different exposure metrics, and inclusion of studies that do not evaluate the end point of interest.

3.4. Reporting and Data Extraction

3.4.1. Data Extraction

Data (such as methods and results) from the individual studies are extracted into a web application (such as Table Builder, Shapiro et al. 2018) in a systematic manner using standardized instructions and questions. The database contains fields that are specific for the various types of extracted information (such as study population characteristics, exposure and outcome assessment, analytical methods, and results). The instructions for data extraction (questions and considerations) describe the specific type of information that should be summarized or entered into each field. The fields from the database are used to populate tables for the monograph.

Typically, for studies in which multiple updates or re-analyses have been published, the reviewer should extract data from the most recently published follow-up or update for each type of cancer included in the study. If there is overlap between study populations, the publication with the most complete or relevant follow-up of the study population usually is reported. Information

(such as exposure data or re-analyses) from relevant publications may also be included in the review if it is needed to assess the study.

Quality assurance of data extraction and database entry are accomplished by: (1) review of the data entry by an independent reviewer and (2) resolution of any discrepancies by mutual discussion with reference to the original data source.

3.4.2. Reporting

Presentation of the data and key study information is crucial in a systematic review, and an important factor in understanding the impact of the conclusions. Systematic reviews generally include these elements:

- A discussion of the key scientific issues (as defined above) for the specific exposure and outcome relationship.
- An overview of study characteristics (e.g., population characteristics, exposure assessment methods, outcomes) included in the review, even if not included in the evidence integration for bias, quality, or other reasons.
- A discussion of biases and limitations for each bias domain across studies, in addition to the rationale for the risk of bias at the study level.
- A scientific narrative of the interpretation of study findings, including a discussion of the confidence in the evidence from each study, heterogeneity across studies (not limited to the potential for biases), and the rationale for the conclusion (e.g., consideration of dose-response relationships, consistency, ruling out of chance, bias, and confounding with reasonably confidence).
- Findings from studies—reported in summary tables and graphed in forest plots.
- Preliminary level-of-evidence conclusions for cancer types of interest.

An evidence-based table captures the overall assessment and is brought forward to the overall evidence integration to reach a preliminary listing recommendation (see Section 7).

Table 3-10. Template Example for Summarizing the Assessment of Key Evidence from Human Cancer Epidemiological Studies

Exposure	Outcome	Evidence Streams	Strength and Limitations	Assessment
Substance	Cancer Type	Number and type of **human cancer studies** Cohort studies Case-control studies Pooled or meta-analyses	Summary of most influential biases (direction, magnitude, impact) across studies by study design or other relevant grouping	Consistency of findings and patterns for factors, such as exposure matrices and levels, cancer subtypes, effect modifiers

References

Amrhein V, Greenland S, McShane B. 2019. Scientists rise up against statistical significance. Nature. 567(7748):305-307. https://doi.org/10.1038/d41586-019-00857-9

Applebaum KM, Malloy EJ, Eisen EA. 2011. Left truncation, susceptibility, and bias in occupational cohort studies. Epidemiology. 22(4):599-606. https://doi.org/10.1097/EDE.0b013e31821d0879

Arroyave WD, Mehta SS, Guha N, Schwingl P, Taylor KW, Glenn B, Radke EG, Vilahur N, Carreón T, Nachman RM, et al. 2021. Challenges and recommendations on the conduct of systematic reviews of observational epidemiologic studies in environmental and occupational health. J Expo Sci Environ Epidemiol. 31(1):21-30. https://doi.org/10.1038/s41370-020-0228-0

Axelson O, Steenland K. 1988. Indirect methods of assessing the effects of tobacco use in occupational studies. Am J Ind Med. 13(1):105-118. https://doi.org/10.1002/ajim.4700130107

Aydin D, Feychting M, Schüz J, Andersen TV, Poulsen AH, Prochazka M, Klaeboe L, Kuehni CE, Tynes T, Röösli M. 2011. Impact of random and systematic recall errors and selection bias in case-control studies on mobile phone use and brain tumors in adolescents (CEFALO study). Bioelectromagnetics. 32(5):396-407. https://doi.org/10.1002/bem.20651

Blair A, Stewart P, Lubin JH, Forastiere F. 2007. Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures. Am J Ind Med. 50(3):199-207. https://doi.org/10.1002/ajim.20281

Bradford Hill A. 1965. The environment and disease: Association or causation? Proc R Soc Med. 58(5):295-300. https://doi.org/10.1177/003591576505800503

Brownstein NC, Louis TA, O'Hagan A, Pendergast J. 2019. The role of expert judgment in statistical inference and evidence-based decision-making. Am Stat. 73(1):56-68. https://doi.org/10.1080/00031305.2018.1529623

Checkoway H, Lees PSJ, Dell LD, Gentry PR, Mundt KA. 2019. Peak exposures in epidemiologic studies and cancer risks: Considerations for regulatory risk assessment. Risk Anal. 39(7):1441-1464. https://doi.org/10.1111/risa.13294

Checkoway H, Pearce NE, Kriebel D. 2004. Research methods in occupational epidemiology. 2nd ed. New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195092424.001.0001

Cherrie JW, Levy L. 2020. Managing occupational exposure to welding fumes: New evidence suggests a more precautionary approach is needed. Ann Work Expo Health. 64(1):1-4. https://doi.org/10.1093/annweh/wxz079

Cocco P, Vermeulen R, Flore V, Nonne T, Campagna M, Purdue M, Blair A, Monnereau A, Orsi L, Clavel J, et al. 2013. Occupational exposure to trichloroethylene and risk of non-Hodgkin lymphoma and its major subtypes: A pooled InterLymph analysis. Occup Environ Med. 70(11):795-802. https://doi.org/10.1136/oemed-2013-101551

Columbia Mailman School of Public Health. 2024. Probabilistic sensitivity analysis of misclassification. New York, NY: Columbia University. [Accessed: May 9, 2024]. https://www.publichealth.columbia.edu/research/population-health-methods/probabilistic-sensitivity-analysis-misclassification

Cooper GS, Lunn RM, Ågerstrand M, Glenn BS, Kraft AD, Luke AM, Ratcliffe JM. 2016. Study sensitivity: Evaluating the ability to detect effects in systematic reviews of chemical exposures. Environ Int. 92-93:605-610. https://doi.org/10.1016/j.envint.2016.03.017

Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. 1977. Bias due to misclassification in the estimation of relative risk. Am J Epidemiol. 105(5):488-495. https://doi.org/10.1093/oxfordjournals.aje.a112408

D'Aloisio AA, Nichols HB, Hodgson ME, Deming-Halverson SL, Sandler DP. 2017. Validity of self-reported breast cancer characteristics in a nationwide cohort of women with a family history of breast cancer. BMC Cancer. 17(1):692. https://doi.org/10.1186/s12885-017-3686-6

dos Santos Silva I. 1999. Cancer epidemiology: Principles and methods. Lyon, France: International Agency for Research on Cancer. https://publications.iarc.fr/Non-Series-Publications/Other-Non-Series-Publications/Cancer-Epidemiology-Principles-And-Methods-1999

Flanders WD. 1992. Inaccuracies of death certificate information. Epidemiology. 3(1):3-5. https://doi.org/10.1097/00001648-199201000-00002

Flegal KM, Brownie C, Haas JD. 1986. The effects of exposure misclassification on estimates of relative risk. Am J Epidemiol. 123(4):736-751. https://doi.org/10.1093/oxfordjournals.aje.a114294

Garshick E, Laden F, Hart JE, Davis ME, Eisen EA, Smith TJ. 2012. Lung cancer and elemental carbon exposure in trucking industry workers. Environ Health Perspect. 120(9):1301-1306. https://doi.org/10.1289/ehp.1204989

Greenland S. 1996. Basic methods for sensitivity analysis of biases. Int J Epidemiol. 25(6):1107-1116. https://academic.oup.com/ije/article/25/6/1107/672425

Health and Safety Executive (HSE). 2019. Change in enforcement expectations for mild steel welding fume. Merseyside, UK: Health and Safety Executive. https://www.hse.gov.uk/safetybulletins/mild-steel-welding-fume.htm

Hernán MA, Hernández-Díaz S, Robins JM. 2004. A structural approach to selection bias. Epidemiology. 15(5):615-625.

Higgins JPT, Thompson SG. 2002. Quantifying heterogeneity in a meta-analysis. Stat Med. 21(11):1539-1558. https://doi.org/10.1002/sim.1186

Higgins JPT, Thompson SG, Deeks JJ, Altman DG. 2003. Measuring inconsistency in meta-analyses. BMJ. 327(7414):557-560. https://doi.org/10.1136/bmj.327.7414.557

Higgins JPT, Thompson SG, Spiegelhalter DJ. 2009. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A Stat Soc. 172(1):137-159. https://doi.org/10.1111/j.1467-985X.2008.00552.x

Honaryar MK, Lunn RM, Luce D, Ahrens W, 't Mannetje A, Hansen J, Bouaoun L, Loomis D, Byrnes G, Vilahur N, et al. 2019. Welding fumes and lung cancer: A meta-analysis of case-control and cohort studies. Occup Environ Med. 76(6):422-431. https://doi.org/10.1136/oemed-2018-105447

Howard BE, Phillips J, Tandon A, Maharana A, Elmore R, Mav D, Sedykh A, Thayer K, Merrick BA, Walker V, et al. 2020. SWIFT-Active Screener: Accelerated document screening through active learning and integrated recall estimation. Environ Int. 138:105623. https://doi.org/10.1016/j.envint.2020.105623

International Agency for Research on Cancer (IARC). 2004. Some drinking-water disinfectants and contaminants, including arsenic. (IARC monographs on the evaluation of carcinogenic risks to humans; vol. 84). Lyon, France: International Agency for Research on Cancer. https://publications.iarc.fr/102

International Agency for Research on Cancer (IARC). 2012. Arsenic, metals, fibres, and dusts. (IARC monographs on the evaluation of carcinogenic risks to humans; vol. 100C). Lyon, France: International Agency for Research on Cancer. https://publications.iarc.fr/120

International Agency for Research on Cancer (IARC). 2018. Drinking coffee, mate, and very hot beverages. (IARC monographs on the evaluation of carcinogenic risks to humans; vol. 116). Lyon, France: International Agency for Research on Cancer. https://publications.iarc.fr/566

International Agency for Research on Cancer (IARC). 2019. Preamble to the IARC monographs. Lyon, France: International Agency for Research on Cancer. https://monographs.iarc.who.int/iarc-monographs-preamble-preamble-to-the-iarc-monographs/

International Agency for Research on Cancer (IARC). 2021. Instructions for authors for the preparation of drafts for IARC monographs. Lyon, France: International Agency for Research on Cancer. https://monographs.iarc.who.int/wp-content/uploads/2018/06/2021-Instructions-for-Authors.pdf

International Agency for Research on Cancer (IARC). 2022. Cancer today [GLOBOCAN]. Lyon, France: International Agency for Research on Cancer. https://gco.iarc.fr/today/home

International Agency for Research on Cancer (IARC). 2024. Statistical methods in cancer research volume V: Bias assessment in case-control and cohort studies for hazard identification. Berrington de González A, Richardson DB, Schubauer-Berigan MK, editors. Lyon, France: International Agency for Research on Cancer. https://publications.iarc.who.int/634

Jager KJ, Zoccali C, MacLeod A, Dekker FW. 2008. Confounding: What it is and how to deal with it. Kidney Int. 73(3):256-260. https://doi.org/10.1038/sj.ki.5002650

James MK, Miller ME, Anderson RT, Worley AS, Longino CF Jr. 1997. Benefits of linkage to the National Death Index in the Longitudinal Study of Aging. J Aging Health. 9(3):298-315. https://doi.org/10.1177/089826439700900302

Jewkes R, Wood K. 1998. Competing discourses of vital registration and personhood: Perspectives from rural South Africa. Soc Sci Med. 46(8):1043-1056. https://doi.org/10.1016/s0277-9536(97)10036-3

Kleinbaum DG, Kupper LL, Morgenstern H. 1982. Epidemiologic research: Principles and quantitative methods. New York, NY: Van Nostrand Reinhold.

Kristensen P. 1992. Bias from nondifferential but dependent misclassification of exposure and outcome. Epidemiology. 3(3):210-215. https://doi.org/10.1097/00001648-199205000-00005

Kristman V, Manno M, Côté P. 2004. Loss to follow-up in cohort studies: How much is too much? Eur J Epidemiol. 19(8):751-760. https://doi.org/10.1023/b:ejep.0000036568.02655.f8

Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. 2014. Good practices for quantitative bias analysis. Int J Epidemiol. 43(6):1969-1985. https://doi.org/10.1093/ije/dyu149

Lawlor DA, Tilling K, Davey Smith G. 2016. Triangulation in aetiological epidemiology. Int J Epidemiol. 45(6):1866-1886. https://doi.org/10.1093/ije/dyw314

Linet MS, Schubauer-Berigan MK, Berrington de González A. 2020. Outcome assessment in epidemiological studies of low-dose radiation exposure and cancer risks: Sources, level of ascertainment, and misclassification. J Natl Cancer Inst Monogr. 2020(56):154-175. https://doi.org/10.1093/jncimonographs/lgaa007

Macaskill P, Walter SD, Irwig L. 2001. A comparison of methods to detect publication bias in meta-analysis. Stat Med. 20(4):641-654. https://doi.org/10.1002/sim.698

MacLehose RF, Ahern TP, Lash TL, Poole C, Greenland S. 2021. The importance of making assumptions in bias analysis. Epidemiology. 32(5):617-624. https://doi.org/10.1097/ede.0000000000001381

Marjerrison N, Jakobsen J, Demers PA, Grimsrud TK, Hansen J, Martinsen JI, Nordby KC, Veierød MB, Kjærheim K. 2022. Comparison of cancer incidence and mortality in the Norwegian Fire Departments Cohort, 1960-2018. Occup Environ Med. 79(11):736-743. https://doi.org/10.1136/oemed-2022-108331

National Cancer Institute (NCI). 2022. Surveillance, Epidemiology, and End Results (SEER) Program. Bethesda, MD: National Institutes of Health, National Cancer Institute. https://seer.cancer.gov/

National Center for Health Statistics (NCHS). 2013. National Death Index: User's guide. Hyattsville, MD: Centers for Disease Control and Prevention, National Center for Health Statistics. [Accessed: April 9, 2025]. https://stacks.cdc.gov/view/cdc/21958

National Toxicology Program (NTP). 2015. Report on Carcinogens monograph on trichloroethylene. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. RoC Monograph 05. https://doi.org/10.22427/ROC-MGRAPH-5

National Toxicology Program (NTP). 2021a. NTP cancer hazard assessment report on night shift work and light at night. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. https://doi.org/10.22427/NTP-CHR-NSWLAN

National Toxicology Program (NTP). 2021b. Substance profile: *o*-Toluidine. In: Report on Carcinogens. 15th ed. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. https://doi.org/10.22427/ntp-other-1003

National Toxicology Program (NTP). 2021c. Substance profile: Tobacco-related exposures: Environmental tobacco smoke. In: Report on Carcinogens. 15th ed. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. https://doi.org/10.22427/ntp-other-1003

National Toxicology Program (NTP). 2021d. Substance profile: Viruses, eight listings. In: Report on Carcinogens. 15th ed. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. https://doi.org/10.22427/ntp-other-1003

National Toxicology Program (NTP). 2023. 15th RoC Dashboard. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. [Last updated: March 7, 2023]. https://ntp.niehs.nih.gov/go/RoC15data

Nohr EA, Liew Z. 2018. How to investigate and adjust for selection bias in cohort studies. Acta Obstet Gynecol Scand. 97(4):407-416. https://doi.org/10.1111/aogs.13319

Patel KV, Eschbach K, Ray LA, Markides KS. 2004. Evaluation of mortality data for older Mexican Americans: Implications for the Hispanic paradox. Am J Epidemiol. 159(7):707-715. https://doi.org/10.1093/aje/kwh089

Pearce N, Checkoway H, Kriebel D. 2007. Bias in occupational epidemiology studies. Occup Environ Med. 64(8):562-568. https://doi.org/10.1136/oem.2006.026690

Pearce N, Richiardi L. 2014. Commentary: Three worlds collide: Berkson's bias, selection bias and collider bias. Int J Epidemiol. 43(2):521-524. https://doi.org/10.1093/ije/dyu025

Picciotto S, Brown DM, Chevrier J, Eisen EA. 2013. Healthy worker survivor bias: Implications of truncating follow-up at employment termination. Occup Environ Med. 70(10):736-742. https://doi.org/10.1136/oemed-2012-101332

Rothman KJ, Greenland S. 2005. Causation and causal inference in epidemiology. Am J Public Health. 95 Suppl 1:S144-S150. https://doi.org/10.2105/ajph.2004.059204

Rothman KJ, Greenland S, Lash TL. 2008. Modern epidemiology. 3rd ed. Philadelphia, PA: Lippincott, Williams, and Wilkins.

Samet JM, Chiu WA, Cogliano V, Jinot J, Kriebel D, Lunn RM, Beland FA, Bero L, Browne P, Fritschi L, et al. 2020. The IARC monographs: Updated procedures for modern and transparent evidence synthesis in cancer hazard identification. J Natl Cancer Inst. 112(1):30-37. https://doi.org/10.1093/jnci/djz169

Savitz DA, Wellenius GA. 2023. Can cross-sectional studies contribute to causal inference? It depends. Am J Epidemiol. 192(4):514-516. https://doi.org/10.1093/aje/kwac037

Schwartzbaum J, Ahlbom A, Feychting M. 2007. Cohort study of cancer risk among male and female shift workers. Scand J Work Environ Health. 33(5):336-343. https://doi.org/10.5271/sjweh.1150

Scott CS, Jinot J. 2011. Trichloroethylene and cancer: Systematic and quantitative review of epidemiologic evidence for identifying hazards. Int J Environ Res Public Health. 8(11):4238-4272. https://doi.org/10.3390/ijerph8114238

Shapiro AJ, Antoni S, Guyton KZ, Lunn RM, Loomis D, Rusyn I, Jahnke GD, Schwingl PJ, Mehta SS, Addington J, et al. 2018. Software tools to facilitate systematic review used for cancer hazard identification. Environ Health Perspect. 126(10):104501. https://doi.org/10.1289/EHP4224

Siddiqui AH, Zafar SN. 2018. Global availability of cancer registry data. J Glob Oncol. 4:1-3. https://doi.org/10.1200/JGO.18.00116

Smith MT, Guyton KZ, Gibbons CF, Fritz JM, Portier CJ, Rusyn I, DeMarini DM, Caldwell JC, Kavlock RJ, Lambert PF, et al. 2016. Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. Environ Health Perspect. 124(6):713-721. https://doi.org/10.1289/ehp.1509912

Smith TJ, Kriebel D. 2010. A biologic approach to environmental assessment and epidemiology. New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195141566.001.0001

Snoep JD, Morabia A, Hernández-Díaz S, Hernán MA, Vandenbroucke JP. 2014. Commentary: A structural approach to Berkson's fallacy and a guide to a history of opinions about it. Int J Epidemiol. 43(2):515-521. https://doi.org/10.1093/ije/dyu026

Soskolne CL, Kramer S, Ramos-Bonilla JP, Mandrioli D, Sass J, Gochfeld M, Cranor CF, Advani S, Bero LA. 2021. Toolkit for detecting misused epidemiological methods. Environ Health. 20(1):90. https://doi.org/10.1186/s12940-021-00771-6

Stayner L, Steenland K, Dosemeci M, Hertz-Picciotto I. 2003. Attenuation of exposure-response curves in occupational cohort studies at high exposure levels. Scand J Work Environ Health. 29(4):317-324. https://doi.org/10.5271/sjweh.737

Steenland K, Schubauer-Berigan MK, Vermeulen R, Lunn RM, Straif K, Zahm S, Stewart P, Arroyave WD, Mehta SS, Pearce N. 2020. Risk of bias assessments and evidence syntheses for observational epidemiologic studies of environmental and occupational exposures: Strengths and limitations. Environ Health Perspect. 128(9):95002. https://doi.org/10.1289/EHP6980

Sterne JAC, Higgins JPT, Reeves BC, on behalf of the development group for ACROBAT-NRSI. 2014. A Cochrane risk of bias assessment tool: For non-randomized studies of interventions (ACROBAT-NRSI), version 1.0.0, 24 September 2014. Bristol, UK: University of Bristol. https://www.bristol.ac.uk/population-health-sciences/centres/cresyda/barr/riskofbias/robins-i/acrobat-nrsi/

Sterne JAC, Tilling K. 2002. G-estimation of causal effects, allowing for time-varying confounding. Stata J. 2(2):164-182. https://doi.org/10.22004/ag.econ.115959

Torre LA, Siegel RL, Ward EM, Jemal A. 2016. Global cancer incidence and mortality rates and trends—An update. Cancer Epidemiol Biomarkers Prev. 25(1):16-27. https://doi.org/10.1158/1055-9965.EPI-15-0578

U.S. Environmental Protection Agency (USEPA). 2022. ORD staff handbook for developing IRIS assessments (2022). Washington, DC: U.S. Environmental Protection Agency, Office of Research and Development. EPA Report No. EPA/600/R-22/268. https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=356370

Vandenbroucke JP, Broadbent A, Pearce N. 2016. Causality and causal inference in epidemiology: The need for a pluralistic approach. Int J Epidemiol. 45(6):1776-1786. https://doi.org/10.1093/ije/dyv341

Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M, STROBE Initiative. 2007. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. PLoS Med. 4(10):e297. https://doi.org/10.1371/journal.pmed.0040297

VanderWeele TJ, Ding P. 2017. Sensitivity analysis in observational research: Introducing the E-value. Ann Intern Med. 167(4):268-274. https://doi.org/10.7326/M16-2607

von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. 2007. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. Lancet. 370(9596):1453-1457. https://doi.org/10.1016/s0140-6736(07)61602-x

Vrijheid M, Deltour I, Krewski D, Sanchez M, Cardis E. 2006. The effects of recall errors and of selection bias in epidemiologic studies of mobile phone use and cancer risk. J Expo Sci Environ Epidemiol. 16(4):371-384. https://doi.org/10.1038/sj.jes.7500509

Wasserstein RL, Schirm AL, Lazar NA. 2019. Moving to a world beyond "p < 0.05". Am Stat. 73 Suppl 1:1-19. https://doi.org/10.1080/00031305.2019.1583913

Yland JJ, Wesselink AK, Lash TL, Fox MP. 2022. Misconceptions about the direction of bias from nondifferential misclassification. Am J Epidemiol. 191(8):1485-1495. https://doi.org/10.1093/aje/kwac035

4. Evaluation of Cancer Studies in Experimental Animals

Summary

The overall cancer hazard evaluation consists of several steps: (1) scoping and mapping the literature to determine the evaluation framework and protocol development, (2) evaluating study informativeness, (3) interpreting the studies and integrating the evidence across studies, and (4) applying the Report on Carcinogens (RoC) listing criteria to reach a conclusion about the level of evidence for carcinogenicity from studies in experimental animals. The study informativeness assessment questions relate to the potential for bias (internal validity) and to study sensitivity, which is the study's ability to detect a true effect. The overall assessment integrates the responses to the questions across domains (e.g., study design, exposure conditions). This integration allows for an overall study interpretation, which considers the findings and the potential for bias. Evidence integration considers relevance to humans (external validity), conclusions about exposure-related effects for observed cancerous tumors, and the level of evidence for a specific type of cancer across studies. The RoC listing criteria for sufficiency of evidence for a cancer outcome is applied to these results.

Introduction and Objective

This section describes the systematic review and evidence integration methods for assessing the level of evidence for the carcinogenicity of a substance (agent, substance, mixture, or exposure circumstance or scenario) from experimental animal studies. Although this section focuses on cancer studies, the study informativeness evaluation questions and guidelines also apply to experimental studies on cancer mechanisms (Section 6 describes the methods for evaluating mechanistic data).

This handbook describes general methods common to all evaluations. In addition, specific protocols will be developed that adapt these methods, identify scientific issues, and develop considerations specifically for each substance. The methods have been updated from the 2015 edition of the handbook; the original methods were informed by input from NTP/NIEHS toxicologists.

The key scientific questions and the major steps in the cancer hazard evaluation are listed below. Subsequent sections provide detailed methods for conducting the evaluation.

Key Questions for Cancer Hazard Evaluations

Primary Question

- What is the level of evidence (i.e., sufficient or not sufficient) for the carcinogenicity of the substance from studies in experimental animals? (See Section 4.3.)

Secondary Questions

- Which experimental animal cancer studies should be included in the review?
- What are key issues for evaluation of the studies?
- What are the methodological strengths and limitations of these studies?

- What are the target tissue sites?
- Are there external validity concerns about the route of administration, human-relevant mechanism of action or other reasons?

Process and Components of the Cancer Hazard Assessment

- Develop the cancer hazard framework for experimental animal studies (Section 4.1).
  - Develop the literature search strategy: search, identify, select, and map literature based on the initial Model, Exposure, Comparison group, and Outcome (MECO) statement and scoping activities.
  - Develop the protocol.
- Evaluate the informativeness of the experimental animal studies (Section 4.2).
- Integrate the evidence and reach health hazard conclusions from the animal cancer studies (Section 4.3).

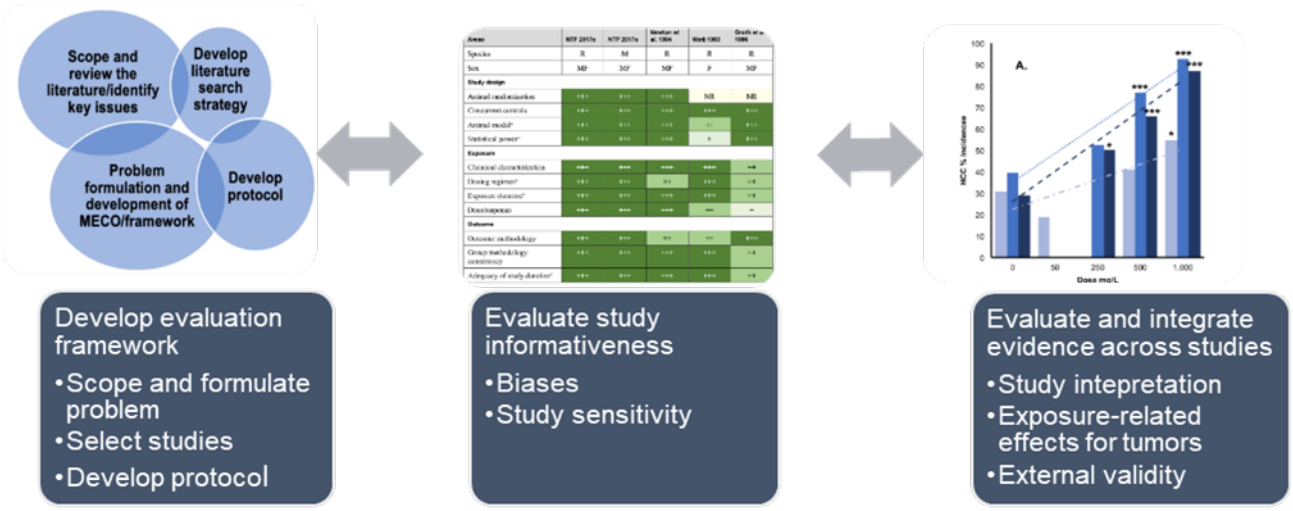


Figure 4-1. Components of the Assessment of Animal Cancer Studies for the Cancer Hazard Evaluation

The first step in the evaluation is to develop a framework for the review. This involves scoping and mapping the literature. Studies are included based on the MECO statement. Next, through a structured approach using questions and guidelines, the included studies are evaluated for their ability to inform the cancer evaluation based on the bias assessment (internal validity) and study sensitivity (ability to detect a true effect). The last step involves evaluating the individual studies and integrating the evidence across studies. Evidence integration considers relevance to humans (external validity), conclusions about exposure-related effects for observed cancerous tumors, and the level of evidence for a specific type of cancer across studies.

4.1. Cancer Hazard Framework Development

Conducting a cancer hazard assessment for a substance begins with scoping and problem formulation activities to develop the research questions and MECO statement (similar to PECO, with "population" replaced by animal "model"), literature search strategies, and protocol for the health hazard evaluation of a specific exposure. Whereas Section 1 discusses scoping and problem formulation activities for the overall cancer hazard evaluation, this section focuses on methods related specifically to the review of experimental animal studies. The process is

necessarily iterative (i.e., there may be several cycles of literature searches and evidence mapping).

4.1.1. Literature Search Strategy, Study Selection, and Evidence Mapping

The cancer evaluation component of the draft monograph evaluates all the relevant cancer studies in experimental animals that have been exposed to a specific substance. As per the RoC process, studies must be peer reviewed and publicly available. Primary studies included in review articles (such as authoritative sources) may be included if there is adequate information to evaluate them. The general approach to identifying and selecting relevant literature is discussed in the search string document; this section discusses the literature search strategy, inclusion/exclusion criteria, and evidence-mapping procedures that are specific to cancer studies in experimental animals.

Searches are conducted in PubMed and at least one other bibliographic database (such as Scopus or Web of Science) using search terms for the substance combined with search terms related to cancer and experimental animal studies. (See Figure 4-1 for examples of search concepts.) Search terms for the substance may be chemical synonyms, which usually are identified from National Library of Medicine databases (e.g., ChemIDplus, PubChem). Relevant literature is also identified from authoritative reviews (e.g., IARC monographs), government websites (e.g., PubChem, CompTox Chemicals Dashboard), other databases (see Appendix B), and citations in retrieved studies. Searches specific for the substance will be developed in the protocol for that substance.

Table 4-1. Examples of Concepts Used in Searches for Cancer Studies in Experimental Animals

PubMed, Scopus, and Web of Science		MeSH Terms Used in PubMed	
Animal Terms	**Cancer Terms**	**Animal Terms**	**Cancer Terms**
Animal	Cancer	Models, animal	Neoplasms
Mouse	Neoplasm	Animal	Carcinogens
Mice	Carcinogens	Experimentation	
Rats	Malignancy	Animal, laboratory	
Hamsters	Oncogene		
“Guinea pig”	Tumor		

Citations retrieved from literature searches are uploaded to web-based systematic review software, such as Health Assessment Workspace Collaborative (HAWC), and screened by two reviewers using predefined inclusion/exclusion criteria (Box 4-1). Studies meeting these criteria often include traditional cancer bioassays and studies in genetically modified animals that are prone to cancer development (e.g., *Tp53* mouse, *RasH2* mouse). In addition, subchronic toxicity studies may be included if there is evidence of cancer or of lesions considered part of a morphologic continuum to neoplasia. In general, subchronic or chronic toxicology (conducted in traditional animal models) studies with durations of less than one year (for rats and mice) with no neoplastic outcomes are excluded from further consideration unless they are in sensitive models (e.g., newborn animals). Studies with no concurrent control group may be excluded from further consideration on a case-by-case basis.

Box 4-1. Selection Criteria for Animal Cancer Studies

Studies are initially included if they meet the MECO:

- Model: Conducted in a relevant animal model
- Exposure: Test the appropriate exposures
- Comparison group: Have appropriate controls (unexposed)
- Outcome: Measure neoplastic endpoints

Supporting studies

- Have non-cancer data that is informative for a cancer assessment, such as reporting preneoplastic lesions
- Describe non-neoplastic lesions that are considered part of a morphologic continuum to neoplasia

Studies meeting the inclusion criteria can be tagged (supplemented by limited data extraction) according to cancer type, species, exposure route, study design, or other relevant issues. These tags (or data extraction) can be used to create interactive systematic evidence map visualizations of the available animal cancer data using visualization programs such as Tableau. Evidence maps inform or refine the MECO statement, protocol, and reporting of the studies. Final selection of studies for inclusion is based on the refined MECO statement.

4.1.2. Protocol Development

The protocol for each substance assessment includes a section for the systematic review of animal cancer studies. It consists of the following sections: (1) developing the framework, which provides information on the objectives, identifying and mapping the evidence, the MECO statement, and substance-specific scientific issues, (2) detailed methods for evaluating study informativeness, such as the potential for exposure and outcome misclassification, confounding, and other potential biases that may be important in evaluating the findings for the hazard evaluation and (3) methods for evaluating and integrating the evidence across studies. Information on data extraction and roles of the evaluation team may also be included. Therefore, protocol development requires background research on the substance and its properties, a basic understanding of the types of studies available, and identification of the key issues and questions to be addressed. A protocol is written after an initial review and selection of the literature.

4.2. Study Informativeness Assessment

This section describes the assessment of the informativeness of the individual studies, including the steps in the process; the responses for each step; signaling questions to evaluate study informativeness, including internal validity (i.e., biases) and study sensitivity; and the overall study judgment to inform the cancer hazard evaluation. Study evaluations are reported in tabular and text format (see Section 4.4).

4.2.1. Domain-based Approach

Each primary study is systematically evaluated for its informativeness by two independent reviewers using a series of signaling questions related to the following study performance domains: study design, exposure conditions, outcome assessment, potential confounding, and analysis. Questions about sensitivity and biases apply to both study design and exposure conditions. Guidance for answering the signaling questions is discussed in Section 4.2.2. These questions highlight concerns toxicologists usually consider when evaluating study informativeness; they are used to increase transparency but are not meant to be a checklist.

Box 4-2. Bias Concern Judgements

***No or minial concern*:** The study design or methodologies are ideal or very close to the ideal study characteristics, and potential bias in unlikely or minor. These studies generally are considered informative for the cancer hazard evaluation.

***Some concern*:** The study design or methodologies indicate a possible low-to-moderate concern for bias. These studies generally are considered informative for the cancer hazard evaluation.

***Moderate or major concern*:** The study design or methodologies suggest a large potential for a specific type of bias. Depending on the direction and distortion of the potential bias, the study still may be informative for cancer hazard evaluation but should be viewed with caution.

***Critical concern*:** The distortion resulting from bias likely makes the study findings unreliable for cancer hazard identification. In most cases, concerns for study sensitivity (unless severe) would not result in excluding the study from the assessment. This category is rare.

Direction of bias:

- ↑Away from the null, or overestimation of the effect.
- ↓Toward the null, or underestimation of the effect.
- Not known (unable to determine).

Magnitude of bias: Minimal, moderate or major, or unknown. In most cases, the magnitude will be unknown or subjective, but, where available, it will be based on the analysis.

The potential for a given bias in a study does not necessarily or automatically mean that the findings of the study should be disregarded. When adequate information is available, the direction of the bias (away or toward the null) and magnitude of the bias should be considered. For example, if the administered substance contained carcinogenic impurities, the degree of the tumor response attributed to the impurity could be calculated (e.g., based on the percent and potency of the impurity). In evaluating whether there is a potential bias or limitation, reviewers provide their judgments by comparing the study elements with those of an *ideal* study for a specific end point (see Box 4-2 for bias concern judgments). *Ideal* study elements pose low-to-minimal concern about potential bias and provide sufficient sensitivity to detect an effect if present. In some cases, a judgment may not be possible, because of the complexity of the issues, and the study informativeness evaluation will be discussed in narrative text.

Differences in reviewer rankings are resolved through mutual discussion with reference to the original data source. A small subset of studies may be used in a "pilot" phase, so that any ambiguity can be discussed and resolved before evaluation of the full set of studies proceeds. If the information needed to evaluate a signaling question is inadequate, the study authors may be contacted.

4.2.2. Study Informativeness Evaluation Questions and Guidance

The study evaluation is used to assess the informativeness of the studies and to inform the interpretation of the study findings (see Section 4.2.2). Signaling questions (e.g., questions to aid the reviewer in the bias evaluation) and considerations for each of the different types of bias and for sensitivity are listed below. Some study elements may overlap between different domains. Study assessments may indicate that a study should not be carried forward to the cancer hazard evaluation, or they may be used to indicate, across the body of evidence, which findings are more informative than others in the hazard evaluation.

Study Design

The study design domain evaluates two questions on biases in the study and one question on the study's sensitivity (Table 4-2). Bias assessment includes questions on randomization and controls. Concurrent controls are the most relevant comparison group for evaluating potential exposure-related tumor effects. Evaluation of study sensitivity integrates study model, statistical power, and study duration.

Table 4-2. Study Design: Questions and Responses

Signaling Questions[a]	Guidance	Response Options[b]
Bias Questions		
Randomization		
Is there concern that the methods by which animals were randomized to groups were inadequate?	Ideally, the randomization method was reported and was based on ensuring that all animals had an equal probability of being assigned to any given control or experimental group.	**No or minor concern** Animals were adequately randomized to control and experimental groups. **Critical concern** There is evidence that animals were not randomized to control and experimental groups and there is convincing evidence that this biases the findings.
Controls		
Is there concern that the concurrent control group was not adequate for evaluating effects across treatment groups? • If no concurrent controls were used, were historical controls reported that could be used in place of concurrent controls?	Concurrent controls are considered to be the most relevant comparison group for evaluating potential exposure-related tumor effects. Ideally, the concurrent control group included at least as many animals as did each treatment group. The absence of an appropriate control group, by itself, may be sufficient for judging a study inadequate for the cancer hazard evaluation. However, in some cases, historical controls of the same animal strain/stock and from the same laboratory may serve in place of concurrent controls. The experimental design of some studies evaluating co-carcinogens may not include untreated (or vehicle) concurrent controls, but generally include single-carcinogen or positive controls, which can result in acceptable study quality.	**No or minor concern** Controls were treated as similarly as possible to the exposed animals but without exposure to the test substance (e.g., appropriate vehicle controls were used). **Critical concern** No concurrent or relevant historical controls (that could be used in place of concurrent controls) were available.

Signaling Questions[a]	Guidance	Response Options[b]
Sensitivity Question		
Is there concern that the study design (i.e., animal model, number of animals/dose group and control group, and study duration) was sensitive enough to adequately detect a neoplastic effect if present? This question considers these factors: • Animal model • Statistical power (number of animals/group) • Study duration	The sensitivity rating integrates the animal model, statistical power, and study design. In some cases, one factor may compensate for limitations in another factor (e.g., a short study duration may be compensated for by a highly sensitive animal model that develops tumors within that duration). The study should use an animal model that is sensitive for detecting tumors (e.g., the background tumor rates for the tumor type are known, and the animal is sensitive to effects via the exposure route). Studies in both sexes are more informative, because a single-sex study may miss cancers that are sex-specific. Outcomes should be measured after an appropriate latency period. Although rodent cancer studies, in general, need to last at least one year, there are exceptions that depend on the animal model and study design. Carcinogenicity studies in transgenic animals may need less than one year for tumor development. Adequate statistical power to detect an effect is based on sufficient numbers of animals in each treatment group surviving to the end of the study. This is particularly important when the incidence of induced tumors is low. The study duration and the rarity of the induced tumor in the animal model may be taken into consideration.	**No or minor concern** The study used an appropriate animal model with a sufficient number of animals and an appropriate study duration. **Major to critical concern** The study used an inappropriate animal model, or too few animals per group, or an insufficient study duration.

[a]For experimental animal cancer studies, a rating response is given to each signaling question (integrating the response from any follow-up questions). Elaborations are meant to add clarification to the signaling questions.
[b]Considerations for responses for other rating categories (e.g., "some" or "major") may be defined in the protocol for the substance(s) under evaluation. Rationales are provided for all ratings. In general, critical concerns apply only to bias and to a few sensitivity questions that would exclude the study from review.

Exposure Conditions

The signaling questions in the exposure domain include one question that addresses the bias and one question on sensitivity (Table 4-3).

The bias question assesses the dose level, and the sensitivity question integrates information related to dose selection and exposure duration. Dose selection is considered as both a bias issue and a sensitivity issue. Aspects of exposure conditions that are specific to the candidate substance are defined in the protocol.

Table 4-3. Exposure: Questions and Responses

Signaling Questions[a]	Guidance	Response Options[b]
Bias Questions		
Dose selection		
Is there concern that the dose level was too high (e.g., exceeded the maximum tolerated dose)?	Ideally, the authors should state their rationale for dose selection. For NTP chronic studies, dose selection is typically based on subchronic (90-day studies), and the high dose should not cause excess toxicity, for the duration of the study.	**No or minor concern** Minimal treatment-related survival effects were seen (other than mortality related to tumors). Tumors at the high dose were the result of a specific treatment-related effect. **Major or critical concern** Severe toxicity was seen in all treatment groups. Toxicity was so high that survival or body weight was greatly reduced. (Reduced survival due to tumors is not a concern.)
Sensitivity Questions		
Is there concern that the conditions of exposure to the test agent did not provide sufficient sensitivity to adequately detect a neoplastic effect, if present?	The sensitivity rating integrates considerations of dose duration and selection. The selection of the dose may depend on the exposure duration. Ideally, exposure would last throughout or for a significant proportion of the animals' lifespan (i.e., 1 to 2 years for rodents) depending on the model and route of exposure. However, the duration depends on the study design. For example, some studies (e.g., transgenic, stop exposure, or *in utero* exposure studies), depending upon the study design or the animals' cancer susceptibility, do not require lifetime exposure. Doses should be high enough (i.e., achieving the maximum tolerable dose such as slightly decreased body weight gains or other signs of clinical toxicities) or based on dose-range finding studies. Evaluation of dose response can contribute to confidence in the study findings and allow for evaluation of potential effects at lower doses. Ideally, studies should use multiple doses; however, for the purpose of hazard identification, multiple doses are not required if the dose selection provides sufficient sensitivity.	**No or minor concern** The study included an appropriately high dose (as evident from signs of toxicity) and an adequate observation period. **Major concern** There is evidence that the combined dose level (i.e., too low) and duration (i.e., short) were not adequate to detect an effect in the animal model.

[a]For experimental animal cancer studies, a rating response is given to each signaling question (integrating the response from any follow-up questions). Elaborations are meant to add clarification to the signaling questions.
[b]Considerations for responses for other rating categories (e.g., "some" or "major") may be defined in the protocol for the substance(s) under evaluation. Rationales are provided for all ratings. In general, critical concerns apply only to bias and sensitivity questions that would exclude the study from review.

Outcome Assessment and Measurement

The outcome domain consists of one signaling question (and related follow-up question) on the adequacy of the methods to assess tumor outcome in exposed and control animals (Table 4-4). This question addresses concerns about both bias and sensitivity. Evaluation of only a few organs for tumors, instead of all organs and tissues, can limit the study's sensitivity. Although blinding generally is considered important to reduce bias in the assessment of subjective outcomes (such as behavior), nonblinding may be preferred for cancer outcomes, to determine normal background histology. The NTP uses an informed approach to histopathological evaluation in its toxicity and carcinogenicity studies (Sills et al. 2019). This principle applies to non-NTP studies, provided that the necropsy and histology methods used were adequate and consistent.

Table 4-4. Outcome: Questions and Responses

Signaling Questions[a] Follow-up Question	Guidance	Response Options[b]
Outcome		
Is there concern that the methods used to assess tumor outcome (necropsy, gross pathology, histology, or diagnosis) were not adequate to allow the effects to be attributed to the exposure? • Is there concern that not all treatment and control groups were assessed in the same way and in balanced blocks, to avoid bias?	Ideally, each study should include full gross necropsies of all tissues and histopathological examination of the majority of them. If details of the histopathological examination (e.g., cell type) are not reported, tumor type (and whether benign or malignant) should be reported. Ideally, the controls and all the treatment groups were treated the same. The control groups should be evaluated at necropsy to the same extent as the treatment groups.	**No or minor concern** Complete necropsies and gross pathology were reported for all tissues, and histopathological examination for most tissues. The control groups were treated exactly the same as the treatment groups except for the presence of the test substance. The conduct of the evaluation by the pathologists was sound. **Major concern** Pathology was assessed on only some tissues. Histopathology was not assessed in tumors. The controls were treated differently from the treatment groups.

[a]For experimental animal cancer studies, a rating response is given to each signaling question integrating the response from any follow-up questions. Elaborations are meant to add clarification to the signaling questions.
[b]Considerations for responses for other rating categories (e.g., "some" or "major") may be defined in the protocol for the substance(s) under evaluation. Rationales are provided for all ratings.

Potential for Confounding

The confounding domain consists of two signaling questions and related follow-up questions and addresses the quality of the chemical characterization and any other potential sources of confounding that could influence the study outcome other than the substance under evaluation (Table 4-5).

Table 4-5. Potential Confounding: Questions and Responses

Signaling Questions[a] Follow-up Questions	Guidance	Response Options[b]
Confounding		
Is there concern about potential confounding? • What is the relative impact of the confounding?	Sources of potential confounding in animal studies are the use of an impure chemical that contains other potential carcinogens, inadequate animal husbandry conditions, and lack of monitoring for pathogens that may be linked to cancer. Food, water, and bedding should also be monitored for potential impurities.	**No or minor concern** The study used a pure testing agent without contaminants and adequate animal husbandry conditions. The test agent is representative of the substance under evaluation. **Major or critical concern** Strong evidence of the presence of carcinogenic contaminants in the testing agent or poor animal husbandry conditions will substantially compromise interpretation of the findings, and there are no data to evaluate the extent of the confounding.
Is there concern that the characterization, dose formulations (e.g., homogeneity, purity, solubility, and stability), or delivery of the test agent (actual vs. desired dose) were not adequate to support attribution of any neoplastic effects to the substance under evaluation?	The purity of the test agent should be reported, and any contaminants listed. The test agent should be stable between making up of new stock solutions. Chemical stability, including in liquid media or feed, should be verified and be taken into account in formulation. Animals should be homogenously exposed to the agent.	The test agent is not representative of the substance under evaluation and/or contains carcinogenic contaminants at levels high enough to compromise the interpretation of the results.

[a]For experimental animal cancer studies, a rating response is given to each signaling question (integrating the response from any. follow-up questions). Elaborations are meant to add clarification to the signaling questions.
[b]Considerations for responses for other rating categories (e.g., "some" or "major") may be defined in the protocol for the substance(s) under evaluation. Rationales are provided for all ratings.

Analysis

The analysis domain evaluates statistical methods and combining of tumor incidences and consists of two bias questions (Table 4-6). These questions address the methods for grouping the outcome (i.e., tumor types) and statistical methods to evaluate the findings. If statistical analysis was not performed, but tumor incidences were reported in enough detail, NIEHS can perform pairwise statistical calculations. Trend analysis across treatment groups (e.g., Cochran-Armitage trend test) can also be performed if there are three or more dose groups. It will be noted whether statistical analyses were performed by NIEHS.

Table 4-6. Analysis: Questions and Responses

Signaling Questions[a]	Guidance	Response Options[b]
Combined tumors		
Is there concern that different types of tumors were inappropriately combined in the analysis?	Analyses of benign and malignant tumors from the same tissue type should be reported both separately and combined. Tumors of the same cellular origin, which may appear at different organ sites (as seen with metastasis), should be combined. Organs that are of the same cellular origin and part of the same organ system can be combined, such as squamous carcinomas of the upper respiratory tract (nasal cavity, pharynx [throat], larynx [voice box], and bronchi) (McConnell et al. 1986).	**No or minor concern** Tumors of the same cellular origin are reported both individually and combined in the analysis. **Major concern** Tumor types of different cellular origins are combined, or tumors are not specified whether they are benign or malignant.
Statistical analysis		
Is there concern that statistical analyses were inadequate or were not conducted to evaluate the results? • If statistical analyses were not conducted, were the results reported in sufficient detail to allow ad hoc analysis?	If statistical analyses were not reported, the study should at a minimum present incidence data for specific tumors, so that statistical tests (e.g., Fisher's exact test for pairwise comparisons) can be conducted. If there is evidence of a decreased survival effect, the studies should use adequate statistical methods, such as the poly-3 test (Bailer and Portier 1988), to control for decreased survival.	***No/minor concerns*** The study reported appropriate methods of analysis using relevant data. Analyses were adjusted for survival (e.g., poly-3 test) where relevant. ***Critical concerns*** There is strong evidence that reporting of data and analytical methods were so limited that the findings are not interpretable.

[a]For experimental animal cancer studies, a rating response is given to each signaling question. (Elaborations are meant to add clarification to the signaling questions.)
[b]Considerations for responses for other rating categories (e.g., "some" or "major") may be defined in the protocol for the substance(s) under evaluation. Rationales are provided for all ratings.

4.2.3. Overall Assessment of Study Informativeness

The overall informativeness of a study considers both bias (i.e., systematic flaws or limitations that may compromise interpretation of the results) and study sensitivity. Studies having elements with major concerns may still be considered in a cancer hazard assessment, but the findings should be interpreted with caution. It should also be noted that some concerns about a study element (such as inadequate observation and/or exposure period or statistical power) would decrease the study's sensitivity to detect an effect. If positive findings were described despite these limitations, these studies would inform a cancer hazard assessment. Studies with critical concerns about important issues generally are inadequate to inform the evaluation.

Box 4-3. Study Informativeness-level Judgment

High: no or minimal concerns about most potential biases, high or moderate sensitivity.

Moderate: low, minimal, or some concerns about most potential biases.

Low: major concerns about several biases, sensitivity rating varies. Depending on the direction and distortion of the potential biases, the study may still be informative for the cancer hazard evaluation but should be viewed with caution.

Inadequate (very rare): critical concern about any bias; sensitivity rating varies. Severe insensitivity.

If a study's information is inadequate for a reviewer to answer a specific question, the impact on overall study quality evaluation depends on the extent and importance of the missing information and is evaluated on a case-by-case basis. The study informativeness-level judgments can be found in Box 4-3. (See Section 4.4 for information on reporting data extraction and study informativeness.) These evaluations are brought forward to the evidence integration section.

4.3. Evidence Evaluation and Integration

This section outlines the approaches to integrating the evidence across studies to identify exposure-related cancer sites (Section 4.3.1), evaluating external validity (Section 4.3.2), applying the RoC listing criteria, and reaching a level-of-evidence conclusion (sufficient or not sufficient) on the carcinogenicity of the substance from experimental animal cancer studies (Section 4.3.3).

The conclusions regarding the assessment of study informativeness are carried forward to the cancer hazard evaluation, and the studies with the greatest utility to inform the cancer hazard evaluation (as described in Section 4.2) are given the most weight. All studies with low, moderate, or high informativeness are brought forward to the evidence integration, we Studies with inadequate ratings are usually not brought forward to the evaluation, but this is rare.

4.3.1. Interpretation of the Evidence from Individual Studies

The findings of each study are interpreted with respect to their limitations and strengths (identified as described in Section 4.3.1). For example, positive findings from studies receiving poor ratings for sensitivity (such as low statistical power or short duration) should not be discounted, because other factors are considered as well when determining whether an effect (e.g., increased incidence of a specific tumor type) is treatment related. The factors considered include statistical significance with respect to controls and dose-related trends, preneoplastic lesions, lesion progression, decreased latency, tumor multiplicity, tumor incidence, historical control range, animal survival, species, sex, strain, and rarity of tumor. For instance, an uncommon tumor type could be deemed treatment-related without a statistically significant increase in incidence. It is important to note that the shape of the dose-response curve may vary (i.e., may not always be monotonic), and various factors (e.g., metabolism and toxicokinetics of the substance or differences in animal survival among the treatment groups) can affect the shape of the curve (IARC 2019). In evaluating potential confounders in an individual study, one should consider the magnitude of the effect, the adequacy of the controls, and whether a potential confounder could modify effects across exposure groups.

4.3.2. External Validity

External validity addresses the extent to which conclusions from one study can be generalized to other situations (i.e., the relevance of experimental animal data to humans). Studies testing only one sex of animal may have limited validity for the human population, and further evaluation may be needed. When interpreting the relevance of experimental animal study findings consideration should be given to the route of exposure, substance disposition, and mode of action. Although the relevance to human exposure is considered, studies using exposure routes that are not common in human exposure are not usually excluded from the cancer hazard assessment; however, this issue may be addressed on a case-by-case basis and would be described in the substance evaluation protocol. Findings of tumors at a similar tissue site by different routes of exposure strengthen the evidence for carcinogenicity.

Neoplasms observed in experimental animals are considered to be relevant to humans unless there is *compelling* evidence indicating that they occur by a mechanism that does not operate in humans. In other words, it is the cancer mechanism that informs the evaluation of potential carcinogenicity in humans. For example, the occurrence of neoplasms in tissues that do not occur in humans (e.g., the rodent forestomach and rat Zymbal's gland) might be relevant to humans. In contrast, some tumors in rodents may occur by mechanisms that may not be relevant to humans, such as kidney neoplasms that occur exclusively from the production of male-rat-specific alpha2u-globulin. Mechanistic and other relevant data are evaluated in a separate section of the monograph and are one of the points to be considered in assessing the human relevance of a tumor outcome.

The following points should be considered in assessing the relevance of an experimental animal cancer study for evaluating the potential for human carcinogenicity:

- Relevance of the route of exposure.
- Relevance of the species, sex, or animals' age.
- Relevance of the mechanism of tumor formation.

4.3.3. Evidence Integration Across Animal Cancer Studies

The final steps in evaluating evidence from experimental animal cancer studies are integrating the evidence for treatment-related tumors across studies, applying the RoC listing criteria (see below), and reaching a level-of-evidence conclusion from studies in experimental animals.

RoC Listing Criteria for Evaluating Carcinogenicity from Studies in Experimental Animals

Sufficient Evidence of Carcinogenicity from Studies in Experimental Animals:
An increased incidence of malignant and/or a combination of malignant and benign tumors

- In multiple species, or
- At multiple tissue sites, or
- By multiple routes of exposure, or
- To an unusual degree with regard to incidence, site, or type of tumor or age at onset.

The first step in evidence integration is to evaluate the evidence across studies for each cancer site of interest. For most databases, heterogeneity in findings is often explained by differences in experimental conditions (e.g., species, sex, strain, doses, duration, route), and few studies have been conducted using exactly the same experimental conditions. As mentioned above, the most informative studies (highest quality and sensitivity) are given the most weight, and positive findings from these studies are considered to provide evidence of treatment-related tumor effects. Moderate- and low-quality studies can also be used in the assessment, especially when it is unlikely that biases (moderate) in the studies would cause false-positive results. Replication of findings across several studies also increases confidence in treatment-related effects.

In general, the RoC criteria for sufficient evidence of carcinogenicity from studies in experimental animals are fulfilled by: (1) two studies (by different exposure routes or in different species) reporting positive findings of malignant or combined malignant and benign tumors or (2) one study reporting positive findings at multiple tissue sites. In addition, positive findings from one robust study can fulfill the criteria if the tumors are rare, have an early onset, or have a high incidence. The spectrum of neoplastic responses, from preneoplastic lesions and benign tumors to malignant neoplasms of a specific tumor type, is relevant for the evaluation of whether benign tumors observed at increased incidences are likely to progress to malignancy.

4.4. Reporting and Data Extraction

This section provides information on reporting (e.g., related to extracting data on the studies and their findings) and documenting the key steps in the cancer hazard evaluation (i.e., documenting the framework, assessing study informativeness, and evaluating and integrating the evidence). The cancer hazard evaluation is captured in text, tabular, and graphical format.

4.4.1. Data Extraction

Data are extracted from individual animal cancer studies into a database or web application (such as Table Builder, Shapiro et al. 2018) in a systematic manner using standardized instructions and questions. The database is organized by study methods, including study design (e.g., species, strain, route), exposure methods (e.g., agent, route, dosing regimens), outcome methods (e.g., tumor incidence, animal survival), information related to evaluating confounding (e.g., animal husbandry methods, chemical purity), and statistical analyses. Extracted study results include the tissue type and histological classification, animal survival, tumor incidence, and statistical significance. If the study authors did not perform statistical analyses, NIEHS will conduct pairwise analysis of neoplasm incidence relative to control group(s) (e.g., Fisher’s exact test) and trend analysis across treatment groups (e.g., Cochran-Armitage trend test) and will note that these analyses were performed by NIEHS.

Quality assurance of data extraction and database entry is done by a reviewer independent of the data extractor. Any discrepant entries are resolved through mutual discussion with reference to the original data source. The extracted data are presented as tables in the monograph and can be downloaded into Excel for additional analyses or visualization.

4.4.2. Reporting

The methods for reporting will be determined in part by the substance(s) under review. The following points provide guidance for reporting cancer studies in experimental animals and presenting preliminary level-of-evidence conclusions for animal studies:

- An overview of study characteristics initially included in the evaluation.
- A scientific discussion of study informativeness across studies, organized by bias and study sensitivity domains.
  - Study informativeness for each study typically is presented in tabular format (e.g., downloaded from web-based software such as Table Builder). (See published RoC monographs for reporting examples.)
  - Reporting of the data for individual studies (e.g., study description and findings) in tabular format downloaded from web-based software such as Table Builder (see Box 4-4 for reporting examples from published RoC monographs).
  - Narrative discussion of the assessment across studies, organized in part by the key questions to be addressed in the protocol (e.g., at what tissue sites did cancer occur?).
  - Discussion highlighting any key issues noted for the type of exposure(s) and tumor outcomes.
  - Findings across studies are presented in tabular or graphical formats.

Box 4-4. Examples of Reporting in Report on Carcinogens Monographs

Haloacetic acids (Section 4 and Appendix C)

Antimony trioxide (Section 5 and Appendix D)

Cobalt and cobalt compounds (Section 5 and Appendix D)

The level of evidence for carcinogenicity from studies in experimental animals (sufficient, not sufficient) and rationales for the conclusions are presented in evidence-based tables (see Table 4-7 for a template and Table 7-2 for an example.)

An evidence-based table captures the overall assessment and is brought forward to the overall evidence integration to reach a preliminary listing recommendation (see Section 7).

Table 4-7. Template Example for Summarizing the Assessment of Key Evidence from Animal Studies

Exposure	Outcome	Evidence Streams	Strength and Limitations	Assessment
Substance	Cancer type or across cancers	Number and type of **animal cancer studies** Animal models (e.g., route, species)	Summary of most influential biases (direction, magnitude, impact) across studies by model, route, or other relevant grouping	Exposure-related cancer sites, common cancer sites across groups of chemicals Information relevant to evidence integration

References

Bailer AJ, Portier CJ. 1988. Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples. Biometrics. 44(2):417-431. https://doi.org/10.2307/2531856

International Agency for Research on Cancer (IARC). 2019. Preamble to the IARC monographs. Lyon, France: International Agency for Research on Cancer. https://monographs.iarc.who.int/iarc-monographs-preamble-preamble-to-the-iarc-monographs/

McConnell EE, Solleveld HA, Swenberg JA, Boorman GA. 1986. Guidelines for combining neoplasms for evaluation of rodent carcinogenesis studies. J Natl Cancer Inst. 76(2):283-289. https://doi.org/10.1093/jnci/76.2.283

Shapiro AJ, Antoni S, Guyton KZ, Lunn RM, Loomis D, Rusyn I, Jahnke GD, Schwingl PJ, Mehta SS, Addington J, et al. 2018. Software tools to facilitate systematic review used for cancer hazard identification. Environ Health Perspect. 126(10):104501. https://doi.org/10.1289/EHP4224

Sills RC, Cesta MF, Willson CJ, Brix AE, Berridge BR. 2019. National Toxicology Program position statement on informed ("nonblinded") analysis in toxicologic pathology evaluation. Toxicol Pathol. 47(7):887-890. https://doi.org/10.1177/0192623319873974

5. Evaluation of Disposition and Toxicokinetic Data

Introduction and Objective

This section describes the approach used to review disposition or ADME (absorption, distribution, metabolism, and excretion) and toxicokinetics (rate and extent of those processes) of the substance(s) under review.

The Report on Carcinogens (RoC) listing criteria note that cancer hazard conclusions (e.g., a listing recommendation for *known or reasonably anticipated to be a human carcinogen*) are based on scientific judgment, with consideration given to all relevant information. Relevant information can include metabolism, toxicokinetics, and other ADME data (Slitt 2019). ADME and toxicokinetic (defined now as ADME) data can impact the listing directly or indirectly. For example, some listed chemical classes are based on metabolism to a *known or reasonably anticipated carcinogen*. ADME can inform the interpretation of all components of the cancer hazard evaluation, such as the quality of exposure biomarkers used in biomonitoring and human cancer epidemiology studies, human relevance of animal cancer studies, and mechanisms of action. It is important to assess the likely effect of ADME and toxicokinetic data as early as possible in the overall evidence evaluation process.

Primary Objectives

- To determine whether ADME evidence is influential for reaching a listing recommendation, and if so, to assess this evidence
- To discuss how ADME informs the mechanism of action and biological plausibility or read-across approaches

Secondary Objectives

- Discuss how the substance is absorbed, distributed, metabolized, and excreted in humans and experimental animals
    - Assess whether ADME data help explain target cancer sites
- Identify the metabolites of the substance and what metabolites are considered in the carcinogenicity assessment
    - Identify the ultimate carcinogenic form—parent compound or metabolite(s)
- Discuss any interspecies differences/similarities in ADME. For example, are they qualitative or quantitative, and could they affect the interpretation of the relevance of animal data to humans?

Components of the Evaluation of ADME and Toxicokinetic Data

- Develop the framework.
    - Develop a literature search strategy, conduct the literature search, and screen, tag, and map the studies.
    - Develop approach and protocol.

- Extract data and evaluate the quality of the selected studies.
- Review and integrate the evidence across studies.

5.1. Evaluation Framework Development

The evaluation of ADME data differs from the human and animal cancer sections of cancer hazard evaluations as it often relies on data from authoritative or other contemporary reviews supplemented by relevant primary studies. The fit-for-purpose scoping activities are stepwise and iterative to determine the extent and depth of ADME knowledge, identify the importance of ADME in the cancer hazard evaluation and listing decisions (e.g., influential questions) and determine the most appropriate approach for reviewing the available literature.

First, the evaluation team searches authoritative sources and citation databases for reviews (see below for methods) on ADME information to determine the relative impact and importance of ADME and toxicokinetic data on reaching hazard identification conclusions (see Box 5-1 for guidance). These reviews also inform whether there are any issues or concerns in the ADME data that would warrant a deeper dive into the literature (e.g., interspecies differences in ADME). Reviews are assessed for their quality for inclusion, detail of reporting, and recency of reviewed studies. Primary studies cited in the reviews may be retrieved to ensure accurate reporting by the reviews.

Box 5-1. Guiding Questions to Inform Evaluation Approach

Is an adequate—recent, sufficiently detailed—authoritative review available?

How critical are the data to the overall cancer hazard listing recommendation?

- Does ADME directly affect the listing conclusion?
- What is the extent that ADME contribute to the cancer mechanisms or interpretation of the cancer studies in experimental animals?
- Is any issue for ADME likely to be an influential mechanistic question?
- What evidence types (human, animal, in vitro) are available and is there evidence of inter-species differences?

If no recent reviews of adequate quality are available, the evaluation team will search the literature for primary studies to identify critical scientific issues and influential questions. Using these scoping and literature searches, the team will determine the adequacy of the review for addressing these issues and what targeted searches are needed. Primary studies cited in the reviews may also be retrieved. The scoping activities also help focus the framework [EECO] used for searching and selecting the literature. (EECO = evidence type [e.g., humans, animals, in vitro], comparison group and outcome or endpoint), and the source (review vs. primary studies) for the EECO. On rare occasions, an RoC monograph may summarize ADME only from authoritative reviews (such as IARC, which undergo quality checks and peer review).

5.1.1. Literature Search Methods

Searches are conducted in PubMed and in at least one other database (such as Scopus or Web of Science), depending on the substance, using the concepts in Table 5-1 below (using the Boolean operator “OR”) in conjunction with terms for the substances or their metabolites (using the Boolean operator “AND”). (See the Handbook appendix on the NTP website for the ADME search strings.)

Citations retrieved from literature searches are uploaded to a web-based systematic review software application (such as Health Assessment Workspace Collaborative [HAWC]) and screened by reviewers using predefined inclusion/exclusion criteria. Studies are initially included in the evaluation if they report ADME or toxicokinetic data in humans or experimental animals for the substance under review, or in a relevant in vitro system for the substance under review.

Studies meeting the inclusion criteria can be mapped by publication type (primary versus secondary), species or sex, route of exposure, specific ADME topic(s), and type of study (in vivo versus in vitro). Primary studies will be included in the assessment to supplement gaps in reviews.

Table 5-1. Examples of Concepts Used in Searches for ADME Studies

PubMed, Scopus, and Web of Science	MeSH Terms Used in PubMed
ADME	Toxicokinetics
Bioavailability	Pharmacokinetics
Tissue-distribution	Metabolism
Bioconcentration	Activation
Metabolite	Metabolic
Excretion	Cytochrome P-450 enzyme system
Elimination	
Metabolism polymorphism	
Toxicokinetics	

Note that these are examples of search terms and not the detailed or fully developed search string used in the actual literature search.

5.1.2. Protocol

The protocol (part of the overall monograph protocol) contains the following sections: (1) framework development, which provides information on literature searches and mapping, influential and other key questions, and (2) the approach, and guidance (if relevant) on assessing study informativeness and reaching conclusions about any influential questions.

5.2. Study Informativeness Assessment

The assessment of study informativeness (e.g., bias analysis and study sensitivity) follows a fit-for-purpose approach and is consistent with the strategy for assessing mechanistic studies (see Section 6.3). The evaluation of study informativeness of primary studies depends on how impactful the literature is for answering questions that may influence the overall cancer hazard classification. ADME studies vary widely in study design and detail, and specific bias questions are not part of our handbook; however, study informativeness considerations for mechanistic studies (see Section 6.3) may also be relevant to the ADME studies. Other sources for evaluating the strengths and weaknesses in the studies include Good Laboratory Practices, OCED guidelines, scientific judgment, and consultation with technical advisors (OECD 2010). ADME data used for RoC cancer hazard evaluations are often from authoritative reviews; study informativeness assessment of individual studies in these reviews is usually not conducted unless needed to inform the ADME assessment such as when the studies report conflicting results. For example, the quality of analytical methods used to measure metabolites or physiologically based pharmacokinetic (PBPK) models may explain potential discrepancies between studies (Barton et al. 2007).

5.3. Evidence Integration Across Studies

For most evaluations, the monograph provides a narrative summary and relevant tables of ADME data in exposed humans and animals, and in vitro (see Section 5.4 for examples of ADME data). For substances with influential questions, we assess the evidence for its informativeness, (see Section 5.2) and consistency to reach conclusions regarding each question and scientific issues (see Box 5-2).

Box 5-2. ADME Topics

Identification of bioactive and carcinogenic metabolites

Extent and nature (e.g., species) of the metabolism (conclusion)

Biological plausibility (conclusion)

Other potential effects of ADME on mechanistic events, e.g., enterohepatic circulation

Biological plausibility encompasses consideration of species similarities and differences and effects of dose level and exposure routes on ADME and observed tumor sites. Studies in human-relevant (humans or human tissues and cells) and biologically relevant (in vivo studies) are considered the most informative, but data from all evidence types can be integrated into the ADME assessment. Data from exposed humans are the most relevant but are usually sparse, and many studies are limited as they measure ADME in only a few volunteers. In addition, these data may not reflect enzyme polymorphisms or other effect modifiers within a larger and more diverse population. Often metabolism studies using human cell lines or tissues help to fill this data gap, support findings from in vivo studies, and provide a valuable link in describing similarities and differences between humans and in vivo or in vitro experimental animal data, or both. Qualitative human cancer hazard assessments consider all pharmacokinetic data, including in vitro information using cells or tissue models from humans and experimental animals. Often human data are not available and integration of the body of evidence considers species differences and/or similarities to human physiology. Conclusions regarding influential questions and other ADME evidence may be integrated with evidence from other data streams and the overall cancer hazard assessment (see Table 5-2 for examples).

Table 5-2. ADME Examples from Report on Carcinogens Evaluations

Example	Evidence Stream	Issue
Antimony trioxide	Animal cancer studies	Particle overload: Antimony trioxide lung concentrations in control and tumor tissue by dose level. Lung tumors in rodents occurred at doses in which lung clearance overload did not occur and were attributed to antimony trioxide toxicity and not inert particles (NTP 2021a).
Trichlorethylene[a]	Mechanistic data	Mechanism: Glutathione conjugation (GSH) pathway results in metabolic and cytotoxic metabolites in the kidneys. This mechanism is considered for trichloroethylene-induced kidney tumors in humans and experimental animals (NTP 2021c).

Example	Evidence Stream	Issue
Diazoaminobenzene[a]	Overall evaluation	Metabolism: Diazoaminobenzene was listed based on metabolism to benzene. The ADME section provided information on the nature and extent of metabolism to benzene (e.g., the metabolic pathway of diazoaminobenzene is similar to benzene and metabolism is quantitative) (NTP 2021b).

[a]See Section 7.

5.4. Reporting

As mentioned above, this section summarizes ADME information from authoritative and/or other contemporary reviews supplemented by key studies. The ADME section is organized into subsections that address the main topics (i.e., absorption, distribution, metabolism, and excretion) or various combinations of these topics. Toxicokinetic data may be presented as a separate subsection or included within the other subsections. If metabolism is a key step in carcinogen formation for read-across methods, a comparison of ADME data for source chemicals to the target chemicals is reported (see Table 5-2).

5.4.1. Data Extraction for Selected Studies

Examples of relevant ADME and toxicokinetic data that may be extracted for the four primary topics are listed in Table 5-3; however, these lists should neither be considered exhaustive nor are they necessarily required elements in all cancer hazard assessments (Krishnan 2019). Data extraction is conducted on a case-by-case basis and may include a narrative summary, Word tables, or web applications such as Table Builder or HAWC. Each substance or substances considered in an RoC monograph will present different challenges depending, in part, on the availability of data to address the key questions listed above.

Table 5-3. Examples of ADME and Toxicokinetic Data

Topic	Types of Data (Human and Animal)	Relevant Toxicokinetic Parameters	Tables and Figures
Absorption	Data for all relevant routes (e.g., oral, inhalation, dermal) • Bioavailability • Permeability coefficients • Low-dose vs. high-dose comparisons • Other factors that affect absorption • Particle deposition in airways	C_{max} T_{max}	**Optional** As needed

Topic	Types of Data (Human and Animal)	Relevant Toxicokinetic Parameters	Tables and Figures
Distribution	General description (e.g., rapid and uniform) • Serum/plasma protein binding • Tissue burdens/storage tissues • Active/passive transport into tissues/cells • Route of exposure comparisons	Vd Blood concentration-time profiles AUC $T_{½}$ **Partition coefficients:** blood:tissue blood:plasma blood:air	**Optional** As needed
Metabolism	Metabolites • Metabolizing enzymes and tissues • Route of exposure/dose effects • Induction/inhibition • Interspecies comparisons • Age/gender differences • Polymorphisms (fast/slow metabolizer) • First pass • Saturation • Glutathione depletion	V_{max} K_m Note: PBPK models, if available, can provide information on relative importance of competing metabolic pathways	**Required** Figure(s) showing metabolism of the substance(s) including all known pathways, metabolites, and enzymes **Optional** Other tables/figures as needed
Excretion	Data for all relevant routes (e.g., urinary, biliary, fecal, exhalation, sweat) • Enterohepatic circulation • Parent vs. metabolites • Mechanisms (active or passive) • Renal tubular secretion and reabsorption • Glomerular filtration rate	Cl (total, renal, nonrenal) $T_{½}$	**Optional** As needed

AUC = area under the curve; Cl = clearance; C_{max} = peak or maximum blood concentration; K_m = Michaelis-Menton constant (chemical concentration at one-half the maximum reaction rate); PBPK = physiologically based pharmacokinetic model; $T_{½}$ = half-life; T_{max} = time to maximum blood concentration; Vd = apparent volume of distribution; V_{max} = maximum reaction rate.

References

Barton HA, Chiu WA, Setzer RW, Andersen ME, Bailer AJ, Bois FY, DeWoskin RS, Hays S, Johanson G, Jones N, et al. 2007. Characterizing uncertainty and variability in physiologically based pharmacokinetic models: State of the science and needs for research and implementation. Toxicol Sci. 99(2):395-402. https://doi.org/10.1093/toxsci/kfm100

Krishnan K. 2019. Chapter 7: Toxicokinetics. In: Klaassen CD, editor. Casarett & Doull's Toxicology: The Basic Science of Poisons. 9th ed. New York, NY: McGraw-Hill Education. https://accesspharmacy.mhmedical.com/content.aspx?bookid=2462&sectionid=202672178

National Toxicology Program (NTP). 2021a. Substance profile: Antimony trioxide. In: Report on Carcinogens. 15th ed. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. https://doi.org/10.22427/ntp-other-1003

National Toxicology Program (NTP). 2021b. Substance profile: Diazoaminobenzene. In: Report on Carcinogens. 15th ed. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. https://doi.org/10.22427/ntp-other-1003

National Toxicology Program (NTP). 2021c. Substance profile: Trichloroethylene. In: Report on Carcinogens. 15th ed. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. https://doi.org/10.22427/ntp-other-1003

Organisation for Economic Co-operation and Development (OECD). 2010. Test no. 417: Toxicokinetics. (OECD Guidelines for the Testing of Chemicals, Section 4). Paris, France: OECD Publishing. https://doi.org/10.1787/9789264070882-en

Slitt AL. 2019. Chapter 5: Absorption, distribution, and excretion of toxicants. In: Klaassen CD, editor. Casarett & Doull's Toxicology: The Basic Science of Poisons. 9th ed. New York, NY: McGraw-Hill Education. https://accesspharmacy.mhmedical.com/content.aspx?bookid=2462&sectionid=194919337

6. Evaluation of Mechanistic Information

Summary

This section of the handbook provides a transparent framework using systematic review and integration methods and guidance, aligning with the Report on Carcinogens (RoC) listing criteria, for reaching a level-of-evidence (LoE) conclusion for carcinogenicity from mechanistic studies. Key elements are (1) using fit-for-purpose approaches, (2) providing a structural framework for evaluating study informativeness and reaching confidence judgments and LoE, (3) using strategies accepted by the scientific community (such as the key characteristics of carcinogens [KCC]), and (4) encompassing scientific advancements and New Approach Methodologies (NAMs) (e.g., read-across approaches).

The first component of the handbook describes steps for conducting broad scoping and evidence-mapping activities to develop a fit-for-purpose health hazard framework and study protocol for the specific substance. The framework identifies influential mechanistic questions and the literature (referred to as study sets) that will be used to answer the questions. These questions are scientific issues that (1) are critical for understanding cancer mechanisms and biological and human relevance, (2) will most likely impact the LoE conclusions (e.g., establishing or refuting biological plausibility), and (3) have an adequate database to assess; examples include the confidence for a biological effect(s) (such as a KCC), mechanism of action (MoA), or prediction of carcinogenicity based on read-across analysis. Each study set is defined by the type of evidence (e.g., exposed humans or animals), exposure, comparison, group, and endpoint (e.g., a specific biomarker or similar biomarkers).

The second component provides study informativeness questions for evaluating mechanistic studies in exposed humans, animals, and cells and has biomarker and assay information for the 10 KCCs to assess the informativeness of selected study sets for each question. The rigor (e.g., streamlined, moderate, in-depth) of the study informativeness and evidence assessment depends on how impactful a study set is to the overall evaluation. The last component of our handbook describes a three-step process for integrating the evidence across studies in each study set, across study sets to reach a confidence judgment for each influential question (e.g., confidence for a KCC, MoA, read-across prediction), and across questions for an LoE conclusion from mechanistic studies. Descriptors and guidance are provided for confidence levels (e.g., high, moderate, low) for each type of mechanistic evidence (including the “ideal” evidence for high confidence for specific KCCs) and LoE (convincing, supporting). “Convincing” (as defined by the RoC listing criteria) refers to LoE from mechanistic studies that allows a substance to be listed as a human carcinogen with insufficient or inadequate evidence from cancer studies in humans or animals, whereas “supporting” LoE refers to robust mechanistic data supporting the findings in human or animal cancer studies.

The LoE from mechanistic studies is integrated with LoE from human cancer epidemiological and animal cancer studies, considering all relevant data to reach a listing recommendation.

Introduction

Mechanistic data—studies on how a substance causes cancer or other diseases—are often crucial for identifying cancer hazards. For the RoC evaluations, the evaluation team identifies and

assesses the relevant mechanistic information to determine the strength of the evidence (i.e., conclusion) for the biological plausibility that a substance may cause cancer in humans, that the animal cancer data are relevant to humans, or both. Using the RoC listing criteria, this conclusion is integrated with the LoE conclusions from human and animal cancer studies to reach a preliminary listing recommendation for the substance. Cancer studies can also inform the evaluation of mechanisms by providing information on type of tumor and can be useful for identifying sensitive populations. When human or animal cancer evidence is limited or inadequate, convincing mechanistic information is needed to fulfill the RoC listing criteria. Mechanistic evidence is also evaluated when there are robust data from human and animal cancer studies. RoC evaluations use the term "supporting mechanism data" for evaluations mainly relying on human or animal evidence, with mechanistic data showing biological plausibility, but by itself may be less than convincing. RoC evaluations may lead to the discovery of new mechanisms for a substance (see Section 6.4.3). Guidance for reaching convincing or supporting LoE is provided in Section 6.4. Understanding cancer mechanisms may sometimes help inform the approach for assessing (e.g., most relevant exposure metrics, latency, and timing of exposure) and interpreting the evidence from human cancer studies (e.g., biological plausibility for specific cancers). Table 6-3 to Table 6-5 provide an example, trichloroethylene review, which assesses the evidence for a given hypothesis (or MoA), B-cell antigen stimulation, was evaluated as a potential mechanism for non-Hodgkin lymphoma, a cancer observed in human and animal studies.

Cancer Mechanism Evaluation Process

The assessment of cancer mechanisms begins with scoping and problem formulation leading to the development of the framework (which is captured in the protocol) for a specific substance under review. Using the methods delineated in the protocol, a project team conducts the evaluation, which includes assessing informativeness and evaluating and integrating the evidence of the selected studies and data to reach the strength or LoE conclusions (see Figure 6-1).

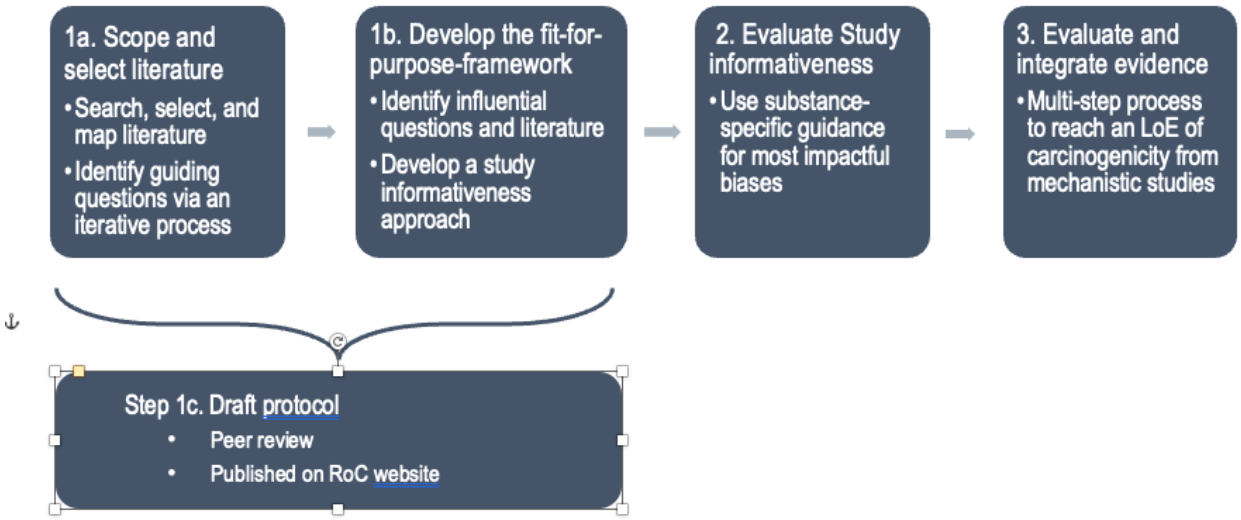


Figure 6-1. Cancer Mechanism Evaluation Process

The schematic depicts the systematic review process, which begins with identifying the substances, searching and mapping the evidence (Step 1a, Section 6.2.1), and developing the protocol (Step 1b, Section 6.2.2). Iterative scoping questions can help inform the framework (e.g., identify the questions and literature [e.g., EECO] most influential to the evaluation) and approach for evaluating study informativeness for a specific substance under the review. The framework is captured in the substance protocol (Step 1c, Section 6.2.6). Using the protocol guidance, we evaluate the informativeness (focusing on the most predominant biases) of the relevant literature for each influential question (Step 2, Section 6.3) and evaluate and integrate the evidence by a multistep process to reach level-of-evidence (LoE) conclusions (Step 3, Section 6.4). EECO = evidence type (e.g., human, animal, in vitro, cell-free system, in silico), **e**xposure, **c**omparison group, and **o**utcome; it is analogous to PECO, with "evidence type" replacing "population").

Cancer Hazard Assessment and Handbook Components

The handbook provides the methods to develop a substance-specific guidance framework (protocol) and guidance for evaluating study informativeness and integrating the evidence to reach an LoE of carcinogenicity from mechanistic studies. To facilitate understanding the methods for assessing mechanistic evidence, we first discuss considerations in method development (Section 6.1), followed by the methods for each step (as per Figure 6-1) in the cancer hazard evaluation process.

- Develop the cancer hazard framework for mechanistic data (Section 6.2).
  - Develop and execute a literature search strategy to search, identify, and map the literature.
  - Develop the framework or protocol.
- Evaluate the informativeness of individual—or groups of similar or influential—mechanistic studies or computational models (Section 6.3).
- Evaluate and integrate the evidence via a multistep process—and reach an LoE conclusion of carcinogenicity from mechanistic studies (Section 6.4).

Clustering and read-across approaches encompass developing and conducting analyses besides scoping and evaluating the evidence, and considerations in the process steps may overlap. Thus, Section 6.5 integrates and discusses the methods across the evaluation process.

6.1. Aims and Considerations for Developing the Methods

The handbook builds on and provides transparency to prior approaches used to evaluate mechanistic information for substances listed in the RoC. Recent advancements in identifying, organizing, and assessing mechanistic data related to cancer [notably as outlined in the 2019 International Agency for Research on Cancer preamble (IARC 2019; Samet et al. 2020)], scientific innovations for generating mechanistic data, and the growing reliance on using these data also informed these methods. In documenting our methods for assessing mechanistic data, we aim to:

- Use *strategies* for evaluating mechanistic data that are accepted by the scientific community (see Section 6.1.1).
    - Incorporate read-across approaches (see Sections 6.1.1 and 6.5) and scientific advancements (see Section 6.7).
- Provide transparency for a structured approach for methods used to reach conclusions based on the RoC listing criteria (see Section 6.1.2).
    - Develop new guidance for determining the informativeness of individual studies (Section 6.3) and collective evidence evaluation and integration (Section 6.4).
- Conduct *fit-for-purpose evaluations* focusing on the influential issues identified by broad, unbiased scoping activities (see Section 6.1.3 and Section 6.2).

6.1.1. Use Strategies for Evaluating Mechanistic Data

Mechanistic data are often derived from ex vivo, in vitro, in chemico, biochemical, or in silico studies, informed by chemical properties. They typically explore early molecular and cellular effects leading to tissue- or organ-level responses or changes in biological functions. Mechanistic information may also be derived from in vivo studies in animals and observational, randomized, or volunteer studies in humans. We define evidence type as testing system (i.e., in vivo, in vitro, in silico, and in exposed humans [including observational and interventional human studies with endpoints other than cancer incidence]). NAMs, such as advanced computational models and multiomics, are being developed and implemented to predict carcinogenicity without relying solely on animal models (see Section 6.7, New Directions). As the scientific consensus for using NAM and multiomics data in cancer hazard assessment advances, we will update the handbook to include methods to assess (e.g., study informativeness), evaluate, and integrate these types of data.

Biological Responses and Pathways

Carcinogenicity involves the acquisition of properties associated with the transition of normal cells to cancer cells. Hanahan and Weinberg describe these functional properties of cancer acquired by human cells as the *hallmarks of cancer* (2011). Since the 2011 publication, new and emerging hallmarks have been identified (Hanahan 2022) (see Box 6-1). Carcinogens can cause cancer through a multistep process. The acquisition of these functional capabilities might be mapped in some fashion to the distinguishable steps of tumor pathogenesis (Hanahan 2022). Sequential biochemical events and perturbations in the biological system resulting in the formation of cancer are described by two terms: mode(s) of action (MoA) and adverse outcome pathway(s) (AOP) (Rowan and Spielmann 2019). An MoA includes the sequence of obligatory (key) biological events from cellular interaction with the chemical to a functional or biological effect within a living system (Kienzler et al. 2017). An AOP is a conceptual model that includes the key cellular and molecular events, beginning with a Molecular Initiating Event (MIE), required to produce an adverse effect from exposure to a toxicant and is not chemical specific (NTP 2023). Mechanism(s) of action refers to a detailed understanding at a biochemical and molecular level of how a substance causes an adverse health effect (Jacobs et al. 2020). In this handbook, we use the term MoA to be either mechanisms or modes of action because a similar approach and guidance is used for both concepts. A MoA can also be thought of as a hypothesis.

Box 6-1. Hallmarks of Cancer

Hallmarks	Emerging Hallmarks[a]
Sustaining proliferative signaling Evading growth suppressor Resisting cell death Enabling replicative immortality Inducing or accessing vasculature Activating invasion and metastasis Deregulating cellular metabolism Avoiding immune destruction Enabling: • Genome instability and mutation • Tumor-promoting inflammation	Unlocking phenotypic plasticity Senescent tumor cells Enabling • Nonmutational epigenetic reprogramming • Polymorphic microbiomes

Source: Hanahan (2022).

[a]Cancer stem cells and the tumor microenvironment are now widely appreciated to play an integral role in tumorigenesis and malignant progression and could be considered emerging hallmarks.

Carcinogens induce tumors through multiple mechanisms and biological pathways (Guyton et al. 2009; Smith et al. 2020), and a complete understanding of all cancer mechanisms for a given substance is rarely, if ever, known. Thus, assessments of mechanistic data focusing *only* on proposed MoA or AOP may miss additional mechanisms or reach erroneous conclusions. Informed by a review of over 100 known human carcinogens, an IARC working group identified shared initiating properties of human carcinogens (i.e., the *key characteristics of carcinogens [KCCs]),* which collectively encompass multiple cancer mechanisms (see Box 6-2). The KCCs provide an unbiased way to find, organize, and evaluate relevant mechanistic literature without reliance on a predetermined cancer formation pathway (Guyton et al. 2018; Smith et al. 2016; Smith et al. 2020). IARC developed a framework for using this approach (captured in the 2019 Preamble) and has applied it to over 50 cancer hazard evaluations (IARC 2019; Smith et al. 2020). With increasing knowledge of cancer mechanisms and testing technology innovations, the

list of KCCs is expected to grow (Tice et al. 2021). In recent years, RoC monographs have also used the KCC approach in their cancer hazard evaluations (e.g., Night shift work, light at night), including those for the RoC (e.g., haloacteic acids found as water disinfection by-products, antimony trioxide) (NTP 2018a; 2018b; 2021a).

Box 6-2. Key Characteristics of Carcinogens

KCC 1. Is electrophilic or can be metabolically activated to an electrophile
KCC 2. Is genotoxic
KCC 3. Alters DNA repair or causes genomic instability
KCC 4. Induces epigenetic alterations
KCC 5. Induces oxidative stress
KCC 6. Induces chronic inflammation
KCC 7. Induces immunosuppression
KCC 8. Modulates receptor-mediated effects
KCC 9. Causes immortalization
KCC 10. Alters cell proliferation, cell death, or nutrient supply

Source: Smith et al. (2016); Smith et al. (2020).

In Silico Clustering and Read-across Approaches

Read-across represents an important data-gap filling technique for chemical safety assessments. Interest in clustering and read-across approaches and other new classification methodologies, as part of NAMs, for identifying hazards has been increasing due to the large number (>350,000) of chemicals (globally) registered for production (Wang et al. 2020) and use with a paucity of human and animal cancer data. Grouping chemicals with similar properties circumvents the time needed, costs, and use of experimental animals. However, few reports of using these methods to predict cancer are available, perhaps due to the complexity of carcinogenesis.

Clustering approaches involve grouping similar compounds based on functional groups, chemical class, precursors, or physicochemical and biological properties. Unsupervised clustering approaches use input data (such as general structural characteristics not specific to an endpoint) without a corresponding predicted variable or endpoint of interest. General unsupervised clustering through structural similarity alone may not provide information on biological activity or function but can be used in scoping activities to inform a read-across plan (see Section 6.5). Typically, supervised classification approaches are used to predict an outcome and involve training a model using input variables (specific for the endpoint), the endpoint of interest or data-gap filling (e.g., outcome variable), and an algorithm to map the input to the output. Supervised learning can also use modern artificial intelligence (e.g., deep neural networks) based on specific biological endpoints.

Read-across based on grouping chemicals with structural and biological similarity and using source chemicals with known cancer hazards (with human or animal cancer data) is utilized to predict the cancer hazard of the target chemical. There are two main approaches to read-across: analogue and category. In the analogue approach, empirical data from one or a group of analogue source chemicals can be used to predict the same endpoint for the target chemical through endpoint specific supervised similarity. The category approach predicts targets starting from a group or a category of multiple sources whose physical-chemical, biological, or toxicological properties are likely to be similar or follow a regular pattern as a result of structural similarity (OECD 2014). For confidence in the read-across prediction, a robust assessment of uncertainty from multiple sources in the read-across process is undertaken (Schultz et al. 2015).

RoC listings include several chemical classes based on metabolism to a carcinogen (e.g., dyes metabolized to benzidine) (NTP 2021b) or a similar MoA (cobalt and cobalt compounds that

release cobalt ions in vivo) (NTP 2021c). We aim to expand our assessment of chemicals by incorporating read-across approaches in our cancer hazard evaluations.

6.1.2. Structured Guidance for Reaching Conclusions

New to the updated handbook are guidance considerations and a structured framework for determining whether the mechanistic evidence is convincing. These new features include guidance for evaluating the quality of the mechanistic studies, confidence guidance for different types of information (e.g., read-across, biological effects or responses), and guiding principles for the overall body of evidence. Our approach also provides transparency for a stepwise approach to evidence integration that builds on methods used in previous evaluations. We aim to find a "sweet spot" between "adequate transparency and consistency across evaluations" and "flexibility for substance-specific issues and scientific judgment" (i.e., to avoid prescriptive algorithms).

The RoC listing criteria provide several means (each of which can be viewed as a criterion) for listing in the RoC based on mechanistic data. To provide transparency and develop a structured framework for evaluating mechanistic evidence, we coined the term "Level of Evidence mechanistic question" and developed guidance for each element of the RoC criteria that is related to the evaluation of mechanistic (and related, such as class approaches) evidence.

The LoE of carcinogenicity from mechanistic data is reached by answering one or more of the LoE questions (see Table 6-1) and considering all relevant data. LoE question 1 (whether the agent acts through mechanisms indicating it would likely cause human cancer) encompasses the concepts of biological plausibility, answered by using mechanistic strategies evaluating (1) biological effects (e.g., KCCs) or MoA or (2) using read-across strategies to predict the carcinogenicity of nonlisted substances. Read-across approaches can also be used to evaluate whether the substances under review are members of similar classes as listed substances (LoE Question 2), and MoA approaches are also used to address whether there is compelling evidence that a substance acts via a mechanism that does not operate in humans (LoE Question 3) (see Table 6-1). In this handbook, we develop strategy-specific (e.g., biological effects, MoA, read-across) confidence judgments (e.g., high, moderate, and low) and guidance to reach "convincing" and "supporting" LoE conclusions (see Section 6.4 for biological effects and MoA and Section 6.5 for read-across).

Table 6-1. The Relationship between LoE Questions and Mechanistic Strategies

Criteria-related Questions	Mechanistic Strategies for Answering the Question
1. What is the level of evidence (LoE) (convincing, supporting) that the agent acts through mechanisms indicating it would likely cause cancer in humans?	
Biological plausibility	Biological effects/KCC, including the relationship between two or more KCCs. One or more structured and reported MoA (e.g., cancer site specific or cancer pathway, such as endocrine disruption)

Criteria-related Questions	Mechanistic Strategies for Answering the Question
Predicted to cause cancer	Read-across methods incorporating mechanistic data
2. Does the agent, substance, or mixture belong to a well-defined, structurally related class of substances whose members are listed in a previous RoC (known to be a human carcinogen or reasonably anticipated to be a human carcinogen)?	Read-across methods incorporating mechanistic data
3. Are there compelling data indicating the agent acts through mechanisms that do not operate in humans and therefore are not "reasonably anticipated" to cause cancer in humans?	MoA for a specific cancer site

6.1.3. Fit-for-purpose Evaluations

One of the primary challenges of a mechanistic data review is the potentially large number of relevant references. An in-depth comprehensive review of all relevant studies is impractical and may not be necessary for hazard assessment. Instead, the goal is to rigorously assess and evaluate the confidence in the evidence for the most "influential mechanistic questions" for the substance under review. These questions are related to the LoE questions and strategies (see Table 6-1) but are substance specific, e.g., an influential question would focus on specific KCCs for which there is an adequate database to review and likely to impact (supporting or arguing against a listing) the overall assessment. A typical cancer hazard evaluation will have several influential mechanistic questions, which are identified through scoping activities—including broad-based comprehensive literature searches and evidence mapping, a series of iterative guiding questions, authoritative and other relevant reviews, and a preliminary review of the human and animal studies, e.g., cancer sites observed in human and animal studies or cancer sites that may not be relevant to humans. For each question, the rigor of the study informativeness assessment (bias evaluation and study sensitivity) depends on how impactful the different groups of studies are to the overall evaluation (see Section 6.2.2). The overall LoE is reached by integrating confidence in the evidence across the questions (see Section 6.3 and Figure 6-2). Although the cancer hazard evaluation (captured in the monograph) focuses on the influential questions, the monograph will briefly summarize all relevant mechanistic data in the evidence integration step. Identifying key questions has been successfully used since 2013 for RoC monographs (see Section 6.2.2, Protocol Development for examples). Although the fit-for-purpose approach provides for flexibility for the specific substance, the major steps (and guidance) in Figure 6-2 are common to all evaluations.

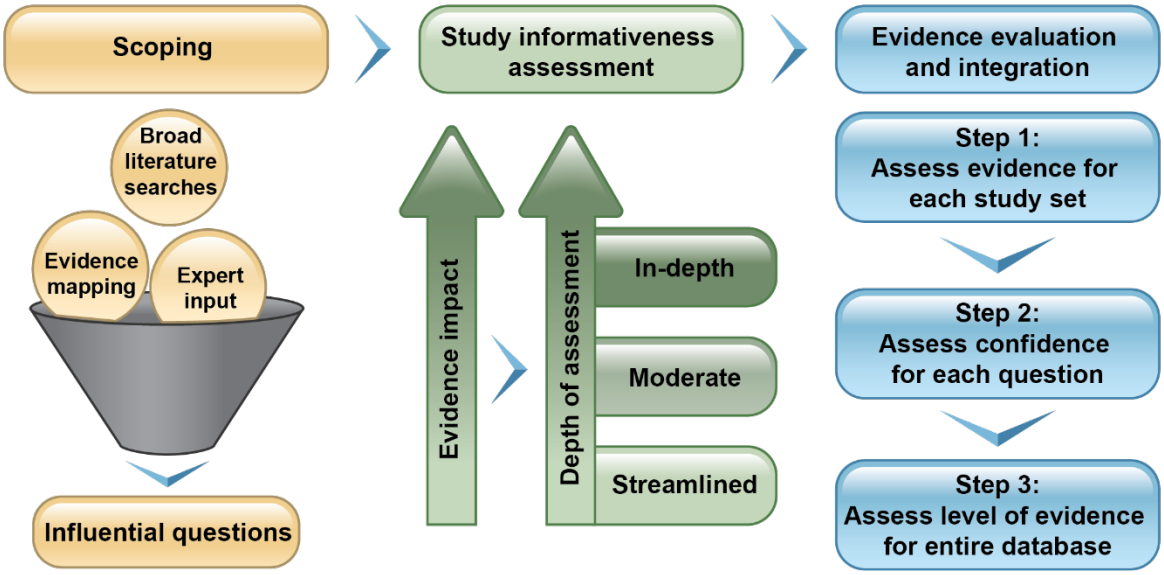


Figure 6-2. Fit-for-purpose Evaluations

Influential mechanistic questions are identified from broad, unbiased literature searches. For each question, the relevant literature (e.g., multiple study sets) is identified and evaluated; the rigor (streamlined to in-depth) of the study informativeness and evidence assessment depends on how impactful or influential (e.g., low to high) a study set is to the overall evaluation (see Sections 6.2 and 6.3, Box 6-3). Each study set is defined by the type of evidence (e.g., exposed humans or animals, in vitro), exposure, comparison, group, and endpoint (e.g., a specific biomarker or similar biomarkers) and are informed by evidence mapping and the influential questions. Lastly, we integrate the evidence across study sets to reach a confidence judgment for each question and across questions for an LoE conclusion.

6.2. Mechanistic Framework Development

To develop the framework, the project evaluation team performs scoping and problem formulation activities, which are used to create a protocol (e.g., methods) to conduct the evaluation. Whereas Section 1 discusses scoping and problem formulation activities for the overall cancer hazard evaluation, this section focuses on methods specific for the review of mechanistic studies. The process is necessarily iterative (i.e., there may be several cycles of literature searches and evidence mapping). The development of the mechanistic framework for read-across approaches is discussed in Section 6.5.

Figure 6-3 provides an overview of the framework development process. When a new substance is identified for potential review, scoping activities begin with a review of authoritative reviews and other scientific information sources, many of which are listed in the general sources found in Appendix B (e.g., EPA CompTox Chemicals Dashboard, NTP Integrated Chemical Environment, OECD eChemPortal, ECHA REACH) and initial literature searches to gather mechanistic information for the substance and develop the initial EECO statement (Evidence type, Exposure, Comparison group, and Outcome), which informs the literature searches and inclusion and exclusion criteria to select the literature. Next, broad-based systematic literature searches and searches for omic and screening biological effects data in recognized data depositories are conducted. Iterative scoping questions and mapping the evidence from literature searches (e.g., by KCC, evidence type, or MoA) and authoritative reports help to focus the evaluation, e.g., identify the influential questions, and literature (i.e., study sets) to answer those

questions. Each study set is defined by the ECCO specific for the question. Finally, the framework is captured in the substance protocol.

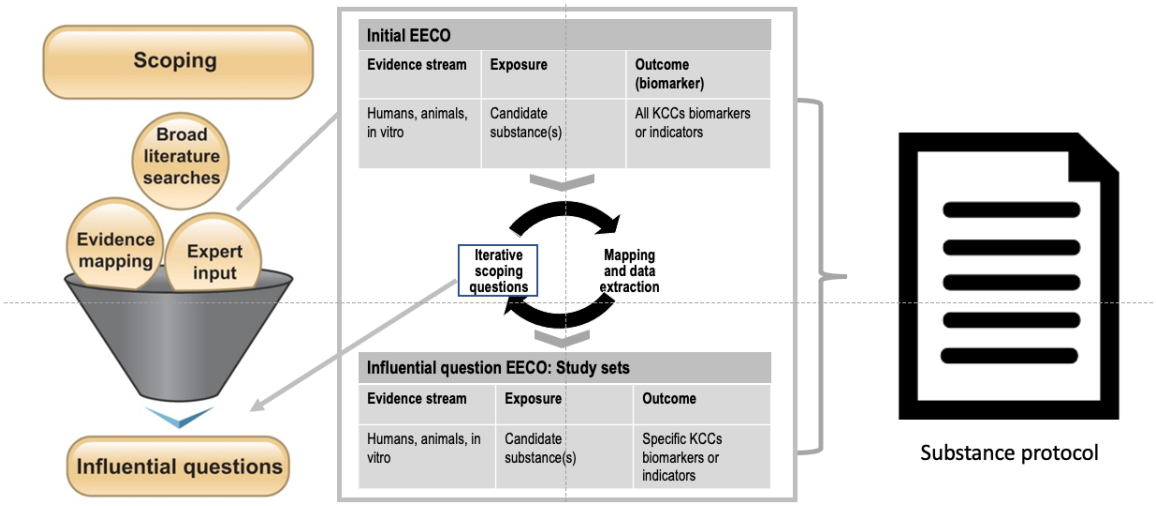


Figure 6-3. Framework Development Schematic

Literature searches and inclusion and mapping of the relevant literature are conducted based on the initial EECO. These scoping activities help to identify the influential questions and the literature used to answer those questions (e.g., the influential question EECO), which are termed study sets.

6.2.1. Literature Search Strategy and Selection

The information specialist conducts a broad-based literature search in PubMed and at least one other citation database (e.g., Scopus or Web of Science) using substance-specific search strings (e.g., exposure) combined with KCC-specific terms (outcome or endpoint) or general mechanism terms (using the Boolean operator "AND"). General mechanism terms capture MoA or other mechanistic data that may be missed by restriction to KCC terms (see the search string document on the NTP website for the search strings for KCCs and general mechanisms). Table 6-2 provides examples of the terms and concepts (not an exhaustive list) covered in the searches. Substance-specific search strings are available in the substance protocol. Using KCCs to search for mechanistic data provides an unbiased literature search that is not targeted to any one mechanism. These searches may be supplemented with terms for substance-specific issues.

Table 6-2. Examples of Concepts Used for Mechanism Searches

Mechanism Searches	Example Concepts	
	PubMed, Scopus, Web of Science	**MeSH Terms**
General Mechanism Search Terms	Biological markers	Causality
	Biomarkers	Tumor markers, biological
	Key events	Etiology [sh]
	Mechanism/mode of action	Oncogene fusion
	Molecular initiating event	Tumor necrosis factors
Characteristics of Carcinogens	See Box 6-2 above and the search string document	
Metabolomics	Metabolome	Metabolomics
	Global metabolism	
	Systemic metabolism	
	Epimetabolites	
	Metabolic fingerprint	
	Metabolomic-wide association study (MWAS)	

Citations are uploaded into a web-based program such as HAWC or SWIFT Active Screener, which has a machine-learning feature for screening and review. Inclusion and exclusion criteria are based on the initial mechanistic EECO (see Figure 6-3). The first round of screening is based on the title and abstract, whereas the second round is based on a full-text review. If needed, publications in other languages will be translated into English. Studies are typically organized and tagged by KCC, agent (if evaluating a class), and evidence type (e.g., exposed humans, exposed animals, in vitro). At times, the broad searches using KCC terms may yield a large database of available literature (e.g., 5,000+ articles). In these situations, a program such as SWIFT Review (a computer tool that assists with literature prioritization) (Howard et al. 2016) may be used to presort the literature by KCC, evidence type, or substance/agent based on key words in the title and abstract. As machine learning and artificial intelligence (AI) technologies improve, we may utilize these programs for evidence mapping and data extraction (Walker et al. 2022).

6.2.2. Evidence Mapping and Scoping Questions

Iterative scoping questions and evidence maps may help with problem formulation and identify the influential, agent-specific mechanistic questions and studies, prioritize publications for evaluations, and inform protocol development. A series of iterative scoping activities—e.g., literature searches and tagging, review of selected studies, streamlined data extraction, authoritative literature, and review of the human and cancer data—informs the development of these questions (see Figure 6-3). New or revised scoping questions are identified based on the answers to the initial questions. Computer tools (e.g., SWIFT Review, Dextr) can facilitate the reviewer-driven categorization of studies and data extraction. Examples of guiding questions from previous RoC reviews are found in Table 6-3. Evidence mapping using an interactive

program such as Tableau can facilitate visualizing the literature, e.g., by agent, KCC, evidence stream, findings, or other relevant categories.

Table 6-3. Guiding Questions for Selected Example Evaluations

Substance	Example Scoping Questions[a]
Cumene (NTP 2013)	Are any tumors in experimental animals thought to be caused by species-specific mechanisms with an established MoA? • Is there an adequate database available to study the MoA and genotoxicity?
Trichlorethylene (TCE) (NTP 2015)	What are the key events for the MoAs for the cancers observed in the human epidemiological studies—kidney cancer and non-Hodgkin Lymphoma (NHL)? • Does TCE cause immunotoxicity in experimental animals and autoimmune diseases in humans? • Are there data to evaluate the proposed mechanisms of immunomodulation and NHL?
Night Shift Work Light at Night (NTP 2021a)	What is the relationship between circadian disruption and cancer? • What intermediate biomarkers of circadian disruption play a role in its pathogenesis? • What types of studies have evaluated circadian disruption and cancer? What circadian biomarkers and KCCs have been measured in night shift workers? Can mechanism information (e.g., biomarker studies on circadian disruption) help contextualize the potential cancer hazard from working night shifts? • What characteristics of night shift work are related to circadian disruption?
Antimony Trioxide (NTP 2018a)	What role do antimony chemical species play in antimony trioxide carcinogenic potential? • Is there a difference in toxicity or carcinogenic potential between the pentavalent and trivalent forms of antimony? • Should mechanistic studies of other trivalent antimony chemicals be included in the evaluation?
Haloacetic Acids (HAA) (NTP 2018b)	What are the major MoA(s) for the carcinogenicity of HAAs? • Do all or some HAAs share an MoA or cause similar key biological events? • How does a potential MoA relate to the bio-physical properties of HAA and the KCC? • How do the animal cancer outcomes relate to HAA mechanistic data or physicochemical properties? Do the mechanistic and other relevant data provide adequate support for considering some or all HAAs as a well-defined, structurally related class?

[a]These questions have been modified from key questions in the protocols or workshop documents. (Note that cumene predates the development of a public protocol.)

6.2.3. Deposited Data Searches

On a case-by-case basis (to potentially supplement traditional mechanistic data based on a focus or specific need), we may search data repositories for omic and other high-throughput data (e.g., ToxCast/Tox 21) related to potential cancer mechanisms and related peer-reviewed publications

(e.g., study design and methods). For example, the NTP cancer hazard evaluation of antimony trioxide included an analysis comparing bioactivity of different antimony compounds (NTP 2018a). In general, these types of data are used to support the evaluation but do not currently influence the evaluation. Examples of repositories and tools for analyzing the data include Gene Expression Omnibus (NCBI 2023) and SpiderSeqR (Sozanska et al. 2020), which searches high-throughput multiomic data repositories. The substance-specific protocol will provide the methods for data analyses.

6.2.4. Identifying Influential Mechanistic Questions and Study Sets

As mentioned in Section 6.1.3 (Fit-for-purpose Evaluations), we aim to focus our evaluations by identifying several "influential" mechanistic questions. These are scientific issues that (1) are critical for understanding cancer mechanisms and biological and human relevance, (2) will most likely impact the LoE conclusions (e.g., establishing or refuting biological plausibility) and (3) have an adequate database to assess. Focusing the literature on these questions does not bias the assessment because the questions are identified based on broad scoping activities (e.g., using KCC terms) and iterative guiding questions. An influential mechanistic question might be related to a single or group of related biological effects or responses (and the connections between these effects), biological pathways (or MoA) reported in the literature, or whether the substances (or subclass) belong to a carcinogen class. Currently, most biological effects of interest are captured by the 10 KCCs, but this may change with knowledge gained by scientific advancements. The set of questions for each evaluation may overlap and would most likely (if the database is adequate) represent several pathways for which a substance can cause cancer; for instance, an evaluation with a question for a specific MoA would also include questions that are pathway agnostic (e.g., based on a KCC). Table 6-4 provides examples of selected influential mechanistic questions from previous evaluations that were developed from the guiding questions in Table 6-3.

The evaluation of mechanistic data may help identify new pathways (e.g., assessment of KCCs and the relationship between KCCs or from integrating data from multiomic studies) or help inform the confidence for cancer sites observed in humans (e.g., cancers with limited evidence, such as the trichloroethylene [TCE] and non-Hodgkin lymphoma [NHL] example) or animal cancer (e.g., human relevance, such as the cumene male rat kidney tumor example).

The protocol will also identify the EECO for the literature (e.g., study sets) or data needed to address each influential mechanistic question (see Figure 6-3). For example, a question related to genotoxicity would include studies from different evidence types evaluating diverse biomarkers. For many evaluations, interactive evidence maps are created, allowing for real-time identification of the exact studies.

Table 6-4. Influential Mechanistic Questions for Example Evaluations

Substance	Example Influential Questions[a]	Cancer Hazard Conclusions
Cumene Reasonably anticipated to be a human carcinogen (NTP 2013)	What is the confidence that cumene causes renal tumors in male rats by MoA that is considered not to be relevant to humans (i.e., alpha(2u)-globulin nephropathy)? What is the confidence for the relevance of mouse lung tumors to humans? What is the confidence that cumene causes genotoxicity, and at what tissue sites?	Sufficient evidence of carcinogenicity from experimental animals[b] • Liver and lung tumors: considered human relevant • Kidney tumors supporting as relevance to humans is unclear Genotoxic in some tissues
Trichlorethylene (TCE) Known to be a human carcinogen (NTP 2015)	What is the confidence level that TCE can cause kidney cancer via a glutathione conjugation (GSH) pathway (e.g., bioactivation of TCE to nephrotoxic and mutagenic (in the kidney) metabolites)? What is the confidence level that TCE can cause lymphoma via an immunomodulation mechanism involving B-cell activation via chronic antigen stimulation (leading to B-cell oncogenic transformation via mutations/genomic recombination during class switching and somatic hypermutation)? • This question differs from a general KCC question regarding whether TCE causes chronic inflammation or immunomodulation as it focuses on a specific type of immunomodulation that is proposed to be linked to non-Hodgkin lymphoma (NHL). • The confidence for this MoA can be combined with limited evidence from cancer epidemiology studies to reach a conclusion related to the level of evidence in humans (e.g., cancer epidemiological studies and mechanistic studies).	Sufficient evidence from studies in humans: kidney cancer • Supporting evidence from toxicological, toxicokinetic (e.g., GSH metabolism), and mechanistic studies (e.g., genotoxic and nephrogenic metabolites in the kidney)[c]. • Limited evidence from studies in humans: Supporting evidence for autoimmunity and immunosuppression, but B-cell activation hypothesis was not adequately tested
Night Shift Work Persistent night shift that causes circadian disruption High confidence for a causal relationship with human cancer (NTP 2021a)	What is the confidence that circadian disruption plays a role in nightshift work carcinogenicity? • What is the confidence that circadian disruption causes cancer? • What is the confidence that night shift work is associated with circadian disruption? What is the confidence that night shift work causes KCCs (specifically, genotoxicity, oxidative stress, immune effect, increased hormones, and epigenetic effects)? • Are KCCs observed in humans the same as the KCCs detected in animal cancer studies?	Sufficient evidence from the collective body of evidence of cancer epidemiological studies (limited by itself) and mechanistic studies in humans and experimental animals[d] • Circadian disruption plays a role in shift-work-mediated carcinogenicity • Night shift work is associated with KCC that were observed in animal cancer studies and with circadian disruption

Substance	Example Influential Questions[a]	Cancer Hazard Conclusions
Antimony Trioxide Reasonably anticipated to be a human carcinogen (NTP 2018a)	What is the confidence that antimony trioxide causes genotoxicity and KCCs (especially oxidative stress and receptor-mediated effects)? What is the confidence that other trivalent compounds cause KCCs (e.g., for which there are no data for antimony trioxide)? • What is the confidence that these can be attributed to antimony trioxide?	Sufficient evidence of carcinogenicity from studies in experimental animals and supporting mechanistic data • Antimony trioxide: Genotoxicity, oxidative stress • Other trivalent compounds: Inhibition of DNA repair, inhibition of cell differentiation • Antimony trioxide may exert its effects through released trivalent antimony ions
Haloacetic Acids (HAA) found as water disinfection contaminants (six) Reasonably anticipated to be a human carcinogen (NTP 2018b)	What is the confidence that HAAs cause genotoxicity and other biological effects (KCCs) (especially oxidative stress and electrophilicity)? • What are the molecular initiating events (MIE) and key steps, and how do they differ across HAAs? • Do the biological effects vary by the number and types of halogens? What is the confidence that HAAs without animal cancer data will likely cause cancer in experimental animals? • Can HAAs be grouped into subclasses based on animal cancer data? • Do individual HAAs (without cancer data) share key biological effects (or are metabolites) as carcinogenic HAAs?	Four HAAs sufficient evidence of carcinogenicity from studies in experimental animals and supporting from mechanistic data • Electrophilic, genotoxic, cause oxidative stress Two HAAs without animal data are listed in the RoC based on metabolism to a carcinogenic HAA and supporting mechanistic data. • Data insufficient to evaluate HAA as a class or subclasses[e]

[a]Questions have been adapted (e.g., made more specific for illustrative purposes) from selected key questions in the monographs for these substances.
[b]See Section 7.2.1, Box 7-1
[c]See Section 7.1.3, Table 7-1
[d]See other supporting information in Section 7.1.3, Table 7-1
[e]See Section 6.5.3, Step 1, Exemplar

6.2.5. Technical Input and New Research

Experts with substance-specific expertise typically provide advice on the framework development or may peer review the evaluation protocol. The type of input—which is related to an influential mechanistic question—varies with the complexity of the evaluation and may range from consultation with individual technical advisors to information groups—experts who meet as a group and provide input on specific guiding questions or issues—to public workshops or webinars with experts giving presentations and focused panel discussions on multiple substance-specific scientific issues. NTP has also conducted additional studies to address research gaps related to influential mechanism questions. Table 6-5 provides examples of technical input and new research in past evaluations.

Table 6-5. Technical Input Strategies and New Research for Some Example Evaluations

Substance	Mechanistic Technical Input or Research
Cumene (NTP 2013)	Information group: Evaluation of the alpha(2u)-globulin nephropathy as the sole MoA for male rat kidney tumors Research: Genotoxicity studies were conducted because of conflicting findings and importance to potential species-specific mechanisms for mouse lung and rat kidney tumors. These studies showed that cumene cause DNA damage in the liver and lung in rodents.
Trichlorethylene (NTP 2015)	Information group: TCE and immune effect studies for evaluating potential cancer mechanisms Rationale: Expert input is needed to evaluate a proposed mechanism (related to B-cell activation, see Table 6-4) for NHL, which is a cancer site of interest from human epidemiology studies.
Night Shift Work/ Light at Night (NTP 2021a)	Public workshop: 12-panel workshop addresses several scientific issues including (but not limited to) circadian disruption, how to define night shift work, and measures of circadian-effective light (Lunn et al. 2017).
HAAs (NTP 2018b)	Information group: Read-across strategies Rationale: Cancer bioassay data were only available for a subset of 13 HAAs considered for the evaluation. Expert input was needed to discuss whether read-across-like-approaches could be used to evaluate the potential carcinogenicity of subclasses of HAAs (e.g., based on the type or number of halogens) or individual HAAs without cancer data).

6.2.6. Protocol Development

The protocol for each assessment includes specific and detailed instructions and considerations for evaluating mechanistic information adapted from the general strategy presented in the RoC handbook. It summarizes the scoping activities (see Section 6.2.1), the evaluation framework (e.g., "influential mechanistic questions" and literature [study sets]), including the impactful literature (see Section 6.3), and the methods for assessing study informativeness and evaluating the evidence, including evidence integration (see Sections 6.3 and 6.4). When relevant, it includes clustering and read-across methods and a summary of technical input (such as an informational group or workshop). The protocol is informed by consultation with substance-specific experts.

6.3. Study Informativeness Assessment

As mentioned in Section 6.1, because there is an abundance of mechanistic studies, a fit-for-purpose approach is used that identifies the issues (e.g., influential questions) and the literature base to answers those questions (as discussed in Section 6.2). The focused approach also applies to the strategy for evaluating study informativeness, which is discussed in Section 6.3.1. Questions and guidance, organized by domains, are provided in Sections 6.3.2 for experimental in vitro and animal in vivo studies and 6.3.3 for exposed human studies. For the endpoint and analysis domains (Table 6-9 and Table 6-11), the questions and guidance are the same across all evidence types.

6.3.1. Strategy

The approach for assessing study informativeness consists of (1) identifying the influential literature and rigor of the evaluation (e.g., gradient ranging from streamlined to in-depth), (2) identifying biases and sensitivity concerns that have the maximum impact on the substance and evidence type under consideration (e.g., influential biases) (Savitz et al. 2019), and (3) assessing those issues across the influential literature and questions identified in the protocol (see Figure 6-4; details are discussed below). The steps in the process are similar across all evaluations, and the rigor and extent of the study evaluation are expected to vary depending on the complexity of the database.

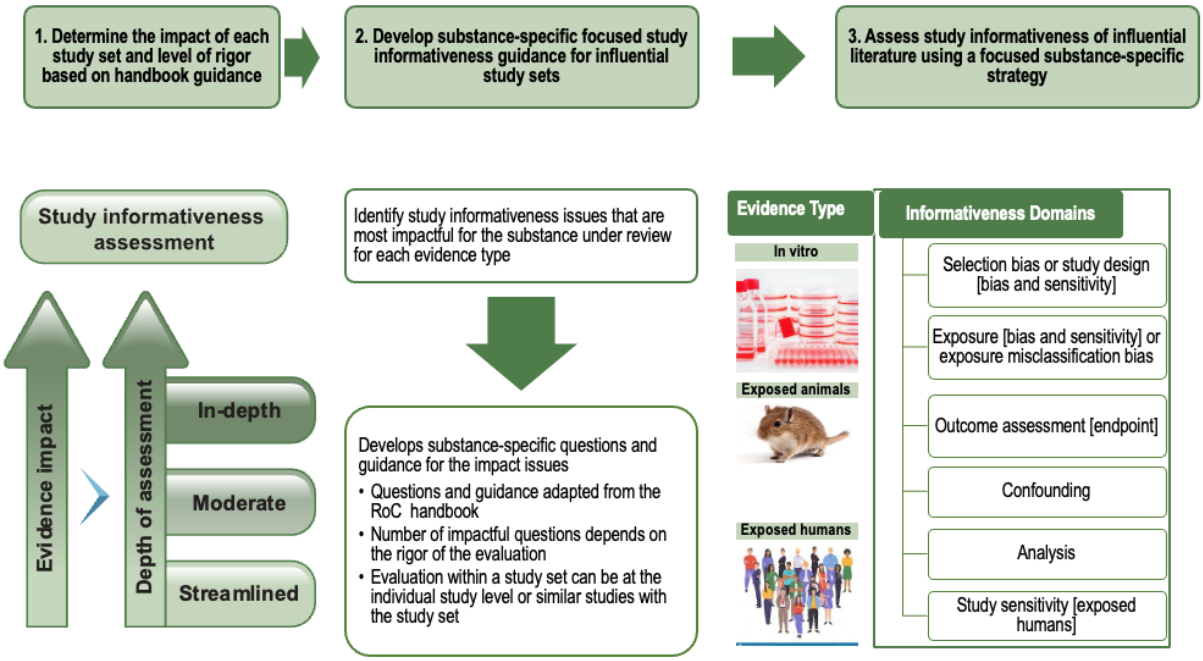


Figure 6-4. Study Informativeness Approach

The fit-for-purpose study informativeness approach includes identifying the most impactful literature (1) and the level of rigor of the assessment and development of study-specific guidance that focuses on the most impactful issues (2). Questions and guidance to conduct the assessment (3) are modified from the general (domain-based) guidance for each evidence type provided in the handbook.

Identifying the Influential Literature and the Rigor of the Evaluation

The fit-for-purpose study informativeness approach includes identifying the most impactful literature. Mechanistic information frequently is found in a large set of studies. The rigor or depth of the evaluation (e.g., data extraction approach and study informativeness assessment) for each influential question can vary along a gradient (ranging from streamlined to in-depth), depending on the literature's impact (or influence) on the overall evaluation (see Box 6-3 for guidance). For less impactful literature, a streamlined review may rely on reporting conclusions from credible or authoritative reviews without a formal study informativeness evaluation, while a moderate review may extract findings from primary studies and assess potential bias, and an in-depth evaluation of influential literature consists of detailed data extraction and formal study informativeness assessment as depicted in Figure 6-4 and discussed below). The protocol will

outline the evaluation approach, including the influential literature and for reporting the findings and conducting the study informativeness assessment.

The light-at-night and night shift evaluations (NTP 2021a) provide an example of using this approach. Because the oncogenic effects of melatonin and clock genes have been well established, the evaluation relied on review articles for this information and conducted a deeper dive into the primary studies of KCCs and biomarkers of night shift work and light at night. (Note: The assessment considered general elements of study quality but did not conduct a structured study informativeness evaluation.) For TCE, immune studies required a more in-depth assessment to determine whether the measured biomarkers were consistent with the proposed MoA of B-cell antigen stimulation leading to lymphoma. Studies using similar methods to evaluate genotoxicity and oxidative stress of multiple HAAs and reporting trend and potency analyses were impactful in the evaluation of HAA subclasses for potential listing in the RoC.

Box 6-3. Guidelines for Identifying Influential Literature for Each Question[a]

How critical is the data to the overall cancer hazard listing recommendation?

- Are there adequate cancer studies in humans or experimental animals, or will the evaluation be primarily based on mechanistic studies?
- How will the mechanistic studies in exposed humans contribute to the overall conclusions regarding whether there is sufficient evidence from studies in humans?
- Is the literature and question associated with the cancer sites of interest?

What is the scientific consensus for the question (considering the evidence type)?

What is the adequacy (e.g., number of studies, human and biological relevance) of the database for each question? Is there an adequate database of biologically- and human-relevant studies?

How consistent is the evidence across a study set?

[a]Influential literature determines the degree of rigor of the review, e.g., study informativeness, reporting, and evaluation ranging from streamline to in-depth.

Guidelines for identifying influential literature are provided in Box 6-3 and consider both the influential question and the adequacy of the database to answer the question. Deposited primary data (with published peer-reviewed methods or analysis guidelines) will be analyzed if the data are more informative (e.g., human relevant) than the published studies, answer a question for which there are no published studies, or inform biological networks or pathways.

Develop Substance-specific Informativeness Guidance

During the scoping process, the study evaluation team identifies several impactful study informativeness issues specific to the exposure under evaluation; the evaluation protocol delineates these questions and guidance for evaluating the questions. The evaluation team adapts these substance-specific questions from the mechanistic-informativeness questions (e.g., exposure conditions or misclassification, endpoint assessment, confounding) for each evidence type (e.g., in vivo animal, in vitro, and exposed humans; see Sections 6.3.2 and 6.3.3). The study informativeness assessment concentrates on the impactful bias questions and does not necessarily provide responses for all the signaling questions in all the domains. Table 6-6 provides examples of impactful bias questions from evaluations of RoC-listed substances.

Table 6-6. Examples of Impactful Bias Questions

Substance	Domain	Evidence Type	Question/Comment
Cumene (NTP 2013)	Exposure/sensitivity	In vitro	Were the assays conducted using closed systems? Cumene is a volatile chemical, and this is a concern for false negatives.
Trichlorethylene (TCE) (NTP 2015)	Confounding	In vitro	Is there potential confounding (most likely causing a false positive) from the presence of stabilizers or using DMSO as a solvent? DMSO can react with some TCE metabolites, raising the pH and distorting the findings.
Styrene (NTP 2008)	Endpoint assessment	Exposed humans	A series of questions regarding the measurement of chromosomal aberrations and micronuclei was systematically assessed across studies.

Evaluation of Bias across Studies

For some evaluations, studies having similar designs, populations, exposures, or endpoint assessments may share similar concerns for potential bias and sensitivity. For example, randomized cross-over studies may be less susceptible to confounding than observational studies but only measure effects from acute exposure to the substance. In these cases, the evaluation may discuss study informativeness and evaluate the evidence for the group of studies (e.g., explore heterogeneity by experimental human studies vs. observational studies). The individual studies in the group are still reviewed, focusing on the shared influential biases and whether they can be binned together for bias assessment and evidence evaluation.

Bias Questions and Guidance

The evaluation team assesses the informativeness of mechanistic studies using a domain-based approach consisting of questions and guidance appropriate for the evidence type. As mentioned in the protocol section, bias assessment is fit for purpose, and the number of "impactful" biases may vary for different bodies of literature. For each question, we evaluate the potential, direction, and magnitude of the bias by comparing the study to an "ideal" study element specific to that bias (e.g., selection of participants). "Ideal" is defined as a study design condition resulting in low concern for bias or sufficient sensitivity to detect an effect if present (see Table 6-7 to Table 6-13 for guidance for each question). The bias assessment may identify a common issue (e.g., DMSO used as a solvent in the TCE studies), the degree of concern (e.g., minimal, some, major), for the bias, or overall informativeness for the issue (e.g., good, adequate, deficient). Because of the heterogeneity of mechanistic studies, guidance for these terms is provided in the substance protocol and not in the handbook. We use triangulation methods (e.g., integrating evidence with different approaches and different types of biases and to exploit these differences to draw qualitative conclusions; see Section 6.4) to explore bias impact across studies.

Two reviewers assess the studies and resolve differences through mutual discussion and reference to the original data source. To facilitate the review process and reduce ambiguity, we conduct a pilot phase using a small set of studies before proceeding with the complete

evaluation. If reporting of the studies is too incomplete to evaluate bias, we will contact the study authors for additional details. See Section 6.6 for information on reporting.

6.3.2. In Vitro and In Vivo Animal Studies

The study informativeness assessment (guidance and questions) was informed by the animal cancer informativeness questions and the SciRAP Tool for evaluating in vitro toxicology studies and other authoritative sources (USEPA 2018). As new tools are being developed to evaluate in vitro studies, we will reassess, and update guidance as needed. The in vitro and in vivo domains (study design, exposure conditions, outcome, confounding, and analysis) are similar; however, the guidance may vary by evidence type. Some study elements may overlap between different domains and will be addressed by one substance-specific influential bias question.

Study Design

Mechanistic studies may pose unique challenges because they often include multiple endpoints and assays with varying exposure conditions; assessment of the study design depends, in part, on the specific exposure and endpoint(s) under investigation.

The study design domain includes signaling questions that address bias and overall study sensitivity (Table 6-7). The in vivo studies have an additional bias question for randomization that is not relevant for in vitro studies. In vitro (and sometimes in vivo) studies generally also include positive (and, to a lesser extent, negative) control groups to verify the assay was conducted properly and was sensitive (or specific) enough to detect an effect and the results are treatment related.

In addition to evaluating bias and study sensitivity, experimental studies should consider reporting and other study quality factors, such as conditions that incorporate considerations of the model (pre-exposure), exposure, and endpoint domains. Ideally, assay conditions and procedures should be described in detail for cell or tissue/organ culture in vitro or ex vivo (before exposure, during exposure, and after exposure) and should follow best practices that approximate those described by Guidance for Good Cell Culture Practice (GCCP) (Pamies et al. 2022) and Good In Vitro Method Practices (GIVIMP) (OECD 2018a; 2018b; Pamies et al. 2022). This may include, but is not limited to, cell density, culture media, and culture conditions. Some of these issues may be confounding factors. However, it is recognized that not all these conditions are reported in the publications.

Core bias question: Is there concern that flaws in study design biased the interpretation of the findings?

Table 6-7. Study Design: Questions and Guidance

Study Design Signaling Question	Guidance: In Vitro	Guidance: In Vivo
Randomization		
Is there concern that the methods by which animals were randomized to groups were inadequate or that animals were not randomized?	Not usually relevant. Note that randomization of replicates is addressed under endpoint assessment.	Ideally, the randomization method should be reported and based on ensuring all animals have an equal probability of being assigned to any given control or experimental group.
Controls		
Is there concern that the concurrent control group/sample is not adequate for evaluating effects across treatment groups?	Control samples should be derived from the same cell line and handled in the same way as treated samples. An untreated or vehicle control should always be included as it is critical for determining treatment-related effects. In most cases, positive controls are important for in vitro studies to ensure the test system is functioning properly.	Concurrent controls are the most relevant comparison group for evaluating potential exposure-related biological effects. Both vehicle-treated and untreated controls should be used, with vehicle-treated being more relevant than untreated if only one type of control group was used. Ideally, the concurrent control group included at least as many animals as did each treatment group. In the absence of controls or the presence of only a few concurrent controls, historical controls from the same testing laboratory can be used. In some cases, positive or negative controls may be used to ensure sensitivity/specificity and detect bias.
Sensitivity Question		
Is there concern that the study design was not sensitive enough to detect a mechanistic event or effect if present?	The overall sensitivity of the study design integrates the combined effects of several sensitivity factors—e.g., sensitive test system used, study duration, adequate replicates, and statistical power. One factor may compensate for limitations in another (e.g., a sensitive test system or positive control endpoint may compensate for low statistical power). Ideally, the in vitro models are biologically (e.g., metabolic capability, primary vs. cell line, proposed mode of action) and human relevant (e.g., human vs. mammalian cells, relevance to the cancer sites of interest).	The sensitivity rating integrates the animal model, statistical power, and study design. In some cases, one factor may compensate for limitations in another factor (e.g., a sensitive animal model may compensate for a low number of animals or short study duration).

Exposure Conditions

The study's informative questions include a bias question related to dose selection and a sensitivity question related to dose levels, exposure duration, and frequency of exposure. Dose (or concentration) selection may affect both bias and study sensitivity. Sensitivity is decreased when doses are too low to induce an effect or too high and are cytotoxic. Bias (potentially resulting in a false positive) could occur if doses are high enough to nonspecifically induce the KCC via toxicity rather than direct (specific) effects on the KCC.

Core bias question: Is there concern that the exposure conditions biased the interpretation of the findings?

Table 6-8. Exposure Conditions: Questions and Guidance

Signaling Question	Guidance: In Vitro	Guidance: In Vivo
Dose Selection		
Is there concern that the dose (concentration) levels were too high to attribute effects to the substance?	Ideally, the concentrations tested should not cause significant cytotoxicity that could affect study results. OECD may provide assay-specific guidance (GIVIMP) on dosing conditions, e.g., on the number and spacing of doses and dose levels (OECD 2018a; 2018b; Pamies et al. 2022).	Exposure should result in tolerable toxicity (e.g., slightly decreased survival and/or body weight gains) at the high dose, but not excess toxicity, for the duration of the study.
Overall Exposure Conditions Sensitivity		
Is there concern that dose/exposure regimen was not sensitive enough to detect a relevant effect?	Exposure conditions sensitivity integrates several factors that may affect the ability of the study to detect an effect (e.g., dose levels that are too low and/or exposure durations that are too short).	The sensitivity rating integrates considerations of dose level, route, exposure, and study duration (e.g., when the samples were collected). The selection of the dose may depend on the exposure duration.
	Ideally, the rationale for dose selection and number of doses should be reported and justified and follow guidelines for the assay. Ideally, exposure to the cells/tissues (e.g., uptake by cells) should be measured.	Ideally the delivered dose was close to the planned target dose (e.g., stability, solubility, volatility of the chemical; changes in eating and drinking patterns of the exposed group that may change the oral dose).
	The stability, solubility, and volatility of the test agent may affect exposure (e.g., lower the delivered dose); adjustments should be made as needed.	The exposure route should consider metabolism of the chemical (e.g., a route that bypasses metabolic activation may not be informative).
	Evaluation of dose response results can contribute to confidence in the study findings and allow for evaluation of potential effects at lower doses.	For many KCCs, long-term or repeated exposure studies are more informative for detecting persistent/chronic effects, which are more relevant for carcinogenicity. Many biomarkers are nonpersistent.
		Evaluation of dose response can contribute to confidence in the study findings and allow for evaluation of potential effects at lower doses.

Endpoint: All Evidence Types

Endpoint assessment includes considerations of how closely the measurement represents the endpoint of interest (e.g., KCC biomarker), how well the endpoint or biomarker is measured, and whether there is potential for observer bias (e.g., in the absence of blinding). Appendix D provides information for evaluating specific KCCs.

Core bias question: Is there concern that the endpoint assessment biased the interpretation of the findings?

Table 6-9. Endpoint Measurement: Questions and Guidance

Signaling Question	Guidance: All Evidence Types
Endpoint Proxy	
Is there concern that the endpoint proxy (e.g., biomarker) did not represent the effect of interest? Is there concern that the biomarker was not measured in relevant tissue or cells?	Ideally, studies should measure the most relevant biomarkers related to the KCC biological effect (e.g., direct vs. indirect measures).[a] In general, local biomarkers measured in local target tissues are more informative than those in circulation (systemic) or in other (nontarget) tissues.
Measurement Methods	
Is there concern that measurement methods (e.g., timepoints, accuracy, precision) are not adequate to generate valid and reliable data? The measurement methods consider consistency, replication, randomization, and sampling (collection methods and timing).	Ideally, each study should use accurate and reliable methods for measuring each endpoint at the appropriate timepoint and should follow applicable protocols and guidelines. For some biological effects, there are OECD guidelines. The timing of sample collection is important to determine whether an observed effect is transient or persistent. The study should include enough replicates or repetitions to generate reliable results for the endpoint of interest and without any serious uncertainties or limitations in the sampling process. Replicate placement used in automated plate systems should be random.
Differential Measurement Error	
Is there concern for differential measurement error (e.g., the treated and control groups were assessed differently)?	All treatment and control groups should follow the same protocols for endpoint assessment and should be clearly reported to properly interpret results. Blinding and randomization should be used during endpoint assessment.
Sensitivity Questions	
Is there concern that the assay or methods were not sensitive enough to detect an effect?	Ideally, appropriate and sensitive assays (e.g., current state-of-the-science assays with sufficient counts, such as the number of cells) should be used to measure direct biological effects (e.g., KCC) in the target tissue or tissue of interest. However, multiple indirect measures may also be informative.

[a]Informativeness for the KCC or biological effect; informativeness with respect to carcinogenicity is considered during evidence integration.

Confounding

Sources of potential confounding in both in vitro and in vivo studies include the chemical and its administration (e.g., vehicle, stabilizer, and conduct of the studies [e.g., cell culture maintenance, animal husbandry]). Chemical contaminants or other conditions that are associated with the endpoint may be potential confounders.

Core question: Is there concern about potential confounding?

Table 6-10. Confounding: Questions and Guidance

Signaling Question	Guidance: In Vitro	Guidance: In Vivo
Chemical Characterization		
Is there concern that the chemical characterization and dose formulations (e.g., source, homogeneity, purity, solubility, volatility, and stability) are not adequate to support attribution of any effects to the substance?	Ideally, the source of the test agent should be identified, the purity and stability documented, and the solubility reported at the concentration used. A nontoxic vehicle should be used, and the test agent should be soluble in the vehicle. Contaminants or vehicles associated with the endpoint (e.g., genotoxic, immunotoxic) are potential confounders.	
Substance Administration		
Is there concern that the vehicle or stabilizers may have confounded the findings?	Vehicles and stabilizers, if used, should be verified not to interfere with exposure or endpoints of interest, even if a vehicle control group is used as the baseline.	
Conduct Issues		
Is there concern about confounding from the conduct of the study or maintenance of the animals or cells?	Examples of conduct that may lead to confounding if associated with the endpoint, in in vitro studies include cell culture maintenance, such as contamination of cell lines with bacteria or other cells,	Examples of sources of potential confounding (if associated with the endpoint) in animal in vivo studies include infection by pathogens, poor animal husbandry (e.g., impurities in the bedding), and decreased consumption of feed and water, leading to less healthy animals (e.g., in feed and water studies that alter smell and taste)

Analysis

Although not necessarily a bias, analyses are considered in the study informativeness evaluation. The study should identify and report the rationale for the statistical/analytical methods. Best practices (e.g., Organisation for Economic Co-operation and Development [OECD] test guidelines and corresponding guidance documents) provide some recommendations for statistical tests and general considerations for statistical analyses of different types of data. If necessary, the evaluation team will consult with statisticians and experts in data analysis. Biostatisticians will be consulted to evaluate omic data.

Core question: Is there concern that the data assumptions and analyses were not adequate?

Table 6-11. Analysis: Questions and Guidance

Signaling Question	Guidance: All Evidence Types
Data Assumptions	
Is there a concern that the data assumptions used in the statistical analysis were inadequate? Is there concern that statistical methods used to investigate the endpoints were inappropriate, were not properly conducted, or were incomplete?	The study should clearly report all statistical methods and conclusions, any other methods, or information that is relevant for determining whether any observed effects are exposure related. For example, are data transformation methods (e.g., log transformation) appropriate? Were outliers removed? (Ideally, outliers should not be removed without statistical justification, as that may be where the effect is strongest.)
Statistical Model and Methods	
Is there concern that statistical methods used to investigate the endpoints were inappropriate or not reliable?	The choice of statistical analyses will depend on the type of study and the nature of the endpoints measured.

6.3.3. Studies in Exposed Humans

Table 6-12 provides a summary of the approach for the study informativeness assessment and additional guidance in exposed humans. Assessment of observational studies in exposed humans follows the domain-based directions found in Section 3.2 (selection bias, exposure assessment, endpoint assessment, analysis, and sensitivity for human cancer studies). Other mechanistic studies in humans are controlled-exposure studies, e.g., randomized controlled trials or experimental studies of volunteers using themselves as the control (before or after exposure). For ethical reasons, experimental studies and interventions in humans are often for real-world exposures (albeit often for shorter times or lower frequencies) that individuals may have in their daily lives. Randomized controlled studies and other experimental studies may suffer from external validity and may not include relevant sensitive populations. For these controlled-exposure studies, bias from exposure conditions rather than exposure assessment is assessed. We have developed additional questions and guidance for the selection bias for these study types. Note that endpoint and analysis guidance are the same for all evidence types and are found above in Table 6-9 and Table 6-11.

Table 6-12. Study Informativeness Strategy for Mechanistic Studies in Exposed Humans

Domain	Strategy	Comment/Additional Guidance
Selection Bias	Observational studies: Human Cancer Studies (see Section 3.2)	Same selection (including attrition) principles as case-control and cohort studies.
	Experimental studies: Human cancer studies and additional questions/guidance on randomization and blinding	**Randomization** Participants should be randomized appropriately for the study design (cross-over or other designs); the effectiveness of the randomization should be evaluated (i.e., are the distributions of important known confounders similar across treatment/exposure groups?) and reported. For cross-over trials, participants should be randomized with respect to the order of exposure and nonexposure periods, and there should also be an adequate wash-out period between exposure/treatment and subsequent nonexposed periods. **Blinding:** Ideally, participants and researchers should be blinded but this may not always be possible depending on the exposure type.
Exposure Assessment	Observational studies: Human Cancer Studies (see Section 3.2)	Exposure biomarkers and personal monitoring are more commonly used in mechanistic studies than cancer studies. Populations with prior, possibly chronic, exposure to the exposure of interest may lead to exposure misclassification and bias the findings toward the null (even if they were controlled-exposure studies). For example, some studies use self as controls.
Exposure Conditions	Experimental studies: New questions and guidance Sensitivity, bias, ethical considerations	Exposure levels should be high enough to detect the relevant effects. Similarly, duration and timing of dose(s) should be appropriate for manifesting the specific endpoint of interest. Ideally, they would be close to real-world exposure levels as practical. For many exposures, long-term or repeated exposure studies are often more informative for detecting certain biological effects (such as chronic inflammation), which are more relevant for carcinogenicity. The exposure conditions should be matched to the endpoint of interest.
Endpoint Assessment	All evidence streams (Section 6.2.2): Bias and sensitivity	Most studies measure circulating biomarkers rather than tissue-specific markers.
Confounding	Human Cancer Studies (see Section 3.2)	Information on potential confounders may be less common in mechanistic studies, and the association of a confounder may be less established. The nature of the exposure (e.g., exposure chambers) may cause stress, which can be an effect modifier.
Analysis	All evidence streams (see Section 6.2.2)	Biological effect biomarkers are often continuous data; thus, different statistical tests are used than those for cancer (ordinal outcome).

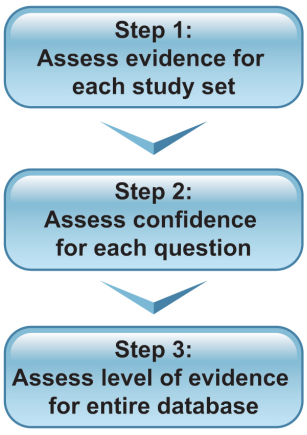
Domain	Strategy	Comment/Additional Guidance
Study Sensitivity	Statistical power and exposure contrast: Human Cancer Studies (see Section 3.2) Endpoint: All evidence streams (Section 6.2.2)	Timing of endpoint measurement replaces latency. Long-term or repeated exposure studies are more informative for detecting persistent/chronic effects, which are more relevant for carcinogenicity.

6.4. Evidence Evaluation and Integration

6.4.1. Overview Strategy

The RoC listing criteria permit (1) listing a substance based on convincing relevant information that the agent acts through mechanisms indicating it would likely cause cancer in humans (e.g., in-depth analysis of mechanistic data related to biological plausibility) or is a member of a well-defined structurally related class of listed substances or (2) not listing a substance with sufficient evidence of carcinogenicity from laboratory animals but with compelling data indicating the agent acts through mechanisms that do not operate in humans (e.g., it is not biologically plausible). However, the listing criteria do not specify the bar for convincing, compelling, or determining whether a substance belongs to a listed class. Historically, we have used the terminology "supporting mechanisms" as part of the rationale for listing a substance in the RoC based on the evidence from human or animal cancer studies. Informed by lessons learned from previous evaluations and input from authoritative sources (e.g., IARC), we define guidelines for LoE conclusions (convincing or supporting) of mechanistic data. Our handbook also describes confidence descriptors and judgments for evaluating evidence for different mechanistic strategies (e.g., KCC, MoA, read-across analyses), which can be applied to the influential questions.

Box 6-4. Evidence Integration Steps



Evidence is integrated using a three-step process (see Box 6-4). After integrating the evidence across studies for each study set (see Section 6.4.2), the evidence is integrated across all study sets to reach a confidence judgment (high, moderate/limited, inadequate) for the influential question; confidence guidance is specific for different types of questions (e.g., KCCs, MoA, and read-across) (see Section 6.4.3). The third step (see Section 6.4.5) is to integrate the evidence across all the influential questions and other relevant data (e.g., absorption, distribution, metabolism, and excretion [ADME] and toxicokinetics) for a level of carcinogenicity (convincing, supporting) from mechanistic studies.

A series of evidence-based tables captures the assessment for each step, with conclusions for each step brought forth to the next-level table (see Section 6.6.3). [This approach was informed by the NTP Cancer Assessment of Night Shift Work and Light at Night (NTP 2021a)]. The evaluation protocol would provide more specific guidance.

6.4.2. Step 1: Assess the Evidence for Each Study Set: Biological Effects

Study sets for each influential mechanistic question, defined by the EECO statement, can also include intermediates that be considered as exposure, and includes many different types of endpoints (e.g., KCC or other intermediates). Night shift work represents a complex data set in which circadian disruption (a key intermediate) can be viewed as an exposure or an outcome in evaluating exposure-outcome associations.

The depth of the assessment of the individual studies depends on the influential literature analysis. Some evidence may be reported from review articles, others at a high level (streamlined extraction) using Tableau or other visualizing tools, while an in-depth assessment will be conducted on the most influential literature. If relevant, a meta-analysis or other quantitative methods can be used to explore heterogeneity, bias, and other key issues in the literature within an evidence stream (e.g., human studies). Additionally, forest plots or other graphical methods stratified by study or other characteristics may be useful in exploring heterogeneity without a quantitative analysis. Meta-regression or other analytical methods could be used to evaluate dose-response relations across studies to support the strength of the evidence. Because of variability in how studies are conducted, and the types of exposure and outcomes reported, it may be difficult or inappropriate to combine studies in a meta-analysis (see Section 3.3.2 for a discussion on strengths and limitations of using meta-analysis in a hazard assessment).

The assessment considers the following factors:

- Consistency of the evidence across studies
    - Assessment of the evidence for a specific KCC from individual studies may integrate the findings of multiple biomarkers or indicators for that effect (e.g., a study-specific assessment of chronic inflammation may include results for proinflammatory cytokines and lymphocyte subsets). We used the term indicators for effects that are not biomarkers per se, such as outcomes (e.g., increased infection for immunosuppression, histopathology, or organ weights).
    - Each KCC-specific biomarker or indicator (e.g., increases or decreases in lymphocytes) may be integrated across the study set.
- Strength of the association, such as magnitude of the effect and exposure-response relationships, and uncertainty
- Study informativeness (e.g., internal validity, sensitivity)
    - Tissue/sample (e.g., local, target tissue, or circulation). In general, evidence for KCC biomarkers in target tissues is more informative than circulating/systemic biomarkers, although it may depend on the KCC and cancer site of interest.
    - Relevance (e.g., duration, dose, timing) of the exposure and measurement of the biological effect.
- The assessment uses a triangulation approach, which integrates evidence from different study designs and methods, as well as different sources of potential biases, to reach conclusions about consistency.

Each evidence-based table captures the integration of the evidence for each study set, the strength and limitations identified by the informativeness evaluation, and the overall assessment (see Section 6.6.2).

6.4.3. Step 2a: Assess the Confidence of the Evidence: Biological Effects

For each influential question, we integrate the assessments across study sets, using triangulation approaches across the mechanistic evidence, to reach a confidence judgment based on the guidance outline in Table 6-13 and discussed below. Key considerations are the consistency and cohesiveness of both the effect and the biomarkers used to measure an effect/KCC, the overall quality of the collective evidence across study sets, and the informativeness of the biomarkers. For most evaluations, a biological effect is one of the KCCs (recognizing that some KCCs have multiple effects); thus, our focus in this handbook is on KCCs. However, this is not a strict requirement, and KCCs may be refined over time, or new biological effects may be identified.

Table 6-13. Confidence Judgment for Biological Effects

Descriptor	Biological Effect
High	*Individual biological effect (KCC)*
	Consistent evidence of informative KCC biomarkers/indicators[b] across studies of sufficient quality and ideally in more than one evidence type
	Coherence across individual KCC biomarkers/indicators
	Relationships between KCCs or multiomic data
	Consistent evidence across studies for an association between two or more biological effect biomarkers/indicators
Moderate/Limited	Less than consistent evidence across studies, KCC biomarkers/indicators, and evidence types
	Consistent evidence for less relevant KCC biomarkers/indicators or less biological or human-relevant evidence[a] types, or few studies (e.g., limited ability to evaluate consistency)
Inadequate	Negative or unexplained inconsistent findings
	Few data or limited study informativeness
	Evidence mainly from evidence types with low biological or human relevance for the substance and specific biological effect under evaluation[a]
	KCC biomarkers/indicators are of limited informativeness for the biological effect and cancer development

[a]Biological and human relevance are evaluated on a case-by-case basis, e.g., some nonmammalian in vivo assays may be relevant for some but not other KCCs, and some species may be more relevant than others.
[b]Indicators are defined as effects that are not biomarkers per se, such as outcomes (e.g., increased infection for immunosuppression, histopathology, or organ weights).

Informativeness of the Study Set

The informativeness of the study set considers the quality of the study set as a whole, including the evidence type, the adequacy of the exposure proxy, and informativeness of the indicators or biomarkers for measuring the biological effect. Our evaluation uses a triangulation approach that considers the source of collective bias and whether the evidence points to the same conclusions. Ideally, the evidence should come from biological (e.g., exposed animals) and human-relevant

studies (exposed humans and human cells); however, each evidence type is associated with different types of potential biases: Human epidemiology studies may be at risk from confounding or exposure measurement error, in vivo experimental studies may have species-specific issues, and human in vitro studies are limited by the physiological relevance to model tissue function (such as inadequate metabolism of the substance to the carcinogenic species or ADME effects that may decrease the delivered dose). Despite these biases, all evidence types can contribute to the evaluation, and consistent findings across evidence types increase the confidence of the evidence conclusion. In cases for which there is heterogenicity across evidence types, substance- and evaluation-specific issues (such as study quality and bias for a particular evidence stream) will be used to consider which evidence type(s) are most informative. A separate confidence call (i.e., robust) is made for studies in exposed humans to facilitate overall evidence integration (human cancer studies, animal cancer studies, and mechanistic studies; see Section 7). Robust means that there is consistent evidence across studies for an informative KCC, which is usually supported by experimental evidence. Studies from all evidence types are not needed to make conclusions.

The exposure proxy (If relevant) for a study set should be adequate for the exposure of interest (see Section 3.2.2, Table 3-4). For example, a question in the antimony trioxide assessment was whether other antimony trivalent compounds were appropriate for evaluating the effects of antimony trioxide. A question in the night shift work assessment was whether differences in the timing of simulated night shift work in experimental studies captured the “real-world” exposures of night shift work.

The KCC biomarkers and indicators should be informative for assessing the KCC. Table 6-14 provides guidance to reach high confidence for the KCCs, and Appendix D provides background information on KCC biomarkers and indicators. Each KCC consists of several subtypes measured by different biomarkers or indicators (see Appendix D tables). For example, the main subtypes of oxidative stress include reactive oxygen species (ROS), ROS modifications to DNA, RNA, protein or lipids, and antioxidants. High confidence for some KCCs (e.g., oxidative stress or chronic inflammation) is reached by integrating across the subtypes, whereas, for other KCCs (such as genotoxicity), evidence for only one subtype (e.g., strong evidence for mutagenicity and negative for chromosomal damage) is sufficient for a high confidence judgment. KCC 10 encompasses several biological effects, each of which can be evaluated independently.

When assessing the relationship between KCCs, we consider consistency across studies and evidence types, dose/response, timing, or sequence of events. Assessing the confidence of biological effects may lead to new mechanistic hypotheses for the substance under evaluation. (These questions differ from influential questions that are for MoA that have been in the literature during the scoping process and are discussed in the next section.) For example, in the RoC review of haloacetic acids found as water disinfection by-products, assessment of the evidence for KCCs led to several proposed mechanisms, suggesting that cancer-initiating events involve electrophilic reactions with proteins and MoA involving oxidative stress resulting in mutations and chronic inflammation and epigenetic changes leading to altered gene expression and DNA repair (Atwood et al. 2019), all of which can lead to cellular proliferation, transformation, and cancer. Proposed biological pathways or MoA (e.g., timing/steps) may be based on actual studies or general knowledge about carcinogenicity. However, the associations between the substance and specific biological changes could reflect the availability of studies and not the level of its contribution to carcinogenicity (Atwood et al. 2019; NTP 2018a; 2018b).

Table 6-14. Informativeness of Key Characteristic of Carcinogens Biomarkers or Indicators

KCC	Ideal Evidence for High Confidence	Other Evidence for High Confidence: Combination of Biomarkers/Indicators[a]	Biomarkers/Indicators and Cancer Prediction
KCC1: Is electrophilic or can be metabolically activated to electrophiles	Covalent DNA adducts in target tissues DNA adducts with higher "hardness" are more relevant for cancer development as they are less likely to be repaired (e.g., O^6 adducts are more informative than N^7 adducts)	Covalent DNA adducts in nontarget tissues Protein adducts; RNA adducts Electrophilic properties, such as potency, measured in vitro or in silico Evidence of electrophilic chemical structure of metabolite or primary chemical	High levels of specific chemical or bulky DNA adducts have been associated with increased risk of several types of cancer (e.g., colon, liver, lung in smokers, prostate, stomach) in prospective studies, albeit the risk estimate may be due to exposure to the substance as adducts are also biomarkers of exposure.
KCC2: Is genotoxic	Positive findings for ≥1 endpoint in exposed animals or humans • Mutations • Chromosomal damage: CA or MN (indicative of CA and changes in chromosome number)	DNA damage: strand breaks in target tissues of exposed humans or animals Mutations or chromosomal damage: in vitro studies using bacteria or eukaryotic cells	MN and CA associated with increased cancer risk in prospective cohort studies. Chemical-specific mutational spectra observed in cancers. Ames-positive chemicals (and chemicals positive for several in vitro mutagenicity assays) and in vivo MN are strong predictors of carcinogenicity.
KCC3: Alters DNA repair (DRC) or causes genomic instability	Direct (e.g., not due to secondary effects) evidence of impaired DRC from the comet DNA repair assays or γH2AX assay Direct evidence of exposure-induced genomic instability (e.g., ideally in nontumor tissue)	Several biomarkers with less specific evidence (Topoisomerases), and relevant biomarkers from other KCCs (e.g., changes in telomere length)	Lower DRC linked to increased cancer risk for combined cancers and several specific cancer types (meta-analyses) using various assays. Most genomic instability assays have been done on tumor cells.
KCC4: Induces epigenetic alterations	DNA methylation in epigenetic clock genes or relevant cancer-related genes OR Epigenetic changes (histone modification, DNA methylation not related to epigenetic clock, and noncoding RNA) in combination	Histone modification Noncoding RNA Global methylation	Accelerated epigenetic aging is associated with an increased risk of cancer development.

KCC	Ideal Evidence for High Confidence	Other Evidence for High Confidence: Combination of Biomarkers/Indicators[a]	Biomarkers/Indicators and Cancer Prediction
	with downstream effects (e.g., transcription)		
KCC5: Induces oxidative stress	Several oxidative stress biomarkers (pro- and antioxidative stress) • Integrative score • Pro-and antioxidative stress biomarkers in the same studies	One or more clinically relevant oxidative stress biomarkers, such as F2-isoP and oxidative damage to DNA *And* Link with other KCCs (e.g., KCC2, KCC6)	Higher levels of 8-OH-dG, hOGG1 (tissue), MDA, or protein carbonyls have been found in urine, tissue, or blood of cancer (urinary bladder, breast, cervical, liver) patients than controls. Studies of oxidative stress biomarkers and cancer risk are limited or unclear.
KCC6: Induces chronic inflammation or immune activation[b]	Evidence of chronic inflammatory (e.g., autoimmune) diseases in exposed humans or animals Histological evidence of chronic cellular inflammation in target tissues (WBC- e.g., lymphocytes) with local increases in proinflammatory biomarkers (e.g., specific cytokines or chemokines) in exposed humans or animals • Ideally, the biomarkers are associated with cancer and are identified in the context of evidence of chronic exposure (or repeated acute exposures)	Histological evidence of chronic cellular inflammation in nontarget tissues with local increases in proinflammatory biomarkers in exposed humans or animals Systemic increases in acute (with evidence of chronic exposure) or chronic inflammatory cells or proinflammatory biomarkers (e.g., cytokines, chemokines, or acute phase proteins) in circulation • Ideally, the biomarkers are associated with cancer and are identified in the context of evidence of chronic exposure (or repeated acute exposures) • Systematic inflammation may be more informative for blood cancers compared to solid tumors	Pre-diagnosis of increased IL-6, IL-8, CRP, SAA, and WBC associated with increased risk for several types of cancers.
KCC7: Is immunosuppressive	Evidence of increased viral infections in exposed humans or animals Impaired immune function, such as decreased antibody responses, NK, CTL, and T cell activation/activity in exposed humans or animals	Evidence of increased nonviral infections in exposed humans or animals and supporting evidence (e.g., alterations in immune components or organs) Severe decreases in WBC/leukocyte subsets and supporting evidence (e.g., decreased cytokines)	

KCC	Ideal Evidence for High Confidence	Other Evidence for High Confidence: Combination of Biomarkers/Indicators[a]	Biomarkers/Indicators and Cancer Prediction
KCC8: Modulates receptor-mediated effects	Direct receptor-mediated binding and transactivation (with effects on gene expression)	Indirect modulation of receptor effects Increased levels of endogenous hormones Overexpression of receptors for hormone sensitivity	ER-, PR-, AR-, AHR-mediated transactivation has been associated with cancer risk. Increased levels of endogenous estrogens, androgens, and thyroid hormones have been associated with cancer risk. Estrogen alone or estrogen-progesterone menopausal therapy and estrogen-progesterone oral contraceptives are known human carcinogens.
KCC9: Causes immortalization	Cell transformation as observed via (1) anchorage independence and loss of contact inhibition (foci formation), (2) de-differentiation, or (3) evidence of invasiveness • Inactivation of tumor suppressor genes or activation of oncogenes increases confidence. Increased telomerase activity in exposed humans or animals	Change in telomere length Inhibition of cellular senescence (assessed by multiple biomarkers)	Formation of cell foci has high concordance with cancer risk in rodents. Increased telomerase activity is a risk factor for cancer.
KCC10[c]: Alters cell proliferation, cell death, or nutrient supply (e.g., angiogenesis)	Hyperplasia or increased DNA synthesis during S-phase Increased pathogenic (sustained) angiogenesis and neoangiogenesis Glycolytic shift in exposed animals	Resistance to apoptotic cell death and other biomarkers of cellular proliferation	KCC represents multiple biomarkers, many of which are for tumor promotion or malignancy progression. Usually considered a consequence of cancer but may also be a cause of cancer.

Source (including citations): Appendix D and Smith et al. (2020).
8-OH-dG = 8-oxoguanine DNA glycosylase; AHR = aryl hydrocarbon receptor; AR = androgen; CA = chromosomal aberration; CTL = cytotoxic T lymphocyte; DNA = deoxyribonucleic acid; DRC = DNA repair capacity; ER = estrogen; h0GG1 = human 8-oxoguanine glycosylase-1; IL = interleukin; KCC = key characteristics of carcinogenicity; MDA = malondialdehyde; MN = micronucleus; NK = natural killer cells; PR = progesterone; RNA = ribonucleic acid; SAA = serum amyloid A; WBC = white blood cells.
[a]In the absence of studies on the ideal biomarkers/indicators, evidence from multiple biomarkers/indicators could be used to reach high confidence. We used the term indicators for effects that are not biomarkers per se, such as outcomes (e.g., increased infection for immunosuppression, histopathology, or organ weights).
[b]Not defined as a KCC by Smith et al. 2020; however, dysregulation/persistent immune activation (such as B-cell antigen activation) can contribute to the development and progression of cancer. The concept of immunomodulation is part of the 2015 Handbook, as recommended by the peer reviewers of that Handbook. See Appendix D for more details.
[c]Primarily from Smith et al. (2020). This will be updated as part of updates to the living document and could be considered multiple KCCs.

6.4.4. Step 2b: Assess the Confidence of the Evidence: Mode of Action and Adverse Outcome Pathway

MoA/AOP (or other biological pathway)-related influential questions are identified after broad literature searches and are unlikely to be the only questions for an evaluation. Confidence judgment is reached by assessing (1) the confidence that the substance causes the predominant biological effects in the proposed MoA/AOP or pathway (e.g., molecular initiating event or several key events) (direction provided in Section 6.4.2, Step 1 above) and (2) how well the data for the substance fit the proposed MoA (see Table 6-15). To illustrate the evaluation process, we discuss two exemplars of RoC-listed substances.

Table 6-15. Confidence Judgment for Mode of Action or Adverse Outcome Pathway

Descriptor	MoA
High	High/moderate confidence that the substance causes the key biological events in the proposed MoA The evidence is a "good fit" for the proposed MoA as demonstrated by: • High confidence for the most critical biological events in the MoA • The biomarker used to measure the specific biological event is the most relevant for the MoA • The sequence/timing and dose/response of events (e.g., semiquantitative) relative to each other and tumor formation are consistent with the proposed MoA
Moderate/Limited	High confidence that the substance causes an event in the proposed MoA, but the evidence is a moderate fit for the proposed MoA Moderate confidence for several key events in the MoA and moderate confidence for a good fit to the MoA
Inadequate	Inadequate confidence for several key events in the MoA

Exemplar: Cumene and Male Rat Kidney Tumors

For the cumene evaluation, a key question was whether the chemical causes renal tumors in male rats by an MoA that is considered not to be relevant to humans (i.e., alpha(2u)-globulin nephropathy). To answer this question, an information group applied criteria consisting of seven factors for 2u-globulin-associated nephropathy (IARC 1999) to the relevant data from cumene toxicity and other studies. The overall conclusion was that cumene exposure induces 2u-globulin-associated nephropathy in male rats; however, it is unclear whether other mechanisms may play a role in carcinogenicity. Not all the specific 2u-globulin-associated nephropathy factors were fulfilled. Cumene is listed as reasonably anticipated to be a human carcinogen based on sufficient evidence of carcinogenicity from animal studies, i.e., it causes lung and liver tumors in mice; kidney tumors in male rats provide supporting evidence (NTP 2013; 2021d).

Exemplar: Trichloroethylene and Non-Hodgkin's Lymphoma

For the TCE evaluation (NTP 2015), a key question was whether the available TCE immune studies supported a B-cell activation MoA leading to somatic hypermutation, proliferation of mutated B cells, and lymphoma. The information group concluded that although trichloroethylene or its metabolites can cause both immunosuppression and autoimmunity in humans and animals, the available studies in humans and animals measuring immune function

were not entirely consistent with this hypothesis. If there was more convincing evidence for this MoA, the collective LoE from studies in humans (human cancer epidemiological and mechanistic studies) might have been sufficient rather than limited (the evidence from human epidemiological studies was limited). Although this evaluation was prior to the 2016 publication on KCCs, conclusions of high confidence for chronic inflammation or immunosuppression may not have translated into an "upgrade" for a cancer-specific LoE conclusion.

6.4.5. Step 3: Assessing the Level of Evidence

The LoE conclusion of carcinogenicity from mechanistic evidence is reached by applying the RoC listing criteria to the body of literature. The final step integrates the confidence judgments across the influential mechanistic questions (Step 2) and considers the cohesiveness of the database to reach a conclusion for one or more LoE questions (Q1 to Q3, see Table 6-16). Each substance evaluation will have several questions, some of which may overlap. For example, evaluating the confidence for a read-across prediction includes assessing the evidence for an MoA as part of its uncertainty analysis (see Section 6.5.2). The review also summarizes the data that was not considered to be influential. The LoE of carcinogenicity from mechanistic studies is integrated with the LoE of carcinogenicity from cancer studies in humans and animals to reach a cancer hazard conclusion (i.e., listing recommendation) (see Section 7).

Considerations for reaching LoE conclusions are discussed below and summarized in Table 6-16, organized by the different LoE questions (see Section 6.1) and the strategies for evaluating those questions. For comprehensiveness, the table (but not the text) also includes read-across and clustering-related questions, which are discussed in Section 6.5.2.

Consideration of All Information

The RoC listing criteria note that cancer hazard conclusions (e.g., listing recommendations for *known or reasonably anticipated to be a human carcinogen*) "are based on scientific judgment with consideration given to all relevant information." These considerations can occur during the assessment of mechanistic evidence or overall cancer hazard evaluation (e.g., evidence integration of human and animal cancer and mechanistic studies; see Section 7). Examples of relevant information include mechanistic data (e.g., biological effects and mechanisms of action) and information related to ADME and toxicokinetics, which can potentially modify a mechanism. For hazard identification, qualitative (or very extreme quantitative) species differences are most relevant. Information on ADME and toxicokinetic/toxicodynamic data can be considered at all three steps in the process: evaluating individual studies or uncertainty in read-across approaches (Step 1), integration across study sets and the confidence judgments for biological pathways/MoA (Step 2), and as part of biological coherence in reaching an LoE conclusion (Step 3).

Substances often cause cancer via their metabolites, and metabolism data may be reviewed at multiple levels, e.g., electrophilicity is a KCC (and can be related to other KCCs), and metabolism to a reactive species could be a key event in an MoA or a part of read-across analyses. The RoC lists chemicals (or classes of chemicals) based on their metabolism to a human or animal carcinogen (e.g., dyes metabolized to benzidine). In these cases, the ADME evaluation (see Section 5) would assess the extent and the nature of the metabolism to the carcinogen, e.g., the amount of metabolite produced, the likelihood the metabolic pathways occur in humans, and data on whether the target chemical metabolism results in the proximate

metabolites of the carcinogen. The overall cancer hazard evaluation (Section 7) integrates the metabolic and relevant mechanistic information with cancer data for the metabolite to reach a listing recommendation.

Biological Effects (KCCs): LoE Q1

Assessing the LoE of mechanistic data based on biological KCCs considers many factors. Although the KCCs, by definition, are properties of carcinogens, they are not necessarily specific to a cancer type, and other adverse health effects share some of these KCCs. Thus, in the absence of other supporting data, high confidence for a single informative KCC is typically not enough to reach c*onvincing* mechanistic evidence for cancer causation but may be adequate for *supporting* evidence. Mechanistic understanding of a substance must be compelling to list a substance with inadequate or limited cancer data. For substances that meet the RoC criteria for listing based on human or animal cancer studies, high confidence for a biological effect for most KCCs may be sufficient to support a moderate LoE from mechanistic studies. Scientific judgment and assessing the coherence of the database are critical for the assessment.

The number, connections, and specificity of the KCCs or other biological effects

A given evaluation may evaluate the evidence for several biological effects, usually KCCs. Information on connections, sequence, and timing of several KCCs may help identify substance-specific biological pathways and increase the certainty of the evidence. The specificity of the KCC toward cancer also contributes to the LoE conclusion as many KCCs are important in nonmalignant disease. Specificity for cancer (e.g., some KCCs are also characteristics of noncancer toxicants) varies across KCCs and may depend on the chemical. For example, inducing oxidative stress and chronic inflammation are less specific for cancer whereas causing immortalization has higher specificity; most others KCCs are in a continuum between these examples.

In addition to considering the specificity of the KCC, the informativeness of the specific biomarker (or several biomarkers) used to measure the KCC is paramount. We evaluate this at all three steps in the process with increasing stringency. For high confidence in biological effects, the biomarker should be highly relevant for the KCC (e.g., mutation, chromosomal damage), and for convincing LoE, it should also predict cancer. For the latter decision, supporting cancer data or strong evidence establishing a cancer mechanism is needed.

Supporting cancer data

In addition to informing the overall cancer hazard evaluation, cancer studies in humans and animals can inform the assessment of mechanistic evidence, such as providing information on the cancer site. Even cancer data not meeting the RoC listing criteria can inform the evaluation of mechanisms. For instance, data on a cancer site may be based on an animal database consisting of a single treatment-related tumor site in one species or a human database of a few studies suggesting a link with a specific type of cancer.

To demonstrate that a specific biomarker predicts cancer or is involved in tumor formation, we ideally would have data that suppression of an adverse biological effect led to suppression of tumor development. For example, a series of studies using different designs of simulated light at night (LAN) in experimental animals demonstrated that exposure to LAN suppresses melatonin (which is protective of cancer) and low melatonin was associated with breast (human implant

studies) or mammary tumor progression (NTP 2021a). Cancer studies measuring KCCs in the same study can also help establish a link between KCCs and substance-specific cancer mechanisms as illustrated in the night shift work evaluation: Several KCCs found in night shift workers were the same as those detected in the animal cancer studies in which animals were exposed to simulated night shift work or jet lag using changes in light patterns (NTP 2021a). Meet-in-the-middle approaches—linking a biomarker (e.g., metabolomics) to both exposure and disease using a prospective epidemiological study design—can also provide evidence of a cancer link in these studies (Chadeau-Hyam et al. 2011). As multiomic approaches advance, these studies may play a greater role in future evaluations. Overall cohesiveness and concordance between a cancer type and the KCC may strengthen the informativeness of the KCCs or other biomarkers.

Biological Pathways/MoA: LoE Q2 and Q3

Typically, these influential mechanistic questions are for cancer site-specific MoA related to biological plausibility or an AOP for an established cancer pathway. LoE conclusions may be cancer site-specific and integrated with human and animal LoE for a specific cancer site in the overall evaluation (see Section 7). Compelling evidence for nonrelevance in humans is needed to discount strong evidence of carcinogenicity from studies in experimental animals. "Compelling" is defined as high confidence for MoA/AOP with scientific consensus that it would not occur in humans, inadequate human cancer data (at any site), and inadequate human-relevant biological effects in the target tissue. Other MoA (Q2) provide evidence of biological plausibility and may be for a specific cancer site (such as those observed in humans) or cancer site agnostic. In general, the confidence judgment for the MoA translates to the LoE conclusion of carcinogenicity of a substance from mechanistic studies (e.g., high confidence to convincing evidence, moderate confidence to supporting evidence).

Table 6-16. Level-of-evidence Guidelines

RoC Criterion	Convincing	Supporting
Q1. Plausibility: Biological effects KCC or other relevant effects, including metabolism Connectivity between KCCs	Coherence in the database (including mechanistic data, toxicokinetics) AND High confidence for one or more *informative* biological effects considering the following[a]: • Number, sequence, connections between KCCs leading to the development of a biological pathway • Specificity and informativeness of KCC biomarkers • Supporting cancer-related data • Suppression of biological effect led to suppression of tumor development • Biological effect measured in cancer studies (typically in vivo) exposed to the substance • Measured in target or cancer site of interest	High or moderate confidence for informative KCCs • If the confidence judgment for KCC is moderate confidence, a greater number and specificity of the KCC is needed • If the confidence judgment is high, a single specific KCC or several less specific KCCs would meet the requirement

RoC Criterion	Convincing	Supporting
Q1: Plausibility: Reported biological pathway/MoA	High confidence for an MoA that is for one of the following: • Cancer site observed in humans or experimental animals • Established cancer pathway	Moderate confidence for an MoA or AOP
Q1 Plausibility: Predicted to cause cancer[b] Q2: Member of a listed class Clustering and read-across approaches	High confidence for read-across/clustering • Low uncertainty encompasses evaluation of MoA for the source chemicals and biological effects of predicted chemicals[d]	Not relevant[c]
Q3. Not relevant to humans/MoA (Compelling)	High confidence for an MoA for a cancer site observed in animal studies Strong scientific consensus that the MoA is not relevant to humans Low confidence for KCCs in the target tissue Inadequate[e] mechanistic data in exposed humans	Not relevant[c]

[a]Scientific judgment is used to integrate the issues in the bullet, and not all are required
[b]See Section 6.5.
[c]"Supporting" is not relevant because these questions are for when there is not an adequate database from human or animal cancer studies, or the evidence is downgrading the studies in experimental animals.
[d]Source chemicals are associated with tumors in experimental animals and possibly in humans and can be listed or nonlisted substances.
[e]"Inadequate" means that either no human data are available or no data in humans supporting biological plausibility are available.

6.5. Read-across Approaches

Because read-across approaches for cancer hazard evaluations are continuously advancing and are evaluation specific, the substance(s) evaluation protocol can more accurately capture the state of the science on read-across approaches than the handbook. The handbook provides high-level concepts for scoping, developing a protocol and evaluating read-across approaches. As a living document, we will update the handbook to reflect advancements in knowledge and technology.

6.5.1. Identification of Read-Across as an Influential Question

As discussed in Section 1, substances are selected for review for the RoC based on initial scoping and problem formulation activities. During this process, substances with cancer data may lead to expanding the scope to evaluate a potential class, subclasses or structurally related chemicals. These chemicals may (e.g., nitro-polycyclic aromatic hydrocarbons) or may not (haloacetic acids found as drinking water contaminants, halogenated organic flame retardants) be related to substances listed in the RoC (Lunn et al. 2022).

Using the methods outline in Section 6.2, RoC staff conduct scoping and problem formulation activities to identify influential questions: scientific issues that will most likely impact the LoE conclusions. If read-across is identified as influential questions for substances with limited human and animal cancer data and that share similarity with substances with cancer data, we

search the literature (using the methods outline in Section 6.2) for published or validated read-across approaches (e.g., analogue or category). If found, the evaluation team may consult with technical advisors to assess its quality and will summarize the analyses in the protocol. If there are no published read-across methods or analyses for carcinogenicity, RoC will conduct further scoping activities to develop a read-across framework (Section 6.5.2).

6.5.2. Read-across Framework Development

Scoping activities for read-across approaches are complex and involve preliminary analyses in addition to information searches (see Section 6.2 for methods for identifying and mapping studies and data) to develop the read-across framework, including the read-across hypothesis and argumentation, the definition of analogues (source and target chemicals), and approach. A series of iterative questions, including questions about the source of uncertainty (Schultz et al. 2015; Schultz et al. 2019) and workflows developed by others (e.g., Ball et al. 2016; OECD 2023b; Patlewicz et al. 2015; Patlewicz et al. 2017), guide the process.

Key events (AOP) or key intermediate steps (MoA) represent a sequence of biologically relevant multiple steps leading to carcinogenicity. Hence, reviewing the quality/informativeness of carcinogenicity studies in humans and/or experimental animals and assessing mechanistic evidence for the source chemicals are critical to evaluate the feasibility of using read-across (see Section 6.5.3). Importantly, the selection of the final list of analogues for read-across ascertains relevant features, such as physical-chemical properties, toxicokinetic-related properties (e.g., bioavailability, metabolism, and degradation pathways), and toxicodynamic properties of chemicals with an emphasis on MoA. Thus, we will also consider endpoint-specific (carcinogenicity and intermediate endpoints) parameters, properties, and chemical biological predictive data beyond 2-D. This will expand the definition of structural for biological similarity and provide a different critical context for analogue selection. This is also intended to create a transparent rationale that supports the selected read-across analogue(s) for the specific endpoint under study (Lester et al. 2023; Moustakas et al. 2022). Analogue selection is discussed further below.

For the purpose of the handbook, we provide general concepts and activities, rather than specific workflows, for scoping and developing a read-across framework plan (see Blackburn and Stuard 2014; ECHA 2017; 2023; Myatt et al. 2018; OECD 2023a; Patlewicz et al. 2018) for more information on read-across tools, flowcharts, and other methods). Given the broad range of complexity and diversity of substances selected for review and the limited, if any, case studies of read-across analysis for cancer, we feel this is appropriate.

Read-Across Purpose

Our read-across predictions require low uncertainty for the cancer prediction and incorporation of biological effects into the model as their purpose is to list or not list substance *in the RoC, a public health congressionally mandated document.*

Identifying the Sources and Negatives

During the scoping activities (see Section 6.2), we search for and extract the findings of the available cancer studies in exposed humans and experimental animals to identify potential positive (sources) and negative chemicals to be proposed in the read-across plan. The quality of the endpoint or outcome data (e.g., cancer) provides one of the fundamental sources of

uncertainty in a read-across strategy. Data with low uncertainty are sought, which may include aspects of the reliability of the data, relevance and specificity to the endpoint of carcinogenicity in exposed humans and experimental animals, as well for human health/environmental effects (Schultz et al. 2019). The final study informativeness and cancer hazard assessment, following the systematic review and evidence integration methods outlined in Sections 3 and 4, informs the uncertainty analysis (see Table 6-17).

Assessing the Mechanistic Evidence and Developing the Read-Across Argumentation

The read-across argumentation encompasses the justification for using read-across based on a developed mechanistic hypothesis. Substantial scoping of mechanistic studies of the source chemicals are done to identify proposed mechanistic frameworks (e.g., MoA, AOP, molecular pathways, metabolism). The read-across hypothesis is based on an identified common biological framework (e.g., the molecular initiating event [MIE] and one or more key events that lead to carcinogenicity). The applicability domain is the chemical, biological, toxicodynamic, or toxicokinetic properties required to identify analogues for read-across (Pestana et al. 2022). Biological, chemical and toxicokinetic similarity, often based on the MoA for cancer, across source chemicals is key to defining the applicability domain. The overall mechanistic assessment will evaluate the confidence in the proposed MoA or biological effects (see Section 6.4), which is an uncertainty factor for read-across (see Table 6-17).

Key descriptors, including properties and parameters of source chemicals related to the MIE and intermediate endpoints, are identified and used to create a data matrix (see Section 6.6.1) for the source chemicals and potential negatives if available. These descriptors may include structures and functional groups, or parameters such as molecular surfaces and volumes, data on relevant chemical biology interactions, ADME properties, or biological activity. Data and descriptors on carcinogenicity and intermediate endpoints can be found in the literature, or from data repositories, or may be calculated. Preliminary examination of the sources (category or not) and available potential negatives can also help deciding which descriptors are more indicative and predictive to discriminate active from inactive substances.

Selecting and Evaluating the Final List of Analogues, Data Matrix and Requirements

Sources and negatives are initially identified through scoping activities from available cancer studies in exposed humans and experimental animals based on the substance under review (discussed above). The final selection of the analogue sources, or clusters of analogue sources, may be redefined with new ones added which may extend outside the chemical class under review and others removed if they are not endpoint specific, based on the assessed mechanistic evidence and developed read-across argumentation (e.g., similar MoA, toxicokinetics), and on the defined appliability domain. We will use authoritative reviews, national and international chemical information and data repositories many of which can be found in the list of general sources in Appendix B (e.g., EPA CompTox Chemicals Dashboard, NTP Integrated Chemical Environment, OECD eChemPortal, ECHA REACH) to identify analogue candidates for read-across within and outside the chemical class under review (Moustakas et al. 2022). To select the final list of analogues, a supervised approach that is based on selected endpoint-specific parameters, properties, predictive data beyond 2-D (for both carcinogenicity and intermediate endpoints) will be used. Target chemicals are analogues that are not sources or potential negatives. The data matrix is expanded to include the identified descriptors for the target chemicals. Like sources, properties, parameters, predictive data for the target chemicals found in

literature, data repositories, or calcuated can be evaluated. The data matrix will identify data rich and data poor targets, the most studied biological effects, ADME, and how well the available data align with the proposed MoA or AOP. Other searches are related to the quality and quantity (i.e., availability) of data for read-across. Data for the data matrix will come from authoritative source (e.g., EPA's CompTox or LeadScope), and will be supplemented with data from primary studies as needed. The data will be evaluated for study informativeness using the methods outlined in Section 6.3.

As part of the evaluation of the final list of analogues, the similarity between targets and sources and fit between the applicability domain is determined using the data collected in the matrix. We also reevaluate the read-across argumentation to determine whether a read-across approach is feasible and to decide which approach (analogue or category) would be used (Patlewicz et al. 2018; Schultz et al. 2019).

Uncertainty Assessment

We assess uncertainty from several sources or types per source—including cancer data for source chemicals, biological plausibility, and similarity justification. These uncertainty sources are part of the pre-read-across implementation during scoping (Table 6-17, Section 6.5.3) and are integrated with other uncertainty types (from the read-across analysis, see Table 6-17) to reach an overall judgment (see Table 6-18, Section 6.5.3). Importantly, we would only develop a read-across plan when we are confident in these factors (Schultz et al. 2019).

Protocol Development

Informed by scoping activities, the protocol will provide the read-across hypothesis and elements in the prioritized read-across plan (see Box 6-5 for examples). Critically, a hypothesis stating the chemical and biological mechanisms underpinning the carcinogenic effect for reading across the candidate substance or group of related substances is needed. Ideally, it will also include alternative plans (e.g., changing from category to an analogue approach, computational to a qualitative noncomputational approach, or adding new chemicals from animal evidence only) if the prioritized plan does not work. For published read-across analyses, the protocol will provide the same information as above and an assessment of the rationale and quality of reported read-across methods.

Box 6-5. Read-Across Plan Elements

Definition and members of the category, including source and target chemicals

Scoping activities and analyses summary

- Biological and chemical evidence maps for the chemicals
- Specific criteria for analogue selection
- Similarity justification of target and source chemicals
- Preliminary assessment of human and animal cancer (e.g., animal cancer data)

Framework, read-across methods

- Read-across argumentation
- Type of approach: category or analogue
- Descriptors, parameters, properties
- Data matrix template and types of input data
- In silico quantitative prediction and clustering methods, classification, visualization, statistical analysis, tools

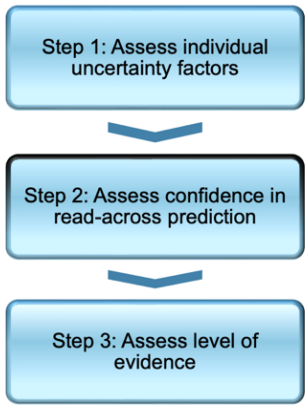
Considerations to evaluate uncertainty

6.5.3. Evidence Evaluation and Integration

Evaluation of the confidence of the read-across prediction for each target chemical (or class of chemicals) considers the strength (e.g., scientific judgment, correlation, nearest neighbor) of the similarity between the target and source chemicals and for those targets with high and moderate prediction, assessing the uncertainty associated with the cancer prediction and read-across process (Schultz et al. 2015). The evaluation may be of a reported read-across/clustering analysis or a new model conducted for the RoC evaluation. For RoC evaluations, first, we rate each uncertainty type (Step 1) and then integrate the ratings across types to reach an overall judgment (Step 2). The confidence judgment for the cancer prediction is integrated with other evidence to reach a level-of-evidence conclusion (Step 3) (see Box 6-6).

Step 1: Rating Uncertainty

Box 6-6. Evidence Integration for Read-Across



Among the many resources for evaluating uncertainty, our evaluation adapts a series of questions proposed by Schultz et al. (2019) for each type and source of uncertainty. Uncertainty types or questions are organized by the read-across process (see Table 6-17) and linked to their source(s): (1) quality of the cancer data for the sources, (2) read-across argumentation, and (3) similarity justification. Read-across argumentation refers to the plausibility of the cancer mechanisms and the completeness, quality, and robustness of the supporting evidence. Similarity justification refers to the similarity (chemical structure, physio-chemical properties, toxicodynamics, and toxicokinetics) between the source and target chemicals. Some uncertainty types are from more than one source.

Table 6-17 provides information—source, evidence base (e.g., study sets, data), and considerations (derived from questions)—for eight uncertainty types, organized by process step. Each uncertainty type (minor, moderate, or major) is rated using scientific judgment and, for some cases, based on conclusions reached for other parts of the evaluation (e.g., confidence conclusions for MoA, biological effects, or LoE of carcinogenicity from animal studies).

Table 6-17. Types and Sources of Uncertainty in the Read-across Analyses for Each Target Chemical

Type[a]	Source	Process Step	Evidence or Data	Considerations
1. Definition and grouping of chemicals as analogues or as a category	Read-across Argumentation Similarity justification	Pre-read-across implementation	Data matrix tables and evidence maps: Physio-chemical properties, mechanistic data	Targets and source chemicals are clearly defined and fit with in the applicability domain
2. Quality of the cancer data for the sources	Source chemicals	Pre-read-across implementation (preliminary) Cancer hazard evaluation (final)	Assessment and conclusions from human and animal cancer studies[b]	LoE conclusions from human and animal cancer studies including study informativeness (bias and study sensitivity) and evidence for tumor formation for source chemicals Presence of presumed negatives control • Study sensitivity for reported negative findings
3. Biological plausibility	Read-across Augmentation	Pre-read-across implementation (preliminary) Mechanistic evidence evaluation (final)	Authoritative reviews and evidence maps Confidence in the biological framework[c] Confidence in biological effects[c]	Sufficiency and completeness of the understanding of the MoA or AOP supported by experimental evidence • Strength: number of key events and number of sources/targets tested for a key event • Consistency within a key event • Selectivity: discriminate between positive and negative
4. Chemistry similarity	Similarity justification	Read-across conduct	Read-across analyses Data matrix	Chemical structures available for the substances Structural and/or biological similarity between targets and sources (e.g., chemical structures, 2D-3D-parameters, and properties) are toxicologically or pharmacologically relevant to the endpoint

Type[a]	Source	Process Step	Evidence or Data	Considerations
5. Toxicodynamic/ toxicokinetic similarity	Similarity justification	Read-across conduct	Read-across analyses Data matrix	Toxicodynamic and toxicokinetic (ADME) similarity (and not dissimilarity) are toxicologically or pharmacokinetically relevant to the endpoint. Metabolism data are available on one or more source chemicals
6. Concordance of the animal cancer data with intermediate effects and potency data (source chemicals)	Source chemicals Read-across Augmentation	Read-across conduct	Animal cancer assessment Read-across analyses output Data matrix including dose-response data	Consistency of cancer potency (if available) across source chemicals Dose-response and temporal relationship between relevant endpoints
7. Strength or robustness of the supporting data sets	Read-across Augmentation	Read-across conduct	Read-across analyses Data matrix Data gap filling	Consistency of the measurement or modeled data of the key events in the studies (in silico, in vitro, in chemico, nonstandard in vivo)

[a]Factors are identified and modified from 12 factors in Table 3 (Schultz et al. 2019). Some factors were combined; documentation and context (which is the RoC purpose) is not included. We access the quality of the cancer evidence as part of the scoping activities, and this was moved up in the workflow.
[b]Conclusions reached following the approach in the evaluation of Cancer Studies in Experimental Animals (Section 4).
[c]Conclusions reached following the approach in the confidence judgment assessment for biological effects and pathway, Section 6.4.2 (Steps 1 and 2).

Step 2: Assess the Confidence for the Read-Across Prediction

Box 6-7 provides considerations for reaching a confidence judgment for read-across analyses. Assessing the strength (e.g., the similarity between target and source chemicals) of the cancer prediction for each target chemical or group of chemicals depends on the nature of the substance(s) under evaluation, and guidance is provided in the read-across plan.

Box 6-7. Confidence Judgments for Read-Across Predictions

Descriptor	Read-across Considerations
High	High in silico prediction (e.g., high-moderate correlation) AND Overall low uncertainty from the read-across analyses
Moderate	High in silico prediction AND Overall medium uncertainty from the read-across analyses
Inadequate	Low in silico prediction High in silico prediction AND Overall high uncertainty from the read-across analysis

In silico predictions can be generated through supervised (e.g., K-nearest neighbors, support vector machines) or unsupervised algorithms (e.g., clustering techniques). More specific statistical analysis will be provided in the provided in the protocol.

Overall uncertainty analysis is conducted for those target chemicals with high in silico predictions and determines the confidence judgment for the cancer prediction (e.g., low uncertainty translates to high confidence). Using weight-of-evidence and triangulation approaches, consideration of substance-specific issues, and scientific expert appraisal, we integrate the ratings across the uncertainty types (Step 1, Table 6-17) to reach an overall judgment. Minor uncertainty from all contributors is not required, and ratings for some factors may compensate for lower ratings for other factors. However, some types, such as the quality of cancer data for the sources or biological plausibility, are more relevant in the assessment. Indeed, to conduct a supervised read-across analysis, high confidence in the (1) category definition, (2) carcinogenicity of the source chemicals, and (3) biological plausibility of the read-across hypothesis is required.

Notably, these are considerations and not criteria for ratings and judgments and allow flexibility for expert opinion and substance-specific issues. Moreover, read-across methodologies are rapidly advancing, and this approach will adapt to include these innovations.

"Qualitative" clustering or read-across analyses

For some assessments, a "qualitative" (is the substance predicted to cause cancer) rather than a quantitative read-across and clustering analysis may be pursued. These analyses involve constructing data matrix tables of pertinent data, including trend analyses, and evaluating the uncertainty of the data, but do not calculate a statistical cancer prediction value (e.g., correlation). The uncertainty analyses are consistent with many of the factors identified for in silico models, such as confidence in the carcinogenicity of source chemicals, biological plausibility (e.g., MoA, toxicokinetics), and consistency of the biological effects and physical-chemical properties of the target chemicals within a category. However, as new read-across methods emerge, this approach may no longer be relevant.

Exemplar: Haloacetic acids found in disinfected drinking water

The NTP review of 13 HAAs applied a read-across data matrix table of physical and biochemical properties, biological effects (KCCs), and animal cancer data to assess data gaps and trends for possible listing of HAAs as a class, subclasses, or individual HAAs without human or animal cancer data (Atwood et al. 2019; NTP 2018b). Given low or inadequate confidence for the category that was due to inconsistency in the carcinogenicity or biological mechanisms for the source chemicals, no overall class, or subclasses (e.g., number and type of halogen) could be identified. However, using a read-across analogue approach based on the confidence of metabolism to a carcinogenic HAA and the strength of the mechanistic data, two additional haloacetic acids without animal or human carcinogenicity data were listed in the RoC. The analogue approach predicted specific cancer sites for the target based on the source chemicals.

Step 3: Assessing the Level of Evidence

The confidence judgment for read-across is integrated with all the relevant evidence to reach an LoE of carcinogenicity from mechanistic studies. A given evaluation may have several influential questions. For listing a substance as *a reasonably anticipated carcinogen in the RoC, a public health congressionally mandated document, l*ow uncertainty of the cancer prediction is required. The uncertainty assessment integrates the confidence of the MoA for the source chemicals and the biological effects of the target chemicals. Thus, high confidence for read-across translates to convincing LoE. However, evidence for biological effects not related to the MoA for a target chemical, depending on the adequacy of the database, may be integrated into the final assessment. As read-across predicts *cancers for substances with little human and animal cancer data, moderate confidence does not translate into a listing recommendation (e.g., supporting LoE) but may identify research gaps.*

Table 6-18. Step 3 Level of Evidence: Read-across Approaches

RoC Criterion	Convincing	Supporting
Q1 Plausibility: Predicted to cause cancer Q2: Member of a listed class	High confidence for read-across • Low uncertainty encompasses evaluation of MoA for the source chemicals and biological effect that are part of the evaluation for the target chemicals	Not relevant

6.6. Reporting

Information on reporting the cancer hazard evaluation is provided below, including methods to capture data extraction and study informativeness (Section 6.6.1), matrix table templates used in the read-across analysis (Section 6.6.2), and evidence-based tables for evaluating biological effects and read-across evidence (Section 6.6.3).

6.6.1. Data Extraction and Study Informativeness

Data extraction and study informativeness are captured in content management systems or software such as HAWC, Table Builder, Excel files, or Word tables. When relevant, findings may be visualized using software such as Tableau. The extent of the data extraction depends on how influential the studies are to the overall assessment (see Section 6.1, Fit-for-purpose).

6.6.2. Data Matrix

Table 6-19. Data Matrix Example for Read-across Analyses

Read-across Parameter	Chemical 1	Chemical 2
Identifiers (e.g., ID, Nrs)		
Chemical Structure		
PC Property 1		
PC Property 2		
3D Parameter		
TK 1: Metabolism		
Biological Endpoint 1		
In Vivo Carcinogenicity		

6.6.3. Evidence-based Tables

Biological Effect Table

Examples of evidence-based tables for Steps 1 and 2 for evaluating biological effects:

Table 6-20. Step 1: Study Sets

Exposure	Endpoint	Evidence Type	Strength and Limitations	Assessment
Chemical X	Mutations	Exposed humans (nine studies)	Potential confounding	Consistent findings Some uncertainty
Chemical X	Mutations	Animal in vivo studies (five studies)	Acute exposure	Consistent findings Some uncertainty

Table 6-21. Step 2: Confidence Judgments

Influential Question	Outcomes	Study Sets	Assessment Summary and Rationale	Confidence Judgment
Strength of evidence for genotoxicity	Mutations Chromosomal damage	In vitro (human cells or mammalian cells) In vivo Exposed humans: mutation	Conclusion and uncertainty for each study set assessment Triangulation-like approach regarding biases across assessments. Informativeness of the KCC biomarkers	Moderate: exposed humans Note calls for individual biomarkers Note ex vivo or in vitro from human cells/tissues

Read-across Evidence

Examples of evidence-based tables for Steps 1 and 2 for evaluating read-across:

Table 6-22. Step 1: Uncertainty Analysis

Influential Question	Uncertainty	Evidence Type	Strengths and Limitations	Rating
Read-across Target 1	Cancer data for source	Assessment and conclusions from animal cancer studies	Source chemicals are predicted to have sufficient evidence.	Minor uncertainty
Read-across Target 1	Chemical similarly	Read-across analysis Data matrix	Similarities between chemicals are toxicologically relevant.	Minor uncertainty

Table 6-23. Step 2: Confidence for Read-across

Influential Question	Evidence	Assessment Summary and Rationale	Confidence Judgment
Read-across prediction: Target 1	In silico prediction Uncertainty contributors	Conclusion for each uncertainty factor Weight of evidence/triangulation	Low uncertainty

6.7. New Directions

NAMs encompass a wide range of innovative techniques, including omics, imaging, in vitro assays, computational modeling, organ-on-a-chip platforms, high-throughput screening, and in silico approaches. Incorporating NAMs into the KCC framework represents a promising avenue for enhancing cancer hazard identification. These methodologies offer a deeper understanding of the intricate molecular mechanisms and pathway interactions involved in carcinogenesis, providing valuable supportive evidence within the KCC framework. In addition, when applied in the context of other evolving features of cancer, which include phenomena such as phenotypic plasticity, immune evasion, and polymorphic microbiomes, NAMs can provide insights into how chemicals influence these pathways, contribute to carcinogenesis, and elucidate their MoA or AOP. NAMs enable the investigation of complex interactions between chemicals and biological systems, including the tumor microenvironment and host immune response. By simulating these interactions in vitro or in silico, significant progress can be made in understanding how chemicals modulate the tumor microenvironment, promote neoangiogenesis and morphological changes, and evade immune surveillance, all of which are critical aspects of cancer development and progression. In the future, this ensures a comprehensive evaluation, that not only detects the early stages of carcinogenesis, but also acknowledges the acquisition of a complete malignant phenotype, providing a system toxicology-oriented and more holistic understanding of the carcinogenic potential of the tested chemicals. However, the successful integration of NAMs faces several challenges, notably the need to demonstrate the relevance, reliability, reproducibility, and predictive capability of NAMs. Achieving this goal would entail identifying pertinent studies that have effectively employed these methods, establishing standardized experimental protocols to ensure consistency, and validating the models to ensure the relevance, accuracy, and reliability of carcinogenic prediction and evaluations.

Furthermore, leveraging artificial intelligence (AI) techniques, such as machine learning and deep learning, within NAMs, enhances analyses of large datasets, discerns patterns, and predicts biological responses to chemical exposure more precisely, i.e., Read-Across Structure-Activity Relationship (RASAR) tool, thereby accelerating the identification of potential hazard (Luechtefeld et al. 2018). These advanced tools can offer valuable help in extrapolating quantitative dose-response relations, thereby strengthening the evidence base for predicting carcinogenicity and identifying mechanistic classes of agents already associated with cancer. Integration of multiple types of omic data from the same patient has been used in clinical research to characterize molecular and clinical features of cancers, which can inform treatment, monitor disease progression, and improve survival (Heo et al. 2021). Multiomic data capture biological responses or permutations caused by environmental or endogenous exposures and often refer to the genome, epigenome, transcriptome, proteome, and metabolome (Wu et al. 2023). Interest in integrating these data with exposure information to elucidate molecular mechanisms and pathways of human diseases (including cancer) is increasing. Omic data can capture multiple molecular mechanisms and represent (e.g., epigenetics) or measure KCCs or they might be molecular responses not fully captured by the 10 KCCs. Biological responses can be measured in samples from model systems (e.g., exposed animals or cells) or humans (epidemiological or clinical studies). New analytical technology in recent years, which can measure thousands of metabolites simultaneously, has advanced the utility of metabolomics in cancer research, including the creation of the Consortium of Metabolomics Studies. Studies (typically called metabolomic-wide association studies) have linked metabolites to exposure, cancer, and occasionally both. Although oncogenic metabolites (often endogenous) are associated to specific cancers and the hallmarks of cancer (Wishart 2022), we are not aware of any formal assessment of metabolomic data as a possible independent characteristic of carcinogens. Analogous to gene expression profiling, metabolites are often linked to biological pathways (Wieder et al. 2021), which may be related to the KCCs. Another growing advancement are microphysiological systems, which offer the potential to mimic tissue dynamics in vitro that could allow characterization of molecular pathway effects and their feed-forward progressions to toxicological phenotypes recognizable by pathologists.

Psychosocial stressors often share the same biological pathways; multiomics can help assess environmental exposure and psychosocial stressor interactions and shed light on contributors to health disparities and sensitive subpopulations. Identification of effect modifiers can help inform the interpretation of human cancer studies. Recent (2020 to 2023) publications on new methods for carcinogenic testing (e.g. Jacobs et al. 2020; NASEM 2023; Oku et al. 2022), and two workshops [Integrating Environmental Exposure Data with Other Omic Data for Cancer Epidemiology and Molecular Signatures of Exposure in Cancer (NIEHS 2023a; 2023b)] may help inform cancer hazard assessments and classifications.

This handbook will be updated as the scientific consensus on validation criteria is reached for using multiomics, NAMs, or AI in cancer hazard assessment advances.

References

Atwood ST, Lunn RM, Garner SC, Jahnke GD. 2019. New perspectives for cancer hazard evaluation by the Report on Carcinogens: A case study using read-across methods in the evaluation of haloacetic acids found as water disinfection by-products. Environ Health Perspect. 127(12):125003. https://doi.org/10.1289/ehp5672

Ball N, Cronin MTD, Shen J, Blackburn K, Booth ED, Bouhifd M, Donley E, Egnash L, Hastings C, Juberg DR, et al. 2016. Toward Good Read-Across Practice (GRAP) guidance. ALTEX. 33(2):149-166. https://doi.org/10.14573/altex.1601251

Blackburn K, Stuard SB. 2014. A framework to facilitate consistent characterization of read across uncertainty. Regul Toxicol Pharmacol. 68(3):353-362. https://doi.org/10.1016/j.yrtph.2014.01.004

Chadeau-Hyam M, Athersuch TJ, Keun HC, De Iorio M, Ebbels TMD, Jenab M, Sacerdote C, Bruce SJ, Holmes E, Vineis P. 2011. Meeting-in-the-middle using metabolic profiling – A strategy for the identification of intermediate biomarkers in cohort studies. Biomarkers. 16(1):83-88. https://doi.org/10.3109/1354750X.2010.533285

European Chemicals Agency (ECHA). 2017. Read-Across Assessment Framework (RAAF). Helsinki, Finland: European Chemicals Agency. ECHA-17-R-01-EN. https://echa.europa.eu/documents/10162/17221/raaf_en.pdf

European Chemicals Agency (ECHA). 2023. Grouping of substances and read-across. Helsinki, Finland: European Chemicals Agency. [Accessed: February 14, 2023]. https://echa.europa.eu/support/registration/how-to-avoid-unnecessary-testing-on-animals/grouping-of-substances-and-read-across

Guyton KZ, Kyle AD, Aubrecht J, Cogliano VJ, Eastmond DA, Jackson M, Keshava N, Sandy MS, Sonawane B, Zhang L, et al. 2009. Improving prediction of chemical carcinogenicity by considering multiple mechanisms and applying toxicogenomic approaches. Mutat Res. 681(2-3):230-240. https://doi.org/10.1016/j.mrrev.2008.10.001

Guyton KZ, Rusyn I, Chiu WA, Corpet DE, van den Berg M, Ross MK, Christiani DC, Beland FA, Smith MT. 2018. Application of the key characteristics of carcinogens in cancer hazard identification. Carcinogenesis. 39(4):614-622. https://doi.org/10.1093/carcin/bgy031

Hanahan D. 2022. Hallmarks of cancer: New dimensions. Cancer Discov. 12(1):31-46. https://doi.org/10.1158/2159-8290.CD-21-1059

Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: The next generation. Cell. 144(5):646-674. https://doi.org/10.1016/j.cell.2011.02.013

Heo YJ, Hwa C, Lee GH, Park JM, An JY. 2021. Integrative multi-omics approaches in cancer research: From biological networks to clinical subtypes. Mol Cells. 44(7):433-443. https://doi.org/10.14348/molcells.2021.0042

Howard BE, Phillips J, Miller K, Tandon A, Mav D, Shah MR, Holmgren S, Pelch KE, Walker V, Rooney AA, et al. 2016. SWIFT-Review: A text-mining workbench for systematic review. Syst Rev. 5:87. https://doi.org/10.1186/s13643-016-0263-z

International Agency for Research on Cancer (IARC). 1999. Species differences in thyroid, kidney and urinary bladder carcinogenesis. Capen CC, Dybing E, Rice JM, Wilbourn JD, editors. Lyon, France: International Agency for Research on Cancer. https://publications.iarc.fr/Book-And-Report-Series/Iarc-Scientific-Publications/Species-Differences-In-Thyroid-Kidney-And-Urinary-Bladder-Carcinogenesis-1999

International Agency for Research on Cancer (IARC). 2019. Preamble to the IARC monographs. Lyon, France: International Agency for Research on Cancer. https://monographs.iarc.who.int/iarc-monographs-preamble-preamble-to-the-iarc-monographs/

Jacobs MN, Colacci A, Corvi R, Vaccari M, Aguila MC, Corvaro M, Delrue N, Desaulniers D, Ertych N, Jacobs A, et al. 2020. Chemical carcinogen safety testing: OECD expert group international consensus on the development of an integrated approach for the testing and assessment of chemical non-genotoxic carcinogens. Arch Toxicol. 94(8):2899-2923. https://doi.org/10.1007/s00204-020-02784-5

Kienzler A, Barron MG, Belanger SE, Beasley A, Embry MR. 2017. Mode of action (MOA) assignment classifications for ecotoxicology: An evaluation of approaches. Environ Sci Technol. 51(17):10203-10211. https://doi.org/10.1021/acs.est.7b02337

Lester C, Byrd E, Shobair M, Yan G. 2023. Quantifying analogue suitability for SAR-based read-across toxicological assessment. Chem Res Toxicol. 36(2):230-242. https://doi.org/10.1021/acs.chemrestox.2c00311

Luechtefeld T, Marsh D, Rowlands C, Hartung T. 2018. Machine learning of toxicological big data enables read-across structure activity relationships (RASAR) outperforming animal test reproducibility. Toxicol Sci. 165(1):198-212. https://doi.org/10.1093/toxsci/kfy152

Lunn RM, Blask DE, Coogan AN, Figueiro MG, Gorman MR, Hall JE, Hansen J, Nelson RJ, Panda S, Smolensky MH, et al. 2017. Health consequences of electric lighting practices in the modern world: A report on the National Toxicology Program's workshop on shift work at night, artificial light at night, and circadian disruption. Sci Total Environ. 607-608:1073-1084. https://doi.org/10.1016/j.scitotenv.2017.07.056

Lunn RM, Mehta SS, Jahnke GD, Wang A, Wolfe MS, Berridge BR. 2022. Cancer hazard evaluations for contemporary needs: Highlights from new National Toxicology Program evaluations and methodological advancements. J Natl Cancer Inst. 114(11):1441-1448. https://doi.org/10.1093/jnci/djac164

Moustakas H, Date MS, Kumar M, Schultz TW, Liebler DC, Penning TM, Salvito DT, Api AM. 2022. An end point-specific framework for read-across analog selection for human health effects. Chem Res Toxicol. 35(12):2324-2334. https://doi.org/10.1021/acs.chemrestox.2c00286

Myatt GJ, Ahlberg E, Akahori Y, Allen D, Amberg A, Anger LT, Aptula A, Auerbach S, Beilke L, Bellion P, et al. 2018. In silico toxicology protocols. Regul Toxicol Pharmacol. 96:1-17. https://doi.org/10.1016/j.yrtph.2018.04.014

National Academies of Sciences, Engineering, and Medicine (NASEM). 2023. Building confidence in new evidence streams for human health risk assessment: Lessons learned from

laboratory mammalian toxicity tests. Washington, DC: National Academies Press. https://doi.org/10.17226/26906

National Center for Biotechnology Information (NCBI). 2023. Gene Expression Omnibus. Bethesda, MD: National Institutes of Health, National Library of Medicine, National Center for Biotechnology Information. [Accessed: February 14, 2023]. https://www.ncbi.nlm.nih.gov/geo/

National Institute of Environmental Health Sciences (NIEHS). 2023a. Integrating environmental data with other omics for cancer epidemiology, Tuesday & Wednesday, February 14 – 15, 2023. Research Triangle Park, NC: National Institutes of Health, National Institute of Environmental Health Sciences. [Last reviewed: May 3, 2023]. https://www.niehs.nih.gov/news/events/pastmtg/2023/environmental_data/index.cfm

National Institute of Environmental Health Sciences (NIEHS). 2023b. Molecular signatures of exposure in cancer: A joint NIEHS and NCI workshop, Thursday & Friday, June 29 – 30, 2023. Research Triangle Park, NC: National Institutes of Health, National Institute of Environmental Health Sciences. [Last reviewed: July 5, 2023]. https://www.niehs.nih.gov/news/events/pastmtg/2023/molecular-signatures/

National Toxicology Program (NTP). 2008. Report on Carcinogens background document for styrene. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. https://ntp.niehs.nih.gov/ntp/roc/twelfth/2010/finalbds/styrene_final_508.pdf

National Toxicology Program (NTP). 2013. Report on Carcinogens monograph on cumene. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. RoC Monograph 02. https://doi.org/10.22427/ROC-MGRAPH-02

National Toxicology Program (NTP). 2015. Report on Carcinogens monograph on trichloroethylene. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. RoC Monograph 05. https://doi.org/10.22427/ROC-MGRAPH-5

National Toxicology Program (NTP). 2018a. Report on Carcinogens monograph on antimony trioxide. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. RoC Monograph 13. https://doi.org/10.22427/ROC-MGRAPH-13

National Toxicology Program (NTP). 2018b. Report on Carcinogens monograph on haloacetic acids found as water disinfection by-products. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. RoC Monograph 12. https://doi.org/10.22427/ROC-MGRAPH-12

National Toxicology Program (NTP). 2021a. NTP cancer hazard assessment report on night shift work and light at night. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. https://doi.org/10.22427/NTP-CHR-NSWLAN

National Toxicology Program (NTP). 2021b. Substance profile: Benzidine and dyes metabolized to benzidine. In: Report on Carcinogens. 15th ed. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. https://doi.org/10.22427/ntp-other-1003

National Toxicology Program (NTP). 2021c. Substance profile: Cobalt. In: Report on Carcinogens. 15th ed. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. https://doi.org/10.22427/ntp-other-1003

National Toxicology Program (NTP). 2021d. Substance profile: Cumene. In: Report on Carcinogens. 15th ed. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. https://doi.org/10.22427/ntp-other-1003

National Toxicology Program (NTP). 2023. Adverse outcome pathways. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. [Last updated: March 29, 2023]. https://ntp.niehs.nih.gov/go/niceatm-aop

Oku Y, Madia F, Lau P, Paparella M, McGovern T, Luijten M, Jacobs MN. 2022. Analyses of transcriptomics cell signaling for pre-screening applications in the integrated approach for testing and assessment of non-genotoxic carcinogens. Int J Mol Sci. 23(21):12718. https://doi.org/10.3390/ijms232112718

Organisation for Economic Co-operation and Development (OECD). 2014. Guidance on grouping of chemicals, second edition. (OECD Series on Testing and Assessment, No. 194). Paris, France: OECD Publishing. https://doi.org/10.1787/9789264274679-en

Organisation for Economic Co-operation and Development (OECD). 2018a. Good cell culture practice (GCCP). In: Guidance Document on Good In Vitro Method Practices (GIVIMP). Paris, France: OECD Publishing. p. 161. https://doi.org/10.1787/9789264304796-16-en

Organisation for Economic Co-operation and Development (OECD). 2018b. Guidance document on Good In Vitro Method Practices (GIVIMP). (OECD Series on Testing and Assessment, No. 286). Paris, France: OECD Publishing. https://doi.org/10.1787/9789264304796-en

Organisation for Economic Co-operation and Development (OECD). 2023a. Grouping of chemicals: Chemical categories and read-across. Paris, France: Organisation for Economic Co-operation and Development. [Accessed: February 21, 2023]. https://www.oecd.org/chemicalsafety/risk-assessment/groupingofchemicalschemicalcategoriesandread-across.htm

Organisation for Economic Co-operation and Development (OECD). 2023b. [Homepage]. Paris, France: Organisation for Economic Co-operation and Development. [Accessed: February 22, 2023]. https://www.oecd.org/

Pamies D, Leist M, Coecke S, Bowe G, Allen DG, Gstraunthaler G, Bal-Price A, Pistollato F, de Vries RBM, Hogberg HT, et al. 2022. Guidance document on Good Cell and Tissue Culture Practice 2.0 (GCCP 2.0). ALTEX. 39(1):30-70. https://doi.org/10.14573/altex.2111011

Patlewicz G, Ball N, Boogaard PJ, Becker RA, Hubesch B. 2015. Building scientific confidence in the development and evaluation of read-across. Regul Toxicol Pharmacol. 72(1):117-133. https://doi.org/10.1016/j.yrtph.2015.03.015

Patlewicz G, Cronin MTD, Helman G, Lambert JC, Lizarraga LE, Shah I. 2018. Navigating through the minefield of read-across frameworks: A commentary perspective. Comput Toxicol. 6:39-54. https://doi.org/10.1016/j.comtox.2018.04.002

Patlewicz G, Helman G, Prachi P, Shah I. 2017. Navigating through the minefield of read-across tools: A review of in silico tools for grouping. Comput Toxicol. 3:1-18. https://doi.org/10.1016/j.comtox.2017.05.003

Pestana C, Enoch SJ, Firman JW, Madden JC, Spînu N, Cronin MTD. 2022. A strategy to define applicability domains for read-across. Comput Toxicol. 22:100220. https://doi.org/10.1016/j.comtox.2022.100220

Rowan A, Spielmann H. 2019. The current situation and prospects for tomorrow: Toward the achievement of historical ambitions. In: Balls M, Combes R, Worth A, editors. The History of Alternative Test Methods in Toxicology. London, UK: Academic Press. p. 325-331. https://doi.org/10.1016/B978-0-12-813697-3.00035-4

Samet JM, Chiu WA, Cogliano V, Jinot J, Kriebel D, Lunn RM, Beland FA, Bero L, Browne P, Fritschi L, et al. 2020. The IARC monographs: Updated procedures for modern and transparent evidence synthesis in cancer hazard identification. J Natl Cancer Inst. 112(1):30-37. https://doi.org/10.1093/jnci/djz169

Savitz DA, Wellenius GA, Trikalinos TA. 2019. The problem with mechanistic risk of bias assessments in evidence synthesis of observational studies and a practical alternative: Assessing the impact of specific sources of potential bias. Am J Epidemiol. 188(9):1581-1585. https://doi.org/10.1093/aje/kwz131

Schultz TW, Amcoff P, Berggren E, Gautier F, Klaric M, Knight DJ, Mahony C, Schwarz M, White A, Cronin MTD. 2015. A strategy for structuring and reporting a read-across prediction of toxicity. Regul Toxicol Pharmacol. 72(3):586-601. https://doi.org/10.1016/j.yrtph.2015.05.016

Schultz TW, Richarz AN, Cronin MTD. 2019. Assessing uncertainty in read-across: Questions to evaluate toxicity predictions based on knowledge gained from case studies. Comput Toxicol. 9:1-11. https://doi.org/10.1016/j.comtox.2018.10.003

Smith MT, Guyton KZ, Gibbons CF, Fritz JM, Portier CJ, Rusyn I, DeMarini DM, Caldwell JC, Kavlock RJ, Lambert PF, et al. 2016. Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. Environ Health Perspect. 124(6):713-721. https://doi.org/10.1289/ehp.1509912

Smith MT, Guyton KZ, Kleinstreuer N, Borrel A, Cardenas A, Chiu WA, Felsher DW, Gibbons CF, Goodson WH 3rd, Houck KA, et al. 2020. The key characteristics of carcinogens: Relationship to the hallmarks of cancer, relevant biomarkers, and assays to measure them. Cancer Epidemiol Biomarkers Prev. 29(10):1887-1903. https://doi.org/10.1158/1055-9965.EPI-19-1346

Sozanska AM, Fletcher C, Bihary D, Samarajiwa SA. 2020. SpiderSeqR: An R package for crawling the web of high-throughput multi-omic data repositories for data-sets and annotation. bioRxiv. https://doi.org/10.1101/2020.04.13.039420

Tice RR, Bassan A, Amberg A, Anger LT, Beal MA, Bellion P, Benigni R, Birmingham J, Brigo A, Bringezu F, et al. 2021. In silico approaches in carcinogenicity hazard assessment: Current status and future needs. Comput Toxicol. 20:100191. https://doi.org/10.1016/j.comtox.2021.100191

U.S. Environmental Protection Agency (USEPA). 2018. Application of systematic review in TSCA risk evaluations. Washington, DC: U.S. Environmental Protection Agency, Office of Chemical Safety and Pollution Prevention. EPA Document No. 740-P1-8001. https://www.epa.gov/sites/default/files/2018-06/documents/final_application_of_sr_in_tsca_05-31-18.pdf

Walker VR, Schmitt CP, Wolfe MS, Nowak AJ, Kulesza K, Williams AR, Shin R, Cohen J, Burch D, Stout MD, et al. 2022. Evaluation of a semi-automated data extraction tool for public health literature-based reviews: Dextr. Environ Int. 159:107025. https://doi.org/10.1016/j.envint.2021.107025

Wang Z, Walker GW, Muir DCG, Nagatani-Yoshida K. 2020. Toward a global understanding of chemical pollution: A first comprehensive analysis of national and regional chemical inventories. Environ Sci Technol. 54(5):2575-2584. https://doi.org/10.1021/acs.est.9b06379

Wieder C, Frainay C, Poupin N, Rodríguez-Mier P, Vinson F, Cooke J, Lai RP, Bundy JG, Jourdan F, Ebbels T. 2021. Pathway analysis in metabolomics: Recommendations for the use of over-representation analysis. PLoS Comput Biol. 17(9):e1009105. https://doi.org/10.1371/journal.pcbi.1009105

Wishart D. 2022. Metabolomics and the multi-omics view of cancer. Metabolites. 12(2):154. https://doi.org/10.3390/metabo12020154

Wu H, Eckhardt CM, Baccarelli AA. 2023. Molecular mechanisms of environmental exposures and human disease. Nat Rev Genet. 24(5):332-344. https://doi.org/10.1038/s41576-022-00569-3

7. Evidence Integration

Introduction

The last step in the cancer hazard evaluation process (Figure 7-1) is to integrate the evidence from the cancer studies in humans and animals with the evidence from mechanistic and other relevant data and apply the Report on Carcinogens (RoC) listing criteria to reach a preliminary listing recommendation. This step is usually captured in the final section of the RoC monograph. Guidelines for integrating evidence across studies of the same evidence stream to reach a level-of-evidence conclusion (LoE) are discussed in the relevant sections of the handbook:

- Human cancer epidemiology studies (see Section 3)
  - LoE: Sufficient, limited, or inadequate
- Animal cancer studies (see Section 4)
  - LoE: Sufficient, not sufficient
- Mechanistic studies (see Section 6)
  - LoE: Convincing, supporting

Section 7 brings forward these assessments and provides guidance for integrating the collective body of evidence as a unit (rather than focusing on individual studies).

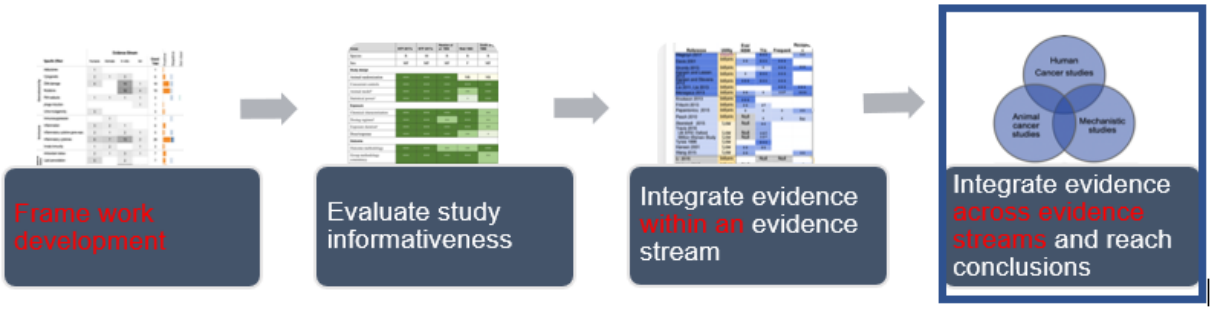


Figure 7-1. Cancer Hazard Evaluation Process

A substance (agent, substance, mixture, or exposure circumstance or scenario) is listed in the RoC as either *known or reasonably anticipated to be a human carcinogen*. The listing recommendation is reached by applying the RoC criteria to the cancer hazard assessment (see Box 7-1). Historically, the RoC criteria have been informed by and overlap with (but are not the same as) the IARC criteria for carcinogen classifications (IARC 2019). Substances not meeting these criteria are not listed in the RoC and the review is captured in an RoC appendix.

Conclusions regarding the carcinogenicity of a substance are based on scientific judgment, considering all relevant data. The listing categories reflect the strength of the evidence or the confidence for these conclusions.

Since the publication of the 2015 *Handbook for Preparing RoC Monographs* (NTP 2015), the RoC Group has developed more transparent, multistep approaches to integrate the evidence across data streams and evaluate its coherence (Section 7.3) and guidance on how the LoE of carcinogenicity from each evidence stream translates to the overall listing recommendation (Section 7.2). Coherence is defined as when the evidence across streams (e.g., human cancer, mechanistic, and animals) is consistent, forms a united whole, and tells a cohesive story of causality.

Box 7-1. Summary of Report on Carcinogens Listing Criteria

Known to be a human carcinogen

- Sufficient evidence of carcinogenicity from studies in humans*

Reasonably anticipated to be a human carcinogen

Meets one of the following:

- Limited evidence of carcinogenicity from studies in humans*
- Sufficient evidence of carcinogenicity from studies in experimental animals
- The substance belongs to a structurally related class of substances that are listed in the RoC
- Convincing relevant information that the agent acts through a mechanism indicating that it would likely cause cancer in humans

* Evidence from studies in humans includes cancer epidemiology studies, or mechanistic studies in exposed humans, or both.

Source: NTP (2023).

7.1. Considerations and Approaches

In evaluating causality, we are cognizant that many factors contribute to cancer hazards and the importance of effect modifiers. Many environmental causes of cancer are neither necessary nor sufficient in the absence of other factors to produce the disease; however, a cause does not have to be either necessary or sufficient for its removal to decrease disease incidence (Rothman and Greenland 2005).

In reaching our listing decisions, we use triangulation approaches, considering Hill's and other causality principles to integrate the evidence across evidence streams. Triangulation refers to integrating evidence from different research or methodological approaches, each of which have different but unrelated sources of potential bias (Lawlor et al. 2016).

7.1.1. Bradford Hill's Considerations of Causality

Bradford Hill published his viewpoints on causality as considerations for determining whether the evidence from observational epidemiological studies is causal (Bradford Hill 1965). His level-of-evidence conclusions from human studies focused on the strength of the association, consistency across studies, evidence of an exposure-response gradient, and temporality of exposure. For evidence integration, we collectively consider and concentrate on the following Bradford Hill factors: strength and consistency across evidence streams, biological plausibility and coherence, and temporality. (Analogy and specificity are less relevant, and we are integrating experimental data). Strong LoE conclusions (e.g., sufficient from human or animal cancer studies, convincing from mechanistic studies) in more than one evidence stream (consistency) increases the strength of the hazard evaluation classification. Table 7-1 provides

examples of evaluations showing evidence concordance between human, toxicological, and mechanistic evidence; the mechanistic studies provide biological plausibility that the association between the exposure and outcome are causal and consistent with biological knowledge.

Table 7-1. Selected Examples of Coherence Biological Plausibility, and Consistency across Evidence Streams

Substance	Toxicological/ Mechanistic Evidence[a]	Epidemiological Evidence
TCE Known human carcinogens	Metabolism and mechanistic studies show that GSH pathway is important in TCE-induced kidney carcinogenicity • Cytotoxic and mutagenic metabolites are formed in or transported to the kidney	A case-control study found an increased risk of kidney cancer among individuals who had genes that produced enzymes involved in GSH but not in individuals without functional genes
NSW High confidence for a causal relationship with human cancer	Biological understanding, animal cancer studies, and human mechanistic studies • Breast cancer biology: higher susceptibility to cancer at different stages of development (prenatal life, infancy, puberty, early adulthood, and timing of first pregnancy) • Simulated night shift work promotes human breast tumors/cells implanted into rodents • Increased estrogen levels associated with night shift work	Epidemiological studies found higher breast cancer risk in women for: • Starting NSW at a younger age in adulthood • Recency of exposure • Hormone-receptor positive tumors

Sources: NTP (2021a; 2021h).
TCE = trichloroethylene; GSH = glutathione conjugation; NSW = night shift work.
[a]All evidence types (exposed humans, animal models, in vitro).

7.1.2. Other Considerations

For some evaluations, substance-specific guidance may be available. Specific lines of evidence from human mechanistic studies were critical in informing the evidence of cancer epidemiology studies of viruses (NTP 2021i). For instance, many epidemiological studies on viruses include cross-sectional exposure assessment (e.g., measured viral antibodies to the virus at the same time as cancer assessment) and findings of monoclonality in cancer cells can help provide evidence of temporality (e.g., suggest that infections take place prior to tumor formation). Several virus experts noted that strict application of the Hill factors was challenging when assessing the cancer evidence for some viruses and developed additional factors to consider when evaluating causality. In the RoC review of several viruses, we synthesized published guidelines from virus experts to guide the cancer hazard evaluation (NTP 2021i).

7.1.3. Evaluating the Coherence of the Evidence: Triangulation Approaches and Evidence-based Tables

To increase transparency and facilitate the overall cancer hazard evaluation, we use a stepwise approach to integrate the evidence across evidence streams. This approach has several advantages: (1) It compiles and brings forward the assessment conclusions from each evidence stream. (2) It uses triangulation approaches to assess the evidence that may have different biases across the different streams. (3) It provides relevant information from each evidence stream to evaluate the coherence of the database. (4) It encompasses a series of tables, with increasing levels of integration, enhancing the transparency of the rationale for our conclusion. (5) It facilitates defining or contextualizing the hazard (if relevant) or identifying research gaps.

Triangulation aims to integrate data from different methodological approaches with different biases, and to exploit these differences to draw qualitative conclusions. For health hazard assessments evaluating causality, triangulation can occur at multiple levels—within a single study, across studies of a specific evidence stream (human and animal cancer and mechanistic studies) or types (epidemiology, in vivo, in vitro) (NASEM 2022). The human cancer section considers the first two levels, whereas this section considers the last level.

The approach typically consists of three progressive steps or evidence-based tables and was informed by the *NTP Cancer Assessment of Night Shift Work and Light at Night* (NTP 2021a) (see Figure 7-2 for an example of three evidence-based tables and how each table informs subsequent tables).

Step 1: Summarize assessments of collective evidence for each data stream

Exposure	Outcome	Type of studies	Strengths & Limitations	Assessment
Light at Night (LAN)	Cancer	Experimental animal studies, mainly initiation and promotion	Some studies used human breast tissue or cells Most studies did not report incidence	Consistent evidence of tumor growth for mammary and other tumors
LAN	Melatonin suppression	Mechanistic experimental animal studies {e.g., human blood exposed to LAN, co-exposure to melatonin)	Human relevance	Co-exposure to melatonin restored human breast cancer inhibitory activity

Step 2: integrate the mechanistic data

Exposure	Outcome	Evidence stream	Conclusions	Assessment
Melatonin suppression	Breast cancer	Experimental studies in humans and animals In vitro studies	Strong evidence that melatonin can reduce tumor growth and for its oncostatic properties	Strong evidence that melatonin suppression plays a role in cancer

Step 3: Integrate evidence across streams

Exposure	Outcome	Evidence stream	Assessment	Overall evaluation
LAN	Cancer	Experimental animal cancer studies Mechanistic studies in experimental animals	Strong evidence LAN promotes tumor growth Strong evidence for melatonin suppression	Strong evidence than LAN causes melatonin suppression, leading to circadian disruption and cancer

Figure 7-2. Series of Progressive Evidence-Based Tables

The three-evidence-based tables are examples from the Light at Night evaluation and correspond to the three-step approach. Note that the entries for each data are only a subset of the evidence, see the Light at Night Cancer Hazard Assessment (NTP 2021a) and Table 7-4 and Table 7-5 for more complete entries of Steps 2 and 3. The purpose of the figure is to show the relationship between the three tables: Step 1 assessments inform Step 2 conclusions or Step 3 assessments. Step 2 assessment integrates the relevant evidence from multiple assessments (e.g., different evidence types) from Step 1. Step 3 overall evaluation integrates the evidence from relevant Step 1 and 2 assessments.

Step 1: Summarize the Assessments of the Collective Evidence for Each Data Stream

The first step in the evidence-integration process brings forward the assessments from each evidence stream, for example, each relevant entry (referred to as study set) in the overall cancer hazard EECO (evidence, exposure, comparison group, outcome/endpoint). For some evaluations, the "exposure" may be a proposed intermediate (e.g., circadian disruption). For each study set, the assessment summary presents the scope and type of evidence, the strengths and limitations of the database (e.g., most impactful bias across studies), and the key findings that may inform evidence concordance (see Table 7-2). Each assessment (e.g., study set) is based on a rigorous review of the consistency and informativeness of the individual studies (see Sections 3, 4, and 6). Here, the strength and limitations and triangulation approaches refer to the collective body of evidence for each study set and were informed by the individual study assessments. These evidence-based summary tables are also included in the relevant section and are brought forward to the evidence integration sections to increase the transparency of the overall evaluation. A human epidemiology study set typically consists of all studies (sometimes stratified by study design or other common characteristic) for a specific cancer site. Depending on the substance, an animal study set may be all studies for a specific tumor type or all tumors, models, or common tumors across similar chemicals. For mechanistic studies, the assessment summary is usually for

an influential question (Step 2 for biological effects; see Section 6.4) rather than at the study set level (Step 1; see Section 6.4). Influential questions (as defined in 6.2.4) are the scientific issues that are: (1) critical for understanding cancer mechanisms and biological and human relevance and (2) will most likely impact the LoE conclusions (e.g., support or arguing against convincing).

Table 7-2. Template Example for Summarizing the Assessment of Each Evidence Stream

Exposure	Outcome	Evidence Streams	Strength and Limitations	Assessment
Substance	Cancer type	Number and type of **human cancer studies** Cohort studies Case-control studies Pooled or meta-analyses	Summary of the most influential biases (direction, magnitude, impact) across studies by study design or other relevant grouping	Consistency of findings and patterns for factors, such as exposure matrices and levels, cancer subtypes, effect modifiers
Substance	Cancer type or across multiple cancers	Number and type of **animal cancer studies** Animal models (e.g., route, species)	Summary of most influential biases (direction, magnitude, impact) across studies by model, route, or other relevant grouping	Exposure-related cancer sites, common cancer sites across groups of chemicals Information relevant to evidence integration
Substance	Biological effect example 1	Number and type of **mechanistic studies** Model (e.g., in vitro, in vivo) Exposed humans	Summary of the most influential biases (direction, magnitude, impact) across studies and relevance of evidence type	Confidence judgment for influential question(s) Information relevant to evidence integration
Intermediate (Biomarker for Proposed **Mechanism**)	Cancer type	Number and type of cancer (human or animal) relevant studies	Summary of most influential biases (direction, magnitude, impact) across studies by relevant grouping	Consistency and patterns for factors, such as exposure matrices, cancer subtypes, effect modifiers

Step 2: Integrate the Mechanistic Evidence

Mechanistic data are often integrated across influential questions to reach LoE conclusions (see Section 6.4.3) or for a specific mechanism (e.g., reported or development of a mode of action [MoA]) before the overall evidence integration. Integration may be multifaceted: multiple evidence types for multiple exposure/outcome pairs (e.g., exposure to intermediate, intermediate to cancer). This step is conducted in the mechanistic section and the evidence is brought forward to Section 7 to provide transparency for the overall evaluation (see Section 6 for guidance on evaluation of the evidence.) Evidence-based tables describe the confidence in the evidence for each exposure/outcome pair or mechanistic question and overall assessment. Table 7-3 provides an example adapted from the light at night (LAN) evaluation of circadian disruption as a cancer mechanism. Notably, the mechanistic evaluation may go beyond an LoE conclusion and provide conclusions for specific facets of the evidence base (e.g., by evidence type). New to this handbook is guidance for evaluating the confidence for the evidence for biological effects and the overall LoE of carcinogenicity from mechanistic studies.

Table 7-3. Intermediate Tables on Specific Mechanisms: Light at Night Example

Exposure or Mechanism	Evidence Streams or Approach	Study Set Conclusions[a]	Assessment
Circadian Disruption Melatonin Suppression	Molecular epidemiology studies measuring nocturnal urinary melatonin levels in night shift workers Some studies measured LAN and melatonin levels among shift workers Experimental studies of LAN and melatonin suppression in humans Experimental studies of LAN, melatonin suppression, and tumor promotion in animals Melatonin studies and cancer in humans and animals Experimental studies: in vivo or in vitro mechanistic studies	Strong evidence for melatonin suppression in night shift workers Database for melatonin suppression in shift work animal models is inadequate Strong evidence that electrical LAN exposure in people's everyday lives (depending on the wavelength, level, duration, and photic history) can cause melatonin suppression Weak evidence that higher melatonin levels (or proxies) are related to decreased cancer incidence Strong evidence that melatonin can reduce tumor growth and for its oncostatic properties, which may protect from all biological effects considered to be hallmarks of cancer Oncostatic properties involve epigenetic mechanisms relevant to cancer, particularly breast cancer	Indirect evidence that melatonin suppression contributes to breast cancer development in night shift workers Strong evidence that melatonin suppression plays a role in LAN-induced breast carcinogenicity in experimental animals Data inadequate to evaluate whether LAN during night work contributes to cancer risk

Source: NTP (2021a).
LAN = light at night.
[a]New to this handbook is formal guidance for reaching conclusions.

Step 3: Integrate the Evidence across All Streams

The final step in the assessment is to integrate all the relevant evidence and apply the RoC listing criteria to this assessment to reach a listing decision. Here we present the evidence-integration assessment, and Section 7.2 discusses applying the RoC listing criteria to that assessment. The overall cancer hazard evaluation uses triangulation approaches for integrating and assessing the coherence of the cancer (human and animal) and mechanistic assessments. Whereas triangulation approaches for epidemiological studies evaluate biases for individual studies (internal validity), triangulation approaches for the overall evidence evaluation consider general biases (identified limitations associated with specific evidence sources) for a collective body of evidence. For example, human cancer studies are the most relevant studies but can be subjected to biases both toward and away from the null due to their observational nature. Animal cancer studies are controlled exposure but are less human relevant, and mechanistic studies are not on the apical endpoint of interest. The strengths and limitations in Table 7-3 (discussed above in Step 2) and Table 7-4 can help facilitate evidence integration. Table 7-4 provides an example adapted from the LAN evaluation, which illustrates the integration of the conclusions for specific questions or datasets from each evidence stream. Each monograph would have similar tables and, when relevant, figures depicting the evidence integration.

Table 7-4. Overall Evidence-Integration: Light at Night and Cancer

Evidence Streams	Conclusion	Overall Integration
Human epidemiological studies of female breast cancer 5 studies of outdoor light 10 studies of light in the sleeping area	Inadequate evidence that LAN (indoor or outdoor) causes breast cancer risk	Strong toxicological and mechanistic than data that exposure to LAN causes melatonin suppression and other types of circadian disruption, which leads to breast or mammary-gland cancer proliferation and growth in experimental animals
Experimental animal studies Primarily initiation-promotion studies of continuous, dim, or interrupted light or bright, blue-enriched light during the day	Strong evidence from studies in experimental animals that exposure to LAN promotes implanted human breast cancer proliferation or growth and mouse mammary-gland tumor growth	LAN induces biological effects in experimental animals associated with: (1) carcinogenicity and (2) melatonin suppression and circadian clock gene deregulation
Mechanistic and biomonitoring data Melatonin suppression hypothesis Circadian disruption theory Biological effects associated with cancer	Bright, blue-enriched light during the day increased the level of nighttime melatonin levels and decreased tumor growth in experimental animals Strong evidence that melatonin suppression plays a role in LAN-induced breast carcinogenicity in experimental animals	Some of these biological effects are observed among night shift workers Exposure to excessive LAN can cause circadian disruption in humans

Source: NTP (2021a).
LAN = light at night.

7.2. Integrating Level-of-evidence Conclusions

New to the handbook are formal considerations for integrating the LoE conclusions of carcinogenicity from human cancer, experimental animal cancer, and mechanistic studies. These considerations were informed by previous RoC evaluations and the 2019 IARC preamble. Although not part of the listing criteria, RoC evaluations often use the term "supporting mechanism data" for evaluations that rely mainly on human or animal mechanistic evidence with data showing biological plausibility. Table 7-5 delineates how LoE from each evidence stream relates to each RoC criterion. The overall listing recommendation also considers the coherence of the body of evidence and all relevant information, as discussed in Section 7.2.

Table 7-5. Evidence-Integration Guidance Table[a]

RoC Criterion	Human Cancer Epidemiology	Animal Cancer	Mechanisms: Overall	Mechanisms: Exposed Humans	Listing[b]
Sufficient Evidence from Studies in Humans	Sufficient	Any[c]	Any[c]	Any[c]	Known
	Limited	Any[c]	Supporting[d]	Robust[e]	Known
Limited Evidence from Studies in Humans	Limited	Any[c]	Any[c]	Not robust	RAHC
	Inadequate	Not sufficient	Convincing	Robust[e]	RAHC

RoC Criterion	Human Cancer Epidemiology	Animal Cancer	Mechanisms: Overall	Mechanisms: Exposed Humans	Listing[b]
Sufficient Evidence from Studies in Experimental Animals	Inadequate	Sufficient	Any[c]	Any[c]	RAHC
Biological Plausibility or Member of a Listed Class	Inadequate	Not sufficient	Convincing[e,f]	Any[c]	RAHC

RAHC = reasonably anticipated to be a human carcinogen.

[a]Descriptors based on the RoC listing criteria and convention. Human cancer studies: sufficient, limited, inadequate (see Section 3); animal cancer studies: sufficient, not sufficient (see Section 4); mechanisms: convincing, supporting (see Section 6).

[b]Also considers the coherence of the database (see Section 7.2.2), which may be especially helpful for integrating human evidence from epidemiological and mechanistic studies but is not required.

[c]Any indicates that the LoE for the evidence stream does not affect the cancer hazard conclusion (e.g., for animal cancer studies, it could be sufficient or not sufficient).

[d] It can also be sufficient. The LoE from mechanistic data depends in part on the human cancer epidemiology studies (e.g., limited evidence can range from bordering on inadequate to bordering on sufficient, and a similar range could be made for the strength of evidence from human mechanistic studies)

[e]See Section 6.4.3, Confidence of the Evidence, informativeness of the study set.

[f]Convincing can be from a mode of action, biological effects, or cancer predictions from clustering/read-across approaches.

7.2.1. Evaluations with Substantial Human and Animal Cancer Databases

The RoC listing criteria requires sufficient—credible association not reasonably explained by chance, bias, or confounding—evidence of carcinogenicity from studies in humans to list a substance as *known to be a human carcinogen*: the evidence demonstrates that the substance causes cancer. Typically, the evidence is from cancer epidemiological studies, but the RoC listing criteria specifies that human evidence also includes other types of studies, such as clinical and mechanistic studies in exposed humans. When the evidence from human cancer epidemiological studies is limited, a substance may be listed as *known to be a human carcinogen* if there is robust human mechanistic evidence (see Box 7-3 for an example of an RoC listing). The overall LoE of carcinogenicity from mechanistic studies can range from supporting (high) to convincing. Convincing mechanistic data are rarely only from human studies, as evidence in cancer models (e.g., in vivo or in vitro) provides context (e.g., proposed mechanisms or pathways for the biological effects) for the effects observed in humans. Current knowledge indicates that a substance would not typically be listed as *known to be a human carcinogen* with inadequate evidence from cancer epidemiological studies because: (1) cancer type often provides relevancy of the mechanistic data and (2) a second evidence stream increases the confidence for the cancer hazard classifications. Examples of notable exceptions include substance(s) that release radioactivity (e.g., neutrons) or are metabolized (e.g., dyes metabolized to benzidine) to a known human carcinogen.

Box 7-3. Ethylene Oxide Example

Known to be a human carcinogen

Limited evidence from cancer epidemiological studies:

- Lymphohematopoietic and breast cancer

Mechanistic evidence: Direct acting alkylating agent leading to adducts, mutation, and DNA and chromosomal damage:

- Genetic damage found in exposed workers and model systems

Evidence integration

- Sufficient evidence from studies in humans: Human cancer epidemiological and mechanistic studies
- Sufficient evidence from studies in experimental animals: hematopoietic system and other sites

Source: NTP (2021g).

Substances are listed as reasonably anticipated rather than known to be a human carcinogen when there is strong evidence but less confidence in it as reflected by the LoE from human studies (limited), less human-relevant evidence (e.g., sufficient from animal cancer studies), or less direct evidence (e.g., noncancer studies, convincing LoE from mechanistic studies) in the absence of human and animal cancer data.

Box 7-2. Cumene Example

Reasonably anticipated to be a human carcinogen

Exposure-related tumors in experimental animals:

- Rats: kidney (males)
- Mice: lung (both sexes), liver (males)

Human cancer epidemiology studies:

- Inadequate

Mechanistic data:

- Human relevance of male rat kidney tumors uncertain

Evidence integration:

- Sufficient LoE from studies in experimental animals: lung and liver tumors in mice
- Supporting mechanistic evidence for lung and liver tumors

Source: NTP (2021e).

Mechanistic evidence can increase or decrease the certainty of the evidence from cancer studies (e.g., limited LoE from human cancer studies to known human carcinogen discussed above for ethylene oxide). If there is compelling evidence that a tumor site observed in experimental animals occurs by a nonhuman relevant mechanism, that cancer site would not be included in the evidence for reaching an LoE conclusion from studies in experimental animals (see Box 7-2 for the cumene example). The nonhuman relevance MoA only applies to the targeted cancer site and not the overall evaluation.

7.2.2. Limited Cancer Databases and Class Evaluations

The RoC criteria permit listing substance(s) with little or no cancer data in humans or experimental animals (see Section 6) as *reasonably anticipated to be a human carcinogen* if the evidence of carcinogenicity from mechanistic studies is convincing or if the substance is a member of a class whose members are listed in the RoC. Mechanistic data can also be used to list a substance as known with limited evidence from human cancer epidemiology studies. Moreover, the listing criteria note that cancer hazard conclusions (e.g., listing recommendations for *known or reasonably anticipated to be a human carcinogen*) "are based on scientific judgment with consideration given to all relevant information," which occur during the assessment of mechanistic evidence (see Section 6.4.3) or overall cancer hazard evaluation. As of the 15th RoC, these criteria have been used primarily to list substance(s) metabolized to a listed carcinogen. The listing decision was reached by integrating mechanistic and metabolism evidence with cancer data of the listed carcinogen or any class members. The listing category for the substance under review (*known* or *reasonably anticipated to be a human carcinogen*) depends on the listed category of the metabolite and the evidence type (e.g., exposed humans, animal model) for the metabolism and mechanistic data. An example of another type of evidence that can be used to list substances as a class is a listing of similar substances that cause cancer by a common mechanism (e.g., cobalt compounds that release cobalt ion in vivo). In this handbook, we expand our assessment of chemicals by incorporating clustering and read-across approaches in our cancer hazard evaluations (see Section 6.5). Table 7-6 provides examples of how different types of evidence can be integrated to reach a listing decision.

Table 7-6. Selected Examples of Substances Listed with Little Cancer Data or Listed from Class Evaluations

Substance(s)	Mechanistic and Relevant Data	Cancer Data	Listing Category
Dyes Metabolized to Benzidine	Release free benzidine in *exposed human*s and other species	Benzidine is a known human carcinogen	Known
	Exposure to dyes is equivalent to exposure to equimolar doses of benzidine (exposed humans)	Three class members cause tumors in experimental animals	
Diazoaminobenzene (DAAB)	Metabolized to benzene (animals and human tissues) and is quantitatively similar to predicted benzene and aniline metabolites (animals)[a]	Benzene is a known human carcinogen	RAHC
	Causes genetic damage in animals and bacteria		

Substance(s)	Mechanistic and Relevant Data	Cancer Data	Listing Category
Cobalt and Cobalt Compounds That Release Cobalt Ions In Vivo	Cobalt ion largely responsible for toxicity and carcinogenicity Similar mechanisms regardless of water solubility (bioavailability and mechanistic studies)	Sufficient evidence of carcinogenicity for cobalt metal and several cobalt compounds from studies in experimental animals regardless of solubility	RAHC

RAHC = reasonably anticipated to be a human carcinogen.
Sources: NTP (2021c; 2021d; 2021f).
[a]DAAB was listed as RAHC instead of known because the mechanistic data were not in exposed humans.

7.3. Supplementary Public Health Considerations

The Report on Carcinogens identifies and lists potential cancer hazards, as mandated by the Public Health Service Act in 1978. However, the RoC does not identify carcinogenic risks (including exposure assessment, dose-response analysis, and risk characterization), and therefore, is not, and should not be interpreted in the context of a risk assessment. Additionally, the document does not provide clinical guidance to individuals. Many factors, including the amount and duration of exposure and an individual's susceptibility to a substance, affect whether a person will or will not develop cancer (NTP 2021b).

The RoC does provide a comprehensive, relevant, and broadly applicable set of cancer hazard evaluations for use in a variety of contexts to ultimately improve public health. As such, our aim is to tailor each evaluation, if possible, to most appropriately characterize and contextualize potential carcinogenic hazards under review in a manner that is data driven, interpretable, and translatable. The evidence-integration stage (within and across data streams) may inform how the hazard under review (e.g., the listing definition) is characterized. Occasionally, the target hazard definition for a given scenario may differ from the substance under review for listing the Report on Carcinogens. When there is compelling scientific evidence, the hazard or chemical class may be defined by a mechanism (e.g., release of cobalt ion [the carcinogenic species] via in vivo metabolism, circadian disruption) or by the human epidemiological evidence (e.g., persistent night shift work, consumption of alcoholic beverages). Contextualizing the hazard may help inform intervention strategies, especially for widespread exposure that is unlikely to be banned (e.g., night shift work). In the absence of compelling evidence, hazard identification is not usually restricted to the exposure conditions of the cancer studies. As we continue to improve the RoC, our goal is to increase the specificity of hazard characterization and the implications of our evaluations, if applicable and relevant. On a case-by-case basis, we aim to improve cancer hazard communication by providing information on interventions, health disparities, and sensitive populations (such as specific lifestages of development) as part of the overall cancer hazard evaluation and other relevant media (Lunn et al. 2022). We also hope to advance cancer hazard classification by identifying and elaborating on research gaps that may encourage future research.

References

Bradford Hill A. 1965. The environment and disease: Association or causation? Proc R Soc Med. 58(5):295-300. https://doi.org/10.1177/003591576505800503

International Agency for Research on Cancer (IARC). 2019. Preamble to the IARC monographs. Lyon, France: International Agency for Research on Cancer. https://monographs.iarc.who.int/iarc-monographs-preamble-preamble-to-the-iarc-monographs/

Lawlor DA, Tilling K, Davey Smith G. 2016. Triangulation in aetiological epidemiology. Int J Epidemiol. 45(6):1866-1886. https://doi.org/10.1093/ije/dyw314

Lunn RM, Mehta SS, Jahnke GD, Wang A, Wolfe MS, Berridge BR. 2022. Cancer hazard evaluations for contemporary needs: Highlights from new National Toxicology Program evaluations and methodological advancements. J Natl Cancer Inst. 114(11):1441-1448. https://doi.org/10.1093/jnci/djac164

National Academies of Sciences, Engineering, and Medicine (NASEM). 2022. Workshops to support EPA's development of human health assessments: Triangulation of evidence in environmental epidemiology, May 9–11, 2022. Washington, DC: National Academy of Sciences. https://www.nationalacademies.org/event/05-09-2022/workshops-to-support-epas-development-of-human-health-assessments-triangulation-of-evidence-in-environmental-epidemiology

National Toxicology Program (NTP). 2015. Handbook for preparing Report on Carcinogens monographs. Research Triangle Park, NC: U.S. Department of Health and Human Services, National Institute of Environmental Health Sciences, Division of the National Toxicology Program, Office of the Report on Carcinogens. https://ntp.niehs.nih.gov/go/rochandbook

National Toxicology Program (NTP). 2021a. NTP cancer hazard assessment report on night shift work and light at night. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. https://doi.org/10.22427/NTP-CHR-NSWLAN

National Toxicology Program (NTP). 2021b. The Report on Carcinogens [fact sheet]. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. https://ntp.niehs.nih.gov/sites/default/files/ntp/pressctr/factsheet/15th_roc_508.pdf

National Toxicology Program (NTP). 2021c. Substance profile: Benzidine and dyes metabolized to benzidine. In: Report on Carcinogens. 15th ed. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. https://doi.org/10.22427/ntp-other-1003

National Toxicology Program (NTP). 2021d. Substance profile: Cobalt. In: Report on Carcinogens. 15th ed. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. https://doi.org/10.22427/ntp-other-1003

National Toxicology Program (NTP). 2021e. Substance profile: Cumene. In: Report on Carcinogens. 15th ed. Research Triangle Park, NC: U.S. Department of Health and Human

Services, Public Health Service, National Toxicology Program. https://doi.org/10.22427/ntp-other-1003

National Toxicology Program (NTP). 2021f. Substance profile: Diazoaminobenzene. In: Report on Carcinogens. 15th ed. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. https://doi.org/10.22427/ntp-other-1003

National Toxicology Program (NTP). 2021g. Substance profile: Ethylene oxide. In: Report on Carcinogens. 15th ed. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. https://doi.org/10.22427/ntp-other-1003

National Toxicology Program (NTP). 2021h. Substance profile: Trichloroethylene. In: Report on Carcinogens. 15th ed. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. https://doi.org/10.22427/ntp-other-1003

National Toxicology Program (NTP). 2021i. Substance profile: Viruses, eight listings. In: Report on Carcinogens. 15th ed. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. https://doi.org/10.22427/ntp-other-1003

National Toxicology Program (NTP). 2023. Report on Carcinogens process & listing criteria. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. [Last updated: March 7, 2023]. https://ntp.niehs.nih.gov/go/rocprocess

Rothman KJ, Greenland S. 2005. Causation and causal inference in epidemiology. Am J Public Health. 95 Suppl 1:S144-S150. https://doi.org/10.2105/ajph.2004.059204

Appendix A. Glossary

Table of Contents

A.1. General Systematic Review Terms

bias analysis: evaluation of internal validity that assesses probability that a specific bias is present, and if so, the direction, magnitude, and impact of bias.

Boolean operator: words (AND, OR, NOT or AND NOT) used as conjunctions to combine or exclude keywords in a search.

Bradford Hill guidelines: a group of nine principles that can be useful in establishing evidence of a causal relationship in epidemiology. The guidelines are: strength of the association, consistency, specificity, temporality, biological gradient, plausibility, coherence, experiment, and analogy.

coherence: when the evidence across streams (e.g., human cancer, animal cancer, mechanistic) is consistent, forms a united whole, and tells a cohesive story of causality.

evidence map: an interactive visual representation of literature generated via broad literature searches to capture the state of the science and identify data gaps; also known as systematic evidence maps or systematic evidence mapping.

evidence type : defines the model or population in which research is conducted. In systematic review, evidence stream can be in humans, animals, in vitro, or in silico.

external validity: addresses the extent to which conclusions from one study can be generalized to other situations, e.g., relevance of experimental animal data to humans.

internal validity: potential for bias within a study.

literature search strategy: method and parameters needed to identify key literature.

PECO statement (Population, Exposure, Comparison group, and Outcome): a framework used to define the type of studies included in the evaluation of human cancer studies. In animal cancer studies, Population is replaced by Model (MECO) and in mechanism studies Population is replaced by Evidence stream (EECO).

scoping review: a structured literature search to determine the extent of the body of literature on a particular topic, as well as key issues and data gaps. It also is a descriptive summary of evidence map results.

study informativeness: ability of the study to inform the cancer hazard evaluation, also known as study utility.

study protocol: a substance specific detailed description of scoping, the conceptual frameworks, the reasoning, methods, and considerations for evaluating study informativeness, and the methods for evidence synthesis and evaluation for one or more monograph sections (e.g., human protocol, animal protocol, mechanism protocol).

study sensitivity: ability of a study to detect a true effect.

systematic review: standardized way of evaluating and reporting published evidence on a specific topic, or substance. A systematic review can include evidence from one or multiple evidence streams and summarizes and interprets the evidence into a refined conclusion.

triangulation: an approach to evidence synthesis that considers the overall literature base and integrated results from several different theoretical approaches, methods, and designs, which have different and unrelated sources of potential bias, to determine if findings converge on one conclusion.

A.2. Human Exposure Terms

biomarkers: measurable substances or characteristics in the human body that can be used to monitor the presence of a chemical in the body, biological responses, or adverse health effects.

biomonitoring: measuring how much of a substance, its metabolites, or its biomarkers are present in the human body.

exposure models: estimates of exposure based on combining information about environmental contaminant concentrations with information about people's activities and locations (e.g., time spent working, exercising outdoors, and sleeping; food consumption).

gray literature: information produced outside the mainstream of published journal and monograph literature that is not controlled by commercial publishers.

NHANES: National Health and Nutrition Examination Survey (NHANES) is a program of studies (a survey designed to assess the health and nutritional status of adults and children in the United States. The survey is unique in that it combines interviews and physical examinations.

occupational exposure: reasonably anticipated skin, eye, mucus membrane, or parenteral contact with blood or other infectious materials that may result from the performance of an employee's duties.

occurrence: measurements of concentrations of pollutants in the environment (e.g., concentrations measured in ambient air, groundwater, etc.).

personal monitoring: measurement of human exposure to environmental contaminants accomplished by an individual wearing a monitoring device during normal day-to-day activities.

sample matrix: a specific type of medium (e.g., surface water, drinking water) in which the analyte of interest may be contained.

A.3. Human Cancer Terms

A.3.1. General Human Cancer Terms

DAG (directed acyclic graph): a tool to visually depict causal relationships in epidemiology to inform study design and statistical analysis.

meta-analysis: statistical method combining results from multiple studies

A.3.2. Study Population: Selection Bias

Berkson's bias: occurs in hospital-based case-control studies when the combination of exposure and disease under investigation increases the risk of hospital admission, leading to a systematically higher exposure rate among the hospital cases than the hospital controls (Dictionary of epidemiology, 2nd edition).

healthy volunteer: a cohort of volunteers recruited from an underlying population or subpopulation, study participants are often less likely to be from marginalized groups, as well as more likely to have higher educational or economic attainment, and/or be in better health than those in the target population; this may distort exposure-disease relationships.

healthy workers hire effect: the HWHE occurs when workers must meet minimal health criteria to begin working and are thus healthier than the general population.

healthy worker survival effect: may occur when healthier workers continue to work, and less healthy workers transfer to jobs with lower exposures, take time off, or leave work prior to the outcome. This may attenuate exposure-response relationships.

left truncation: when study subjects who are risk for the outcome do not remain observable for a later start of follow-up.

prevalent hire: when workers recruited into a cohort have a date of hire prior to the start of the study (baseline) and are still working at the start of follow-up.

A.3.3. Information: Exposure and Outcome

Berkson type error: occurs when a group's average is assigned to each individual suiting the group's characteristics; will not bias effect estimates, but will make them less precise.

classical error measurement: occurs when a quantity is measured by some device and repeated measurements vary around the true value and will bias effect estimates to the null.

detection bias: when exposure groups vary systematically in the measurement or detection of the outcome.

differential misclassification: when classification of either exposure or outcomes differs by study group.

differential recall bias: occurs due to differences in accuracy of recall between cases and non-cases or of differential reporting of a health outcome between exposed and unexposed which can often lead to an overestimate of effect. Due to their health concerns or exposures, cases or exposed persons may have greater incentive to recall past exposures or symptoms

non-differential misclassification: when there is equal misclassification of either exposure or outcomes by study group

observer bias: a systematic difference in measured exposure or outcome due to variation in the observer.

population-based cancer registry: registries that consolidate data from many sources and strive to provide an unbiased estimate of cancer incidence in the population.

r**everse causality:** reverse causality describes the event where an association between an exposure and an outcome is not due to direct causality from exposure to outcome, but rather because the defined "outcome" actually results in a change in the defined "exposure."

A.3.4. Confounding

confounder: a factor that distorts the association between and exposure and outcome that is associated with both exposure (causally or non-causally) and the outcome of interest (causally) and is not an intermediate in the disease pathway.

A.4. Animal Cancer Terms

cancer bioassay: provides information on the possible carcinogenic effects likely to arise from repeated exposure over a considerable part of the lifespan of the species used (e.g., at least 1 year for rodents).

neoplastic (benign, malignant) end points: tissues with non-reversible neoplastic changes.

Poly 3 test: a statistical method to determine the significance of pair-wise comparisons or overall exposure-related trends. The test is a survival-adjusted quantal-response procedure that modifies the Cochran-Armitage linear trend test. It takes survival differences into account by more closely approximating the total number of animal years at risk. Survival of individual animals are given a weight of one or less depending on how long they survived compared to the total study duration, raised to the power of three.

preneoplastic lesions: tissues with reversible neoplastic changes.

subchronic toxicity study: provides information on the possible health hazards likely to arise from repeated exposure over 28 to 90 days.

trend analysis: analysis of trends in neoplasm incidence across at least three treatment groups (e.g., the Cochran-Armitage trend test).

A.5. ADME Terms

absorption: diffusion or uptake by an organism into the blood, tissue or other system.

ADME: short for absorption, distribution, metabolism, excretion to describe the disposition of a xenobiotic substance and how it is processed by a living organism.

bioavailability: the proportion of a substance that enters the blood stream in an organism.

cytochrome P450 enzymes: superfamily of enzymes found in the liver responsible, as well as other cells throughout the body responsible for metabolism and detoxification of xenobiotics and essential for the production of cholesterol, steroids, prostacyclin and thromboxane A_2.

distribution: the movement or transfer of a substance from one location to another within the body.

enzyme polymorphisms: genetic variations in metabolic enzymes which can lead to changes in the rate of metabolism of a substance, e.g., fast/slow metabolizer.

excretion: the removal of a substance or its metabolites from the body.

Good Laboratory Practices: a system which non-clinical laboratory studies are planned, performed, monitored, recorded, reported, and archived.

half-life: the time it takes for half of a radioactive isotope to decay.

metabolism: the process of changing a substance within an organism to provide energy, cell maintenance, and detoxify and eliminate xenobiotics.

parent compound: in organic chemistry, the simplest member of a class of compounds, usually a basic linear or ring structure with no added chemical groups.

PBPK: physiologically based pharmacokinetic model based on concentrations in blood and tissue compartments.

toxicodynamics: the dynamic interactions of a substance or xenobiotic with a biological target and its biological effects.

toxicokinetics: the description of both what rate a substance or xenobiotic will enter the body and what occurs to excrete and metabolize the compound once it is in the body.

ultimate carcinogenic form; ultimate carcinogen: chemical or metabolite that initiates a carcinogenic process.

xenobiotic: chemical substances foreign to animal life.

A.6. Mechanistic Terms

A.6.1. General Mechanistic Terms

adverse outcome pathway: a structured representation of biological events leading to adverse effects, such as cancer, as is considered relevant to risk assessment.

epigenomics: the method of analyzing all modifications of DNA and associated proteins throughout the entire cell, tissue, or organism simultaneously.

epitranscriptomics: the method of analyzing all changes in RNA transcription and gene expression throughout the entire cell, tissue, or organism simultaneously under epigenetic regulation.

ex vivo: an experiment or study conducted outside a living organism from samples or tissues taking from a living organism.

genomics: the method of analyzing all of the DNA throughout an entire cell, tissue, or organism simultaneously.

in chemico: the use of abiotic chemical reactivity methods as replacements for animal/ in vivo assays.

in silico: research or experiment conducted or produced by means of computer modeling or simulation.

in vitro: a process or study performed outside a living organism.

in vivo: a process or study that takes place in a living organism (such as in animals).

in vivo apical endpoint: an observable outcome in a whole organism, such as a clinical sign or state, that indicative of an outcome that can results from exposure to a substance.

influential literature: the key literature identified that is integral to answering the influential and level of evidence mechanistic questions.

influential questions: questions that are substance-specific and related to the level of evidence questions and strategies; these questions are rigorously assessed in the mechanistic section of a monograph.

key event: an empirically observable step or its marker, which is a necessary element of the mode of action critical to the outcome; key events are measurable and reproducible.

mechanism/mode of action: a functional change resulting from exposure to a substance.

metabolomics: the method of analyzing all changes in metabolites throughout an entire cell, tissue, or organism simultaneously.

molecular initiating event: the initial interaction between a molecule and a biomolecule or biosystem that can be causally linked to an outcome via a pathway.

multi-omics: the method of analyzing all changes of a given set of cellular components (DNA, RNA, proteins, or metabolites) throughout an entire cell, tissue, or organism simultaneously.

omics: the method of analyzing all changes of a given cellular component (DNA, RNA, proteins, or metabolites) throughout an entire cell, tissue, or organism simultaneously.

proteomics: the method of analyzing all changes in protein expression throughout the entire cell, tissue, or organism simultaneously.

quantitative structure-activity relationship (QSAR) models: regression or classification models used in chemical or biological sciences.

transcriptomics: the method of analyzing all changes in RNA transcription and gene expression throughout the entire cell, tissue, or organism simultaneously.

A.6.2. Read-across Terms

analogues: a list of selected sources or group of sources and selected targets based on endpoint specific supervised similarity to be used in read-across.

analogue-based read-across: empirical data from one or a group of analogue source chemicals can be used to predict the same endpoint for the target chemical through endpoint specific supervised similarity.

applicability domain: the response and chemical structure spaces in which the model makes predictions with a given reliability.

category-based read-across: predicts targets starting from a group or a category of multiple sources whose physical-chemical, biological, or toxicological properties are likely to be similar or follow a regular pattern as a result of structural similarity.

data matrix: a spreadsheet or database containing the chemicals used for the prediction along with their data for selected parameters, profiles, and endpoint tree positions.

negative chemical: chemical that has been investigated in human or animal cancer studies but have no cancer hazards (with human or animal cancer data).

read-across: a technique for predicting an endpoint (such as carcinogenicity) for a target substance by using information on this endpoint from a source substance.

source chemical: chemical with known cancer hazards (with human or animal cancer data) used to predict the cancer hazard of the target chemical using read-across.

supervised classification: used to predict an outcome and involve training a model using input variables (specific for the endpoint), the endpoint of interest or data-gap filling (e.g., outcome variable), and an algorithm to map the input to the output.

target chemical: chemical with unknown cancer hazards that are predicted from sources using read-across.

uncertainty: extent of the interpretability and defensibility of a read-across hypothesis, justification, and prediction, which are described in a transparent manner.

unsupervised clustering: uses input data (such as general structural characteristics not specific to an endpoint) without a corresponding predicted variable or endpoint of interest.

Appendix B. General and Exposure-specific Authoritative Sources

Table of Contents

B.1. Authoritative Reviews and Reports

- Agency for Toxic Substances and Disease Registry (ATSDR) Toxicological Profiles (https://www.atsdr.cdc.gov/toxprofiledocs/index.html)
- California Environmental Protection Agency Proposition 65 hazard identification documents (http://www.oehha.ca.gov/prop65/prop65_list/Newlist.html)
- U.S. Environmental Protection Agency (EPA) Integrated Risk Information System (IRIS) (http://cfpub.epa.gov/ncea/iris/search/index.cfm?keyword=)
- European Chemicals Agency (ECHA) Risk Assessments (https://echa.europa.eu/)
- Health Canada Environmental Health Assessments (https://www.canada.ca/en/health-canada.html)
- International Agency for Research on Cancer (IARC) Monographs (http://monographs.iarc.fr/ENG/Monographs/PDFs/index.php)
- New York State Department of Health — Health Topics A to Z (http://www.health.ny.gov/healthaz/)
- National Academy of Sciences reports and publications (http://www.nationalacademies.org/publications/)
- NTP publications, including, but not limited to, technical reports, nominations for toxicological evaluation documents, RoC, RoC background documents or monographs, and NTP Office of Health Assessment (OHAT) (formerly CERHR) monographs (http://ntp.niehs.nih.gov; search NTP)
- World Health Organization (WHO)/United Nations Environment Programme (UNEP) International Programme on Chemical Safety (IPCS) INCHEM-related documents (http://www.inchem.org/)

B.2. Databases or Web Resources

- Carcinogenic Potency Database (https://cebs.niehs.nih.gov/cebs/) and (https://www.lhasalimited.org/products/lhasa-carcinogenicity-database.htm)
- U.S. Environmental Protection Agency (EPA) CompTox Chemical Dashboard (https://comptox.epa.gov/dashboard/)
- European Chemical Agency (ECHA) Registration, Evaluation, Authorisation, and Restriction of Chemicals (REACH) (https://www.oecd.org/chemicalsafety/risk-assessment/echemportalglobalportaltoinformationonchemicalsubstances.htm)
- European Food Safety Authority (http://www.efsa.europa.eu/en/publications.htm)
- International Labour Organization (http://www.ilo.org/global/publications/lang--en/index.htm)
- International Uniform Chemical Information Database (http://iuclid.eu/)
- National Institute for Occupational Safety and Health (NIOSH) Publications (http://www2.cdc.gov/nioshtic-2/)

- U.S. NTP Integrated Chemical Environment (ICE) (https://ice.ntp.niehs.nih.gov/)
- United Nations Environment Programme (www.unep.org)
- U.S. National Library of Medicine (NLM) TOXNET

  TOXNET has moved. As part of a broader NLM reorganization, most of NLM's toxicology information services have been integrated into other NLM products and services. See https://www.nlm.nih.gov/toxnet/index.html for guidance on how to access these sources.

- OECD eChem Portal (https://www.oecd.org/chemicalsafety/risk-assessment/echemportalglobalportaltoinformationonchemicalsubstances.htm)

B.3. Exposure-specific General Sources

The initial step of the exposure section literature search strategy and evidence mapping process is to identify relevant exposure information for the candidate substance in the following sources:

- American Conference of Governmental Industrial Hygienists (ACGIH) Threshhold Limit Value/Biological Exposure Indices (TLV/BEI) documentation (available for purchase) (https://portal.acgih.org/s/store#/store/browse/cat/a0s4W00000g02f8QAA/tiles)
- ATSDR Public Health Assessment Guidance Manual (PHAGM) (https://www.atsdr.cdc.gov/pha-guidance/resources/)
- ATSDR Toxicological Profiles (https://www.atsdr.cdc.gov/toxprofiledocs/index.html)
- U.S. EPA AP-42, Compilation of Air Pollutant Emission Factors (https://www.epa.gov/air-emissions-factors-and-quantification/ap-42-compilation-air-emissions-factors

  https://cfpub.epa.gov/webfire/

  https://cfpub.epa.gov/webfire/index.cfm?action=fire.detailedSearch)

- U.S. EPA Chemical Data Reporting (https://chemview.epa.gov/chemview)
- U.S. EPA CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard/)
- U.S. EPA EJView Database (https://ejscreen.epa.gov/mapper/)
- U.S. EPA Enforcement and Compliance History Online (ECHO) database (https://echo.epa.gov/)
- U.S. EPA Locating and Estimating (L&E) Documents — Locating and Estimating Air Toxic Emissions from Sources of (source category or substance) (https://www.epa.gov/air-emissions-factors-and-quantification/locating-and-estimating-le-documents)
- U.S. EPA/Office of Pesticide Programs (OPP) National Pesticide Information Retrieval System (http://npirspublic.ceris.purdue.edu/ppis/)

- U.S. EPA Toxics Release Inventory (https://enviro.epa.gov/triexplorer/tri_release.chemical)
- U.S. Food and Drug Administration (FDA) Orange Book: Approved Drug Products with Therapeutic Equivalence Evaluations (http://www.fda.gov/cder/ob/default.htm)
- U.S. FDA Pesticide Program Residue Monitoring (http://www.fda.gov/Food/FoodborneIllnessContaminants/Pesticides/UCM2006797.htm)
- U.S. FDA Total Diet Study http://www.fda.gov/Food/FoodScienceResearch/TotalDietStudy/ucm184293.htm
- Kirk-Othmer Encyclopedia of Chemical Technology (online access through the NIEHS Library)
- Material Safety Data Sheets (MSDS) (https://chemicalsafety.com/sds-search/)
- MedlinePlus (https://medlineplus.gov/)
- National Health and Nutrition Examination Survey (NHANES) https://www.cdc.gov/nchs/nhanes.htm

  https://www.cdc.gov/exposurereport/

  https://www.cdc.gov/biomonitoring/biomonitoring_summaries.html)
- NIOSH Health Hazard Evaluations (https://www2a.cdc.gov/hhe/search.asp)
- NIOSH-sponsored Research Publications and Products (https://www.cdc.gov/niosh/pubs/)
- NIOSH Worker Health Charts (https://wwwn.cdc.gov/niosh-whc/)
- NIOSH Workplace Safety and Health Topics (https://www.cdc.gov/niosh/topics/)
- NLM TOXNET: ChemIDplus, Hazardous Substances Data Bank (HSDB), Haz-Map, Consumer Product Information Database (formerly Household Products Database), TOXMAP (TOXNET has moved. As part of a broader NLM reorganization, most of NLM's toxicology information services have been integrated into other NLM products and services. See https://www.nlm.nih.gov/toxnet/index.html for guidance on how to access these sources.)
- Sphera CyberRegs (https://www.cyberregs.com)
- Ullmann's Encyclopedia of Industrial Chemistry (https://onlinelibrary.wiley.com/doi/book/10.1002/14356007)
- U.S. Air Force Defense Meteorological Satellite Program (https://catalog.data.gov/dataset/defense-meteorological-satellite-program-dmsp)
- U.S. Coast Guard National Response Center (https://nrc.uscg.mil/)
- U.S. Department of Agriculture Pesticide Recordkeeping Program (https://www.ams.usda.gov/rules-regulations/pesticide-records)
- U.S. Department of Labor, Bureau of Labor Statistics (BLS) (https://www.bls.gov/)

- U.S. Geological Survey Minerals Yearbook (https://www.usgs.gov/centers/national-minerals-information-center/publications) and Commodity Sheet Summaries (https://www.usgs.gov/centers/national-minerals-information-center/mineral-commodity-summaries)
- U.S. International Trade Commission (USITC) Interactive Tariff and Trade DataWeb (import/export data) (https://dataweb.usitc.gov/); Schedule B Codes for USITC Database Query (https://www.census.gov/foreign-trade/schedules/b/index.html)
- U.S. Patent and Trademark Office Patent Search (https://www.uspto.gov/patents/search); Trademark Electronic Search System (TESS) (https://tmsearch.uspto.gov/bin/gate.exe?f=tess&state=4804:91j259.1.1)
- WHO/UNEP IPCS INCHEM-related documents (https://inchem.org/#/)

Appendix C. Systematic Review-related Tools Used by the Report on Carcinogens

Table of Contents

Tables

The RoC program uses the following computer-based tools to gather, evaluate, and visualize data during candidate substance reviews. Not all tools are utilized for each review but may be used based on the scope and nature of the substance specific evaluation.

Table C-1. Selected Systematic Review and Other Related Tools

Tool Name	Description
Adobe Acrobat DC	Software for the creating, editing, and full text searching of PDFs.
Causaly	Online database for flexible searching to find relationship between topics.
ChemDraw	Software for drawing high quality images of chemical structures for publications.
CRAB	Fully integrated text mining tool that searches the PubMed database and identifies relevant data categories for topics including human cancer, animal tumors, and mechanistic studies.
Dexter	Semi-automated data extraction tool for abbreviated data extractions to support scoping and systematic evidence mapping products.
Distiller SR	Web-based screening tool with capabilities including AI classifiers and prioritization ranking for screening with prediction of when to stop.
DoCTER	Web-based "Document Classification and Topic Extraction Resource" software application that identifies and assigns references to topic clusters through machine learning or supervised clustering. DoCTER also helps identify the principal themes in a body of literature (Varghese et al. 2018).
HAWC	Content-management system that facilitates all aspects of environmental human health assessments. The system allows a team to store, evaluate, share, and present multiple literature searches and to record the review (e.g., quality) of the literature (Shapiro et al. 2018).
LeadScope	Computer software that links chemical and biological data, allowing researchers to visualize and interactively explore large sets of chemical compounds, their properties, and biological activities.
LION	Website that helps cancer researchers form hypotheses by providing a graph-based view of the research literature using searches from PubMed.
Lit-EmCee	Suite of tools designed to help researchers in the initial stages of literature reviews.
PubMed Abstract Sifter	Microsoft Excel based application that enhances existing search capabilities of PubMed by 1) allowing research to search effectively and triage results and 2) keep track of articles of interest. Users can operate PubMed Abstract Sifter as a stand-alone Excel program or as an integrated tab in EPA's CompTox Chemicals Dashboard.
QInsight	Online database that finds relationships between topics in various forms of data and information.
R	Open-source statistical software.
Stata	Statistical software for data science.
SWIFT Active Screener	Web-based, collaborative systematic review software application used during the document screening phase of systematic review and scoping activities (Howard et al. 2020).

Tool Name	Description
SWIFT Review	Freely available interactive workbench which provides numerous tools to assist with problem formulation and literature prioritization.
Table Builder	Web-based application to store, organize data, and output data (in tabular format) for systematic review literature-based assessment. Access to Table Builder is behind the NTP firewall (Shapiro et al. 2018).
Tableau	Software to visualize and explore data. The desktop app allows users to develop visualizations, and upload of visualizations is web-based and shareable via a URL.
WebPlot Digitizer	Web-based tool to create accurate estimation of data points from plots and graphs.

C.1. References

Howard BE, Phillips J, Tandon A, Maharana A, Elmore R, Mav D, Sedykh A, Thayer K, Merrick BA, Walker V, et al. 2020. SWIFT-Active Screener: Accelerated document screening through active learning and integrated recall estimation. Environ Int. 138:105623. https://doi.org/10.1016/j.envint.2020.105623

Shapiro AJ, Antoni S, Guyton KZ, Lunn RM, Loomis D, Rusyn I, Jahnke GD, Schwingl PJ, Mehta SS, Addington J, et al. 2018. Software tools to facilitate systematic review used for cancer hazard identification. Environ Health Perspect. 126(10):104501. https://doi.org/10.1289/EHP4224

Varghese A, Cawley M, Hong T. 2018. Supervised clustering for automated document classification and prioritization: A case study using toxicological abstracts. Environ Syst Decis. 38(3):398-414. https://doi.org/10.1007/s10669-017-9670-5

Appendix D. Background Information on Key Characteristics of Carcinogens Biomarkers and Indicators

Table of Contents

Tables

Appendix D provides background information on the most commonly used biomarkers and indicators for each key characteristic of carcinogens (KCC). We used the term indicators for effects that are not biomarkers per se, such as outcomes (e.g., increased infection for immunosuppression), histopathology, or organ weights. Information for each KCC is captured in a table. This information can inform both the assessment of the informativeness of the mechanistic studies (see Section 6.3) and reaching conclusions of the confidence of the evidence for each KCC (see Table 6-14, Section 6.4, which provides the ideal type of evidence to reach high confidence for a given KCC).

The 2020 publication by Smith et al. was a starting point for the tables and was supplemented by additional publications. We are aware of international activities related to evaluating KCCs (e.g., the International Agency for Cancer workshop in July 2023, Key Characteristics of Cancer Assay Mapping Workgroup). The Handbook for the Report on Carcinogens (RoC) is a living document, which will be updated based on new publications and additional information on current assays (e.g., those in the KCC tables or additional ones), advancements in research related to the KCC and multi-omic data, and lessons learned from cancer hazard evaluations following this framework.

D.1. Is Electrophilic or Can Be Metabolically Activated to Electrophiles (KCC1)

Electrophiles are electron-poor atoms or molecules that form covalent bonds with electron-rich nucleophiles. Reversible electrophilic interactions with nucleophiles mediate many important biological functions; however, irreversible adduction of cellular macromolecules (e.g., DNA) with an electrophilic xenobiotic molecule is often an initiating step for many toxicological modes of action and pathogenic processes, including cancer. Table D-1 is organized based on relevance of the subtypes for reversible interactions; within each subtype, biomarkers are arranged by specificity (e.g., specific adducts followed by total measures), and then alphabetically.

Table D-1. Background Information on Common Biomarkers or Indicators of Electrophilicity (KCC1)

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Persistence	Comments or Guidance
DNA adducts	Base adducts • Adenine adducts: *N*7, N^6, *N*3, *N*1, *C8* • Cytosine adducts: *N*3, N^4, and O^2 • Guanine adducts (most targeted): *N7, N3,* N^2*, N1,* O^6*, C8* • Thymine adducts: *N*3, O^2, and O^4	Exposed humans or animals: Lymphocytes, urine (humans), tissue Ex vivo, in vitro, cell-free	Day(s)-month(s)	**Endpoint considerations**: • Most nucleophilic sites: N3 and N7 positions of guanine and adenine • Types of adduct formation depend on reactive chemicals, the nature of electrophiles, the ability to intercalate the DNA, and steric factors. • Animal studies on certain carcinogens (e.g., PAHs) suggest that DNA adducts detected in target tissue correlate with those found in surrogate tissue/cells (e.g., leukocytes). **Endpoint association with cancer:** • High levels of chemical-specific or bulky DNA adducts have been associated with increased risk of several types of cancer (e.g., colon, liver, lung in smokers, prostate, and stomach) in prospective studies, albeit risk estimate could be measuring risk due to exposure rather than early biological effects because adducts are also exposure biomarker.

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Persistence	Comments or Guidance
				• Animal studies have found that DNA adducts ranging from 53 to 5543 adducts/10^8 nucleotides in rat or mouse liver are associated with a 50% tumor incidence. **Commonly used assays: S**pecific adducts, ^{32}P-postlabeling (not chemical-specific), immunoassays (moderate specificity), LC-MS or GC-MS (specific)
-	Phosphate adducts	-	Longer persistence than DNA base adducts	**Endpoint considerations**: Challenging direct measurement due to their structural complexity
-	Adductomics (DNA)	-	-	**Endpoint considerations**: Totality of adducts (untargeted)
Protein adducts	Serum specific adducts Albumin Hemoglobin Most common is cysteine, but also aspartate, histidine, valine, tryptophan, glutamate, and lysine	Exposed humans or animals: Blood, RBC Ex vivo, in vitro, cell-free	Week(s)-month(s) (longer for hemoglobin)	**Endpoint considerations**: • Relative hardness (polarizability) of low to high: thiol group of cysteine, at a sulfur atom of methionine, primary amino groups (e.g., lysine, arginine), secondary amino groups of histidine • In humans, substance-specific hemoglobin and albumin adducts may also serve as exposure biomarkers. • Animal studies indicate that initial levels of DNA and protein adducts in animals administered a genotoxic agent are proportional to one another. **Common specific assays:** Immunoassay, fluorescence (e.g., laser-induced fluorescence), LC-MS, GC-MS

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Persistence	Comments or Guidance
-	Adductomics (protein)	Exposed humans	-	**Endpoint considerations**: Untargeted approach, usually for human serum albumin and hemoglobin **Commonly used assay**: LC-MS
-	Chemoproteomics	In situ, in vitro, ex vivo Exposed humans or animals	-	**Endpoint considerations**: Activity-based protein profiling using various probes
Protein modification	Protein thiol groups and other nucleophilic sites: Glutathione depletion	Rat or human liver microsomes	-	**Endpoint considerations:** Also used as a biomarker of oxidative stress (KCC5)
Electrophilic reactivity	E_{LUMO} or E_{HOMO} Band gap (E_{HOMO} - E_{HOMO})	In silico or in vitro cells	Not relevant	-
-	Chemical reactivity assays	In chemico with reference nucleophiles	-	-
-	Electrophilicity index	In silico	-	**Endpoint considerations**: Power of a chemical species to accept random numbers of electrons

Sources: (Carlsson et al. 2019; Hwa Yun et al. 2020; La Barbera et al. 2022; LoPachin and Gavin 2016; Ma et al. 2019; Nesnow et al. 1993; Pfohl-Leszkowicz 2008; Poirier 2016; Poirier et al. 2000; Schwöbel et al. 2011; Shalini et al. 2017; Smith et al. 2020; Törnqvist et al. 2002; Veglia et al. 2008).
GC-MS = gas chromatography-mass spectrometry, EHOMO = energy of the highest occupied molecular orbital, ELUMO = energy of the lowest unoccupied molecular orbital, LC-MS = liquid chromatography-mass spectrometry.
We used the term indicators for effects that are not biomarkers per se, such as outcomes.

D.2. Is Genotoxic (KCC2)

Genotoxicity typically refers to a substance's ability to cause gene mutations, DNA damage, structural chromosome aberrations, and aneuploidy (numerical chromosome aberrations) and is directly linked to carcinogenicity. OECD provides procedures for testing many of these endpoints. Genotoxicity overlaps with KCC1 (is electrophilic [e.g., DNA adducts]), KCC3 (alters DNA repair or causes genomic instability), and KCC5 (induces oxidative stress [e.g., oxidative damage to DNA). Subtypes in Table D-2 are categorized by increasing structure hierarchy, followed by older or less commonly used tests; within each subtype, biomarkers are arranged alphabetically.

Table D-2. Background Information on Common Biomarkers or Indicators of Genotoxicity (KCC2)

Subtype	Biomarkers/Indicators/Assay	Evidence Type: Biospecimen	Persistence/Induction	Comments or Guidance
DNA damage	DNA damage • Comet assays [in vitro and in vivo including specialized lesions] • γH2AX (Phosphorylation of H2AX) • TGx-DNA Damage Induced Transcriptomic Biomarker • Transcription factor p53 activation	Exposed humans or animals: Lymphocytes, exfoliated cells (exposed humans), target tissues In vitro: Various cell lines (e.g., human lymphoblastoid TK6 cells)	Hours	**Recommended assay:** OECD Test No. 489: In Vivo Mammalian Alkaline Comet Assay • Timing for in vivo comet assay depends on substance-specific metabolism and DNA repair kinetics Multiplexed fluorescence staining assays for DNA damage
Mutations	All biomarkers			**Endpoint association with cancer**: • Ames-positive and in vivo MN-positive chemicals are strong predictors of carcinogenicity. • Chemical-specific mutational spectra observed in cancers.

Subtype	Biomarkers/Indicators/Assay	Evidence Type: Biospecimen	Persistence/Induction	Comments or Guidance
	Bacterial reverse mutation tests • Base-pair substitution/frame shifts	In vitro: Bacteria (Ames) In vitro: Panel of *Salmonella* and some *E. coli* strains; positive result in any strain is relevant Exposed humans: Can use urine to test mutagenicity	Persistent (cell life)	**Recommended assay**: OECD Test No. 471: Bacterial Reverse Mutation Test
	Forward gene mutations: Reporter locus • *HPRT* • *XPRT* • *tk* (broader range)	In vitro Various cell lines (e.g., Chinese hamster As52) In vitro: Mouse lymphoma assay; TK6 cells	Persistent (cell life) Days to weeks Days	**Recommended assay**: OECD Test No. 476: In Vitro Mammalian Cell Gene Mutation Tests Using the *HPRT* and *XPRT* Genes **Recommended assay**: OECD Test No. 490: In Vitro Mammalian Cell Gene Mutation Tests Using the *Thymidine Kinase* Gene
	Glycophorin A *HPRT* mutational frequency	Exposed humans: blood (erythrocytes) Exposed humans (usually): lymphocytes		
	Pig-a gene mutation assay	Exposed humans or animals: Blood (erythrocytes)	Persistent (cell life)	**Recommended assay**: OECD Test No. 470: Mammalian Erythrocyte Pig-a Gene Mutation Assay
	Somatic or germ cell transgenic rodent assays (e.g., Big Blue mouse or rat)	Exposed rodents: almost every organ or tissue	Days (fast dividing) to weeks (slow dividing)	**Recommended assay**: OECD Test No. 488: Transgenic Rodent Somatic and Germ Cell Gene Mutation Assays
	Ultra-accurate, error-corrected DNA sequencing approaches (not locus-dependent)	In vitro, exposed animals or humans (blood, cells from urine)	Persistent	**Commonly used assays**: Duplex sequencing, PacBio HiFi sequencing
Chromosomal damage	Chromosomal aberration (structural) [CA] test with or without FIS)	In vitro or ex vivo: Primary cells or cell lines (e.g., lymphocytes)	Persistent (cell life)	**Endpoint association with cancer**: MN and CA associated with increased cancer risk in prospective cohort studies **Recommended assay**: OECD Test No. 473: In Vitro Mammalian Chromosomal Aberration Test

Subtype	Biomarkers/Indicators/Assay	Evidence Type: Biospecimen	Persistence/Induction	Comments or Guidance
		Exposed humans or animals: bone marrow, whole blood, lymphocytes, exfoliated cells (humans), target tissues	Days to weeks	**Recommended assay**: OECD Test No. 475: Mammalian Bone Marrow Chromosomal Aberration Test
	Micronucleus [MN] test: Structural and numerical (CBMN, centromere/kinetochore analysis)	In vitro cells Exposed humans or animals: Erythrocytes and other proliferating cells	Days to weeks	**Recommended assay**: OECD Test No. 487: In Vitro Mammalian Cell Micronucleus Test **Recommended assay**: OECD Test No. 474: Mammalian Erythrocyte Micronucleus Test
Older tests or less common: Mutations	Rodent dominant lethal Sex-linked recessive lethal (drosophila), assays in yeast	Exposed animals Exposed non-mammalian systems		**Recommended assay**: OECD Test No. 478 **Endpoint considerations**: No longer recommended; good indicators of genotoxicity; however, its relevance to humans is unclear
Older tests or less common: Chromatid damage	Sister-chromatid exchanges	In vitro Ex vivo: Cells from exposed humans/animals		**Endpoint considerations**: No longer recommended; findings do not correlate well with rodent carcinogenicity

Sources: (Battershill et al. 2008; Bonassi et al. 2011; Eastmond et al. 2009; European Commission 2008; Kirkland et al. 2005; Ladeira and Smajdova 2017; Li et al. 2019; Myers and Grant 2014; Norppa et al. 2006; Olsen et al. 1996; OECD 2015; 2016a; 2016b; 2016c; 2016d; 2016e; 2016f; 2016g; 2020; 2022; Smith et al. 2020).
CBMN = cytokinesis-blocked micronucleus, FISH = fluorescence in situ hybridization, HPRT = hypoxanthine-guanine phosphoribosyl transferase, MN = micronucleus, OECD = Organisation for Economic Co-operation and Development, TK = thymidine kinase, XRPT = xanthine phosphoribosyl transferase.
We used the term indicators for effects that are not biomarkers per se, such as outcomes.

D.3. Alters DNA Repair or Causes Genomic Instability (KCC3)

Genome integrity includes (1) nucleotide instability (NIN) or DNA repair capacity (DRC), (2) microsatellite instability (MSI or MIN), and (3) Chromosomal instability (CIN) and chromosome structure instability (CSI). It is maintained by DNA damage response pathways that include the following five major DNA repair pathways: (1) base excision repair (BER), (2) nucleotide excision repair (NER), (3) mismatch repair (MMR), (4) homologous recombination, and (5) nonhomologous end joining (Chatterjee and Walker 2017). Nucleotide instability results from replication errors and impairment of BER and NER pathways in nuclear DNA or BER in mitochondrial DNA. Within the DRC subtype, direct biomarkers of DNA damage (related to repair) are presented first, organized alphabetically, followed by measures of enzyme activities.

Table D-3. Background Information on Common Biomarkers or Indicators of DNA Repair or Genomic Instability (KCC3)

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Comments or Guidance
DNA repair capacity (DRC) or nucleotide instability (NIN)	All		**Endpoint considerations**: NIN results from replication errors and impairment of BER and NER pathways in nuclear DNA, or BER in mitochondrial DNA. **Endpoint association with cancer**: Inherited DNA repair-deficiency syndromes are linked to an increased risk of cancer. Lower DRC is linked to elevated cancer risk for all cancer combined and several specific cancer types (meta-analyses) using various assays.
	Comet assay: DNA damage	In vitro (kinetics): cells treated at various times	**Assay considerations**: • OECD recommends over UDS, OECD TG 489 (in vivo comet assay) • Assays can be done in vivo but are usually not done due to expense (multiple time points)

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Comments or Guidance
	Comet: in vitro DNA repair assay	Exposed humans or animals: Tissues and lymphocytes Possible in vitro	**Endpoint consideration**: Extracts of tissues/cells from exposed animals/humans mixed in vitro with substrate DNA (e.g., from cells exposed to certain agents); artificially induce BER- or NER-related lesions
	Challenge assay	Ex vivo: lymphocytes from humans	**Endpoint consideration**: DNA strand breaks (comet) or chromosome aberrations after treatment with IR, UVR, or other agents; validated for its sensitivity and specificity
	γH2AX assay: double-strand breaks and repair proteins	Ex vivo: lymphocytes from humans Exposed humans	**Endpoint association with cancer**: Increased γH2AX foci were associated with significantly increased risks for cancer.
	Host cell reactivation: site-specific DNA lesions	Ex vivo: lymphocytes from humans	**Assay considerations**: • Newer test: fluorescence-based/multiplex; different types of DNA damage • Older tests require a significant amount of blood and have high background rates
	Topoisomerases (I and II) activity	In vitro: Cell extracts (nuclear and cytosolic) from cultured cells, or cell or tissue extracts from exposed humans or animals	**Endpoint association with cancer:** Topoisomerase inhibitors are anti-cancer drugs, and some are also human carcinogens.
	UDS	In vitro	**Assay considerations**: • Low sensitivity: OECD TG 482 (retired in 2014)
		Exposed animals: mammalian liver	• Low sensitivity for some types of chemicals; OECD TG 486

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Comments or Guidance
Microsatellite instability (MSI or MIN)	MSI markers	Usually in tumors	**Endpoint association with cancer**: Random insertion and expansion of microsatellites can lead to a hypermutable phenotype. Result from defects in MMR and are features of several types of MMR-deficient cancers. **Assay considerations**: • Commonly used assays: NGS, PCR with HPLC or fluorescence or radioactive probes, and gel or capillary electrophoresis • NGS allows for a greater number of MSI markers to be analyzed.
Chromosomal instability (CIN) and chromosome structure instability (CSI)	Inter-/intra-chromosomal translocations: spectral karyotyping, WGS/NGS Copy number variations: duplications, deletions, amplifications, and insertions, WGS/NGS, aCGH	Exposed humans or animals, in vitro: Karyotyping: single-cell level Exposed humans or animals, in vitro: WGS/NGS: single- or multi-cell Exposed humans or animals, in vitro: Other assays: usually multi-cell Cells: in vitro or from exposed humans or animals DNA from cancer or normal cells (controls)	**Endpoint considerations**: Rate of gain or loss of segmental and whole chromosomes during cell division; consequences are aneuploidy and LOH **Endpoint association with cancer:** • CIN is characteristic of 90% of solid tumors and is associated with increased tumor progression and invasiveness. • Inherited CSI and CIN syndromes are associated with increased cancer risk in humans.

Sources: (Azqueta et al. 2014; Baudrin et al. 2018; Figueroa-González and Pérez-Plasencia 2017; Gantchev et al. 2022; Kaina et al. 2018; Nitiss et al. 2012; OECD 2016f; Owiti et al. 2021; Smith et al. 2020; Thompson and Compton 2011; Wu et al. 2022).
aCGH = array comparative genomic hybridization, BER = base excision repair, CIN = chromosomal instability, CSI = chromosome structure instability, DRC = DNA repair capacity, HPLC = high preference liquid chromatography, IR = ionizing radiation, LOH = Loss of heterozygosity, MMR = mismatch repair, MIN = microsatellite instability, NER = nucleotide excision repair, OECD = Organisation for Economic Co-operation and Development, WGS/NGS = whole genome sequencing/next-generation sequencing, XRCC1 = X-ray repair cross complementing 1, UDS = unscheduled DNA synthesis, UVR = ultraviolet radiation.
We used the term indicators for effects that are not biomarkers per se, such as outcomes.

D.4. Induces Epigenetic Alterations (KCC4)

Epigenetics encompasses various mechanisms that regulate gene expression without altering the underlying DNA sequence and plays a key role in normal mammalian development (Marczylo et al. 2016). The epigenome is a term that encompasses the complete epigenetic status of a cell at any given time. Environmentally-induced epigenetic toxicity is a rapidly emerging field and there is increasing evidence that epigenetic alterations play an important role in chemically-induced carcinogenesis (Chappell et al. 2016). The following sections provide guidance on rating the confidence in the body of evidence that a substance induces epigenetic changes and that these changes are key mechanistic events leading to carcinogenicity. Subtypes are organized by increasing structure hierarchy.

Table D-4. Background Information on Common Biomarkers or Indicators of Epigenetic Effects (KCC4)

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Persistence	Comments or Guidance
DNA methylation: 5-methylcytosine (5mC), Promoters/CpG islands	Gene-specific: Clock genes or genes related to cancer mechanisms or pathways (e.g., tumor suppressor, DNA repair, immune function, metabolism, etc.)	Exposed humans or animals: Body fluids or cells (e.g., blood, sputum, urine, buccal cells), dried blood, fresh or formalin-fixed tissues, tissue (tumor or non-tumor) circulating cell-free DNA In vitro	Rapid Potentially reversible and can be inherited Stable	**Endpoint association with cancer**: Biological age can be assessed by measuring methylation at selected CpG sites in clock genes. Accelerated epigenetic aging measured in WBC is associated with an increased risk of cancer development. **Assay considerations**: Two-step process: (1) separate methyl from non-methyl (e.g., bisulfite conversion, restriction enzyme) and (2) amplify and measure (e.g., immunostaining, GS-MS, HPLC-MS); Immunoassays are less reliable.
	Global methylation content (level of 5mC relative to total cystine) • Direct measurement • Repeated elements LINE-1 or Alu	Same as above	Same as above	**Endpoint association with cancer or considerations:** Cancer cells usually exhibit global hypomethylation and promoter hypermethylation at CpG islands of specific genes. • DNA methylation can serve as a biomarker of exposure, biological age (not chronological), cancer risk, or early cancer detection.

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Persistence	Comments or Guidance
				• Most methylation occurs in repetitive elements (30% of the genome): Line-1/Alu methylation levels in cancerous tissue correlate with global methylation content, and the correlation varies across cell types (usually measured by pyrosequencing). **Assay considerations**: Same as gene-specific assays
	Genome-scans or panel of genes			**Commonly used assays**: microarrays, next-generation sequencing
Chromatin: histone modifications	Circulating, variants, or post-translational modifications (PTM) usually at the N- and C-terminal tails	Exposed humans or animals: Body fluids or cells (e.g., blood, sputum, urine, buccal cells), dried blood, fresh or formalin-fixed tissues, tissue (tumor or non-tumor) circulating cell-free DNA In vitro	Potentially reversible Lysine methylation and acetylation is reversible	**Endpoint association with cancer**: Lower or higher levels of PTM are associated with cancer progression and are used in cancer diagnosis.

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Persistence	Comments or Guidance
	Gene-specific: PTM Global genomic content: direct measurement of circulating histones			**Endpoint considerations**: • 4 core histones (H2A, H2B, H3, and H4) and one linker (H1) • Main modifications: acetylation, methylation, phosphorylation, and ubiquitination • Histone methylation occurs at lysine and arginine residues, mainly at H3, followed by H4. Acetylation occurs at lysine residues. Monoubiquitination most commonly occurs on H2A and H2B (MS). • Functional consequence: Transcription (e.g., RNA seq) **Assay considerations**: • Immunoassays (e.g., ELISA), MS • ChIP qPCR (parallel sequencing technologies coupled to chromatin immunoprecipitation) • Many genes: ChIP-on-Chip (chromatin immunoprecipitation with DNA microarray analysis)
Chromatin accessibility	Tn5 transposase activity			**Assay considerations**: Transposase-accessible chromatin using sequencing (ATAC-seq)

Sources: (Chen et al. 2022; García-Giménez et al. 2017; Mehrmohamadi et al. 2021; Nowacka-Zawisza and Wiśnik 2017; O'Brien et al. 2018; Pajares et al. 2021; Park 2020; Smith et al. 2020; Sun et al. 2019; Vryer and Saffery 2017; Xu and Taylor 2014; Zhao and Shilatifard 2019).
Line-1 = long interspersed numerical elements, Alu are short interspersed numerical elements, ELISA = enzyme-linked immunosorbent assay, WBC = white blood cells
We used the term indicators for effects that are not biomarkers per se, such as outcomes

D.5. Induces Oxidative Stress (KCC5)

Oxidative stress occurs when there is an imbalance in the redox status within target tissues that favors formation of reactive oxygen species (ROS) and/or reactive nitrogen species (RNS) at the expense of their detoxification (Smith et al. 2020). This imbalance can lead to oxidative damage to DNA, proteins, and lipids and other effects that are directly related to several other KCCs. These include genotoxicity (KCC2), altered DNA repair (KCC3), chronic inflammation (KCC6), and altered cell proliferation, cell death, or nutrient supply (KCC10). Although oxidative stress is a KCC, it is not specific to carcinogens as many non-carcinogens can also induce oxidative stress. Table D-5 generally follows the steps in the biological process of oxidative stress. Within each subtype, biomarkers are generally organized alphabetically.

Table D-5. Background Information on Common Biomarkers or Indicators of Oxidative Stress (KCC5)

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Persistence/ Induction	Comments or Guidance
Oxidants: ROS, RNS, ROM	ROS: H_20_2, OH^-, ROO^-, or O_2^- RNS: $ONOO^-$, NO_2	Cell-free Exposed humans or animals: WBC or other cells, cellular components In vitro (real/time live cells)	Very transient (nsec to sec) to longer lived[a]	**Endpoint considerations**: Urinary H_2O_2 can be an indicator of whole-body oxidative stress but is confounded by diet **Assay considerations**: • Measurement instrumentation: electron spin resonance, fluorescent probes, biosensors • Fresh samples are needed; oxidants are unstable • Not recommended for ex vivo tissue homogenates
	ROM: ROOH	Exposed humans or animals: Serum/plasma		**Assay considerations**: Criticisms of the reliability of the d-ROM test
ROS Modifications: lipid, DNA, protein	All			**Endpoint considerations**: Systemic or tissue-specific oxidative stress
Lipid peroxidation	All	Exposed humans or animals: Body fluid (e.g., urine, serum, plasma), exhaled breath cells, tissues	Minutes to hours	**Endpoint considerations**: May directly affect the function of target molecules or enzymes or indicate local degrees of oxidative stress

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Persistence/ Induction	Comments or Guidance
	F2-isoP		Minutes (serum), hours (urine)	**Endpoint association with cancer**: Conflicting findings found for breast cancer risk **Assay considerations**: • IsoP in serum and urine correlate with in vivo oxidative stress in humans and animals (unaffected by diet). • Preferred method and recognized by EFSA.
	MDA/TBARs			**Endpoint or assay considerations**: • MDA/TBARs is a nonspecific biomarker and is susceptible to methodological bias, yet it may have clinical relevance (induces IL-7 producing cells). However, MDA/TBARS is not recommended as the only test of lipid peroxidation. • MDA measured by MS is useful.
	Others: HNE, LOOH, oxLDL			**Assay considerations**: oxLDL recognized by EFSA
Oxidative damage to DNA/RNA	8-OHdG	Exposed humans or animals: Urine, plasma, serum, tissue	Minutes	**Endpoint considerations**: Rapidly repaired, usually measured in urine, may serve as an indicator of whole-body oxidative stress **Endpoint association with cancer**: Some evidence shows that pre-diagnosis frequencies are associated with increased breast cancer risk in postmenopausal women and lung cancer in non-smokers. Often carries greater weight than other biomarkers
	Comet assay modified with lesion-specific repair endonucleases (e.g., OGG1, FPG, Endonuclease III)	Cells (e.g., leukocytes), ex vivo		**Assay considerations**: Accepted by EFSA

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Persistence/ Induction	Comments or Guidance
	Repair enzymes: hOGG APE			
	Oxidized guanine/guanosine (OxGua)	Exposed humans or animals: Urine		**Endpoint considerations**: OxGua molecules are derived from repair products of the oxidatively generated DNA/RNA lesions, 8-OH-dGuo from DNA, and 8-OHGuo from RNA. **Endpoint association with cancer**: Pre-diagnosed levels are associated with increased risks of all cancer in non-smokers and possibly men; colorectal cancer in women, non-smokers, and non-obese people; and prostate cancer in non-smokers.
Oxidative stress: Proteins	Carbonylated proteins (CPs), AOPP s-glutathionylation 3-nitrotyrosine	Exposed humans or animals: Plasma/serum	Days	**Endpoint considerations**: • CPs: irreversible; a hallmark of oxidative stress and is biologically significant and clinically relevant • Prone to methodological artifacts • Circulating levels of the biomarkers (3-nitrotyrosine) are not equivalent to tissue levels
ROS generating enzymes	MPO, XO	Exposed humans or animals: serum, urine, tissues Ex vivo: neutrophils	Hours	**Endpoint considerations**: MPO released from neutrophils can also be an indicator of inflammation (KCC6). **Endpoint association with cancer**: MPO is associated with cancer progression.

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Persistence/ Induction	Comments or Guidance
Inflammation/oxidative stress biomarkers	COX-2	Exposed humans or animals: tissues, serum In vitro	Hours to days	**Endpoint association with cancer**: COX-2 inhibitors can prevent the carcinogenesis of colorectal cancer. **Endpoint considerations**: It can also be considered as a proinflammatory biomarker (KCC6).
Antioxidant status	Enzymes: SOD, catalase, GST, GPx	Exposed humans or animals: Serum, erythrocytes (catalase)	Minutes	
	GSH; Vitamin A, C, E		Minutes	**Assay considerations:** Unstable, difficult to analyze
	Nrf2-ARE response pathway	Exposed humans or animals: Leukocytes, tissue		**Assay considerations:** Difficult to analyze
	Total antioxidant capacity			
Indices	GSH/GSSG ratio Oxy score (damage – protection)			

Sources: (Andries et al. 2021; Basu 1998; Brenner et al. 2014; Dai et al. 2009; Frijhoff et al. 2015; Gào et al. 2019; Gryszczyńska et al. 2017; Harris et al. 2008; Ho et al. 2013; Ito et al. 2017; Katerji et al. 2019; Lee et al. 2017; Lim and Thomas 2013; Loft et al. 2013; Loft et al. 2006; Marrocco et al. 2017; Menzel et al. 2021; Murphy et al. 2022; Smith et al. 2020; Tas and Erturk 2017; Valadez-Cosmes et al. 2022).

8-OH-dG = 8-hydroxy-2'-deoxyguanosine, APE = apurinic/apyrimidinic endonuclease, ARE = antioxidant responsive element, AOPP = advanced oxidation protein products, COX-2 = cyclooxygenase-2, FPG = formamidopyrimidine (fapy)-DNA glycosylase, GPx = glutathione peroxidase GSH = glutathione (reduced), GSSG = oxidized glutathione, GST = glutathione S-Transferase, H_20_2 = hydrogen peroxide, HNE = 4-hydroxy-2-nonenal, hOGG = human 8-oxoguanine-DNA-glycosylase, Iso-P = isoprostanes, LOOH = lipid hydroperoxides, oxLDL = oxidized low density lipoproteins, MDA = malondialdehyde, MPO = myeloperoxidase, NO_2 = nitrogen dioxide, O_2^- = superoxide, OH- = hydroxyl radical, OGG1 = 8-oxoguanine DNA glycosylase, ONOO- = peroxynitrite, RNS = reactive nitrogen species, ROM = reactive oxygen metabolites, ROO- = peroxyl radicals, ROOH = hydroperoxides, ROS = reactive oxygen species, SOD = superoxide dismutase, TBARS = thiobarbituric acid reactive substances, XO = xanthine oxidase, WBC = white blood cells.

We used the term indicators for effects that are not biomarkers per se, such as outcomes.

[a]Radical electrons/ionization charge species are very transient, others such as H_2O_2 are longer lived.

D.6. Induces Chronic Inflammation (KCC6) or Immune Activation

Many protein biomarkers (e.g., cytokines) can indicate chronic or acute inflammation depending on the exposure conditions. Thus, evidence of chronic or persistent/repeated exposure and/or the timing/duration of response is critical in determining whether the study is measuring chronic inflammation. The RoC review also considers immune activation (e.g., by B-cell antigens) which may be linked to chronic inflammation and both chronic inflammation and dysregulation/persistent immune activation can contribute to the development and progression of cancer. The concept of immunomodulation is part of the 2015 handbook, as recommended by the peer reviewers of that handbook. Table D-6 is organized inflammation-related diseases and pathology, with the remaining subtypes listed alphabetically. Within each subtype, biomarkers are organized alphabetically as well.

Table D-6. Background Information on Common Biomarkers or Indicators of Chronic Inflammation (KCC6) or Immune Activation

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Persistence	Comments or Guidance
Chronic inflammatory diseases (e.g., autoimmune diseases)	Increased risk or incidence of autoimmune diseases that have been linked to cancer	Exposed humans Animal models of autoimmune disease	Months to years	
Chronic inflammation with WBC infiltration	Histology Local evidence of infiltration of acute (with evidence of acute exacerbations from repeated exposures) or chronic inflammatory cells	Exposed animals or (possibly) humans: tissue	Timing/ persistence can vary	**Endpoint considerations**: • Specific cell types can indicate chronic inflammation. Granulocytes (including neutrophils) in tissue indicate acute inflammation. Lymphocytes, plasma cells, and monocytes/macrophages in tissue indicate chronic inflammation. • Pathologists can diagnose acute vs. chronic. • Histological evaluation is limited in exposed humans.

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Persistence	Comments or Guidance
Cytokines and chemokines (some proinflammatory)	Chemokines: MCP-1, MIP-2 Interferon: IFNγ Interleukins[a]: IL-1α, IL-1β, IL-2, IL-6, IL-8, IL-12, IL-15, IL-23 Transforming growth factor: TGFβ Tumor necrosis factor: TNFα	Exposed humans or animals: serum/plasma, tissue, body fluids, exhaled breath Ex vivo, in vitro	Minutes: IL-1β, IL-8, TNFα Hours: IL-6	**Endpoint considerations:** • IL-6 is involved in inflammation, autoimmunity, and B-cell malignancies. • The associated cytokine/chemokine receptors are also critical and ideally should be considered together with ligands. **Endpoint association with cancer**: • Pre-diagnosed elevated circulating (systemic) levels of IL-6 and IL-8 are associated with increased lung cancer risk. Increased IL-6 levels are also associated with all cancers combined and CRC in several studies or meta-analyses. • Experimental (rodent) models for the role of TNFα, TGFβ, IL-1, IL-6, and IL-23 in cancer development or progression.
Acute phase proteins	Amyloid A (serum, SAA) CRP	Ex vivo, in vitro Exposed humans or animals: serum/plasma, tissue	24 to 48 hr Hours to days	**Endpoint considerations**: CRP is nonspecific and the most sensitive acute phase protein in humans. **Endpoint association with cancer**: • Pre-diagnosed elevated circulating CRP is associated with increased cancer incidence/mortality for all cancers combined, and several cancer types, such as lung, CRC, breast, and ovarian. • SAA is associated with an increased risk of several cancers, such as lung and colon, and is correlated with CRP.
	ESR		Weeks	**Endpoint considerations and association with cancer:** • ESR is the most widely used laboratory test for evaluating inflammation status in clinical practice, including infection, autoimmunity, and cancer.

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Persistence	Comments or Guidance
Prostaglandin endoperoxide synthase	COX-2	Exposed humans or animals: tissue, serum In vitro	Hours to days	**Endpoint association with cancer**: COX-2 inhibitors can prevent the carcinogenesis of colorectal cancer. **Endpoint considerations**: It can also be considered as an oxidative stress biomarker (KCC5).
Transcription factors	JAK/STAT NF-κβ	Exposed humans or animals: cells/tissue Ex vivo, in vitro	Minutes Hours	**Endpoint considerations and association with cancer:** NF-κβ activation is essential for inflammation and is activated in several types of cancer.
WBC (circulating)	Increases in total WBC or leukocyte subsets: lymphocytes, monocytes, granulocytes Ratios (NLR, PLR, LMR) and SII Increased bone marrow hematopoiesis	Exposed humans or animals: blood	Days to weeks	**Endpoint considerations**: • Decreased systemic WBC can also indicate increased inflammation via extravasation into tissue (local inflammation). • In general, lymphocytes are chronic inflammation indicators, and granulocytes are acute inflammation indicators. • NLR represents the imbalance between the innate and adaptive immune response. • SII is based on peripheral lymphocyte, neutrophil, monocyte, and platelet counts. **Endpoint association with cancer**: • Pre-diagnosed elevated lymphocytes, monocytes, neutrophils, basophils, and NLR are associated with increased lung cancer risk. • Pre-diagnosed elevated leukocytes are associated with an increased risk of all cancers combined, lung cancer, or CRC.

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Persistence	Comments or Guidance
Immune cell activation	B cell stimulation/antigens (antibody production)	Exposed humans or animals: cells/tissue Ex vivo		**Endpoint association with cancer**: B-cell stimulation (by the by itself due to autoimmunity or by foreign antigens due to immunosuppression) leads to DNA damage from genomic recombination and mutation during class/isotype switching and somatic hypermutation and possibly increased risk of B-cell lymphoma.
	Macrophage and granulocyte phagocytosis, ROS production	Exposed humans or animals: cells/tissue Ex vivo, in vitro		**Endpoint considerations**: • ROS from immune cells contribute to oxidative stress (immune-regulated ROS can be considered under inflammation or oxidative stress- KCC5). • Persistent immune cell activation can be a driver of chronic proinflammatory responses. • It can be difficult to definitively determine if evidence/endpoints of immune activation are linked to chronic inflammation.

Sources: (Allin et al. 2016; Brenner et al. 2017; Brenner et al. 2014; Chauhan and Trivedi 2020; Germano et al. 2008; He et al. 2022; Hirano 2021; Ji et al. 2022; Kakourou et al. 2015; Kang et al. 2019; Liu et al. 2021; Michels et al. 2021; Puar et al. 2018; Qian et al. 2019; Smith et al. 2020; Van Hemelrijck et al. 2011; Wong et al. 2020; Zhou et al. 2012; Zhou et al. 2014; Zhu et al. 2022).
COX-2 = cyclooxygenase-2, CRP = C-reactive protein, ESR = erythrocyte sedimentation rate, JAK/STAT = Janus kinase/signal transducers and activators of transcription, INFγ = interferon gamma, LMR = lymphocyte to monocyte ratio, MCP-1 = macrophage chemoattractant protein-1, MCP-2 = macrophage chemoattractant protein-2, NF-κβ = nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κβ), NLR = neutrophil to lymphocyte ratio, PLR = platelet to lymphocyte ratio, ROS = reactive oxygen species, SAA = serum amyloid A, SII = systemic immune-inflammation index, TGFβ = transforming growth factor beta, TNFα = tumor necrosis factor alpha, WBC = white blood cell.
We used the term indicators for effects that are not biomarkers per se, such as outcomes.
[a]Some interleukins (like IL-8) are considered to be chemokines.

D.7. Is Immunosuppressive (KCC7)

Immunosuppression is characterized by a reduction in the capacity of the immune system to respond effectively to foreign antigens, including surface antigens on tumor cells. Potentially neoplastic cells may escape immune surveillance which facilitates the survival of tumor cells. T cells and natural killer cells are critical components in anti-tumor immunity. Table D-7 prioritizes the most clinically relevant and direct measures of immunosuppression, followed by indirect measures (or subtypes). Biomarkers within each subtype are organized alphabetically.

Table D-7. Background Information on Common Biomarkers or Indicators of Immunosuppression (KCC7)

Subtype or Assay	Biomarkers/Indicators	Evidence type: Biospecimen	Comments or Guidance
Increased infections	Increases in the incidence of opportunistic infections (especially viral)	Exposed humans (observational studies) or animals	**Endpoint considerations**: Strong evidence for immunosuppression
Immune function (challenge to foreign antigen)	Decreases in primary or secondary antibody response to vaccinations or natural antigens	Exposed humans (controlled clinical or observational studies) or animals	**Endpoint considerations**: • Strong evidence for immunosuppression but indicator may not be relevant for immunosurveillance and cancer risk. • Not usually conducted in humans because of ethical reasons. • Observational studies may evaluate whether environmental exposure affects vaccination antibody response.
	Decreases in NK function, phagocytosis/bacterial killing by PMNLs, antigen presentation	Exposed humans or animals Ex vivo	
Immune function (humoral or cell-mediated immunity)	Decreases in antibody production (e.g., T cell dependent, antigen-specific) including specific subclasses/isotypes, NK or CTL activity, T cell activity	Exposed humans or animals Ex vivo	**Endpoint association with cancer**: • Impact on CTL or NK activity and memory T cells may be most relevant for cancer (e.g., immunosurveillance) with B cell/antibody production less relevant. • Low CTL activity is associated with increased cancer risk.

Subtype or Assay	Biomarkers/Indicators	Evidence type: Biospecimen	Comments or Guidance
Immune components	Cytokines: IL-10, TGFβ	Exposed humans or animals	**Endpoint considerations**: Not sensitive or predictive alone to predict immunosuppression but may be used to support experimental animal data.
	Immunoglobulins (T cell-dependent or -independent)		
	Lymphocyte phenotyping (decreased NK, NKT, CD4+ T, CD8+ T; increased CTLA4+ T, Tregs)		
Immune components (hematology)	Altered WBC and leukocyte subsets	Exposed humans or animals	**Endpoint considerations**: Only severe changes are sufficient evidence of immunosuppression.
Immune organs (histopathology and organ weights)	Lymph node or splenic germinal centers, bone marrow suppression of hematopoiesis	Exposed humans or animals	**Endpoint considerations**: • Reduced organ weight may be secondary to general toxicity or stress. • Extensive histopathology may support the weight of evidence for immunosuppression. • Evaluation is limited in exposed humans.

Source: (Imai et al. 2000; IPCS 2012; Lebrec et al. 2016; Ponce et al. 2014; Sharma et al. 2017; Smith et al. 2020).
CTL = cytotoxic T lymphocyte, NK = natural killer cell, PMNL = polymorphonuclear leukocyte, TGFβ = transforming growth factor beta, WBC = white blood cell.
We used the term indicators for effects that are not biomarkers per se, such as outcomes.

D.8. Modulates Receptor-mediated Effects (KCC8)

Some receptors promote cell proliferation (e.g., hormone nuclear receptors such as estrogen [ER], androgen [AR], and progesterone [PR]). Activation of the aryl hydrocarbon receptor (AhR) can lead to immunosuppression (KCC7), in addition to effects on cell proliferation and survival (KCC10) (Smith et al. 2020). Table D-8 prioritizes the most relevant and direct measures of receptor-mediated effects, followed by indirect measures (or subtypes). Biomarkers within each subtype are organized alphabetically.

Table D-8. Background Information on Common Biomarkers or Indicators of Receptor-mediated Effects (KCC8)

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Persistence	Comments or Guidance
Receptor-mediated genomic action (NRs, AhR)	All			**Endpoint considerations**: Nuclear and cytosolic receptors are more relevant than peptide-regulated growth factor receptors.
	Activates or antagonizes receptors, e.g., ER, PR, AR, AhR	Cell-free Cell culture (in vitro): including non-cancerous and cancerous, endogenously expressing or transiently/stably transfected Exposed humans or animals	Months	**Assay considerations**: • Commonly used assays: ER, AR, AhR transactivation assays (OECD TG-455, TG-458, US EPA METHOD 4435, JIS K 0463) • Measures both receptor binding and function
	Interacts with receptors: ER and AR binding	Cell-free In vitro cells (usually human)	Months	**Assay considerations:** • Recommended assay: OECD TG 493 • Receptor binding-only assays do not indicate downstream effects
Alters hormone/ligand synthesis (levels), distribution, and transport	Binding to enzymes and transport proteins f h	Cell-free In vitro cells (human H295 cancer cells), human recombinant microsomes	Months	**Recommended assays**: Steroidogenesis with H295R (OECD TG-456), Aromatase assays (US EPA 890.1200/1550), Radioligand binding and displacement (e.g., displacement of 125I-T4)

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Persistence	Comments or Guidance
	Endogenous hormone levels (e.g., E2, T, T4) Sex hormone-binding globulins levels and modulation	Exposed humans or animals: serum/plasma, urine, other fluids, tissue		**Endpoint considerations**: Agents that modulate ligand transport, distribution, and clearance can also indirectly modulate receptor-mediated effects. **Recommended assays**: Usually hormone immunoassays **Endpoint association with cancer**: • Increased levels of endogenous estrogens and androgens with subsequent increase in receptor-mediated activation is associated with breast cancer risk in pre-and postmenopausal women. • Estrogen (endometrial, ovary, and limited for breast cancer) and estrogen-progestogen menopausal therapy (breast and endometrial cancer), and estrogen-progestogen oral contraceptives (breast, cervical, and liver cancer) are known human carcinogens. • Pre-diagnosed circulating androgens and estrogens are associated with an increased risk of non-serous ovarian cancer in postmenopausal women. • Thyrotropin levels that are suggestive of hyperthyroid function are associated with increased cancer risk, specifically, with an increased risk of lung and prostate cancer.
Alters nuclear receptor expression	ER, PR, AR	Exposed humans or animals: serum/plasma	Months	**Recommended assays**: Immunohistochemistry or western blotting with animal or human tissue **Endpoint association with cancer**: The overexpression of ER in benign breast epithelium is associated with estrogen sensitivity and breast cancer risk.

Sources: (Diel et al. 2000; Endogenous Hormones Breast Cancer Collaborative Group et al. 2011; Endogenous Hormones Breast Cancer Collaborative Group et al. 2013; Hellevik et al. 2009; Hogervorst et al. 2013; IARC 2012; 2019; Jacobs et al. 2020; Key et al. 2002; Khan et al. 1998; McIver et al. 2013; Meerts et al. 2000; NTP 2021; Smith et al. 2020; Trabert et al. 2019).
NR = nuclear receptor, AR = androgen receptor, ER = estrogen receptor, PR = progesterone receptor, AhR = aryl hydrocarbon receptor, OECD = Organisation for Economic Co-operation and Development,
We used the term indicators for effects that are not biomarkers per se, such as outcomes.

D.9. Causes Immortalization (KCC9)

Cancer cells can become immortal by reversing the normal telomere shortening process (which occurs with age) and lengthening their telomeres. Telomere shortening can lead to senescence or apoptosis. Table D-9 prioritizes direct measures of immortalization, followed by indirect measures (or subtypes). Subtypes are organized alphabetically.

Table D-9. Background Information on Common Biomarkers or Indicators of Immortalization (KCC9)

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Induction Period	Comments or Guidance
Cell transformation	Cell foci formation	In vitro: most common, include Syrian Hamster Embryo (SHE) and Balb/c 3T3 cells; others are C3H10T1/2 and Bhas 42 cells Exposed animals: Transfer transformed cells into an immune-deficient host such as nude mice	Weeks to months Months	**Endpoint considerations**: • Validated testing protocols to detect rodent carcinogens have been established for SHE and Balb/c 3T3 cell transformation assays, with SHE assays being the most reproducible between laboratories. Note that SHE cells are primary cells and less likely to undergo spontaneous transformations, whereas the other cell types noted are immortalized cell lines. • Inactivation of tumor suppressor genes or activation of oncogenes or increases in cyclin D1 expression in foci-derived or other cell assays adds confidence to any positive findings.
	De-differentiation (epithelial-mesenchymal transition)	In vitro: human or animal epithelial cells (e.g., lung, prostate)	Months	• Loss of E-cadherin, increased expression of fibronectin, increased growth in serum-free media
	Increased invasiveness	In vitro: human or animal epithelial cells	Months	• Increased invasiveness into extra cellular matrix or increased colony formation in soft agar

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Induction Period	Comments or Guidance
Cellular senescence inhibition	Alterations in senescence biomarkers: beta-galactosidase, cell cycle check-point inhibitors, or tumor suppressor gene expression Continued cell proliferation and inhibition of change to senescent cell phenotype	In vitro: embryonic stem cells, human fibroblasts, and other cell lines Exposed humans or animals: tissue		**Endpoint considerations**: Multiple biomarkers increase the validation of whether senescence has occurred or is inhibited (a single biomarker is insufficient).
Stem cell gene alterations	Morphological changes Transcription factors expression*: c-Myc, Oct3/4, Klf4, Sox2* Loss of tumor suppressor genes in stem cells: p16, p53, MYC	In vitro: Somatic cells Exposed animals: genetically modified, chimeric mice Exposed humans		**Endpoint considerations**: Induced pluripotent stem cells as evidenced by morphological effects (e.g., cell size, cell component ratios, colony ratio) and induction of transcription factors that can reset the somatic cell epigenome

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Induction Period	Comments or Guidance
Telomere and telomerase activity	Changes in telomere length (shorter or longer) Telomerase activity	Exposed humans or animals: Whole blood or peripheral blood leukocytes (PBLs), tissue (normal, embryonic) In vivo or ex vivo: Stem cells	Days	**Endpoint association with cancer**: • Increased telomerase activity is a risk factor for cancer. • Change in telomere length (e.g., shorter or longer) varies with different cancer types. More assay validation and standardization are needed across studies on telomere length. **Recommended assays**: • Telomerase activity assays: telomere repeat amplification protocols (TRAPs) or detection of telomerase-synthesized DNA; Telomerase activity is not detected in most human somatic cells. • Telomere length assays: quantitative polymerase chain reaction (qPCR) standardization to a single copy gene and may be used in tandem with TRAPs. **Endpoint considerations**: Telomere length, a marker of biological age, correlates with telomerase activity.

Sources: (Ahmadzai et al. 2012; Axelrad et al. 2013; Chiba et al. 2017; Creton et al. 2012; Dimri et al. 1995; González-Gualda et al. 2021; Mascolo et al. 2018; Masuda et al. 1997; Mender and Shay 2015; Ohnishi et al. 2014; Person et al. 2013; Schmidt and Plath 2012; Smith et al. 2020; Tokar et al. 2005; Tokar et al. 2010; Walcher et al. 2020; Wentzensen et al. 2011; Zhang et al. 2017).
We used the term indicators for effects that are not biomarkers per se, such as outcomes.

D.10. Alters Cell Proliferation, Cell Death, or Nutrient Supply (KCC10)

The ability of carcinogens to modify cell proliferation, cell death, and nutrient supply is central to their role in cancer development. By disrupting these essential processes, carcinogens create an environment where cells can grow uncontrollably, evade pathways that prevent tumor formation, and sustain their growth through enhanced nutrient and oxygen supply. Several of these biomarkers measure tumor progression. This table is mainly based on Smith et al. (2020) and will be expanded as part of our periodic updates based on lessons learned. Subtypes are organized based on the definition of the KCC. Within each subtype, biomarkers are organized alphabetically.

Table D-10. Background Information on Common Biomarkers or Indicators of Cell Proliferation, Cell Death, or Nutrient Supply (KCC10)

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Comments or Guidance
Sustained cellular proliferation	Cell numbers	Exposed animals or humans, in vitro	**Recommended assays**: radiolabeled 3H-thymine incorporation and non-radioactive bromodeoxyuridine (BrdU) immunoassay
	Colony formation	In vitro	
	DNA synthesis during S phase	In vitro Exposed animals Ex vivo: human tumor tissue	
	Hyperplasia	Ex vivo: animal and human tissue or biopsies	
	Metabolic activity	In vitro	
	Nuclear antigen (e.g., proliferating cell nuclear antigen or cell cycle biomarkers)	In vitro Exposed humans or animals	
Evasion or reduction of apoptosis	Apoptosis biomarkers: Expression of pro- and anti-apoptotic factors (e.g., caspase-3)	In vitro Exposed humans or animals: Tissue (including tumor) and blood	**Assay considerations**: Detection of apoptosis *in vivo* is challenging because of the short half-life of apoptotic cells, thus timing is critical. **Recommended assays**: Immunoassays, DNA arrays
	DNA fragmentation	In vitro	**Recommended assay**: TUNEL
	Loss of phospholipid asymmetry		**Recommended assay**: Annexin V with nuclear staining

Subtype	Biomarkers/Indicators	Evidence Type: Biospecimen	Comments or Guidance
Pathogenic (sustained) angiogenesis and neo-angiogenesis	Endothelial cell activation: proliferation, migration, and differentiation	In vitro or ex vivo: endothelial cells, umbilical vein endothelial cells, endothelial cell spheroid, stem cells,	**Recommended migration and invasion assays**: Cell culture wound closure, trans-well migration and invasion assays • **Other endothelial cell angiogenic assays**: endothelial cell proliferation in response to stimulus (e.g., VEGF, PDGF, bFGF); • chemotaxis in response to attractant gradient; • haptotaxis or tubular differentiation on matrix proteins; • proteolytic enzyme production; • tube-forming assays (e.g., on Matrigel).
	Endothelial sprouting, migration, and differentiation into capillaries	Ex vivo: umbilical artery	**Recommended assay**: Human arterial ring assay
	Vascularity: Active neovascularization	Exposed humans or animals: tissue; ex vivo	**Recommended assays**: Histological assessment of vessels, increased tumor levels of pro-angiogenic factors for endothelial cell activation (e.g., VEGF, PDGF, bFGF), tissue blood flow rates
	Vascularity: Tumor vascularity	Exposed humans or animals: tumor tissue; ex vivo	**Endpoint considerations**: Angiogenic antigens (e.g., endogen, nestin)
Glycolytic (Warburg) shift	Cellular respiration	In vitro	
	Glucose rate	Exposed humans or animals: tumors	**Endpoint association with cancer**: Promotes tumorigenesis and malignancy progression; it remains uncertain whether it is a direct cause or a consequence of cancer

Source: (Beresford et al. 2006; Brown et al. 2016; Devic 2016; Eccles et al. 2016; Nishida et al. 2006; Nowak-Sliwinska et al. 2018; Pang et al. 2016; Seano and Primo 2016; Smith et al. 2020; Tozer et al. 2016; Ward et al. 2008; Zippel et al. 2016).
TUNEL = terminal deoxynucleotidyl transferase-mediated dUTP nick end labeling VEGF = vascular endothelial growth factor; PDGF = platelet-derived growth factor; bFGF = basic fibroblast growth factor.
We used the term indicators for effects that are not biomarkers per se, such as outcomes.

D.11. References

Ahmadzai AA, Trevisan J, Fullwood NJ, Carmichael PL, Scott AD, Martin FL. 2012. The Syrian hamster embryo (SHE) assay (pH 6.7): Mechanisms of cell transformation and application of vibrational spectroscopy to objectively score endpoint alterations. Mutagenesis. 27(3):257-266. https://doi.org/10.1093/mutage/ger086

Allin KH, Bojesen SE, Nordestgaard BG. 2016. Inflammatory biomarkers and risk of cancer in 84,000 individuals from the general population. Int J Cancer. 139(7):1493-1500. https://doi.org/10.1002/ijc.30194

Andries A, Rozenski J, Vermeersch P, Mekahli D, Van Schepdael A. 2021. Recent progress in the LC-MS/MS analysis of oxidative stress biomarkers. Electrophoresis. 42(4):402-428. https://doi.org/10.1002/elps.202000208

Axelrad MD, Budagov T, Atzmon G. 2013. Telomere length and telomerase activity; a yin and yang of cell senescence. J Vis Exp. (75):e50246. https://doi.org/10.3791/50246

Azqueta A, Slyskova J, Langie SAS, O'Neill Gaivão I, Collins A. 2014. Comet assay to measure DNA repair: Approach and applications. Front Genet. 5:288. https://doi.org/10.3389/fgene.2014.00288

Basu S. 1998. Metabolism of 8-iso-prostaglandin $F_{2\alpha}$. FEBS Lett. 428(1-2):32-36. https://doi.org/10.1016/s0014-5793(98)00481-5

Battershill JM, Burnett K, Bull S. 2008. Factors affecting the incidence of genotoxicity biomarkers in peripheral blood lymphocytes: Impact on design of biomonitoring studies. Mutagenesis. 23(6):423-437. https://doi.org/10.1093/mutage/gen040

Baudrin LG, Deleuze JF, How-Kit A. 2018. Molecular and computational methods for the detection of microsatellite instability in cancer. Front Oncol. 8:621. https://doi.org/10.3389/fonc.2018.00621

Beresford MJ, Wilson GD, Makris A. 2006. Measuring proliferation in breast cancer: Practicalities and applications. Breast Cancer Res. 8(6):216. https://doi.org/10.1186/bcr1618

Bonassi S, El-Zein R, Bolognesi C, Fenech M. 2011. Micronuclei frequency in peripheral blood lymphocytes and cancer risk: Evidence from human studies. Mutagenesis. 26(1):93-100. https://doi.org/10.1093/mutage/geq075

Brenner DR, Fanidi A, Grankvist K, Muller DC, Brennan P, Manjer J, Byrnes G, Hodge A, Severi G, Giles GG, et al. 2017. Inflammatory cytokines and lung cancer risk in 3 prospective studies. Am J Epidemiol. 185(2):86-95. https://doi.org/10.1093/aje/kww159

Brenner DR, Scherer D, Muir K, Schildkraut J, Boffetta P, Spitz MR, Le Marchand L, Chan AT, Goode EL, Ulrich CM, et al. 2014. A review of the application of inflammatory biomarkers in epidemiologic cancer research. Cancer Epidemiol Biomarkers Prev. 23(9):1729-1751. https://doi.org/10.1158/1055-9965.EPI-14-0064

Brown RM, Meah CJ, Heath VL, Styles IB, Bicknell R. 2016. Tube-forming assays. In: Martin SG, Hewett PW, editors. Angiogenesis Protocols. 3rd ed. New York, NY: Humana Press. p. 149-157. https://doi.org/10.1007/978-1-4939-3628-1_9

Carlsson H, Rappaport SM, Törnqvist M. 2019. Protein adductomics: Methodologies for untargeted screening of adducts to serum albumin and hemoglobin in human blood samples. High Throughput. 8(1):6. https://doi.org/10.3390/ht8010006

Chappell G, Pogribny IP, Guyton KZ, Rusyn I. 2016. Epigenetic alterations induced by genotoxic occupational and environmental human chemical carcinogens: A systematic literature review. Mutat Res Rev Mutat Res. 768:27-45. https://doi.org/10.1016/j.mrrev.2016.03.004

Chatterjee N, Walker GC. 2017. Mechanisms of DNA damage, repair, and mutagenesis. Environ Mol Mutagen. 58(5):235-263. https://doi.org/10.1002/em.22087

Chauhan R, Trivedi V. 2020. Inflammatory markers in cancer: Potential resources. Front Biosci (Schol Ed). 12(1):1-24. https://doi.org/10.2741/S537

Chen L, Ganz PA, Sehl ME. 2022. DNA methylation, aging, and cancer risk: A mini-review. Front Bioinform. 2:847629. https://doi.org/10.3389/fbinf.2022.847629

Chiba K, Lorbeer FK, Shain AH, McSwiggen DT, Schruf E, Oh A, Ryu J, Darzacq X, Bastian BC, Hockemeyer D. 2017. Mutations in the promoter of the telomerase gene TERT contribute to tumorigenesis by a two-step mechanism. Science. 357(6358):1416-1420. https://doi.org/10.1126/science.aao0535

Creton S, Aardema MJ, Carmichael PL, Harvey JS, Martin FL, Newbold RF, O'Donovan MR, Pant K, Poth A, Sakai A, et al. 2012. Cell transformation assays for prediction of carcinogenic potential: State of the science and future research needs. Mutagenesis. 27(1):93-101. https://doi.org/10.1093/mutage/ger053

Dai Q, Gao YT, Shu XO, Yang G, Milne G, Cai Q, Wen W, Rothman N, Cai H, Li H, et al. 2009. Oxidative stress, obesity, and breast cancer risk: Results from the Shanghai Women's Health Study. J Clin Oncol. 27(15):2482-2488. https://doi.org/10.1200/JCO.2008.19.7970

Devic S. 2016. Warburg effect - a consequence or the cause of carcinogenesis? J Cancer. 7(7):817-822. https://doi.org/10.7150/jca.14274

Diel P, Schulz T, Smolnikar K, Strunck E, Vollmer G, Michna H. 2000. Ability of xeno- and phytoestrogens to modulate expression of estrogen-sensitive genes in rat uterus: Estrogenicity profiles and uterotropic activity. J Steroid Biochem Mol Biol. 73(1-2):1-10. https://doi.org/10.1016/s0960-0760(00)00051-0

Dimri GP, Lee X, Basile G, Acosta M, Scott G, Roskelley C, Medrano EE, Linskens M, Rubelj I, Pereira-Smith O, et al. 1995. A biomarker that identifies senescent human cells in culture and in aging skin in vivo. Proc Natl Acad Sci U S A. 92(20):9363-9367. https://doi.org/10.1073/pnas.92.20.9363

Eastmond DA, Hartwig A, Anderson D, Anwar WA, Cimino MC, Dobrev I, Douglas GR, Nohmi T, Phillips DH, Vickers C. 2009. Mutagenicity testing for chemical risk assessment:

Update of the WHO/IPCS Harmonized Scheme. Mutagenesis. 24(4):341-349. https://doi.org/10.1093/mutage/gep014

Eccles SA, Court W, Patterson L. 2016. In vitro assays for endothelial cell functions required for angiogenesis: Proliferation, motility, tubular differentiation, and matrix proteolysis. In: Martin SG, Hewett PW, editors. Angiogenesis Protocols. 3rd ed. New York, NY: Humana Press. p. 121-147. https://doi.org/10.1007/978-1-4939-3628-1_8

Endogenous Hormones Breast Cancer Collaborative Group, Key TJ, Appleby PN, Reeves GK, Roddam AW, Helzlsouer KJ, Alberg AJ, Rollison DE, Dorgan JF, Brinton LA, et al. 2011. Circulating sex hormones and breast cancer risk factors in postmenopausal women: Reanalysis of 13 studies. Br J Cancer. 105(5):709-722. https://doi.org/10.1038/bjc.2011.254

Endogenous Hormones Breast Cancer Collaborative Group, Key TJ, Appleby PN, Reeves GK, Travis RC, Alberg AJ, Barricarte A, Berrino F, Krogh V, Sieri S, et al. 2013. Sex hormones and risk of breast cancer in premenopausal women: A collaborative reanalysis of individual participant data from seven prospective studies. Lancet Oncol. 14(10):1009-1019. https://doi.org/10.1016/S1470-2045(13)70301-2

European Commission. 2008. Risk assessment methodologies and approaches for mutagenic and carcinogenic substances. Brussels, Belgium: European Commission, Health and Consumer Protection Directorate-General. https://ec.europa.eu/health/ph_risk/committees/04_scher/docs/scher_o_107.pdf

Figueroa-González G, Pérez-Plasencia C. 2017. Strategies for the evaluation of DNA damage and repair mechanisms in cancer. Oncol Lett. 13(6):3982-3988. https://doi.org/10.3892/ol.2017.6002

Frijhoff J, Winyard PG, Zarkovic N, Davies SS, Stocker R, Cheng D, Knight AR, Taylor EL, Oettrich J, Ruskovska T, et al. 2015. Clinical relevance of biomarkers of oxidative stress. Antioxid Redox Signal. 23(14):1144-1170. https://doi.org/10.1089/ars.2015.6317

Gantchev J, Ramchatesingh B, Berman-Rosa M, Sikorski D, Raveendra K, Amar L, Xu HH, Villarreal AM, Ordaz DJG, Litvinov IV. 2022. Tools used to assay genomic instability in cancers and cancer meiomitosis. J Cell Commun Signal. 16(2):159-177. https://doi.org/10.1007/s12079-021-00661-z

Gào X, Holleczek B, Cuk K, Zhang Y, Anusruti A, Xuan Y, Xu Y, Brenner H, Schöttker B. 2019. Investigation on potential associations of oxidatively generated DNA/RNA damage with lung, colorectal, breast, prostate and total cancer incidence. Sci Rep. 9(1):7109. https://doi.org/10.1038/s41598-019-42596-x

García-Giménez JL, Seco-Cervera M, Tollefsbol TO, Romá-Mateo C, Peiró-Chova L, Lapunzina P, Pallardó FV. 2017. Epigenetic biomarkers: Current strategies and future challenges for their use in the clinical laboratory. Crit Rev Clin Lab Sci. 54(7-8):529-550. https://doi.org/10.1080/10408363.2017.1410520

Germano G, Allavena P, Mantovani A. 2008. Cytokines as a key component of cancer-related inflammation. Cytokine. 43(3):374-379. https://doi.org/10.1016/j.cyto.2008.07.014

González-Gualda E, Baker AG, Fruk L, Muñoz-Espín D. 2021. A guide to assessing cellular senescence in vitro and in vivo. FEBS J. 288(1):56-80. https://doi.org/10.1111/febs.15570

Gryszczyńska B, Formanowicz D, Budzyń M, Wanic-Kossowska M, Pawliczak E, Formanowicz P, Majewski W, Strzyżewski KW, Kasprzak MP, Iskra M. 2017. Advanced oxidation protein products and carbonylated proteins as biomarkers of oxidative stress in selected atherosclerosis-mediated diseases. Biomed Res Int. 2017:4975264. https://doi.org/10.1155/2017/4975264

Harris RE, Beebe-Donk J, Alshafie GA. 2008. Similar reductions in the risk of human colon cancer by selective and nonselective cyclooxygenase-2 (COX-2) inhibitors. BMC Cancer. 8:237. https://doi.org/10.1186/1471-2407-8-237

He MM, Lo CH, Wang K, Polychronidis G, Wang L, Zhong R, Knudsen MD, Fang Z, Song M. 2022. Immune-mediated diseases associated with cancer risks. JAMA Oncol. 8(2):209-219. https://doi.org/10.1001/jamaoncol.2021.5680

Hellevik AI, Asvold BO, Bjøro T, Romundstad PR, Nilsen TIL, Vatten LJ. 2009. Thyroid function and cancer risk: A prospective population study. Cancer Epidemiol Biomarkers Prev. 18(2):570-574. https://doi.org/10.1158/1055-9965.EPI-08-0911

Hirano T. 2021. IL-6 in inflammation, autoimmunity and cancer. Int Immunol. 33(3):127-148. https://doi.org/10.1093/intimm/dxaa078

Ho E, Karimi Galougahi K, Liu CC, Bhindi R, Figtree GA. 2013. Biological markers of oxidative stress: Applications to cardiovascular research and practice. Redox Biol. 1(1):483-491. https://doi.org/10.1016/j.redox.2013.07.006

Hogervorst JG, Fortner RT, Mucci LA, Tworoger SS, Eliassen AH, Hankinson SE, Wilson KM. 2013. Associations between dietary acrylamide intake and plasma sex hormone levels. Cancer Epidemiol Biomarkers Prev. 22(11):2024-2036. https://doi.org/10.1158/1055-9965.EPI-13-0509

Hwa Yun B, Guo J, Bellamri M, Turesky RJ. 2020. DNA adducts: Formation, biological effects, and new biospecimens for mass spectrometric measurements in humans. Mass Spectrom Rev. 39(1-2):55-82. https://doi.org/10.1002/mas.21570

Imai K, Matsuyama S, Miyake S, Suga K, Nakachi K. 2000. Natural cytotoxic activity of peripheral-blood lymphocytes and cancer incidence: An 11-year follow-up study of a general population. Lancet. 356(9244):1795-1799. https://doi.org/10.1016/S0140-6736(00)03231-1

International Agency for Research on Cancer (IARC). 2012. Pharmaceuticals. (IARC monographs on the evaluation of carcinogenic risks to humans; vol. 100A). Lyon, France: International Agency for Research on Cancer. https://publications.iarc.fr/118

International Agency for Research on Cancer (IARC). 2019. Tumour site concordance and mechanisms of carcinogenesis. Baan RA, Stewart BW, Straif K, editors. Lyon, France: International Agency for Research on Cancer. https://publications.iarc.fr/Book-And-Report-Series/Iarc-Scientific-Publications/Tumour-Site-Concordance-And-Mechanisms-Of-Carcinogenesis-2019

International Programme on Chemical Safety (IPCS). 2012. Guidance for immunotoxicity risk assessment for chemicals. Geneva, Switzerland: World Health Organization. Harmonization Project Document No. 10. https://iris.who.int/handle/10665/330098

Ito F, Ito T, Suzuki C, Yahata T, Ikeda K, Hamaoka K. 2017. The application of a modified d-ROMs test for measurement of oxidative stress and oxidized high-density lipoprotein. Int J Mol Sci. 18(2):454. https://doi.org/10.3390/ijms18020454

Jacobs MN, Colacci A, Corvi R, Vaccari M, Aguila MC, Corvaro M, Delrue N, Desaulniers D, Ertych N, Jacobs A, et al. 2020. Chemical carcinogen safety testing: OECD expert group international consensus on the development of an integrated approach for the testing and assessment of chemical non-genotoxic carcinogens. Arch Toxicol. 94(8):2899-2923. https://doi.org/10.1007/s00204-020-02784-5

Ji M, Du L, Ma Z, Xie J, Huang Y, Wei X, Jiang X, Xu J, Yin R, Wang Y, et al. 2022. Circulating C-reactive protein increases lung cancer risk: Results from a prospective cohort of UK Biobank. Int J Cancer. 150(1):47-55. https://doi.org/10.1002/ijc.33780

Kaina B, Izzotti A, Xu J, Christmann M, Pulliero A, Zhao X, Dobreanu M, Au WW. 2018. Inherent and toxicant-provoked reduction in DNA repair capacity: A key mechanism for personalized risk assessment, cancer prevention and intervention, and response to therapy. Int J Hyg Environ Health. 221(7):993-1006. https://doi.org/10.1016/j.ijheh.2018.07.003

Kakourou A, Koutsioumpa C, Lopez DS, Hoffman-Bolton J, Bradwin G, Rifai N, Helzlsouer KJ, Platz EA, Tsilidis KK. 2015. Interleukin-6 and risk of colorectal cancer: Results from the CLUE II cohort and a meta-analysis of prospective studies. Cancer Causes Control. 26(10):1449-1460. https://doi.org/10.1007/s10552-015-0641-1

Kang J, Chang Y, Ahn J, Oh S, Koo DH, Lee YG, Shin H, Ryu S. 2019. Neutrophil-to-lymphocyte ratio and risk of lung cancer mortality in a low-risk population: A cohort study. Int J Cancer. 145(12):3267-3275. https://doi.org/10.1002/ijc.32640

Katerji M, Filippova M, Duerksen-Hughes P. 2019. Approaches and methods to measure oxidative stress in clinical samples: Research applications in the cancer field. Oxid Med Cell Longev. 2019:1279250. https://doi.org/10.1155/2019/1279250

Key T, Appleby P, Barnes I, Reeves G, Endogenous Hormones Breast Cancer Collaborative Group. 2002. Endogenous sex hormones and breast cancer in postmenopausal women: Reanalysis of nine prospective studies. J Natl Cancer Inst. 94(8):606-616. https://doi.org/10.1093/jnci/94.8.606

Khan SA, Rogers MA, Khurana KK, Meguid MM, Numann PJ. 1998. Estrogen receptor expression in benign breast epithelium and breast cancer risk. J Natl Cancer Inst. 90(1):37-42. https://doi.org/10.1093/jnci/90.1.37

Kirkland D, Aardema M, Henderson L, Müller L. 2005. Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens: I. Sensitivity, specificity and relative predictivity. Mutat Res. 584(1-2):1-256. https://doi.org/10.1016/j.mrgentox.2005.02.004

La Barbera G, Nommesen KD, Cuparencu C, Stanstrup J, Dragsted LO. 2022. A comprehensive database for DNA adductomics. Front Chem. 10:908572. https://doi.org/10.3389/fchem.2022.908572

Ladeira C, Smajdova L. 2017. The use of genotoxicity biomarkers in molecular epidemiology: Applications in environmental, occupational and dietary studies. AIMS Genet. 4(3):166-191. https://doi.org/10.3934/genet.2017.3.166

Lebrec H, Brennan FR, Haggerty H, Herzyk D, Kamperschroer C, Maier CC, Ponce R, Preston BD, Weinstock D, Mellon RD. 2016. HESI/FDA workshop on immunomodulators and cancer risk assessment: Building blocks for a weight-of-evidence approach. Regul Toxicol Pharmacol. 75:72-80. https://doi.org/10.1016/j.yrtph.2015.12.018

Lee JD, Cai Q, Shu XO, Nechuta SJ. 2017. The role of biomarkers of oxidative stress in breast cancer risk and prognosis: A systematic review of the epidemiologic literature. J Womens Health (Larchmt). 26(5):467-482. https://doi.org/10.1089/jwh.2016.5973

Li HH, Yauk CL, Chen R, Hyduke DR, Williams A, Frötschl R, Ellinger-Ziegelbauer H, Pettit S, Aubrecht J, Fornace AJ Jr. 2019. TGx-DDI, a transcriptomic biomarker for genotoxicity hazard assessment of pharmaceuticals and environmental chemicals. Front Big Data. 2:36. https://doi.org/10.3389/fdata.2019.00036

Lim MY, Thomas PS. 2013. Biomarkers in exhaled breath condensate and serum of chronic obstructive pulmonary disease and non-small-cell lung cancer. Int J Chronic Dis. 2013:578613. https://doi.org/10.1155/2013/578613

Liu C, Chu D, Kalantar-Zadeh K, George J, Young HA, Liu G. 2021. Cytokines: From clinical significance to quantification. Adv Sci (Weinh). 8(15):e2004433. https://doi.org/10.1002/advs.202004433

Loft S, Olsen A, Møller P, Poulsen HE, Tjønneland A. 2013. Association between 8-oxo-7,8-dihydro-2'-deoxyguanosine excretion and risk of postmenopausal breast cancer: Nested case-control study. Cancer Epidemiol Biomarkers Prev. 22(7):1289-1296. https://doi.org/10.1158/1055-9965.EPI-13-0229

Loft S, Svoboda P, Kasai H, Tjønneland A, Vogel U, Møller P, Overvad K, Raaschou-Nielsen O. 2006. Prospective study of 8-oxo-7,8-dihydro-2'-deoxyguanosine excretion and the risk of lung cancer. Carcinogenesis. 27(6):1245-1250. https://doi.org/10.1093/carcin/bgi313

LoPachin RM, Gavin T. 2016. Reactions of electrophiles with nucleophilic thiolate sites: Relevance to pathophysiological mechanisms and remediation. Free Radic Res. 50(2):195-205. https://doi.org/10.3109/10715762.2015.1094184

Ma B, Stepanov I, Hecht SS. 2019. Recent studies on DNA adducts resulting from human exposure to tobacco smoke. Toxics. 7(1):16. https://doi.org/10.3390/toxics7010016

Marczylo EL, Jacobs MN, Gant TW. 2016. Environmentally induced epigenetic toxicity: Potential public health concerns. Crit Rev Toxicol. 46(8):676-700. https://doi.org/10.1080/10408444.2016.1175417

Marrocco I, Altieri F, Peluso I. 2017. Measurement and clinical significance of biomarkers of oxidative stress in humans. Oxid Med Cell Longev. 2017:6501046. https://doi.org/10.1155/2017/6501046

Mascolo MG, Perdichizzi S, Vaccari M, Rotondo F, Zanzi C, Grilli S, Paparella M, Jacobs MN, Colacci A. 2018. The transformics assay: First steps for the development of an integrated approach to investigate the malignant cell transformation in vitro. Carcinogenesis. 39(7):955-967. https://doi.org/10.1093/carcin/bgy037

Masuda A, Kondo M, Saito T, Yatabe Y, Kobayashi T, Okamoto M, Suyama M, Takahashi T, Takahashi T. 1997. Establishment of human peripheral lung epithelial cell lines (HPL1) retaining differentiated characteristics and responsiveness to epidermal growth factor, hepatocyte growth factor, and transforming growth factor beta1. Cancer Res. 57(21):4898-4904.

McIver CR, Shaw IC, Gin S, Ellis MJ. 2013. Can thyroid hormone mimics affect thyroid hormone measurement by immunoassay? Clin Biochem. 46(13-14):1302-1304. https://doi.org/10.1016/j.clinbiochem.2013.06.030

Meerts IATM, van Zanden JJ, Luijks EAC, van Leeuwen-Bol I, Marsh G, Jakobsson E, Bergman A, Brouwer A. 2000. Potent competitive interactions of some brominated flame retardants and related compounds with human transthyretin in vitro. Toxicol Sci. 56(1):95-104. https://doi.org/10.1093/toxsci/56.1.95

Mehrmohamadi M, Sepehri MH, Nazer N, Norouzi MR. 2021. A comparative overview of epigenomic profiling methods. Front Cell Dev Biol. 9:714687. https://doi.org/10.3389/fcell.2021.714687

Mender I, Shay JW. 2015. Telomerase Repeated Amplification Protocol (TRAP). Bio Protoc. 5(22):e1657. https://doi.org/10.21769/bioprotoc.1657

Menzel A, Samouda H, Dohet F, Loap S, Ellulu MS, Bohn T. 2021. Common and novel markers for measuring inflammation and oxidative stress ex vivo in research and clinical practice—Which to use regarding disease outcomes? Antioxidants (Basel). 10(3):414. https://doi.org/10.3390/antiox10030414

Michels N, van Aart C, Morisse J, Mullee A, Huybrechts I. 2021. Chronic inflammation towards cancer incidence: A systematic review and meta-analysis of epidemiological studies. Crit Rev Oncol Hematol. 157:103177. https://doi.org/10.1016/j.critrevonc.2020.103177

Murphy MP, Bayir H, Belousov V, Chang CJ, Davies KJA, Davies MJ, Dick TP, Finkel T, Forman HJ, Janssen-Heininger Y, et al. 2022. Guidelines for measuring reactive oxygen species and oxidative damage in cells and in vivo. Nat Metab. 4(6):651-662. https://doi.org/10.1038/s42255-022-00591-z

Myers NT, Grant SG. 2014. The blood-based glycophorin A (GPA) human in vivo somatic mutation assay. In: Keohavong P, Grant SG, editors. Molecular Toxicology Protocols. 2nd ed. Totowa, NJ: Humana Press. p. 223-244. https://doi.org/10.1007/978-1-62703-739-6_18

National Toxicology Program (NTP). 2021. Substance profile: Estrogens, steroidal. In: Report on Carcinogens. 15th ed. Research Triangle Park, NC: U.S. Department of Health and Human

Services, Public Health Service, National Toxicology Program. https://doi.org/10.22427/ntp-other-1003

Nesnow S, Ross J, Nelson G, Holden K, Erexson G, Kligerman A, Gupta RC. 1993. Quantitative and temporal relationships between DNA adduct formation in target and surrogate tissues: Implications for biomonitoring. Environ Health Perspect. 101 Suppl 3:37-42. https://doi.org/10.1289/ehp.93101s337

Nishida N, Yano H, Nishida T, Kamura T, Kojiro M. 2006. Angiogenesis in cancer. Vasc Health Risk Manag. 2(3):213-219. https://www.tandfonline.com/doi/full/10.2147/vhrm.s23213

Nitiss JL, Soans E, Rogojina A, Seth A, Mishina M. 2012. Topoisomerase assays. Curr Protoc Pharmacol. 57:3.3.1-3.3.27. https://doi.org/10.1002/0471141755.ph0303s57

Norppa H, Bonassi S, Hansteen IL, Hagmar L, Strömberg U, Rössner P, Boffetta P, Lindholm C, Gundy S, Lazutka J, et al. 2006. Chromosomal aberrations and SCEs as biomarkers of cancer risk. Mutat Res. 600(1-2):37-45. https://doi.org/10.1016/j.mrfmmm.2006.05.030

Nowacka-Zawisza M, Wiśnik E. 2017. DNA methylation and histone modifications as epigenetic regulation in prostate cancer (Review). Oncol Rep. 38(5):2587-2596. https://doi.org/10.3892/or.2017.5972

Nowak-Sliwinska P, Alitalo K, Allen E, Anisimov A, Aplin AC, Auerbach R, Augustin HG, Bates DO, van Beijnum JR, Bender RHF, et al. 2018. Consensus guidelines for the use and interpretation of angiogenesis assays. Angiogenesis. 21(3):425-532. https://doi.org/10.1007/s10456-018-9613-x

O'Brien KM, Sandler DP, Xu Z, Kinyamu HK, Taylor JA, Weinberg CR. 2018. Vitamin D, DNA methylation, and breast cancer. Breast Cancer Res. 20(1):70. https://doi.org/10.1186/s13058-018-0994-y

Ohnishi K, Semi K, Yamamoto T, Shimizu M, Tanaka A, Mitsunaga K, Okita K, Osafune K, Arioka Y, Maeda T, et al. 2014. Premature termination of reprogramming in vivo leads to cancer development through altered epigenetic regulation. Cell. 156(4):663-677. https://doi.org/10.1016/j.cell.2014.01.005

Olsen LS, Nielsen LR, Nexø BA, Wassermann K. 1996. Somatic mutation detection in human biomonitoring. Pharmacol Toxicol. 78(6):364-373. https://doi.org/10.1111/j.1600-0773.1996.tb00220.x

Organisation for Economic Co-operation and Development (OECD). 2015. Test no. 478: Rodent dominant lethal test. Paris, France: OECD Publishing. https://doi.org/10.1787/9789264243118-en

Organisation for Economic Co-operation and Development (OECD). 2016a. Test no. 473: In vitro mammalian chromosomal aberration test. (OECD Guidelines for the Testing of Chemicals, Section 4). Paris, France: OECD Publishing. https://doi.org/10.1787/9789264264649-en

Organisation for Economic Co-operation and Development (OECD). 2016b. Test no. 474: Mammalian erythrocyte micronucleus test. (OECD Guidelines for the Testing of Chemicals, Section 4). Paris, France: OECD Publishing. https://doi.org/10.1787/9789264264762-en

Organisation for Economic Co-operation and Development (OECD). 2016c. Test no. 475: Mammalian bone marrow chromosomal aberration test. (OECD Guidelines for the Testing of Chemicals, Section 4). Paris, France: OECD Publishing. https://doi.org/10.1787/9789264264786-en

Organisation for Economic Co-operation and Development (OECD). 2016d. Test no. 476: In vitro mammalian cell gene mutation tests using the Hprt and xprt genes. (OECD Guidelines for the Testing of Chemicals, Section 4). Paris, France: OECD Publishing. https://doi.org/10.1787/9789264264809-en

Organisation for Economic Co-operation and Development (OECD). 2016e. Test no. 487: In vitro mammalian cell micronucleus test. (OECD Guidelines for the Testing of Chemicals, Section 4). Paris, France: OECD Publishing. https://doi.org/10.1787/9789264264861-en

Organisation for Economic Co-operation and Development (OECD). 2016f. Test no. 489: In vivo mammalian alkaline comet assay. (OECD Guidelines for the Testing of Chemicals, Section 4). Paris, France: OECD Publishing. https://doi.org/10.1787/9789264264885-en

Organisation for Economic Co-operation and Development (OECD). 2016g. Test no. 490: In vitro mammalian cell gene mutation tests using the thymidine kinase gene. (OECD Guidelines for the Testing of Chemicals, Section 4). Paris, France: OECD Publishing. https://doi.org/10.1787/9789264264908-en

Organisation for Economic Co-operation and Development (OECD). 2020. Test no. 471: Bacterial reverse mutation test. (OECD Guidelines for the Testing of Chemicals, Section 4). Paris, France: OECD Publishing. https://doi.org/10.1787/9789264071247-en

Organisation for Economic Co-operation and Development (OECD). 2022. Test no. 470: Mammalian erythrocyte Pig-a gene mutation assay. (OECD Guidelines for the Testing of Chemicals, Section 4). Paris, France: OECD Publishing. https://doi.org/10.1787/4faea90e-en

Owiti NA, Nagel ZD, Engelward BP. 2021. Fluorescence sheds light on DNA damage, DNA repair, and mutations. Trends Cancer. 7(3):240-248. https://doi.org/10.1016/j.trecan.2020.10.006

Pajares MJ, Alemany-Cosme E, Goñi S, Bandres E, Palanca-Ballester C, Sandoval J. 2021. Epigenetic regulation of microRNAs in cancer: Shortening the distance from bench to bedside. Int J Mol Sci. 22(14):7350. https://doi.org/10.3390/ijms22147350

Pang JM, Jene N, Fox SB. 2016. Assessing tumor angiogenesis in histological samples. In: Martin SG, Hewett PW, editors. Angiogenesis Protocols. 3rd ed. New York, NY: Humana Press. p. 3-33. https://doi.org/10.1007/978-1-4939-3628-1_1

Park HL. 2020. Epigenetic biomarkers for environmental exposures and personalized breast cancer prevention. Int J Environ Res Public Health. 17(4):1181. https://doi.org/10.3390/ijerph17041181

Person RJ, Tokar EJ, Xu Y, Orihuela R, Ngalame NNO, Waalkes MP. 2013. Chronic cadmium exposure in vitro induces cancer cell characteristics in human lung cells. Toxicol Appl Pharmacol. 273(2):281-288. https://doi.org/10.1016/j.taap.2013.06.013

Pfohl-Leszkowicz A. 2008. Chapter 7: Formation, persistence and significance of DNA adduct formation in relation to some pollutants from a broad perspective. In: Fishbein JC, editor. Advances in Molecular Toxicology: Volume 2. Amsterdam, The Netherlands: Elsevier. p. 183-239. https://doi.org/10.1016/s1872-0854(07)02007-3

Poirier MC. 2016. Linking DNA adduct formation and human cancer risk in chemical carcinogenesis. Environ Mol Mutagen. 57(7):499-507. https://doi.org/10.1002/em.22030

Poirier MC, Santella RM, Weston A. 2000. Carcinogen macromolecular adducts and their measurement. Carcinogenesis. 21(3):353-359. https://doi.org/10.1093/carcin/21.3.353

Ponce RA, Gelzleichter T, Haggerty HG, Heidel S, Holdren MS, Lebrec H, Mellon RD, Pallardy M. 2014. Immunomodulation and lymphoma in humans. J Immunotoxicol. 11(1):1-12. https://doi.org/10.3109/1547691X.2013.798388

Puar YR, Shanmugam MK, Fan L, Arfuso F, Sethi G, Tergaonkar V. 2018. Evidence for the involvement of the master transcription factor NF-κB in cancer initiation and progression. Biomedicines. 6(3):82. https://doi.org/10.3390/biomedicines6030082

Qian S, Golubnitschaja O, Zhan X. 2019. Chronic inflammation: Key player and biomarker-set to predict and prevent cancer development and progression based on individualized patient profiles. EPMA J. 10(4):365-381. https://doi.org/10.1007/s13167-019-00194-x

Schmidt R, Plath K. 2012. The roles of the reprogramming factors Oct4, Sox2 and Klf4 in resetting the somatic cell epigenome during induced pluripotent stem cell generation. Genome Biol. 13(10):251. https://doi.org/10.1186/gb-2012-13-10-251

Schwöbel JAH, Koleva YK, Enoch SJ, Bajot F, Hewitt M, Madden JC, Roberts DW, Schultz TW, Cronin MTD. 2011. Measurement and estimation of electrophilic reactivity for predictive toxicology. Chem Rev. 111(4):2562-2596. https://doi.org/10.1021/cr100098n

Seano G, Primo L. 2016. Human arterial ring angiogenesis assay. In: Martin SG, Hewett PW, editors. Angiogenesis Protocols. 3rd ed. New York, NY: Humana Press. p. 191-203. https://doi.org/10.1007/978-1-4939-3628-1_13

Shalini A, Tandon H, Chakraborty T. 2017. Molecular electrophilicity index – A promising descriptor for predicting toxicological property. J Bioequivalence Bioavailab. 9(6):518-527.

Sharma P, Kumar P, Sharma R. 2017. Natural killer cells - Their role in tumour immunosurveillance. J Clin Diagn Res. 11(8):BE01-BE05. https://doi.org/10.7860/JCDR/2017/26748.10469

Smith MT, Guyton KZ, Kleinstreuer N, Borrel A, Cardenas A, Chiu WA, Felsher DW, Gibbons CF, Goodson WH 3rd, Houck KA, et al. 2020. The key characteristics of carcinogens: Relationship to the hallmarks of cancer, relevant biomarkers, and assays to measure them. Cancer Epidemiol Biomarkers Prev. 29(10):1887-1903. https://doi.org/10.1158/1055-9965.EPI-19-1346

Sun Y, Miao N, Sun T. 2019. Detect accessible chromatin using ATAC-sequencing, from principle to applications. Hereditas. 156:29. https://doi.org/10.1186/s41065-019-0105-9

Tas F, Erturk K. 2017. Elevated erythrocyte sedimentation rate is associated with metastatic disease and worse survival in patients with cutaneous malignant melanoma. Mol Clin Oncol. 7(6):1142-1146. https://doi.org/10.3892/mco.2017.1440

Thompson SL, Compton DA. 2011. Chromosomes and cancer cells. Chromosome Res. 19(3):433-444. https://doi.org/10.1007/s10577-010-9179-y

Tokar EJ, Ancrile BB, Cunha GR, Webber MM. 2005. Stem/progenitor and intermediate cell types and the origin of human prostate cancer. Differentiation. 73(9-10):463-473. https://doi.org/10.1111/j.1432-0436.2005.00047.x

Tokar EJ, Diwan BA, Waalkes MP. 2010. Arsenic exposure transforms human epithelial stem/progenitor cells into a cancer stem-like phenotype. Environ Health Perspect. 118(1):108-115. https://doi.org/10.1289/ehp.0901059

Törnqvist M, Fred C, Haglund J, Helleberg H, Paulsson B, Rydberg P. 2002. Protein adducts: Quantitative and qualitative aspects of their formation, analysis and applications. J Chromatogr B Analyt Technol Biomed Life Sci. 778(1-2):279-308. https://doi.org/10.1016/s1570-0232(02)00172-1

Tozer GM, Prise VE, Cunningham VJ. 2016. Quantitative estimation of tissue blood flow rate. In: Martin SG, Hewett PW, editors. Angiogenesis Protocols. 3rd ed. New York, NY: Humana Press. p. 265-281. https://doi.org/10.1007/978-1-4939-3628-1_18

Trabert B, Michels KA, Anderson GL, Brinton LA, Falk RT, Geczik AM, Harris HR, Pan K, Pfeiffer RM, Qi L, et al. 2019. Circulating androgens and postmenopausal ovarian cancer risk in the Women's Health Initiative Observational Study. Int J Cancer. 145(8):2051-2060. https://doi.org/10.1002/ijc.32157

Valadez-Cosmes P, Raftopoulou S, Mihalic ZN, Marsche G, Kargl J. 2022. Myeloperoxidase: Growing importance in cancer pathogenesis and potential drug target. Pharmacol Ther. 236:108052. https://doi.org/10.1016/j.pharmthera.2021.108052

Van Hemelrijck M, Holmberg L, Garmo H, Hammar N, Walldius G, Binda E, Lambe M, Jungner I. 2011. Association between levels of C-reactive protein and leukocytes and cancer: Three repeated measurements in the Swedish AMORIS study. Cancer Epidemiol Biomarkers Prev. 20(3):428-437. https://doi.org/10.1158/1055-9965.EPI-10-1190

Veglia F, Loft S, Matullo G, Peluso M, Munnia A, Perera F, Phillips DH, Tang D, Autrup H, Raaschou-Nielsen O, et al. 2008. DNA adducts and cancer risk in prospective studies: A pooled analysis and a meta-analysis. Carcinogenesis. 29(5):932-936. https://doi.org/10.1093/carcin/bgm286

Vryer R, Saffery R. 2017. What's in a name? Context-dependent significance of 'global' methylation measures in human health and disease. Clin Epigenetics. 9:2. https://doi.org/10.1186/s13148-017-0311-0

Walcher L, Kistenmacher AK, Suo H, Kitte R, Dluczek S, Strauß A, Blaudszun AR, Yevsa T, Fricke S, Kossatz-Boehlert U. 2020. Cancer stem cells-origins and biomarkers: Perspectives for

targeted personalized therapies. Front Immunol. 11:1280. https://doi.org/10.3389/fimmu.2020.01280

Ward TH, Cummings J, Dean E, Greystoke A, Hou JM, Backen A, Ranson M, Dive C. 2008. Biomarkers of apoptosis. Br J Cancer. 99(6):841-846. https://doi.org/10.1038/sj.bjc.6604519

Wentzensen IM, Mirabello L, Pfeiffer RM, Savage SA. 2011. The association of telomere length and cancer: A meta-analysis. Cancer Epidemiol Biomarkers Prev. 20(6):1238-1250. https://doi.org/10.1158/1055-9965.EPI-11-0005

Wong JYY, Bassig BA, Loftfield E, Hu W, Freedman ND, Ji BT, Elliott P, Silverman DT, Chanock SJ, Rothman N, et al. 2020. White blood cell count and risk of incident lung cancer in the UK Biobank. JNCI Cancer Spectr. 4(2):pkz102. https://doi.org/10.1093/jncics/pkz102

Wu HC, Kehm R, Santella RM, Brenner DJ, Terry MB. 2022. DNA repair phenotype and cancer risk: A systematic review and meta-analysis of 55 case-control studies. Sci Rep. 12(1):3405. https://doi.org/10.1038/s41598-022-07256-7

Xu Z, Taylor JA. 2014. Genome-wide age-related DNA methylation changes in blood and other tissues relate to histone modification, expression and cancer. Carcinogenesis. 35(2):356-364. https://doi.org/10.1093/carcin/bgt391

Zhang X, Zhao Q, Zhu W, Liu T, Xie SH, Zhong LX, Cai YY, Li XN, Liang M, Chen W, et al. 2017. The association of telomere length in peripheral blood cells with cancer risk: A systematic review and meta-analysis of prospective studies. Cancer Epidemiol Biomarkers Prev. 26(9):1381-1390. https://doi.org/10.1158/1055-9965.EPI-16-0968

Zhao Z, Shilatifard A. 2019. Epigenetic modifications of histones in cancer. Genome Biol. 20(1):245. https://doi.org/10.1186/s13059-019-1870-5

Zhou B, Liu J, Wang ZM, Xi T. 2012. C-reactive protein, interleukin 6 and lung cancer risk: A meta-analysis. PLoS One. 7(8):e43075. https://doi.org/10.1371/journal.pone.0043075

Zhou B, Shu B, Yang J, Liu J, Xi T, Xing Y. 2014. C-reactive protein, interleukin-6 and the risk of colorectal cancer: A meta-analysis. Cancer Causes Control. 25(10):1397-1405. https://doi.org/10.1007/s10552-014-0445-8

Zhu M, Ma Z, Zhang X, Hang D, Yin R, Feng J, Xu L, Shen H. 2022. C-reactive protein and cancer risk: A pan-cancer study of prospective cohort and Mendelian randomization analysis. BMC Med. 20(1):301. https://doi.org/10.1186/s12916-022-02506-x

Zippel N, Ding Y, Fleming I. 2016. A modified aortic ring assay to assess angiogenic potential in vitro. In: Martin SG, Hewett PW, editors. Angiogenesis Protocols. 3rd ed. New York, NY: Humana Press. p. 205-219. https://doi.org/10.1007/978-1-4939-3628-1_14