Protocol: Methods for Preparing the Draft Report on Carcinogens Monograph on "Cobalt and Certain Cobalt Compounds"

Background information

Cobalt is a naturally occurring element that is present in several different forms. Elemental cobalt is a hard, silvery grey metal which is found in the environment combined with other elements, e.g., with oxygen (cobalt oxide), sulfur (cobalt sulfate) or arsenic (cobalt arsenide). The most common oxidation states of cobalt are +2 and +3; for most simple cobalt compounds, the valence is +2, designated as cobalt(II). Cobalt compounds can be organic or inorganic as well as water-soluble or -insoluble.

Cobalt was initially nominated for review for possible listing in the Report on Carcinogens due to the recent publication of a 2-year inhalation study in rodents (NTP 2013). A concept for the review of cobalt was presented to the NTP Board of Scientific Counselors on April 17, 2014. The first step in the review would be to define the candidate substance, i.e., cobalt compound(s), for evaluation. Based on input from several invited scientific experts at a Cobalt Information Group Meeting held at NIEHS on October 7, 2104, the scope of the evaluation was defined as "cobalt and certain cobalt compounds"; certain refers to cobalt metal, those compounds that release cobalt ion or are cobalt particles. The review will exclude cobalt compounds that have confounding exposures, such as cobalt alloys and radioactive forms of cobalt.

Although a variety of cobalt-containing compounds have been tested in experimental animals, limitations of the available database for most cobalt compounds make it difficult to assess cancer risk for each substance individually. However, mechanistic data described in the literature for several cobalt compounds suggests a common mode of action involving the cobalt ion, which supports evaluating cobalt compounds as a class or group.

To facilitate the evaluation of cobalt as a class, a structured approach will be used that compares the physiochemical properties, such as form (ionic, particulate) and solubility information, toxicokinetics (such as cellular intake and absorption), the key biological effects in the proposed mechanism (including early events such as DNA damage and later events such as hypoxia response) and findings from cancer from experimental animal studies available on the different cobalt compounds. This information will be visualized as a table or matrix.

In summary, 'cobalt and certain cobalt compounds' was selected for review for possible listing in the Report on Carcinogens (RoC) because of documented exposure to cobalt in the United States and an adequate database to evaluate the potential carcinogenicity of certain cobalt compounds (NTP 2013, 1998; IARC 2006). One cobalt compound, cobalt sulfate, has been listed since 2004 in the RoC as reasonably anticipated to be a human carcinogen.

Cancer hazard evaluation

The purpose of the cancer evaluation component is to assess the available scientific evidence, apply the RoC listing criteria[1] to this evidence, and reach a preliminary level of evidence

[1] For the formal Report on Carcinogens listing criteria, see *http://ntp.niehs.nih.gov/go/15209*.

conclusion. Based on this conclusion, the evaluation also provides a preliminary recommendation for a listing status in the RoC.

Briefly, the RoC listing criteria for the two listing categories are as follows:

Known to be a human carcinogen

- Sufficient evidence of carcinogenicity from studies in humans.

Reasonably anticipated to be a human carcinogen:

- Limited evidence of carcinogenicity from studies in humans, or
- Sufficient evidence of carcinogenicity from studies in experimental animals, or
- Substance belongs to a structurally related class of substances that are listed in the RoC, or
- Convincing relevant information that the agent acts through a mechanism indicating it would likely cause cancer in humans.

The goal of the cancer hazard evaluation component of the draft RoC monograph is to conduct an assessment of the scientific literature, utilizing information from the primary literature as well as from authoritative and other reviews. The monograph will evaluate the level of evidence from studies of cancer in humans and in experimental animals, as well as from studies reporting on mechanistic and related data such as metabolism and genotoxicity. The evaluation will be informed by the utilization of an information table (described above) that will provide pertinent available data on cobalt-containing compounds in an easily comparable format.

Conclusions regarding the carcinogenicity in humans or experimental animals, as well as mechanistic and related data, are based on scientific judgment with consideration of all relevant data.

Protocol components

This protocol discusses the methods that will be used to prepare the cancer evaluation component of the draft monograph on cobalt and certain cobalt compounds.

- Part A: Preliminary Outline of the Draft RoC Monograph
- Part B: Literature Search Strategy
- Part C: Methods for Evaluating Human Exposure Information
- Part D: Methods for Evaluation Disposition and Toxicokinetics
- Part E: Methods for Evaluating Human Cancer Studies
- Part F: Methods for Evaluating Cancer Studies in Experimental Animals
- Part G: Methods for Evaluating Mechanistic Studies and Other Relevant Effects

Part A: Preliminary Outline of Draft RoC Monograph

The draft RoC monograph on cobalt focuses on the relationship between exposure to certain cobalt-containing compounds and cancer. The cancer evaluation component of the monograph is organized by topic and includes several sections, as described below. Appendices to the monograph contain additional information, including descriptions of the quality evaluation of the human cancer studies and carcinogenesis studies in experimental animals, as well as genotoxicity data tables.

The major sections in the cancer evaluation component are as follows:

1 Properties and Human Exposure

This section includes details on the chemical properties of cobalt and cobalt compounds and the current available information on human exposure to cobalt in the United States.

2 Disposition and Toxicokinetics

This section discusses the available information on absorption, distribution, metabolism, and excretion of cobalt and cobalt compounds in humans and experimental animals.

3 Human Cancer Studies

This section reviews and assesses the quality and utility of the available studies of cancer in humans and exposure to cobalt exposure to inform the cancer hazard evaluation. It also interprets the study findings, integrates the evidence across studies, and applies the RoC listing criteria to the body of evidence to reach a preliminary level of evidence conclusion for the carcinogenicity of cobalt in humans.

4 Studies in Experimental Animals

This section reviews and assesses the quality and utility of the available studies of cancer in experimental animals exposed to cobalt compounds to inform the cancer hazard evaluation. It also integrates the evidence across studies and applies the RoC listing criteria to reach a preliminary level of evidence conclusion for the carcinogenicity of cobalt in experimental animals.

5 Mechanisms and Other Relevant Effects

This section assesses the strength of mechanistic and related evidence for effects resulting from exposure to cobalt. It includes relevant studies on cytotoxicity and genotoxicity studies conducted *in vivo* or *in vitro*.

6 Overall Cancer Evaluation

The final section of the draft RoC monograph integrates the evidence from the cancer studies in experimental animals and humans with mechanistic and relevant data and applies the RoC listing criteria to reach the NTP preliminary listing recommendation for cobalt and certain cobalt compounds in the RoC. The section will also discuss the evidence for grouping cobalt and certain cobalt compounds as a class using the structural approach outlined in the introduction.

Appendices

The appendices in the Draft RoC Monograph will contain important supplementary information, such as the literature search strategy, study quality tables, and study descriptions and results for some sections.

Part B: Literature Search Strategy

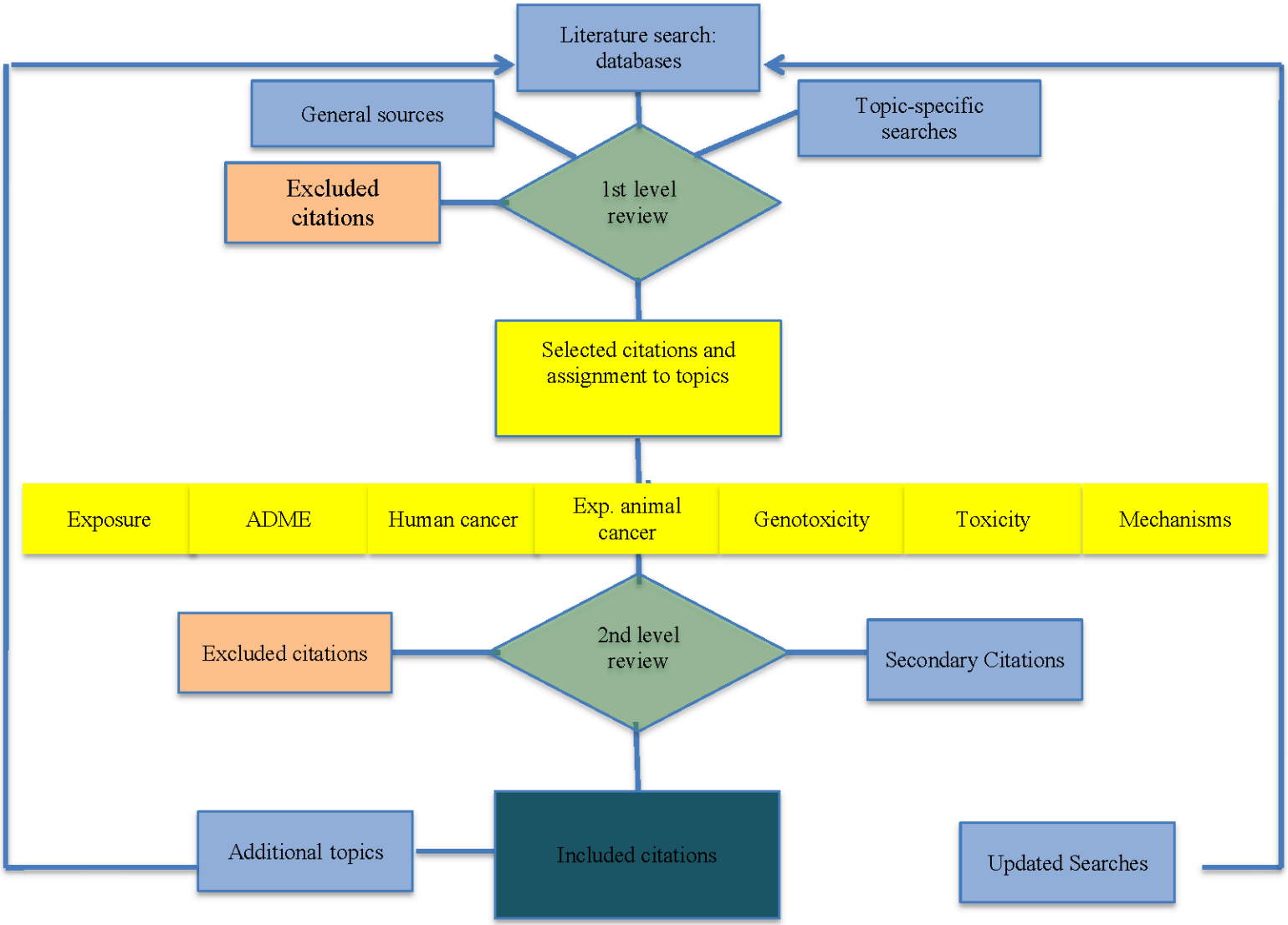
Introduction and objective

The objective of the literature search strategy is to identify the published literature that is relevant for evaluating the potential carcinogenicity of cobalt and certain cobalt compounds. This section discusses the general approach for the literature search; specific details, such as search terms and inclusion/exclusion criteria, are described in relevant sections for the following topics:

- Properties
- Human Exposure (focusing on the U.S. population)
- Disposition (ADME) and Toxicokinetics
- Human Cancer Studies
- Studies of Cancer in Experimental Animals
- Mechanistic Data and Other Relevant Effects
    - Genotoxicity and Related Effects
    - Mechanistic Considerations

The methods for identifying the relevant literature, including the literature search strategy (Section 1) and the review of citations using web-based systematic review software (Section 2), are discussed below.

Figure B-1. Literature Search Strategy



1 Literature search strategy

Relevant literature is identified using multiple approaches (see Figure B-1) including:

- **Database searches** (typically PubMed, Scopus, and Web of Science). This is the major source for identifying relevant paper on the relevant topics (see above) and is described in detailed below.
- **General data search (see Section 3):** Examples include authoritative reviews and exposure-related data searches (see Part B), which cover a broad range of general data sources for information relevant to many candidate substances.
- **Exposure-related data search:** This search covers a broad range of potential sources for exposure-related information and physical-chemical properties (see Table 2).
- **Focused searches for specific scientific issues**
- **Secondary citations:** Citations identified from authoritative reviews or from primary references located by literature search.
- **Quosa library:** Full text searches of library of specific type of studies. These searches are performed to identify studies where the candidate substance may not be identified in the title or abstract. Currently, a library created by the ORoC for occupational case-control studies of cancer using QUOSA scientific literature management software is used to identify human epidemiologic studies of specific occupational exposures and cancer.

Database searches

Database searching involves selecting search terms and databases used in the searches and conducting the searches.

Literature searches of several databases are generally conducted using search terms for cobalt and certain cobalt compounds, combined with search terms for cancer and/or specific topics, such as human exposure, cancer studies in animals, epidemiological studies, mechanistic studies, etc. Titles, abstracts, and key words are searched in these databases. For example, in Part C, literature searches for exposure scenarios or settings are also used when cobalt exposure could occur in a specific occupational setting or through use of a specific consumer product.

A critical step in the process involves consultation with an information specialist to develop cancer- and topic-specific search terms for cobalt. These terms are used to search databases such as PubMed, Scopus, and Web of Science. Literature searches are updated by creating monthly alerts in the appropriate databases.

2 Screening and selecting literature

Citations retrieved from literature searches (and other sources) are uploaded to an EndNote library and any duplicates removed. Next, the EndNote library is uploaded to web-based systematic review software DistillerSR for multi-level screening using inclusion and exclusion criteria. Each level of screening is done by two scientists (ideally at least one NTP staff person).

The citations are first screened (Level 1) using the title and abstract (where available) by two screeners to eliminate papers that do not contain information on the candidate substance or on any of the key topics or questions (exposure, cancer studies in humans and animals, toxicokinetics, genotoxicity, toxicity, and mechanisms of action) covered by a RoC monograph.

The initial screen is designated as "liberal," i.e., it is intended to retrieve a PDF if there is any reasonable possibility that it contains information that could be useful for the review process; a positive response by only one of the reviewers is sufficient to pass a publication on to the next review level. The initial reviewers assign (or tag) the citation to one or more of the topic(s) covered by a RoC (see above).

PDFs are obtained for all citations not excluded at Level 1 for Level 2 screening. Topic-specific experts (e.g., writers for each monograph section) screen the citations using inclusion/exclusion criteria. Similar to Level 1, two scientists screen the literature (the 2nd screening is typically by a topic-specific expert who is the reviewer of the section). In general, these exclusion and inclusion criteria are somewhat similar to the first level (e.g., information on the candidate substance and topic); however, Level 2 can make a more informed judgment about the citations than the Level 1 screeners because they have the full text (PDF). Depending on the topic, more specific inclusion and exclusion criteria may be developed, which will be delineated in the protocol, which can be part of the Level 2 screening or a Level 3 screening. Third level reviews are generally limited to the human cancer and animal tumor studies sections, and should identify all studies included in the monograph sections on those topics. Examples of third level screening may be to exclude case reports (human cancer) or studies with very poor reporting. Citations can also be redistributed to other topics (writers) if that topic(s) was not identified by the reviewer at Level 1.

3 Data sources

The following is a list of general data sources that are searched for information on a specific candidate substance. The list includes authoritative reviews or study reports and web-based resources and/or databases.

Authoritative reviews and reports

- National Toxicology Program (NTP) Technical Reports, nominations for toxicological evaluation documents, and RoC background documents or monographs
- Agency for Toxic Substances and Disease Registry (ATSDR) Toxicological Profiles
- U.S. Environmental Protection Agency Integrated Risk Information System (EPA IRIS)
- International Agency for Research on Cancer (IARC) Monographs
- WHO/UNEP IPCS INCHEM-related documents
- Environmental Health Criteria (EHC) monographs
- Concise International Chemical Assessment Documents (CICADS)

Databases or web resources

- Carcinogenic Potency Database
- CCRIS (Chemical Carcinogenesis Research Information System)
- European Chemicals Agency: http://echa.europa.eu/
- International Uniform Chemical Information Database (IUCLID)

Part C: Methods for Evaluating Human Exposure Information

Objectives

Cobalt exposure to humans in the United States will be assessed in this section of the monograph by reviewing available data on production, use, sources, environmental releases, and measured levels, in the general population as well as occupational exposure.

Key questions

- How are people in the United States exposed to cobalt and cobalt compounds?
- How do we measure exposure?
- What are the non-occupational sources and levels of exposure?
- What are the occupational settings and levels of exposure?
- Has exposure changed over time?
- What federal regulations and guidelines limit exposure to cobalt?
- Are a significant number of people residing in the Untied States exposed to cobalt and cobalt compounds?

Approach

The information present in this section will rely primarily on secondary literature sources. The sources of the information used for exposure assessment will include comprehensive review documents by authoritative, publicly available agencies, e.g., the Agency for Toxic Substances and Disease Registry (ATSDR), US Environmental Protection Agency (US EPA), International Agency for Research on Cancer (IARC), and World Health Organization (WHO). The data will be used to (1) identify how cobalt is currently being used and (2) to determine what processes are emission sources. The document will discuss both occupational and non-occupational sources of exposure, including occurrence of cobalt in the environment, consumer products, food, industrial settings, etc., and levels of cobalt when measured in exposed populations.

This information will include a discussion of current versus past uses, as available.

A preliminary review of the literature indicated several exposure scenarios from sources including the EPA Toxics Release Inventory (TRI), Chemical Data Reporting (CDR) rule, NIOSH Health Hazard Evaluation (HHE) studies, U.S. Geological Survey, and the Cobalt Development Institute. These scenarios have been identified including gold, copper, or nickel ore mining, hazardous waste treatment and disposal, nonferrous metal (except aluminum) smelting and refining, chemicals, petrochemical manufacturing (e.g., catalysts), paint/ink/varnish dryers, alloy production (e.g., hard metals), and medical uses (e.g., replacement joints, dental prostheses). The carcinogenicity of alloys will not be reviewed in the document, but the exposure to cobalt for workers and the general public that results from the production and use of cobalt-containing alloys will be discussed.

Additionally, recent primary literature will be reviewed (from literature search results) to identify any other exposure scenarios in which cobalt is being used in some quantity with data showing exposure, e.g., studies showing cobalt tissue levels for mining and metal processing workers.

Finally, available information (as described above from publicly available sources, review articles, and possibly primary literature) will be integrated and used to draft sections on chemical identification and properties, production, use, sources of exposure, measured environmental levels, general population exposure, and occupational exposure. This information will be used to substantiate that a significant number of people living in the United States are exposed to cobalt and certain cobalt compounds.

Literature search strategy

Part B of this protocol discusses general procedures used to identify and select relevant literature for preparing the RoC monograph on cobalt and certain cobalt compounds. This section discusses the literature search strategy and inclusion and exclusion criteria specific for identifying studies relevant for evaluating human exposure information.

Literature searches of three scientific databases – PubMed, Scopus, and Web of Science – are conducted using a pre-determined range of search terms. For potential cobalt exposure, terms related to cobalt exposure scenarios are included; the specific search terms used for this section are listed in the table below. Terms in the columns are combined with "AND" in the databases.

Literature search strategy for evaluating human exposure information

Substance-specific terms	Exposure search terms[a]	Exposure media search terms[a]
Cobalt NOT (cobalt-60 OR Co 60 or 60Co OR radioactive OR gamma ra* OR radiotherapy OR radiation therapy OR radiation treatment)	environmental pollut* OR environmental expos* OR occupational expos*) OR (expos* OR occur* OR oral OR dermal OR skin OR contact OR absorb* OR eat* OR consum* OR ingest* OR swallow* OR inject* OR administ*)	air OR water OR food OR soil OR blood OR urine

[a]Search terms were developed in consultation with an information specialist.

Part D: Methods for Evaluating Disposition and Toxicokinetics

Objectives

The purpose of this section is to provide background information important for understanding potential mechanisms of carcinogenicity for cobalt. This section provides an overview of absorption, distribution, metabolism, and excretion (ADME) in experimental animals and humans. The specific roles of disposition and toxicokinetics in the carcinogenicity of cobalt compounds will be discussed in the mechanistic section (see Part G).

Key questions

- How is cobalt absorbed, distributed, metabolized, and excreted (ADME)?
- What, if any, are the qualitative and/or quantitative species or sex differences for ADME?
- What is known about the form of cobalt (particulate, ion) from ADME studies, in exposed tissue, particularly in the lung?
- How can toxicokinetic models (if any) inform biological plausibility, interspecies extrapolation, or other mechanistic questions for cobalt?

Approach

The initial approach for preparing this section is to identify comprehensive agency reviews (e.g., NTP, EPA, IARC, ATSDR, WHO, NIOSH) and summarize the relevant data they provide. The overall assessments/conclusions from the reviews will be evaluated to identify primary issues, controversies, and any inconsistencies in the data.

The sub-sections will be written in a review style rather than with study-by-study descriptions and will be supplemented with tables and figures as needed. The goal is a comprehensive review of all pertinent topics and not necessarily a comprehensive review of the literature. The primary focus will be on recent reviews and all studies that provide cogent responses to the key questions.

The literature search (described below) for this section will focus on literature published after the above reviews and will be supplemented with relevant primary literature cited in the agency reviews. Additional searches will be performed for new studies or those reporting on specific issues identified as critical. Secondary sources will generally be cited for concepts that are widely accepted while primary literature is cited when greater depth is required to adequately address a topic or aid in understanding discrepancies in the data.

Literature search strategy

Part B of the protocol discusses general procedures used to identify and select relevant literature for preparing the RoC monograph on cobalt and certain cobalt compounds. This section discusses the literature search strategy and inclusion and exclusion criteria specific for identifying studies relevant for disposition and toxicokinetics. Literature searches of three scientific databases – PubMed, Scopus, and Web of Science – are conducted using a pre-determined range of search terms. The specific search terms used for this section are listed in the table below. Terms in the columns are combined with “AND” in the databases.

Literature search strategy for evaluating disposition and toxicokinetics

Substance-specific terms	ADME & toxicokinetics search terms[b]	Species-related search terms[b]
Cobalt NOT (cobalt-60 OR Co 60 or 60Co OR radioactive OR gamma ra* OR radiotherapy OR radiation therapy OR radiation treatment)	(Absorption OR distribution OR "tissue distribution" OR bioavailab* OR "biological availability" OR metaboli* OR biotransform* OR activat* OR bioactivat* OR detoxif* OR excret* OR clearance OR eliminat* OR kinetic* OR pharmacokinetic* OR toxicokinetic* OR cytochrome P450 OR cytochrome P-450) OR MeSH terms (PubMed only): (("Pharmacokinetics"[Mesh]) OR "Metabolism"[Mesh]) OR "Cytochrome P-450 Enzyme System"[Mesh]	in vivo OR animal* OR mouse OR mice OR rat OR hamster OR guinea pig OR rabbit OR monkey OR dog) OR (person* OR people OR individual* OR subject* OR participant*

[a]ADME.
[b]Search terms were developed in consultation with an information specialist.

Part E: Methods for Evaluating Human Cancer Studies

Introduction

Objectives

The cancer evaluation component of the draft monograph on cobalt and certain cobalt-containing compounds [hereafter referred to as "cobalt"] will evaluate all relevant epidemiologic studies on cobalt exposure and cancer, including those studies previously reviewed for the RoC.

The available studies on exposure to cobalt and human cancer consist primarily of (1) cohort studies of exposed workers in hard-metal factories, electrochemical factories producing cobalt, stainless steel and metallic alloys production factories, nickel refineries, and in porcelain factories; (2) clinical and population-based case-control studies of levels of cobalt in biological tissue; (3) studies of cobalt in lung tissue from deceased copper smelter workers with and without lung cancer; and (4) geographically based studies of environmental exposure to cobalt. Studies of exposure to cobalt found in the literature investigated the effects of occupational exposure to several types of cobalt compounds, although in none of these studies were the specific compounds measured separately. The case-control studies also did not evaluate exposure to specific compounds; the case-control studies were biomonitoring studies that measured levels of cobalt in human tissue of cancer cases and controls – e.g., toenails, lung tissue.

The objective of this section is to reach a preliminary level of evidence conclusion [sufficient, limited, or inadequate] for the carcinogenicity of cobalt from studies in humans by applying the RoC listing criteria to the body of evidence.

RoC listing criteria for evaluating carcinogenicity from studies in humans

Sufficient evidence of carcinogenicity from studies in humans: indicates 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.

Key questions

- Which epidemiologic studies should be included in the review?
- What are the methodological strengths and limitations of these studies?
- What are the potential confounders for cancer risk for the tumor sites of interest in these studies?
- Is there a credible association between exposure to cobalt and cancer?
- If so, can the relationship between cancer endpoints and exposure to cobalt be explained by chance, bias, or confounding?

Approach: Steps in the cancer hazard evaluation process

The steps for conducting the human cancer hazard evaluation are outlined below. The procedures and guidelines for conducting each step are described in Sections 1 through 4 of Part B of this protocol.

1. Selection of the literature included in the cancer hazard evaluation of the human studies (Section 1)
2. Systematic extraction of data from the epidemiologic studies (Section 2)
3. Assessment of the quality of the individual epidemiologic studies (Section 3)
4. Assessment of the level of evidence of carcinogenicity (sufficient, limited, or inadequate) of cobalt from studies in humans (Section 4)
5. Integration of the scientific evidence across human cancer studies

1 Selection of the literature included in the human cancer evaluation

Part B of this protocol discusses general procedures used to identify and select relevant literature for preparing the RoC monograph on cobalt and certain cobalt compounds. This section discusses the literature search strategy and inclusion and exclusion criteria specific for identifying studies relevant for the human cancer evaluation of exposure to cobalt including the primary epidemiologic studies, which form the basis for the cancer evaluation, and supporting literature (e.g., included supporting citations) that may be relevant for interpretation of the studies.

Cobalt has multiple uses and is or has been used widely in industries such as the production of alloys and manufacture of cobalt salts, including those used as pigments for coloring glass, ceramics, and paint, as well as by some new applications of cobalt for producing technologies such as re-chargeable batteries in camcorders and cell phones, and electric car batteries and solar panels, an iterative strategy will be used to explore the use of search strings with the best yield for studies of human exposure and cancer outcomes.

The following approaches for identifying literature are employed:

1. Search terms used to search the three scientific databases (PubMed, Scopus, and Web of Science for potential cobalt exposure and cancer studies in humans are provided in the table below. The search strategy used terms related to cobalt or relevant exposure scenarios combined (using “and”) with search terms for epidemiologic studies and with search terms for the outcome, i.e., cancer.

Literature search strategy for human cancer studies

Substance-specific search terms	Epidemiologic search terms	Cancer endpoint search terms	Exposure scenario search terms
PubMed cobalt NOT (cobalt-60 OR Co 60 or 60Co OR radioactive OR gamma ra* OR radiotherapy OR radiation therapy OR radiation treatment) Scopus or Web of Science cobalt NOT (cobalt-60 OR Co 60 or 60Co OR radioactive OR gamma ra* OR radiotherapy OR radiation therapy OR radiation treatment)	PubMed ("epidemiology"[Subheading] OR "epidemiologic studies"[Mesh] OR "case reports"[publication type] OR "epidemiologic factors"[mh] OR "epidemiologic methods"[mh]) OR (epidemiologic* OR workers OR case-control OR cohort OR case-report OR case-series) Scopus or Web of Science: (epidemiologic* OR workers OR case-control OR cohort OR case-report OR case-series)	PubMed, Scopus or Web of Science [cancer OR mortality OR follow-up OR incidence]	Mining for gold, copper, or nickel Nonferrous meal smelting or refining (except aluminum) Hazardous waste treatment and disposal Petroleum refining and hydrotreating or desulfirization Terephthalates production Pigments and dyes Electroplating Superalloys

2. Full-text searches of a database of case-control studies assembled by RoC for the purpose of consolidating information on occupational exposures conducted over many decades and created using QUOSA scientific literature management software. Terms for cancer and cobalt were used to search this database.
3. Citation searches from articles, reports, and reviews identified above to identify any additional primary studies or other relevant literature.

Citations are screened for primary epidemiologic studies and supporting literature using the procedures outlined in Part B of this protocol. Studies are initially included in the study if they meet the following inclusion criteria for Levels 1 (titles and abstracts) and 2 (full text):

- Primary studies (including analytical epidemiologic studies, descriptive studies), pooled analyses, and meta-analyses on exposure to cobalt and human cancer
- Studies providing supporting information for topics that are relevant to the evaluation of the human epidemiologic evidence including but not limited to reviews, letters to editors, exposure-assessment or validation studies (for use in epidemiologic studies), relevant epidemiologic studies of biomarkers, and information on co-exposures or potential confounders.

Primary studies were included in the review if they met the following inclusion criteria (Level 3)

- The publication is a peer-reviewed, primary research study on potential exposure to cobalt and human cancer.

- The study reports a risk estimate (or information to calculate a risk estimate) for these cancers; descriptive studies will not be included in the evaluation.
- The publication is a peer-reviewed, primary research study that provides information specific for potential exposure to cobalt at the individual level.

2 Systematic extraction of data from the epidemiologic studies

Information on the methods and findings is selected from the individual studies and entered into *Table Builder*, a proprietary database created for the purpose of entering data in a systematic manner using standardized instructions and questions. The database contains "fields" that are specific for the different types of extracted information (e.g., study population characteristics, exposure and disease assessment, analytical methods, and results). Questions and guidelines are available to describe the specific type of information that should be summarized or entered into each field; and the fields are used to populate tables in the monograph.

The latest published follow-up or update for each of the cohort, nested case-control, and case-control studies, is extracted for each cancer endpoint included in the study. Additional relevant information (such as exposure data or re-analyses) from earlier and related publications on the same or overlapping study population is also included if these publications provide unique or additional data to inform the cancer evaluation of the primary study under review.

One reviewer initially extracts the data from the individual studies into *Table Builder*. Quality assurance and quality control of the data extraction and database entry are accomplished by a second independent reviewer who (1) double-checks each data entry; and (2) flags discrepant entries. Resolution is achieved through mutual discussion between the two reviewers using the original data source as reference.

3 Assessment of the utility of the individual epidemiologic studies

3.1 Overview of approach

Biases in observational studies are often classified into three major categories: (1) selection bias, (2) information bias, and (3) confounding (Rothman 2008). In addition, studies should have adequate reporting methods and apply appropriate analytical methods for calculating effect estimates. Finally, studies with greater sensitivity to detect an effect (e.g., adequate statistical power, levels, duration, range, window of exposure, length of follow-up) are also considered to be more informative for the evaluation, although studies with lesser sensitivity may not suffer from bias *per se.*

Each primary study is systematically evaluated by two independent reviewers for its ability to inform the cancer hazard evaluation using five domains related to study quality and one domain related to study sensitivity. The evaluation for the potential for bias for each domain is captured by a core question. Core questions are largely similar (with some exceptions) to those being developed for the U.S. EPA Integrated Risk Informational System (IRIS) Toxicological Reviews. A series of signaling and follow-up questions are used to address specific issues related to the core question similar to those used for the review of other chemicals (e.g., trichloroethylene (TCE), ortho-toluidine (http://ntp.niehs.nih.gov/NTP/roc/thirteenth/Protocols/ortho-ToluidineProtocol_508.pdf), and pentachlorophenol and byproducts of its synthesis

(http://ntp.niehs.nih.gov/NTP/roc/thirteenth/Protocols/PCPHumanStudies20130815_508.pdf). These questions are concerns that epidemiologists usually think about for each of the types of bias and are not meant to be a checklist (See Figure E-1). Some of these concerns (such as healthy worker effect) overlap or could be considered in more than one domain but will only be evaluated in one domain in the cancer hazard evaluation. The overall evaluation of the study utility is based on integration of the assessment for each domain level judgment. The response for answering the questions of whether there is a potential bias is based on a comparison of the study element with that of the "ideal" study for a specific endpoint and exposure (see Section 4.2). 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 (over or underestimate of the effect estimate, or unknown) and the potential magnitude of the distortion of the bias on the effect estimate.

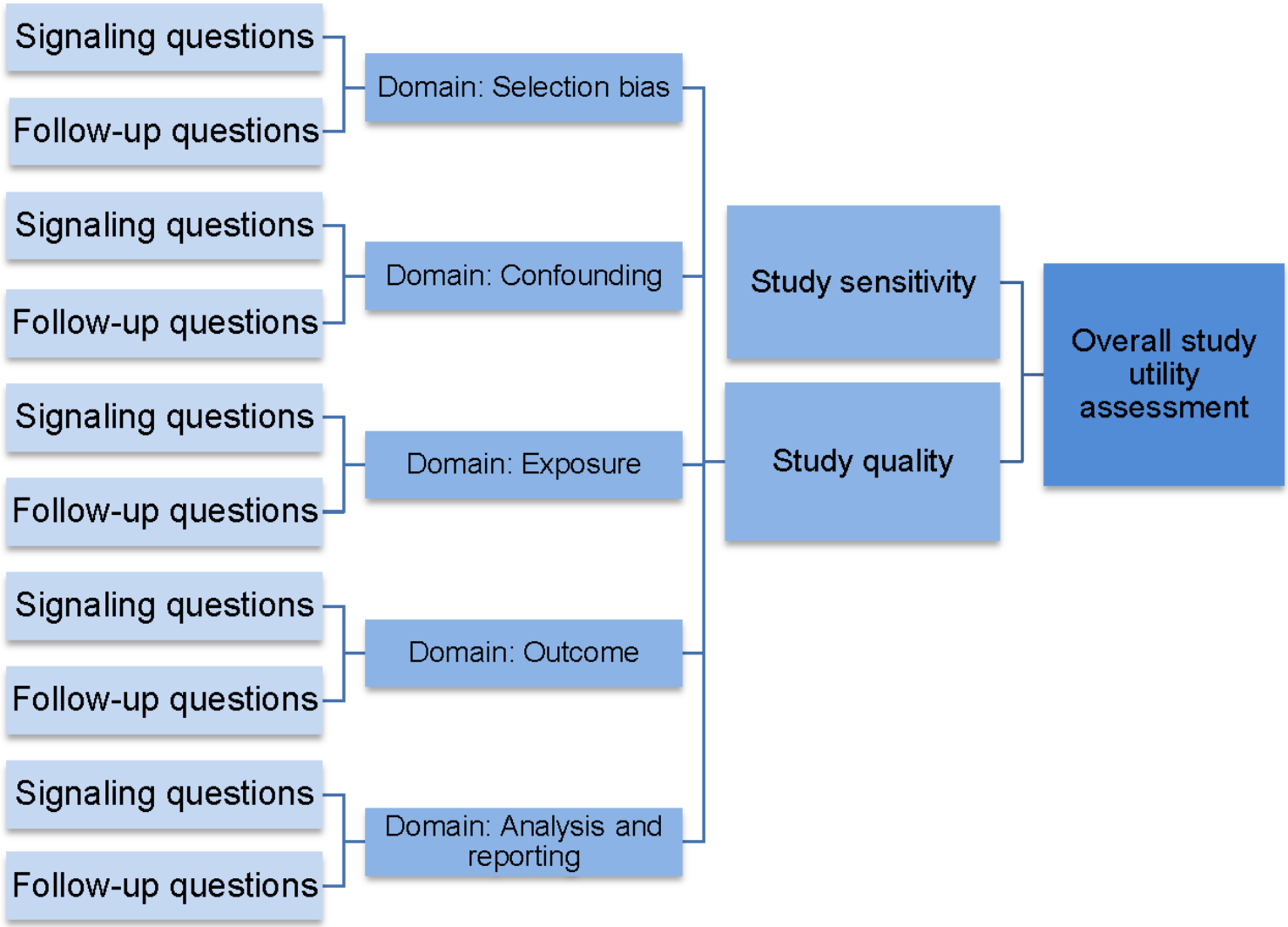
This step is done prior to interpreting the individual study's findings and assessing the level of evidence across studies. Differences are resolved by mutual discussion with reference to the original data source. Study authors may be contacted if there is inadequate information to evaluate a signaling question. The approach of using signaling and follow-up questions for evaluating the quality of different type of biases and reaching conclusion of the quality of the studies, as well as the study domains for quality are somewhat similar to those used by other systematic review methodologies (e.g., Acrobat-NRSI). Terms used in the evaluation are defined as follows, and the evaluation of the specific domains follows Figure E-1. The overall evaluation of the utility of the study (including the judgment terms) is discussed in Section 4.4.

Domain-level judgment terms: Responses for core questions

The signaling and follow-up questions are used to provide transparency for the domain level judgment and the responses to the questions are captured in the rationale for the response to the core question (e.g., domain level judgment) rather than answered for each signaling or follow-up question. When adequate information is available, a judgment is made for the direction and distortion of each bias.

- Low/minimal concern: Information from the study design and methodology indicate that the study is close to ideal and that the potential for bias is unlikely, recognizing general limitations of observational studies.
- Some concern: The study design or methodology is less than ideal, and there are some concerns about potential bias.
- Major concern: The study design or methodology suggests that the potential for a specific type of bias is likely.
- Critical concern or inadequate: Distortion of estimates due to bias makes the study unreliable for hazard identification.
- Inadequate information.

Figure E-1. Schematic of approach for systematic review of study utility



3.2 Evaluation of biases: Signaling and follow-up questions and guidance

The questions, domain level ratings, and guidelines for evaluating the different components of study utility are described in Sections 3.3 to 3.5.

Previous RoC and IARC monographs included epidemiologic studies that potentially could contribute data on the evaluation of cobalt. These primary epidemiologic studies were used to identify these specific cancer endpoints (lung cancer and other aerodigestive cancers) as cancer sites of interest and thus helped inform the development of the guidelines for evaluation the potential for bias or confounding.

3.2.1 Selection and attrition bias

Selection bias is a concern for any observational study, as it arises when study participants are not selected from the same underlying source population and selection is related to both exposure and disease. This occurs when the relationship between the exposure and disease is different for those who participate and for those who should have been eligible for the study, including those who did not participate (e.g., in case-control studies) or did not remain in the cohort (e.g., in cohort studies) (Rothman 2008). Selection bias can be a concern in any type of epidemiologic study, but is most often a factor in case-control studies.

Core Question:

Is there a concern that selection into the study or out of the study was related to both exposure and to outcome?

Signaling and follow-up questions

Case-control studies

- Is there a concern that cases and controls may not have been selected from the same underlying population during a similar time period?
- Are there concerns that eligibility criteria (inclusion/exclusion), recruitment strategies and participation of cases and controls may be related to exposure and disease status?
- If there is a concern about the potential for bias, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?

Cohort studies

- Is there a concern that follow-up is incomplete?
    - If so, is there a concern that completeness of follow-up is related to both exposure and disease?
- Is there a concern for Healthy Worker Hire Effect (HWE) or that non-exposed subjects may not have been selected from the same underlying population during a similar time period?
    - If there is a concern for HWE, were appropriate analyses (e.g., use of internal analyses) used to control it or provide information about it?
- Are there concerns for a Healthy Worker Survival Effect (HWSE), prevalent hires, or left truncation?
    - If there is a concern about HWSE? Were appropriate analyses performed to address the potential bias?

- If there is a concern about the potential for selection or attribution bias, what is the predicted direction or distortion of the estimate (if there is enough information)?

Domain level ratings

Low/minimal concerns: (***) rating

Case and controls selected from the same population using similar methods and criteria. No evidence that selection of the subjects is related to both exposure and disease.

Cohort is clearly defined (e.g., includes the relevant exposed, non-exposed or referent group for a specific time period/location) with no evidence that follow-up differ between exposed and non-exposed. No evidence of HWE or appropriate measures were used to address the potential bias.

Critical concerns: Inadequate rating

Strong evidence that selection or attrition of subjects is clearly related to both exposure and disease

Guidelines for domain level judgments

In nested and population-based case-control studies, controls and cases should be selected from the same underlying population (or cohort) and should be representative of the population (or cohort) from which they were selected. Controls should be free of any diseases related to cobalt exposure; the use of controls with diseases related to cobalt exposure would bias the effect estimate toward the null. Ideally, participation rates should be high and should be similar for cases and controls, although it is recognized that participation rates in population-based case-control studies are sometimes lower than those in a hospital-based or nested case-control study.

Selection bias is usually less of a concern in cohort studies with complete recruitment and follow-up, as the cohort itself acts as the source population (Checkoway *et al.* 2007). In cohort studies, the exposed and unexposed groups should ideally be similar in all respects except for exposure to cobalt compounds. Occupational cohorts should consist of all potentially exposed employees within a given plant or exposure setting (employed over a specified period of use of cobalt compounds) compared with similar unexposed employees from within the same plant or setting (i.e., internal controls) to minimize the healthy worker effect and other differences between exposed and unexposed groups. When external referents are used in, e.g., SMR or SIR studies, local (or regional) mortality or incidence rates are generally, but not always, preferable to national rates.

In occupational studies, there is a concern for the healthy worker effect, the healthy worker survival effect, and left truncation, all of which can be considered as types of selection bias or types of confounding. The HWE can be partially controlled for by comparing the exposed workers with unexposed workers in a nested case-control study instead of with the general population. The HWSE arises from the fact that long-term, healthier workers (prevalent hires) are over-represented in the cohort and thus may underestimate exposure-response relationships. Overlapping with HWSE is “Left Truncation.” Left truncation can occur when workers hired prior to the start of the study and who are still working at baseline are followed over time. Some methods to control for HWSE, prevalent hires and left truncation include controlling for (time-related) employment status, “g-estimation”, inverse probability of treatment weighted (IPTW), and use of censoring weights (Rothman 2008; Applebaum 2011; Naimi 2000; Chevrier 2014).

In occupational cohort studies, including the cohort studies of cobalt, the concern for attrition bias occurs when loss to follow-up is related to both exposure and disease status. In cancer studies, there is often overlap between evaluation of completeness of follow-up and outcome since similar methods are often used for both (e.g., cancer registries, death certificates). The evaluation of attrition bias should consider methods used to obtain vital status and the number of cases/deaths, but not the methods of diagnosis. Ascertainment of vital status relies upon either death certificate data, medical records, and/or cancer registry data, with medical records being the least preferred. In the United States and other industrialized countries death certificate data (e.g., SSN or National Death Index file [U.S.]) are considered to be mostly complete. Completeness of cancer registry incidence data can vary by collection methods, region, and calendar period. One of the case-control studies for cobalt relied on the Washington State Cancer Surveillance System, which has been in existence since 1974 and was one of the earliest registries in the SEER program (NIH SEER); although from 1974 to 1992 reporting of cancer took place through voluntary agreements with hospitals, radiation and surgery centers, and pathology laboratories. Starting in 1992, cancer diagnoses were required by state law to be reported to the Washington State Department of Health.

Systematic biases may be introduced if the length and completeness of follow-up differ between exposed and unexposed groups and are related to the outcome of interest. Ideally, the total loss to follow-up should be less than approximately 5% over the duration of the study observation period. Overall, studies should have more than 80% to 90% total follow-up, although incidence studies may have greater loss to follow-up than mortality studies. Statistical power may also be reduced in studies having a high percentage of all subjects (regardless of exposure and disease status) lost to follow-up.

3.2.2 Information bias: Potential exposure misclassification

One of the most important aspects of a study is the ability to characterize exposure at the individual level.

Core Question

Is there a concern that the exposure assessment methods do not distinguish between exposed and non-exposed subjects or exposure categories?

Signaling and follow-up questions

- Is there a concern that the subjects are misclassified with respect to ever exposure?
    - If so, did any misclassification vary by exposure category?
- Is there a concern that the exposure classification does not capture the variability of exposure?
- Is there a concern that the exposure assessment did not capture the relevant window or metric of exposure?
- Is there a concern that knowledge (e.g., observation or recall bias) or presence of the outcome (e.g., reverse causality) for exposure may potentially bias the findings?
- Are there concerns that missing exposure data (including methods used to input data) may result in exposure misclassification?

- If there is a concern about exposure misclassification, is it differential or non-differential, and if so, what is the predicted direction of the bias (if there is adequate information to evaluate)?

Domain level ratings

Low/minimal concerns: (***) rating

Exposure assessment methods have good sensitivity and specificity leading to reliable classification (or discrimination) with respect to both ever exposure, exposure level, timing, or other relevant metrics (see guidelines for characteristics of ideal exposure assessments). Alternatively, exposure assessment methods may be less than ideal, but detailed information on exposure assessment allows for discrimination between exposed and non-exposed and exposure category.

Critical concerns: Inadequate rating

Exposure is not at the individual level or not likely to reflect individual exposure. Study has poor sensitivity and specificity resulting in poor discrimination between exposed and non-exposed and exposure category.

Guidelines for reaching domain level ratings

The ideal study would include quantitative estimates of exposure to various cobalt compounds alone and relevant co-exposures for each individual that are based on monitoring data (e.g., ambient or personal air levels or biological monitoring). In occupational studies, exposure assessment is often based on a job- or job-task exposure matrix (JEM or JTEM) or expert assessments that link the subject's occupational history (e.g., job or department titles, task descriptions, duration of employment, calendar years worked) with data on relevant exposure monitoring or on production methods or applications that are calendar-year specific. Exposure estimates using multiple metrics (such as cumulative, peak, average intensity) improve the quality of the assessment.

The sensitivity and specificity of the method for measuring the biomarker in different biological media (e.g., urine, plasma, fat, soft tissue) is important to know, together with the limit of detection. Some markers are non-selective and can also be markers of other compounds. Knowledge of the time period over which the biomarker integrates dose is also important for determining the exposure period over which the biomarker is a valid indicator. Because the induction period for most cancers is many years or decades, it is desirable that a biochemical marker reflect long-term exposure that was prior to diagnosis.

Available studies of unexposed humans have reported cobalt blood levels of 0.05 to 0.19 µg/dL and urinary cobalt levels of 0.04 to 2 µg/dL (Alexandersson 1988; Ichikawa *et al.* 1985). Occupational exposure to 0.1 mg/m^3 cobalt resulted in blood levels of cobalt ranging (95% CI) from 0.57 to 0.79 µg/dL, compared with 0.19 µg/dL in unexposed workers, and urinary levels from 59 to 78 µg/L, compared with 2 µg/L in unexposed workers (Ichikawa *et al.* 1985). Correlations between recent exposure and cobalt levels in the blood or urine are more consistent for soluble cobalt compounds (metal, salts, and hard metals), while blood and/or urinary cobalt levels are less reflective of recent exposure for less soluble compounds (cobalt oxides) (Lison *et al.* 1994).

In contrast to trace element levels in serum, plasma, or whole blood which vary from day to day, toenail clippings from all ten toes are likely to reflect exposure integrated over the previous 3 to 12 months and possibly longer (Longnecker *et al.* 1993), which is likely to be long after any induction event, even if the samples were taken at the exact time of diagnosis. Sabbioni *et al.* (1994) reported in a study of workers with asthma and/or lung fibrosis exposed to hard metal dusts that concentrations of cobalt in pubic hair and toenails could be valuable indicators of exposure, although these metrics are limited in that they cannot establish quantitative levels of cobalt accumulation. In this study, levels of cobalt in blood and urine, and blood and pubic hair were significantly correlated; however, levels of toenail cobalt were not correlated with pubic hair or with blood. High reproducibility over time of a biomarker known to reflect exposure suggests that the marker has sufficiently low within-person variability such that a single measurement will usefully reflect long-term exposure. In a paper investigating trace element levels from toenails, Garland *et al.* (1993) found that cobalt had intermediate to high within-person variability. Finally, Gerhardsson *et al.* (1985) reported that cobalt in lung tissue does not appear to decline with duration of exposure, indicating that it has a long half-life in the body. These findings, the relatively short integrated exposure time reflected in toenail measurements, and the lack of any clear data on cobalt exposures suggest that these studies may be less than useful for establishing a causal relationship between cobalt and cancer.

Misclassification of exposure in cohort studies is almost always non-differential and usually results in a bias towards the null, i.e., an underestimate of the true risk estimate. In general, exposure is better characterized in most occupational cohort studies than in geographical cohort studies or population-based case-control studies. Potential misclassification of exposure can be reduced in studies conducted in geographical areas with industries associated with cobalt and that assess exposure using a JEM or expert assessment of information on tasks and jobs collected via detailed occupational questionnaires and interviews. In-person interviews are preferred over mailed or phone interviews, and information obtained from the subject is preferred over information from proxies. Ideally, exposure assessment investigators and interviewers should be blinded to the status of cases and controls. Of these, the blinding of the investigators conducting exposure assessments is considered the most important; blinding in-person interviewers may not be feasible, depending on, e.g., the health of the subject with cancer.

Recall bias in case-control studies in which occupational exposure is assigned based on job titles is less likely to be a concern than in studies using self-assessment of chemical-specific exposures (e.g., use of questionnaires with exposure check-lists). With the possible exception of recall bias from self-reported exposure, misclassification of exposure is usually non-differential.

3.2.3 Informational bias: Potential outcome misclassification

Studies are evaluated for their adequacy in measuring disease outcomes, including missing data and the probability of misclassification of disease. Similar to exposure misclassification, the effects of non-differential misclassification of a binary endpoint will produce bias toward the null, provided that the misclassification is independent of other errors. Also, when the risk of disease is low in both exposed and non-exposed (< 10%), the odds ratio will remain biased towards the null although the bias will be small (Rothman *et al.* 2008). However, when both exposure and disease are non-differentially misclassified but the classification errors are dependent, it is possible to obtain substantial bias away from the null (Chavance *et al.* 1992; Kristensen 1992).

Core question

Is there a concern that the outcome measure does not reliably distinguish between the presence or absence (or degree of severity) of the outcome?

Signaling questions

- Is there a concern that diagnosis of disease is incomplete?
- If mortality data is used, does it adequately reflect incidence?
- Is there a concern that the disease was not accurately diagnosed?
    - Does misclassification vary across study groups?
- Is there a concern that the non-diseased group might have the disease?
- Is there concern about observation bias?
- Is there a concern that any outcome misclassification is either differential or non-differential?
    - What is the predicted direction or distortion of the bias (if known) on the effect estimate?

Domain level ratings

Low/minimal concerns: (***) rating

Outcome methods clearly distinguish diseased and non-diseased subjects and follow-up and diagnoses are conducted independent of exposure status.

Critical concerns: Inadequate rating

Strong evidence that the methods do not discriminate between disease and non-disease subjects and/or follow-up and diagnoses are likely related to exposure status.

Guidelines for reaching domain level judgments

Diagnosis of the type of cancer generally relies upon either death certificate data, medical records, and/or cancer registry data and thus the completeness of overlaps with ascertainment of vital status and loss to follow-up (see Selection bias). Incidence data from population-based cancer registry sources or hospital pathology data are generally more detailed and accurate from medical records and cancer registry data than death certificates. Ideally, cases of cancer should be histologically confirmed and/ or undergo independent pathology review (or a subset of the cases) by the study investigator, which is more likely to be performed in case-control studies.

As mentioned, the two major cancer sites of interest are lung and upper aerodigestive cancers. While cancer incidence data may be considerably more informative than mortality data (depending on ascertainment, reporting, and diagnostic accuracy) for cancers with relatively longer survival times and good treatment prognoses, lung cancer mortality rates or esophageal mortality rates may be comparable to their respective incidence rates given their lower five-year survival rate (16.8%, 17.8%, respectively), and typically late diagnosis (85%) (SEER Cancer Statistics Review, 1975-2011). In contrast, laryngeal and oral cancers have relatively high five-year survival rates (60% and 62.7%, respectively) suggesting incidence data is more informative since mortality analysis would miss cases that do not result in death.[2]

[2] Age-adjusted annual incidence or mortality rates (per 100,000 males or females) in the United States from 2007-2011 (U.S. SEER Statistics) for the cancer sites of interest are as follows: (1) Lung - 72.2 (male) and 51.1 (female) for incidence

Length of follow-up is also critical in identifying cases or deaths from long latency cancer endpoints. For longer latency but lower survival cancers such as lung cancer, both incidence and mortality data may be of similar utility, assuming an adequate length of follow-up. (The evaluation of length of follow-up is usually considered in the study sensitivity domain.)

Non-differential (not related to exposure status) misclassification of cancer would most likely result (if not related to exposure status) in loss of statistical power and an underestimation of the risk estimate.

3.2.4 Potential confounding bias

A key question in the evaluation of the level of evidence from observational studies is whether any association between exposure and the potential carcinogen can be explained by confounding. The evaluation of whether actual confounding occurs is a multi-step process and involves consideration of the findings (see Section 4 for a description of the steps). This section only discusses signaling and follow-up questions related to the first step in the process, which is the evaluation of the study methods to consider potential confounding and whether there is other information provided by the authors that might inform the evaluation.

The control of confounding in a study can be done in the design phase, through matching or restriction, for example; or during the analysis phase through a variety of statistical techniques such as stratification and multivariable methods. The adequate control of any confounding factor is predicated on that factor being accurately measured and quantified in the study, as confounders which are themselves misclassified may only partially control for the bias, and may themselves introduce bias (Rothman *et al.* 2008).

The healthy worker effect is both a special type of confounding and a type of selection bias and is evaluated and described under Selection Bias (Section 3.3).

Core question

Is there a concern that either the methods are inadequate or there is inadequate information to evaluate potential confounding?

Signaling and follow-up questions

- Is there concern that co-exposures, or other non-occupational risk factors are not adequately measured in the study?
    - If no data are provided about potential confounders, are surrogate data on potential confounders available?
- Is there concern that the design or analysis may not adequately address important confounding through matching, stratification, multivariable analysis, or other approaches?
- Is there additional information available to evaluate potential confounding?

and 48.4 (male) and 61.1 (female) for mortality, 2) esophageal- 7.7 (male) and 1.8 (female) for incidence and 7.5 (male) and 1.6 (female) for mortality, (3) laryngeal - 5.9 (male) and 1.2 (female) for incidence and 2.0 (male) and 0.4 (female) for mortality, and (4) oral cancer - 16.5 (male) and 6.2 (female) for incidence and 3.8 (male) and 1.4 (female) for mortality.

Domain level ratings

Low/minimal concerns: (***) rating

Studies measured all relevant potential confounders and/or used appropriate statistical analyses or designs such as in the selection of the study participants to address them. Final statistical models should, however, only include "actual" confounders and not variables that have minimal effect on the risk estimate.

Critical concerns: Inadequate rating

Strong evidence that the effects of the exposure cannot be distinguished from potential confounders.

Guidelines for reaching domain level ratings

For occupational **co-exposures**, studies should ideally provide quantitative exposure data for each potential confounder as part of a job-exposure matrix or expert assessment for each worker, although this is rare. However, some studies provide quantitative or qualitative data on co-exposures for subsets of workers, which can be used to evaluate potential confounding. Knowledge of cobalt manufacturing processes or patterns of use in ceramics workers, electrochemical cobalt production workers, and hard-metal production workers may also be helpful in understanding the potential co-exposures likely to be present in the workplace. PAH, nickel, tungsten carbide, silica, asbestos, and chromium compounds are the most likely co-exposures of concern.

Ideally, quantitative information on other **non-occupational exposures or lifestyle factors** should be assessed, and preferably by in-person interview by interviewers blinded to the status of the respondent in cancer incidence studies, rather than via proxy respondents, work records, or other indirect methods. Residual confounding is more likely when only limited qualitative information on a given risk factor (dichotomous yes/no) is available. Smoking is an important risk factor for lung and esophageal cancers. While few or little data are available on smoking or alcohol (a risk factor for oral and esophageal cancers) in the available historical cohort studies of cobalt, some will provide information on these risk factors from subsamples that can be helpful. In addition, when available, data on nonmalignant respiratory diseases (e.g., hard-metal pneumoconiosis or pulmonary fibrosis), or smoking-related diseases (e.g., chronic bronchitis, emphysema), or alcohol-related diseases (e.g., esophageal cancer) may provide indirect information about risk factors for specific cancer endpoints of concern.

3.2.5 Potential bias from selective reporting and analysis

Studies are evaluated for biases that result during the reporting and analysis phase of the studies, and domain level judgments are made separately for each of these elements. When selective or partial *reporting* is based on the direction, magnitude, or statistical significance of exposure-effect estimates, then reporting bias can occur. Selective outcome *reporting* occurs when the effect estimate for an outcome measurement was selected from among analyses of multiple outcome measurement instruments and reflected the most favorable result or subcategories (Sterne *et al.* 2014). *Selective analysis* reporting occurs when results are selected from exposure effects estimated in multiple ways, but reported for only one (or a subset) of the outcomes.

Studies with adequate data should evaluate exposure-response, and latency or conduct subgroup analyses (especially those exposed to higher levels or longer durations). Analysis bias may also

arise due to the inappropriate data assumptions or use of models or statistical methods used to evaluate the overall findings, exposure-response relationships, latency, and confounding. Controlling for variables in the pathway between exposure and response or for variables unrelated to both the exposure and outcome could create bias (Sterne et al. 2014).

Core question – Selective reporting

Is there a concern that the study does not provide results for all relevant measures and participants biasing its interpretation?

Signaling questions – Selective reporting

- Is there a concern that, while multiple types of data were collected, only a subset were reported or were only reported for some subgroups?
- Is there a concern that while multiple analyses may have been performed of the exposure-disease relationship, only one or a subset of analyses may have been reported?

Core question – Analysis bias

Is there a concern that the data assumptions and analysis are not adequate or the study does not conduct relevant analysis on available data?

Signaling questions – Analysis bias

- Is there a concern that the data assumptions used in the statistics are adequate (e.g., log transformed, etc.)?
- If the study has adequate data, does it evaluate exposure-response and latency or conduct subgroup analyses (especially for those exposed to higher levels or longer durations)?
- Is there concern about the adequacy of the models used to evaluate the overall findings, exposure-response relationships, latency, and confounding? Is there evidence of over-controlling for confounding?

Domain level ratings

Low/minimal concerns: (***) rating

Selective reporting: No evidence that reporting of the data or analyses were limited to only a subset of the data that was collected.

Analysis: Study used relevant data, appropriate assumptions and methods of analysis.

Critical concerns: Inadequate rating

Strong evidence that selective reporting of data or analyses compromised the interpretation of the study.

Strong evidence that the study analytical methods were so limited that the findings were uninterpretable or distorted.

Guidelines for domain level ratings

Direct evidence of selective reporting requires congruence between outcome measurements and analyses specified in a statistical analysis plan, which is typically unavailable. Indirect evidence that selective reporting may not be a problem can be gleaned from the consistency between clearly defined outcome measurements and analyses that are internally consistent across the Methods and Results in the paper, and externally consistent with other papers reporting the

study. Inconsistency (internally or externally) in outcome measurements, analyses or analysis cohorts (e.g., a large difference between the size of cohort of eligible participants and the size of the cohort analyzed) should indicate a risk of selective reporting, especially if all reported results are statistically significant, and there is no additional information regarding the varying proportions of the original cohort (Sterne *et al.* 2014).

Authors may be contacted to determine the findings for other analyses if there appears to be selected reporting or analysis, or simply missing data that would be helpful in interpreting the findings. If other data or analyses are available, then they would be reported.

Evidence for selective analysis can come from the selection of analyses of a subgroup from a larger cohort; for example, the cohort for analysis may have been selected from a larger cohort for which data were available on the basis of a more interesting finding. Subgroups defined in unusual ways (e.g., an unusual classification of subgroups by dose or dose frequency) may provide evidence of such selective analysis.

Analysis bias may arise due to the inappropriate use of models or statistical methods used to evaluate the overall findings, exposure-response relationships, latency, and confounding. Controlling for variables in the pathway between exposure and response or for variables unrelated to both exposure and outcome could create bias. The evaluation of exposure-response relationships would ideally be based on a range of at least two or more exposure categories (intensity or duration), together with sufficient numbers of exposed persons or controls in each exposure subgroup. Analysis of exposure-response relationships (using either linear models or exposure categories, or other methods to evaluate the shape of the exposure-response curve) and calculation of trends using quantitative exposure assessments adds more information than analysis by simple binary exposure categories, as does analysis of tumor site by average, cumulative, peak, or duration of exposure, time since first exposure, calendar periods of exposure, and exposure lags. Evaluating the shape of the exposure-response relationships curve is considered to be a positive attribute of a study.

3.2.6 Evaluation of study sensitivity

Factors that increase the ability of a study to detect an effect (if present) include moderate to large numbers of exposed and non-exposed participants or cases and controls, evidence of substantial exposure (e.g., level, duration, frequency, probability) during an appropriate window, adequate range in exposure levels or duration that allows for evaluation of exposure-response relationship and adequate length of follow-up in cohort studies (or follow-back in case-control studies). When both exposure and disease are rare, power is largely determined by the number of exposed cases (Thomas 2009 as cited by NRC 2014). The assessment of study sensitivity requires an integration of the different factors: e.g., a study evaluating effects from low levels of exposure most likely will need larger numbers of exposed subjects than studies of higher exposed subjects. Some of these factors may overlap with exposure assessment, outcome assessment, and analysis; however, it is possible to have a well-designed study that may not be informative for cancer evaluation because of low sensitivity (e.g., cohort studies evaluating rare cancers). Poor study sensitivity may make it harder to detect an effect (if present) and may also help explain heterogeneity across studies. A systematic review of these factors will help increase transparency of the interpretation of the findings.

Core question

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

Signaling questions

- Is the number of exposed cases adequate to detect an effect in the exposed population and/or subgroups of the exposed population?
- Are the levels, duration, window, or range of exposure of the population at risk in cohort and case-control studies sufficient or adequate to detect an effect of exposure?
- Is the follow-up adequate to allow for a cancer induction period of 20 years or greater?
- Is there concern that any analyses of exposure lagging might not have been adequate for detecting cancers with longer latency?

Domain level ratings

High utility (***) rating

Study has adequate exposed subjects, with substantial (level, duration, or range) of exposure with adequate duration of follow-up for latency.

Inadequate utility

Moderate or small study with few exposed subjects and/or exposure is unlikely to be substantial (based on other knowledge) to detect an effect

Guidelines for domain level ratings

Detection of cancer endpoints requires a relatively large cohort and/or higher exposure prevalence for an adequate ability to detect an effect although statistical power is greater for more common cancers, such as lung cancer, compared to esophageal cancer, for example. Thus, studies with larger numbers of exposed cases in cohort studies and/or controls in case-control studies are considered to be more informative.

Studies of workers in industries or occupations with higher levels of exposure, or workers with longer duration of exposure and sufficient variability in exposure are generally the most informative for evaluating cancer risk. Studies reporting minimum latency estimates for cancer from cobalt based on direct observation of latencies would be ideal. In this case, the population studied must be large enough to develop a reasonable estimate of the lower bound of the distribution of latencies, which is the estimate of the minimum latency (Howard 2013).

Studies evaluating exposure groups in which the majority of workers classified as "exposed" have very low exposure, very short duration of employment, or limited evidence of actual exposure may be inadequate to detect an effect due to a dilution effect. Further, the ability to evaluate exposure-response relationships depends on an adequate range of exposure (in intensity or duration) among the study participants, and adequate numbers of subjects in each exposure category.

Inadequate duration of follow-up may bias findings toward the null for cancer endpoints with longer latencies. Minimum latency estimates have been reported in the literature for lung cancer associated with exposure to asbestos, 19 years (Selikoff 1980; Magnani 2008; Harding 2009), and to chromium, 5 years (Harding 2009). The minimum interval between the onset of gastro-esophageal reflux disease (GERD) and diagnosis of esophageal cancer (latency) has been reported to be 20 years (den Hoed *et al.* 2011). Estimates of minimum latency for

nasopharyngeal cancer are 15 years (Hauptmann *et al.* 2004) and for cancer of the pleura, 30 years (Magnani *et al.* 2008) based on statistical modeling in epidemiologic studies of associations between an exposure and cancer.

Ideally, follow-up periods should exceed 15 to 20 years to permit adequate determination of these and other solid tumors with longer latencies, particularly in mortality studies. In addition, if cohort studies are sufficiently large to permit lagged analyses to be conducted to allow for a latency period, such analyses would contribute additional strength to the study and increase the study's sensitivity (Richardson *et al.* 2011).

3.2.7 Study-level judgment for overall utility for health hazard evaluation

The overall study utility is based on consideration of both the potential for biases (i.e., study quality) and consideration of study sensitivity. Serious concerns about study quality would result in lower utility ranking; however, a well-designed study with low study sensitivity (such as few exposed/expected cases for a specific endpoints) could be given a lower utility ranking. When adequate information is available, a judgment is made for the direction and distortion from the overall biases for a study or whether it has low sensitivity to detect an effect. Studies with critical concern for bias in a domain are usually considered to have inadequate utility and are not brought forward to the cancer evaluation. This evaluation occurs prior to the cancer assessment (e.g., interpreting the finding of the study).

- High (low/minimal concerns for bias and high sensitivity rating)
- Moderate (low/minimal or some concerns for bias, high or moderate sensitivity rating)
- Moderate/low (major to low concerns for bias, sensitivity rating varies)
- Low (major concerns, sensitivity rating varies)
- Inadequate (critical concerns for bias, sensitivity rating varies)

4 Cancer hazard assessment

This section outlines the approaches for reaching a level of evidence (e.g., sufficient, limited, or inadequate) for the carcinogenicity of cobalt from studies in humans. The conclusions regarding the assessment of study utility are carried forward to the cancer assessment, which consist of two phases: the evaluation of the evidence from the individual studies (Section 4.1) and integration of the evidence across studies to reach a preliminary level of conclusion (Section 4.2). Studies with the highest utility (e.g., lowest risk of bias and greatest sensitivity to detect an effect) are given the most weight in the assessment. The identification of the potential for specific bias, confounding, and study sensitivity utility is also used to interpret the findings from studies and to help explain heterogeneity across studies.

The application of the RoC listing criteria to the body of studies on cobalt includes evaluating (1) whether there is credible evidence for an association between exposure to cobalt and cancer, and (2) whether such an observed association can be explained by chance, bias, or confounding. Several existing considerations – strength of the association, consistency across studies, evidence of an exposure-response gradient, and temporality of exposure– are used to help guide the evaluation of these questions.

4.1 Evaluation of the evidence from individual studies

Conclusions are made on the strength of the evidence for each study; that is where we have confidence in the study findings given its strength and limitations. This section provides the integrated methods for evaluating confounding, which is considered in several steps of the evaluation and guidelines and terms for rating the strength of the evidence for a specific study.

4.1.1 Evaluation of confounding

A key question in the evaluation of the level of evidence from the observational studies of cobalt is whether any association between cobalt and cancer can be explained by confounding. Confounding occurs when the comparison groups under study (the exposed versus the unexposed groups in a cohort and the case versus control groups in a case-control study) have different background risks of disease (Checkoway *et al.* 2007; Rothman *et al.* 2008), in effect mixing the association of interest with the effects of other factors. Potential confounders include any exposures or risk factors that could be associated with both exposure and the disease outcome(s) of interest and that are not part of the disease pathway.

The evaluation of potential confounding takes the following into consideration:

- Identification of potential confounders, including co-exposures, demographics, and lifestyle factors, identified from literature reviews and a review of the studies (see below).
- Assessment of analytical or statistical methods to control for variables with evidence of confounding, or the use of other methods (e.g., directed acyclic graphics (DAGs) (Greenland *et al.* 1999), or other information included in the publication that provide information about potential confounders (see Confounding Bias, Section 3.4).
- The magnitude of the effect estimate or the strength of exposure-response relationships for specific cancer endpoints.

Identification of potential confounders

Confounders or co-exposures for cobalt are identified from the study authors and authoritative sources or literature reviews. Depending on the exposure scenario, several co-exposures of interest should be taken into consideration when evaluating the risk of cancer from exposure to cobalt.

The major occupational co-exposures that are potential or known risk factors in humans for lung cancer include arsenic, asbestos, beryllium, cadmium, coke oven fumes, chromium compounds, coal products, nickel refining, foundry substances, radon, soot, tars, silica, vinyl chloride, diesel exhaust, and radioactive ores such as uranium (ACA 2015). Among hard metal workers, lung carcinogens most likely to be present in the occupational setting include PAHs, silica, asbestos, nickel, chromium, cadmium, cobalt-tungsten, benzene, and nitrosamines, and diesel engine exhaust; among electrochemical workers producing cobalt, carcinogens include nickel and arsenic; among porcelain painters, carcinogenic co-exposures include nickel and silica.

Information on non-occupational risk factors are more often available in case-control studies than in cohort studies, and for studies of lung cancer, may include tobacco smoking and second hand smoke, radon, ionizing radiation, X-radiation. Tobacco smoking is of greatest concern

based on its potential to be related to exposure status. In the case of esophageal cancer, alcohol consumption, and tobacco smoking are risk factors (NIH NCI Cancer statistics).

Impact of potential confounders on study findings

For most of the non-occupational risk factors, unless there is a reason to suspect that they are related not only to the outcome, but also to exposure to cobalt they would not be considered as confounders or effect modifiers. Tobacco smoking and alcohol consumption may have the greatest potential to be related to exposure status.

Ideally, all potential confounders should be both quantified and subject to consideration for analysis for confounding, using appropriate statistical models; or confounding should be controlled for using other methods such as in the selection of the study participants (see Section 3.4). In the absence of information on confounders, analyses using internal referents similar to the exposed subjects can help reduce potential confounding.

In cohort studies, an indirect evaluation of the impact of confounding from co-exposures may be assessed by considering (1) the relative levels of exposure to cobalt compared with exposure to the potential confounder, (2) the strength of the association of the potential confounder with the endpoint of interest, and (3) the magnitude of the risk estimate or strength of exposure-response relationship for cobalt and specific cancer endpoints. As noted above, indirect information on the relationship between the estimated levels of exposure (albeit based on crude approximations) to the confounder compared with the estimated level of exposure to cobalt may be available from exposure monitoring or biomonitoring studies of subsets of workers.

4.1.2 Strength of the evidence conclusions

The presence of potential bias (such as selection and information bias) or confounding in a study does not mean that the study should be disregarded. Conclusions of the evidence from each study should consider the strengths and weaknesses of the study, the direction and distortion of the biases, as well as the strength of the findings – including the magnitude of the effect estimate, the strength of any exposure-response relationships, associations with adequate latency period, and internal consistency. The presence of biases (such as selection and information bias) and confounding may help explain heterogeneity across studies.

The following terms and guidelines are used to describe the level of evidence for the study (i.e., confidence in the reported effect estimate). This evaluation requires scientific judgment and the guidelines are NOT strict criteria or checklists.

- Evidence of an association (increased or decreased): Statistically significant risk, evidence of an exposure-response relationship, *or* patterns of internal consistency from a well designed study with limited potential (or small distortion) from biases, or biases mainly towards the null hypothesis or underestimate of the risk estimate (for a positive association). Methods used to assess confounding or information on potential confounder indicates potential confounding is unlikely to account for all the excess or reduced risk.
- Some evidence of an association: Evidence of an association but strength of association not likely to account for potential confounding or bias.
- Null: Effect estimates close to 1.0 but most bias towards the null or study has low sensitivity to detect an effect.

- Inconclusive: Study findings vary but unclear whether all the excess or decreased risk can be explained by bias or confounding and/or direction of bias unknown.

4.2 Integration of the scientific evidence across human cancer studies

For cobalt, a qualitative evaluation will be made that integrates the evidence across studies, giving the most weight to the studies that have the highest utility, and applies the RoC listing criteria to reach a listing recommendation. The application of the RoC listing criteria to the body of studies on a specific substance includes evaluating (1) whether there is credible evidence for an association between exposure to the substance and cancer, and whether such an observed association can be explained by chance, bias, or confounding. This evaluation considers temporality, consistency of findings across studies with adequate methodologies, strengths of the observed relationship between the substance and cancer, and evidence for an exposure-response gradient (see definitions below) (Hill 1965). While these factors are important for summarizing evidence across studies, it should be noted that they do not constitute absolute criteria. With the exception of temporality, causality can be demonstrated without each and every element being satisfied. Rather, the emphasis in integrating scientific evidence across human cancer studies should be placed on evaluating the extent to which biases, or confounding by co-exposures that may also cause cancer, could explain observed increases in cancer risk. No published quantitative assessments (meta-analyses) were identified for cobalt.

- *Temporality*. Exposure must occur before disease outcome.
- *The consistency of findings across studies*. Consistency is evaluated in the context of variations in outcome definitions, exposure assessment methodologies, exposure levels or duration of exposure of the population, exposure windows, length and completeness of follow-up, or other differences in population characteristics or study methodologies (e.g., those factors affecting study sensitivity).
- *The strength of observed associations between exposure to the substance and cancer*. The strength of the association can be important in evaluating whether specific confounders or biases can explain the observed association; however, the fact that an association is weak does not necessarily rule out a causal relationship. There are many examples of weak associations between an exposure to a substance and an endpoint that are nevertheless considered to be causal (e.g., environmental tobacco smoking and lung cancer).
- *Evidence for an exposure-response gradient*. A positive exposure-response relationship (which does not necessarily need to be monotonic) generally provides more convincing evidence of a causal association than a simple excess of disease. However, 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. Exposure-response relationships can also be evaluated across studies if adequate information on exposure levels (or duration) is available, in addition to looking for evidence in individual studies.
- *Evidence for associations with appropriate latency*.
- *Alternative explanations of chance, bias, or confounding*. Chance bias and confounding can also be evaluated across studies in addition to considering it at the individual study level. The finding of consistent, elevated, positive associations 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 can explain the associations observed across the body of studies.

Part F: Methods for Evaluating Cancer Studies in Experimental Animals

Objectives

This section describes the procedures used to prepare the cancer studies in the experimental animal section of the draft RoC monograph, and to reach a level of evidence conclusion on the carcinogenicity of cobalt and certain cobalt compounds.

Studies in experimental animals were found for the following cobalt compounds: cobalt chloride, cobalt acetate, cobalt sulfate (including hydrate and heptahydrate forms), cobalt oxide, cobalt metal, and cobalt nanoparticles.

Although most of the animal studies identified for consideration did evaluate exposure to a specific cobalt compound, in some experiments cobalt was administered in the presence of factors that modify carcinogenic effects (e.g., initiation-promotion or co-carcinogenicity studies). These studies can provide useful information such as whether the chemical can alter normal growth process leading to the promotion of tumors.

The objective of the evaluation of the studies in experimental animals is to reach a preliminary level of evidence conclusion [sufficient, not sufficient] for the carcinogenicity of cobalt and certain cobalt compounds from studies in experimental animals by applying the RoC listing criteria to the body of evidence.

RoC listing criteria for sufficient evidence of carcinogenicity from studies in experimental animals

"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."

Key questions

- What is the level of evidence (sufficient or not sufficient) of carcinogenicity of cobalt from animal studies?
- What are the methodological strengths and limitations of the studies?
- What are the tissue sites?

Approach

A preliminary search in PubMed was conducted for cobalt compounds (excluding studies of cobalt-containing alloys or radioactive cobalt compounds) with carcinogenicity studies, and articles were reviewed for relevance. Based on expert input at the Cobalt Information Group Meeting, the listing was suggested to generally be limited to cobalt compounds that produce cobalt ions *in vivo*. This could include compounds that are soluble in water and biological fluids (e.g., lung) or that are insoluble in water but may be solubilized and release cobalt ions at the lower pH within lysosomes in lung or other tissue (e.g., cobalt oxide and cobalt sulfide).

The steps for conducting the cancer evaluation include:

- Selection of the literature included in the cancer hazard evaluation of the experimental animal studies (Section 1)
- Systematic extraction of the data from experimental animal studies (Section 2)
- Assessment of the quality of the individual studies (Section 3)
- Assessment of the level of evidence of carcinogenicity from the studies based on applying the RoC criteria to the body of studies. (Section 4)

Details on the approach to be used to identify and report on relevant studies, including the factors used for study quality assessment, is provided following the brief description of the literature search.

1 Selection of the literature included in the cancer evaluation of experimental animals studies

Part B of this protocol discusses general procedures used to identify and select relevant literature for preparing the RoC monograph on cobalt and certain cobalt compounds. This section discusses procedures to identify and select literature relevant for the cancer evaluation, including the literature search strategy and the inclusion and exclusion criteria specific for identifying studies relevant for cancer in experimental animals. The available literature includes studies of several cobalt-containing compounds in experimental animals (mostly rodents) by different routes of exposure. As per the RoC process, studies must be peer reviewed and publicly available.

The identification of the relevant literature includes developing strategies for searching for citations and inclusion/exclusion questions for selecting the relevant citations from the searches. Animal studies for cobalt and cobalt compounds have previously been reviewed by NTP for cobalt sulfate (NTP 2002) and cobalt tungsten-carbide hard metals (NTP 2009), which included a list of studies of cobalt and cobalt compounds (but did not conduct a cancer hazard evaluation for possible listing in the RoC). The literature search for animal studies is limited to publications from 2007 (to allow for some overlap with the literature search for the cobalt tungsten-carbide hard metals review) to date.

The following approaches for identifying literature are employed:

- Literature searches of three scientific databases – PubMed, Scopus, and Web of Science – are conducted for this section using the specific search terms for the substance (see Table below) combined with search terms for cancer and experimental animals listed in the table below.
- For cobalt, sources searched for studies that might not be included in the scientific databases listed above were the Carcinogenicity Potency Data Base, PHS149 (*Survey of Compounds Which Have Been Tested for Carcinogenic Activity*), CCRIS, IARC monographs, and the RoC background document on Cobalt Tungsten-Carbide Hard Metals (NTP 2009).

Literature search strategy for cancer studies in experimental animals

Substance-specific search terms	Cancer search terms[a]	Animal species terms
cobalt NOT (cobalt-60 OR Co 60 OR 60Co OR radioactive OR gamma ra* OR radiotherapy OR radiation therapy OR radiation treatment	cancer OR neoplasm* OR tumor*	mouse OR mice OR rat OR hamster OR rabbit OR guinea pig OR swine OR pig

[a]Search terms were developed in consultation with an information specialist.

Citations are screened for primary studies on cancer in experimental animals using the procedures outlined in Part B of the protocol. Studies are initially included in the study if they meet the following inclusion criteria:

- Measure neoplastic (benign, malignant) endpoints.
- Have non-cancer data that are informative for a cancer assessment, such as reporting preneoplastic lesions.
- Describe non-neoplastic lesions that are considered part of a morphologic continuum to neoplasia.
- Provide information on chronic study dose selection (such as a subchronic or short-term toxicity study used for chronic study dose selection).
- Co-carcinogen studies

Studies that have no concurrent control group or poor reporting of study design or of results may be excluded from further consideration.

2 Data extraction

Two independent reviewers extract data (such as methods and findings) from the individual studies into a database (Table Builder) in a systematic manner using standardized instructions and questions. The database contains "fields" that are specific for the different types of extracted information (such as species, strain, sex, route, dosing regimen, duration, and results). The instructions for data extraction (questions and guidelines) describe the specific type of information that should be summarized or entered into each field. The fields are used to populate tables used in the monograph.

Data for the study results include neoplasm location, neoplasm histotype, survival, incidence, and significance. If the study authors did not perform statistical analysis, NTP will calculate pair-wise analysis of neoplasm incidence using Fisher's Exact Test and will note that this was calculated by NTP.

Quality assurance and quality control of data extraction and database entry are accomplished by (1) double-checking of each data entry by the two independent reviewers and (2) flagging any discrepant entries and resolving them by mutual discussion with reference to the original data source.

3 Assessment of the utility of the individual studies in experimental animals

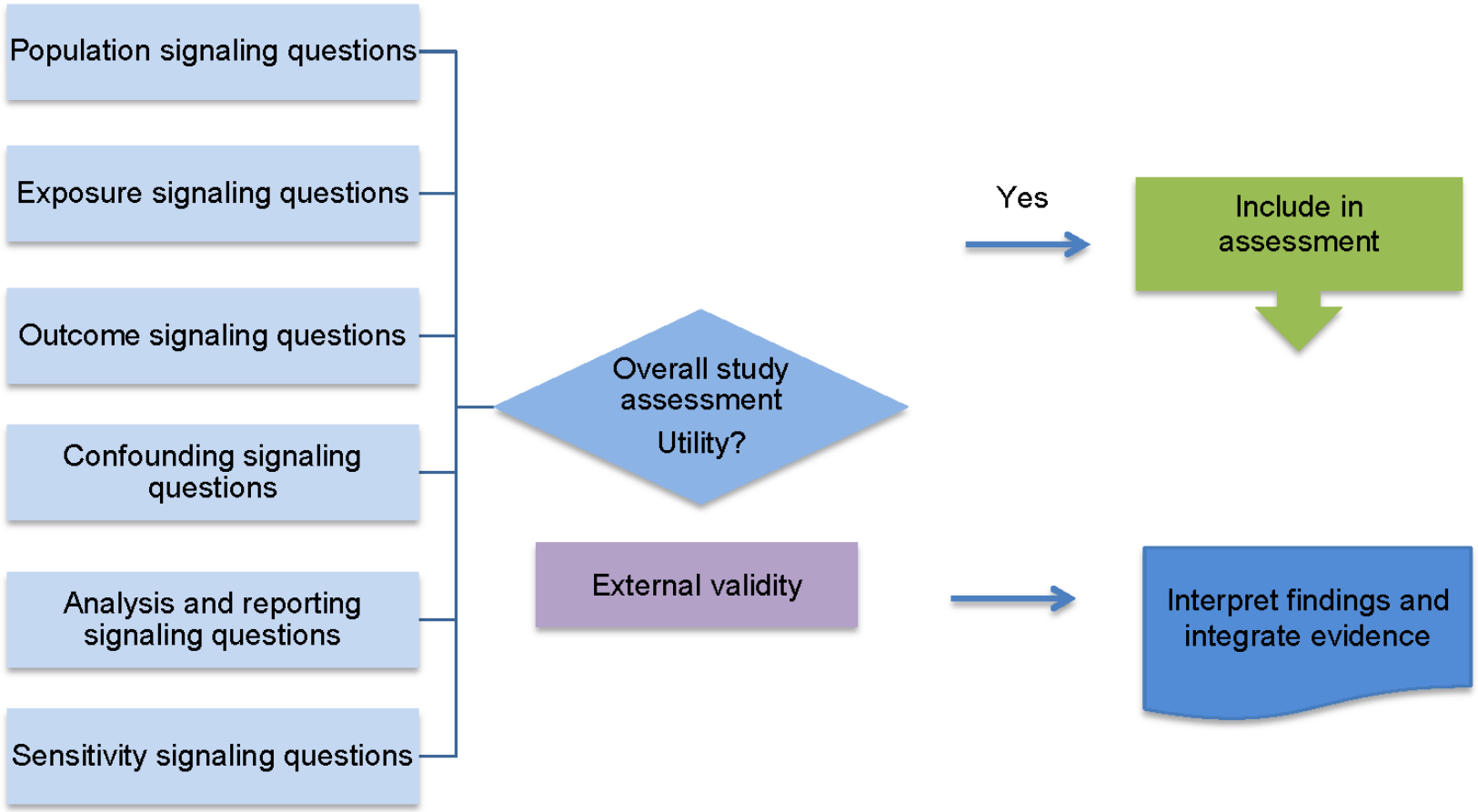
This section describes the assessment of the quality and utility of the individual studies including the steps in the process, responses for each step, signaling questions to evaluate study quality (internal validity), external validity, and the overall judgment of the utility of the study to inform the cancer hazard evaluation. This step is completed prior to interpreting the studies and assessing the evidence across studies.

Each primary study is systematically evaluated by two independent reviewers for its ability to inform the cancer hazard evaluation using a series of signaling questions related to the following study performance elements: population, exposure conditions, outcome assessment, potential confounding, and statistics and reporting. Studies are also evaluated for elements relevant to external validity (interpreting the findings for relevance to humans). The response for answering the signaling question of whether there is a potential bias or limitation is based on a comparison of the study element with that of the "ideal" study for a specific endpoint and exposure to the candidate substance. Differences are resolved by mutual discussion with reference to the original data source. Study authors may be contacted if there is inadequate information to evaluate a signaling question. The overall evaluation of the utility of the study is based on an integration of the responses to the signaling questions, and the judgment terms are similar to those for the human cancer studies (see Section 4). This step is done prior to interpreting the study findings and assessing the level of evidence across studies.

Signaling questions: Response

- Minimal concerns: Information from study designs and methodologies indicate that they are close to the ideal study characteristics and that the potential for bias is unlikely or minimal (+++).
- Some concerns: Study designs or methodologies are less than ideal, indicating possible bias (++).
- Major concerns: Study designs or methodologies suggest that the potential for a specific type of bias is likely (+).
- Inadequate: Study designs or methodologies suggest that the bias is critical and would make the study not informative for cancer hazard evaluation.
- No information: Inadequate information in the study to evaluate the level of concern.

Figure F-1. Schematic of approach for systematic review of study utility for cancer studies in experimental animal studies



3.1 Study quality and sensitivity evaluation

3.1.1 Quality of the selection of the study population: Guidelines and questions

Study elements that are critical for informing the cancer hazard evaluation include randomization of the animals to dose groups and the use of an appropriate comparison group (e.g., ideally unexposed, sham-treated concurrent controls). The absence of an appropriate control group, by itself, is sufficient for judging a study to be inadequate for cancer evaluation. However, a study that reports the incidence of a historical control group, but not a concurrent control group is not usually considered inadequate. The experimental design of some studies evaluating co-carcinogens may not include untreated concurrent controls but they generally include positive controls, which can result in acceptable study quality. The availability of historical control data from the testing laboratory can be helpful in assessing the significance of a finding, especially in the case of rare tumors, lower powered studies, or assessment of background tumor incidences, in particular when background rates are high. In addition, the treatment of animals in each dose group should be identical except for exposure status (e.g., control, dose).

- Are there concerns that the concurrent control group was not adequate for evaluating the study?
- Are historical control data reported?
- Are there concerns that the study design did not include randomization of animals to dose groups or take appropriate steps to ensure that dose groups are identical except for dosing status?

3.1.2 Quality of the exposure conditions: Guidelines and questions

Ideally, a study should use a chemical preparation that is representative of the candidate substance (in terms of purity and stability) so that any observed effects can be attributed to the candidate substance and the identity has been confirmed. Inhalation studies should also consider the impact of an aerosol generation system on the purity/stability and particle size of the substance. The animals should be exposed to high enough doses (resulting in tolerable toxicities) for a sufficiently long duration to assess carcinogenicity (usually approaching the lifetime of the animal for non-persistent substances). Ideally, this dose should not limit survival of the animals over the exposure period, unless due to tumor formation. Although not a requirement, the use of more than one dose level is ideal because evaluating exposure-response relationships may help to inform the carcinogenicity evaluation.

- Are there concerns that the chemical characterization and dose formulations (e.g. confirmation, homogeneity, purity, solubility, and stability) and delivery of the chemical (actual vs. desired dose) are not adequate for attributing any neoplastic effects to the substance?
- Are there concerns that the dosing regimen (dose selection and dose groups, or other factors) or the exposure duration are either not (1) adequate for detecting a neoplastic effect or (2) for attributing any tumor effects to the substance?
- Are there concerns that survival or body weight change(s) over time for treatment and/or control groups could affect attributing the study findings to the exposure?

3.1.3 Quality of measuring the endpoint (outcome): Guidelines and questions

Ideally, each study should perform full gross necropsies of all tissues and histopathological examinations on the majority of them. Pathology and/or diagnostic procedures and tissues examined should be accurately reported. Studies only examining tissues of interest may be informative for that tissue/organ but the evaluation should note limitations for other organs or tissues. All treatment groups and controls should be assessed at the same time and in the same way (such as sectioning of specific tumors).

- Are there concerns that the methods to assess tumor outcome and the pathology procedures (necropsy, histology, or diagnosis) are not adequate for attributing the effects to the exposure?

3.1.4 Potential for confounding

Some sources for potential confounding in animal studies are the use of an impure chemical that contains potential carcinogens and inadequate animal husbandry conditions and monitoring. Potential confounding may arise from carcinogens present in the animal feed, water, bedding, or the presence of disease or parasites or housing with animals exposed to other potential carcinogens.

- Are there concerns about the potential for confounding?

3.1.5 Analysis and reporting: Guidelines and questions

Each study should adequately report incidence data and use appropriate statistical methods. If statistical tests are not reported, the study should at a minimum present incidence data for specific tumors so that statistical tests can be run. If there is evidence of a decreased survival effect, the studies should use adequate statistical methods (such as poly 3) to control for these effects. Ideally, studies using multiple dose groups would conduct trend analysis to evaluate dose-response relationships. Analyses of benign and malignant tumors from the same tissue type should be reported both separately and combined; tumors of the same cellular origin may be combined (McConnell *et al.* 1986).

- Are there concerns that reporting of the data and statistical analysis are inadequate for evaluating the results? Are there concerns that different types of tumors are not accurately combined in the analysis?

3.1.6 Sensitivity: Guidelines and questions

The study should use an animal model that is sensitive for detecting tumors and does not have high background rates for the observed tumors or is well characterized with respect to background tumor rates, survival, and growth rates. Studies in both sexes are more informative than those testing only one sex. Adequate statistical power to detect an effect is based on the number of animals used in a study and their survival to study termination, the incidence of tumors in control vs. treated group(s), and the rarity of the tumor. As mentioned above, ideally the animal should be exposed for almost their lifetime; however, in cases with shorter exposure, it is recommended that animals be observed close to their lifetime. Studies less than one year that do not observe or report neoplasms are not informative. Poor animal husbandry conditions (such as feeding/bedding, housing, care, environmental conditions) and treatment-related survival

effects (those in which the animal does not survive long enough to develop tumors) may lead to high mortality and decrease statistical power.

- Are there concerns about the animal model (source, species, strain, or sex) that could affect study interpretation?
- Are there concerns that the study does not have adequate statistical power (number of animals per exposure and control group) to detect a neoplastic effect, if present? Are there concerns that survival-related effects or high mortality due to poor husbandry conditions have decreased statistical power?
- Are there concerns that the study duration (observation period) is not adequate to detect a neoplastic effect, if present?

3.2 Overall assessment of study utility

An overall assessment of study utility is based on consideration of both the potential for bias (limitations) and consideration of study sensitivity. Studies having elements that are judged to have major concerns may still be considered in the evaluation or can be considered to provide support to the more informative studies. Studies having critical concerns for important issues will generally be considered to be inadequate to inform the evaluation. It should also be noted that some concerns about a study element (such as inadequate observation and/or exposure period and statistical power) would decrease the sensitivity of a study to detect an effect; however, if despite these limitations positive findings were described, these studies would inform a cancer assessment. In some cases there is inadequate information to answer a specific question; the interpretation of how inadequate information affects the overall study quality evaluation depends on the extent and importance of the missing information and will be evaluated on a case-by-case basis. Some studies, such as co-carcinogen studies, have less utility for determining whether a substance is a cancer hazard but may provide utility regarding mechanism of action or other issues and thus utility would be rated based on the purpose of the study.

- High utility (low concerns for most potential biases)
- Moderate utility (some concerns for many potential biases)
- Low utility (major concerns for several biases)
- Inadequate utility (critical concerns for some potential biases)

3.3 External validity or interpretation

Some issues relevant to interpretation of the study findings in experimental animals for evaluating potential human carcinogenicity include the route of exposure and mechanisms of action (which would include other relevant information such as substance disposition). Studies of different exposure routes may not be as relevant to human exposures but are not usually excluded from the cancer assessment; route of exposure is evaluated on a case-by-case basis (see Section 5). Neoplasms observed in experimental animals are considered to be relevant to humans unless there is *compelling* evidence that indicates they occur by a mechanism that does not operate in humans. Other relevant data, such as mechanisms of carcinogenicity, are evaluated in a different monograph section, and conclusions are brought forward in the overall cancer evaluation section of the monograph, which integrates mechanistic evidence with human and experimental animal studies, to reach a preliminary listing recommendation.

- Are there concerns that the route of exposure is not appropriate for evaluating the potential for human carcinogenicity?

4 Cancer assessment

First, the findings from each study are interpreted with respect to their limitations and strengths (as identified in Section 4). The following factors are taken into consideration in determining whether an effect is treatment related: statistical significance with respect to concurrent controls and dose-related trends, non-neoplastic lesions, lesion progression, decreased latency, tumor multiplicity and survival, historical control range, animal species and strain, and rarity of tumor. Next, the evidence will be integrated across studies for each cobalt compound or type of compounds and exposure-related sites will be identified.

Because the candidate substance is defined as a class rather than as individual cobalt compounds, the application of the RoC listing criteria and preliminary level of evidence conclusions from studies in experimental animals will not be discussed until the overall cancer hazard conclusions (Section 6), which is after the evaluation of the mechanistic data. The overall cancer evaluation will also consider any mechanistic data related to biological plausibility. Findings in experimental animals are considered to be relevant to humans unless there is *compelling* evidence to suggest otherwise. Although the relevance of the route of the exposure to humans is considered, the findings of a similar tumor site by multiple routes of exposure strengthens the evidence for carcinogenicity.

Part G: Methods for Evaluating Mechanistic Studies and Other Relevant Data

Objectives

The purpose of this section is to discuss and assess the potential mechanisms of carcinogenicity for cobalt and certain cobalt compounds. It will include relevant discussions on the roles of ADME, toxicokinetics, and genotoxicity that are potentially involved in the carcinogenicity of cobalt (see Part D).

Key questions

- What are the genotoxic effects due to cobalt exposure? Does genotoxicity vary by cobalt compound?
- What are the cytotoxic or toxic effects of cobalt exposure? Does cytotoxicity or toxicity vary by cobalt compound?
- What are the major mechanistic modes of action for the carcinogenicity of cobalt?
    - What are the common key steps or mode(s) of action of toxicity or carcinogenicity across different cobalt compounds? What role and contribution does cobalt ion play in the proposed mechanism? What are the effects from exposure to particulate cobalt?
    - What factors influence biological or carcinogenic effects? How does particle size, solubility, and cellular uptake of a cobalt compound affect biological or carcinogenic effects?
    - Is there evidence that support grouping cobalt and certain compounds together in the assessment?

Approach

One of the major issues to be considered in this review of cobalt is the mechanism of carcinogenicity for cobalt and certain cobalt compounds with available mechanistic data. The approach will involve the assessment of whether cobalt compounds induce effects via a common mechanism, i.e., reactivity of the cobalt ion. Discussions in this section will include the reported effects of cobalt particles versus cobalt ions, including cell uptake, particle properties, and the cytotoxicity observed for each of these forms of cobalt. The proposed modes of action to be considered and described include genotoxic effects such as inhibition of DNA repair, oxidative stress, and the consequences of cobalt-induced HIF-1 alpha activation.

To evaluate potential mechanistic effects, there will be a comprehensive review to identify pertinent studies followed by a focused evaluation to identify those studies that provide specific information for the key steps in the potential mechanism(s) of carcinogenicity of cobalt and certain cobalt compounds. Key studies include those that provide information on similarities or differences in the biological effects of different cobalt compounds. The evaluation will provide specific information on key mechanistic events in pertinent studies, but it is not intended as a comprehensive summary of all available data. Both secondary sources, such as authoritative reviews, and primary publications will be used in this section. For genotoxicity, IARC published a comprehensive review s in 2006 (IARC 2006), and no issues involving genotoxic effects were

identified for cobalt compounds, thus their report will be relied upon to provide the basis of the evaluation of genotoxicity in the current evaluation. Literature searches will be performed to identify relevant, recently published primary studies to be included in the evaluation.

As described in the introduction, a table of available specific information will be constructed to provide a direct visual comparison of properties and effects of exposure for several cobalt-containing compounds.

Literature search strategy

Literature searches of three scientific databases – PubMed, Scopus, and Web of Science – are conducted using a pre-determined range of search terms listed in the table below.

The strategy for the literature search is to identify published information on cobalt, which will potentially include cobalt metal and cobalt compounds, both inorganic and organic. Substances excluded from the search include radioactive cobalt, cobalt-tungsten carbide hard metals, and cobalt alloys; a rationale will be provided for the exclusion of these forms of cobalt. The cobalt compounds for which we currently have the most information include cobalt metal, cobalt sulfate, cobalt chloride, and cobalt oxide. Based on preliminary information on these forms of cobalt, publications on these forms and on any cobalt compounds that can release cobalt ions in biological fluids will be included in the literature search for this evaluation. (See the approach for information related to selecting primary studies for inclusion in the review.)

The specific search terms used for this section are listed in the table below. Terms in the columns are combined with "AND" in the databases.

Literature search strategy for mechanistic studies and other relevant data

Substance	Key mechanistic search terms[a]	Species-related search terms[a]
Cobalt NOT (cobalt-60 OR Co 60 or 60Co OR radioactive OR gamma ra* OR radiotherapy OR radiation therapy OR radiation treatment)	(mode OR mechanism*) AND action) OR (carcinogen OR genetic OR epigenetic OR inhibit* OR promot* OR interact* OR activate* OR detoxific* OR "oxidative damage" OR alkylat* OR adduct) OR "hypoxia-inducible factor 1, alpha subunit" [MeSH] OR "reactive oxygen species" [MeSH] OR "alpha subunit hypoxia-inducible factor 1" OR "hif 1 alpha" OR "oxidative stress" [MeSH] OR DNA repair inhibition OR cytotoxicity OR "lung toxicity"	In vivo OR animal OR animals OR mouse OR mice OR rat OR hamster OR "guinea pig" OR rabbit OR monkey OR dog OR pig) OR (person* OR people OR individual* OR subject* OR participant*

[a]Search terms were developed in consultation with an information specialist. This list contains the key terms and not the detailed search strategy.

REFERENCES

1 Alexandersson R. 1988. Blood and urinary concentrations as estimators of cobalt exposure. *Arch Environ Health*. 43(4):299-303.

2 American Cancer Society 2015. http://www.cancer.org/acs/groups/content/@nho/documents/document/occupationandcancerpdf.pdf

3 Anderson LA, Johnston BT, Watson RG, Murphy SJ, Ferguson HR, Comber H, McGuigan J, Reynolds JV, Murray LJ. 2006. Nonsteroidal anti-inflammatory drugs and the esophageal inflammation-metaplasia-adenocarcinoma sequence. *Cancer Res*. 66(9):4975-82.

4 Applebaum KM, Malloy EJ, Eisen EA. 2011. Left truncation, susceptibility, and bias in occupational cohort studies. *Epidemiol.* 22(4):599-606.

5 ATSDR, 2004. *Toxicological profile for cobalt*. Atlanta GA.: Agency for Toxic Substances and Disease Registry. http://www.atsdr.cdc.gov/toxprofiles/tp.asp?id=373&tid=64

6 Chavance M, Dellatolas G, Lellouch J. 1992. Correlated nondifferential misclassifications of disease and exposure: application to a cross-sectional study of the relation between handedness and immune disorders. *Int J Epidemiol*. 21(3):537-46.

7 Checkoway H, Pearce N, Kriebel D.2007. Selecting appropriate study designs to address specific research questions in occupational epidemiology. 2007. *Occup Environ Med.* 64(9):633-8.

8 Chevrier J, Brown D, Picciotto S, Costello S, Neophytou A, Eisen E. 2014. The healthy worker survivor effect dissected: addressing component parts. *Occup Environ Med*. 71 Suppl 1:A120. doi: 10.1136/0348.

9 Collecchi P, Esposito M, Brera S, Mora E, Mazzucotelli A, Oddone M. 1986. The distribution of arsenic and cobalt in patients with laryngeal carcinoma. *J Applied Toxic.* 6(4):287-289.

10 Coyle YM1, Minahjuddin AT, Hynan LS, Minna JD. 2006. An ecological study of the association of metal air pollutants with lung cancer incidence in Texas. *J Thorac Oncol*. 1(7):654-61.

11 den Hoed CM, van Blankenstein M, Dees J, Kuipers EJ. 2011. The minimal incubation period from the onset of Barrett's oesophagus to symptomatic adenocarcinoma. *Br J Cancer*. 2011;105:200-205.

12 Garland M, Morris JS, Rosner BA, Stampfer MJ, Spate VL, Baskett CJ, Willett WC, Hunter DJ. 1993. Toenail trace element levels as biomarkers: reproducibility over a 6-year period. *Cancer Epidemiol Biomarkers Prev*. 2(5):493-7. Erratum in: *Cancer Epidemiol Biomarkers Prev.* 1994 Sep;3(6):523.

13 Gerhardsson, L, Brune D, Nordberg GF and Wester PO. 1986. Selenium and other trace elements in lung tissue in smelter workers relationship to the occurrence of lung cancer. *Acta Pharmacol Toxicol* (Copenh). 59 Suppl 7:256-259.

14 Gerhardsson, L, Brune D, Nordberg GF and Wester PO. 1988. Multielemental assay of tissues of deceased smelter workers and controls. *Sci Total Environ.* 74:97-110.

15 Gerhardsson, L, Brune D, Nordberg GF and Wester PO. 1985. Protective effect of selenium on lung cancer in smelter workers. *Br J Ind Med.* 42:617-26.

16 Gerhardsson L and Nordberg GF. 1993. Lung cancer in smelter workers--interactions of metals as indicated by tissue levels. *Scand J Work Environ Health*. 19 Suppl 1:90-94.

17 Gerhardsson L, Wester PO, Nordberg GF and Brune D. 1984. Chromium, cobalt and lanthanum in lung, liver and kidney tissue from deceased smelter workers. *Sci Total Environ*. 37:233-246.

18 Greenland S, Pearl J, Robins JM. 1999. Causal diagrams for epidemiologic research. *Epidemiol.* 10:37-48.

19 Grimsrud TK, Berge SR, Haldorsen T, Andersen A. 2005. Can lung cancer risk among nickel refinery workers be explained by occupational exposures other than nickel? *Epidemiol*. 16(2):146-154.

20 Harding AH, Darnton A, Wegerdt J, McElvenny D. 2009. Mortality among British asbestos workers undergoing regular medical examinations (1971-2005). *Occup Environ Med*. 66:487-495.

21 Hauptmann M, Lubin JH, Stewart PA, Hayes RB, Blair A. 2004. Mortality from solid cancers among workers in formaldehyde industries. *Am J Epidemiol*. 159:1117-1130.

22 Hill AB. 1965. The environment and disease: association or causation? *Proc R Soc Med.* 58 (5):295-300.

23 Howard J. 2013. *Minimum latency and types or categories of cancer*. Report from the World Trade Center Health Program, Revision: May 1, 2013. http://www.cdc.gov/wtc/pdfs/wtchpminlatcancer2013-05-01.pdf

24 IARC, 1991. *Chlorinated drinking-water; chlorination by-products; some other halogenated compounds; cobalt and cobalt compounds*. IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Humans. Lyon, France: International Agency for Research on Cancer.

25 IARC. 2006. *Cobalt in Hard-metals and Cobalt Sulfate, Gallium Arsenide, Indium Phosphide and Vanadium Pentoxide*. IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Humans. Lyon, France: International Agency for Research on Cancer. 330 pp.

26 Ichikawa Y, Kusaka Y, Goto S. 1985. Biological monitoring of cobalt exposure, based on cobalt concentrations in blood and urine. *Int Arch Occup Environ Health*. 55(4):269-76.

27 Jensen AA, Tüchsen F. 1990. Cobalt exposure and cancer risk. *Crit Rev Toxicol*. 20(6):427-37.

28 Kristensen P. 1992. Bias from nondifferential but dependent misclassification of exposure and outcome. *Epidemiol*. 3(3):210-5.

29 Lison D, Buchet JP, Swennen B, Molders J, Lauwerys R. 1994. Biological monitoring of workers exposed to cobalt metal, salt, oxides, and hard metal dust. *Occup Environ Med.* 51(7):447-50.

30 Lauwerys R, Lison D. 1994. Health risks associated with cobalt exposure--an overview. *Sci Total Environ*. 30;150(1-3):1-6.

31 Longnecker MP, Stampfer MJ, Morris JS, Spate V, Baskett C, Mason M, Willett WC. 1993. A 1-y trial of the effect of high-selenium bread on selenium concentrations in blood and toenails. *Am J Clin Nutr*. 57(3):408-13.

32 Magnani C, Ferrante D, Barone-Adesi F, *et al.* 2008. Cancer risk after cessation of asbestos exposure: a cohort study of Italian asbestos cement workers. *Occup Environ Med*. 65:164-170.

33 Marsh GM, Esmen NA, Buchanich JM, Youk AO. 2009. Mortality patterns among workers exposed to arsenic, cadmium, and other substances in a copper smelter. *Am J Ind Med*. 52(8):633-44. doi: 10.1002/ajim.20714.

34 McConnell EE, Solleveld HA, Swenberg JA, Boorman GA. 1986. Guidelines for combining neoplasms for evaluation of rodent carcinogenesis studies. *JNCI.* 76(2):283-289.

35 Moulin JJ, Wild P, Mur JM, Fournier-Betz M, and Mercier-Gallay M. 1993. A mortality study of cobalt production workers: an extension of the follow-up. *Am J Indust Med*. 23:281-288.

36 Moulin JJ, Wild P, Romazini S, Lasfargues G, Peltier A, Bozec C, Deguerry P, Pellet F, Perdrix A. 1998. Lung cancer risk in hard-metal workers. *Am J Epidemiol.* 148(3):241-248.

37 Mur JM, Moulin JJ, Charruyer-Seinerra MP, Lafitte J. 1987. A cohort mortality study among cobalt and sodium workers in an electrochemical plant. *Am J Ind Med.* 11(1):75-81.

38 Moulin JJ, Clavel T, Buclez B, Laffitte-Rigaud G. 2000. A mortality study among workers in a French aluminium reduction plant. *Int Arch Occup Environ Health*. 73(5):323-330.

39 Naimi AI, Cole SR, Hudgens MG, Brookhart MA, Richardson DB. 2013. Assessing the component associations of the healthy worker survivor bias: occupational asbestos exposure and lung cancer mortality. *Ann Epidemiol*. 23(6):334-341. doi: 10.1016/j.annepidem.2013.03.013.

40 National Cancer Registry of Ireland (http://www.ncri.ie/data/performance-indicators)

41 NTP. 2013. *DRAFT Toxicology Studies of Cobalt Metal (CAS NO. 7440-48-4) in F344/N Rats and B6C3F1/N Mice and Toxicology and Carcinogenesis Studies of Cobalt Metal in*

F344/NTac Rats and B6C3F1/N Mice (Inhalation Studies). Technical Report Series No. 581. NIH Publication No. 14-5923. Research Triangle Park, NC: National Toxicology Program. 354 pp. http://ntp.niehs.nih.gov/NTP/About_NTP/TRPanel/2013/October/DRAFT_TR-581.pdf#search=cobalt%20metal.

42 NTP. 1998. *Toxicology and Carcinogenesis Studies of Cobalt Sulfate Heptahydrate in F344/N Rats and B6C3F1 Mice*. Technical Report Series No. 471. NIH Publication No. 98-3961. Research Triangle Park, NC: National Toxicology Program. 268 pp. http://ntp.niehs.nih.gov/ntp/htdocs/lt_rpts/tr471.pdf.

43 NTP. 2002. *Report on Carcinogens Background Document for Cobalt Sulfate*. National Toxicology Program. http://ntp.niehs.nih.gov/ntp/newhomeroc/roc11/CoSO4Pub_no_appendices_508.pdf.

44 NTP. 2009. *Report on Carcinogens Background Document for Cobalt-Tungsten Carbide: Powders and Hard Metals*. National Toxicology Program. http://ntp.niehs.nih.gov/NTP/roc/twelfth/2010/FinalBDs/HardMetalsBD20100408_508.pdf.

45 NTP. 2011. *Report on Carcinogens, Twelfth Edition*. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. 499 pp. http://ntp.niehs.nih.gov/go/roc12.

46 O'Rorke MA, Cantwell MM, Abnet CC, Brockman JD, and Murray LJ on behalf of the FINBAR Study Group. 2012. Toenail trace element status and risk of Barrett's oesophagus and oesophageal adenocarcinoma: results from the FINBAR study. *Int J Cancer.* 131(8):1882-1891.

47 Qayyum MA and Shah MH. 2014. Comparative assessment of selected metals in the scalp hair and nails of lung cancer patients and controls. *Biol Trace Elem Res*. 158:305-322.

48 Richardson DB, Cole SR, Chu H, Langholz B. 2011. Lagging exposure information in cumulative exposure-response analyses. *Am J Epidemiol*. 174(12):1416-22. doi: 10.1093/aje/kwr260. Epub 2011 Nov 1.

49 Rogers MAM, Thomas DB, Davis S, Vaughan TL, Nevissi AE. 1993. A case-control study of element levels and cancer of the upper aerodigestive tract. *Cancer Epidemiol, Biomarkers & Prev.* 2:205-312.

50 Rothman K, Greenland S, Lash T. 2008. *Modern Epidemiology*, 3rd Edition. New York: Lippincott, Williams, and Wilkins, pp 851.

51 Sabbioni E, Minoia C, Pietra R, Mosconi G, Forni A, Scansetti G. 1994. Metal determinations in biological specimens of diseased and non-diseased hard metal workers. *Sci Total Environ*. 150(1-3):41-54.

52 Selikoff IJ, Hammond EC, Seidman H. 1980. Latency of asbestos disease among insulation workers in the United States and Canada. *Cancer*. 46:2736-2740.

53 Sterne JAC, Higgins JPT, Reeves BC on behalf of the development group for ACROBAT-NRSI. *A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI), Version 1.0.0*, 24 September 2014. Available from http://www.riskofbias.info [accessed {date}].

54 Thomas 2009 as cited by NRC 2014.

55 Tüchsen F, Jensen MV, Villadsen E, Lynge E. 1996. Incidence of lung cancer among cobalt-exposed women. *Scand J Work Environ Health.* 22(6):444-450.

56 U.S. National Institutes of Health. National Cancer Institute. Cancer statistics. http://www.cancer.gov/statistics/

57 U.S. National Institutes of Health. National Cancer Institute. SEER Cancer Statistics Review, 1975-2011.

58 U.S. National Institutes of Health. National Cancer Institute. Surveillance, Epidemiology, and End Results Program. http://seer.cancer.gov/registries/sps.html. Seattle-Puget Sound cancer registry; http://www.fredhutch.org/en/labs/phs/projects/cancer-surveillance-system.html.

59 U.S. National Institutes of Health. National Cancer Institute. Surveillance, Epidemiology, and End Results Program. http://seer.cancer.gov/registries/list.html.

60 Wild P, Perdrix A, Romazini S, Moulin JJ, Pellet F. 2000. Lung cancer mortality in a site producing hard metals. *Occup Environ Med.* 57:568-573.