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

PROTOCOL FOR A SYSTEMATIC REVIEW OF

MOUNTAINTOP REMOVAL MINING: IMPACTS ON HEALTH IN THE SURROUNDING COMMUNITY

July 27, 2016

Office of Health Assessment and Translation (OHAT)

Division of the National Toxicology Program

National Institute of Environmental Health Sciences

PROTOCOL FOR A SYSTEMATIC REVIEW OF MOUNTAINTOP REMOVAL MINING: IMPACTS ON HEALTH IN THE SURROUNDING COMMUNITY

Project Leader: Abee L. Boyles, Office of Health Assessment and Translation (OHAT), DNTP

Summary: OHAT is conducting a systematic review to evaluate potential impacts of mountaintop removal (MTR) mining on the health of people living in the surrounding communities and identify key research needs. Depending on the extent and nature of the literature available, level of evidence conclusions for hazard identification may be reached.

BACKGROUND AND SIGNIFICANCE

Background

Since its introduction in the 1960s, MTR mining has become a major method of coal mining in and around Central Appalachia (including parts of Kentucky, Ohio, Pennsylvania, Tennessee, Virginia, and West Virginia) because it is typically faster, cheaper, and less labor intensive than underground mining (Holzman 2011). This mining method involves clearing the area of trees and topsoil and using explosives to blast apart the mountain rock to access coal seams (Palmer *et al.* 2010). The excess rock (i.e. mine spoil) is often pushed into adjacent valleys (i.e. valley fill). The air, water, and soil in the surrounding area are impacted by these mining practices and have the potential to adversely impact human health and the environment (Simmons *et al.* 2008, Palmer *et al.* 2010, Acton *et al.* 2011). Exposures associated with MTR include particulate matter, polycyclic aromatic hydrocarbons, metals, and other potentially harmful substances (Palmer *et al.* 2010).

Significance

This OHAT evaluation will review the available literature on MTR mining and potential human health effects. If the literature base is sufficient, this review will reach level of evidence conclusions for hazard identification. Even if the database is too limited to support hazard conclusions, the final evaluation product will include critical appraisal of the identified studies that may strengthen the design and conduct of future studies. Identification and summarization of studies of chemical exposures associated with mountaintop removal mining will provide necessary context and identify areas of future research on potential human health effects.

OVERALL OBJECTIVE AND SPECIFIC AIMS

The overall objective of this evaluation is to understand the human health impacts of MTR mining by conducting a systematic review of published studies of MTR mining and community health, occupational studies of MTR mining, and any available animal and *in vitro* experimental studies investigating the effects of exposures to MTR mining-related mixtures.

Primary Review Aims:

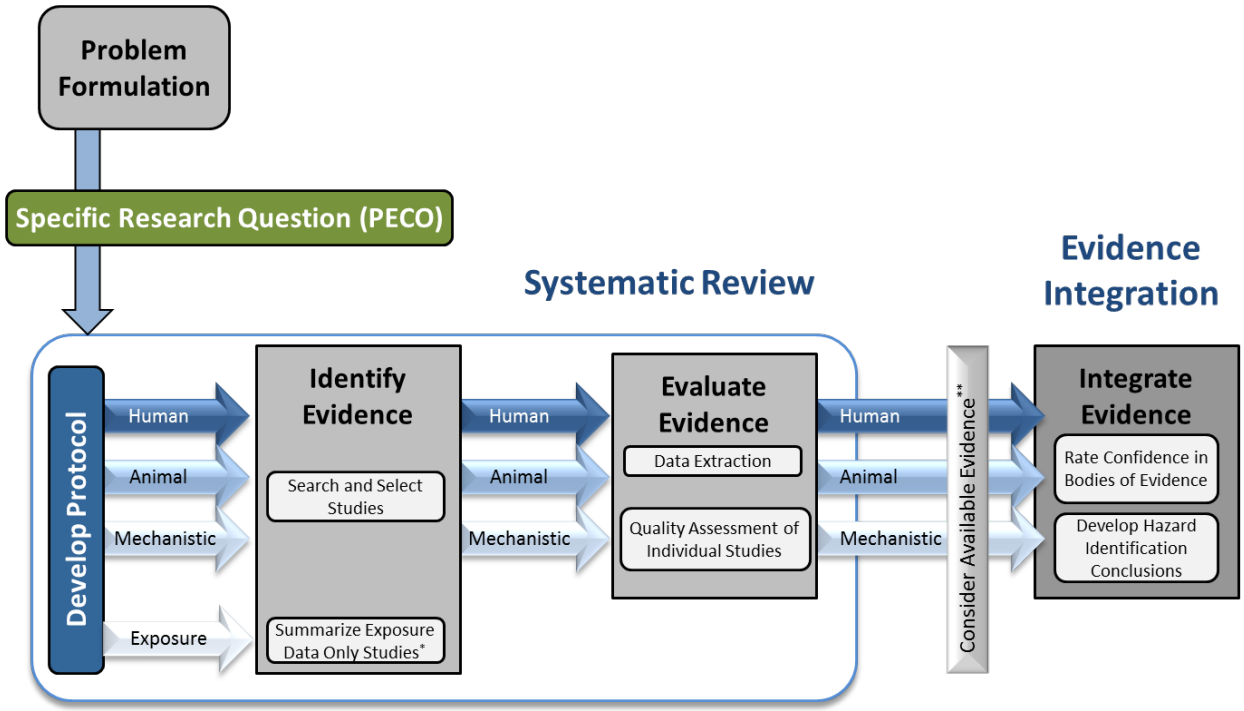
- Identify literature reporting the health effects of exposure to MTR including related open coal mining practices in Appalachia in human, experimental animal, and *in vitro* model systems with no restriction on the type of health outcome.
- Extract data on potential health effects from relevant health effects studies.
- Assess the internal validity (risk of bias) of individual health effects studies.
- Summarize the extent and types of health effects evidence available.
- Collect and summarize the available information on human exposure to chemicals associated with mountaintop removal mining including measurements in air, water, and soil.
- Provide context for main exposures identified by referencing authoritative reviews and established human health benchmarks.
- Describe limitations of the systematic review, limitations of the evidence base, identify data gaps and key research needs, and describe findings in the context of human exposure levels.

Dependent on the extent and nature of the available evidence:

- Synthesize the evidence, including meta-analyses if appropriate, considering limitations on data integrating such as study design heterogeneity. Conduct analyses to consider variation in community/occupational, MTR/unspecified mining, and pre-1980/post-1990 effect estimates.
- Rate confidence in the body of evidence for human and animal studies separately according to one of four statements: (1) High, (2) Moderate, (3) Low, or (4) Very Low/No Evidence Available.
- Translate confidence ratings into level of evidence of health effects for human and animal studies separately according to one of four statements: (1) High, (2) Moderate, (3) Low, or (4) Inadequate.
- Combine the level of evidence ratings for human and animal data and consider the degree of support from mechanistic data to reach one of five possible hazard identification conclusions: (1) Known, (2) Presumed, (3) Suspected, (4) Not classifiable, or (5) Not identified to be a hazard to humans.

The evaluation will be conducted based on a 7 step process outlined in the OHAT approach (Rooney *et al.* 2014). The "Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration" provides standard operating procedures for the implementation of systematic review in OHAT evaluations (http://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf). The standard operating procedures are based on (1) lessons learned from developing protocols for two case studies for implementing systematic review, (2) consideration of public comments received on systematic review during the past two years, and (3) discussions with experts at other organizations and agencies working on applying methods of systematic review to environmental health and toxicology. The handbook is a living document and will be updated as methodological practices are refined and evaluated and strategies are identified that improve the reliability, ease, and efficiency of conducting systematic reviews. The schematic of application of the OHAT approach to this evaluation of MTR Mining is summarized in **Figure 1**.

Figure 1. Schematic of Review Process



*Studies with data only on MTR mining exposures (no health outcomes, including primary and secondary sources) provide critical context for the evaluation of studies with human, animal and mechanistic outcomes

**The extent and types of evidence available, will determine if the review will proceed to the Evidence Integration steps

To address our overall objective we developed a PECO statement (**P**opulation, **E**xposure(s), **C**omparator(s), and **O**utcome(s)) (Table 1) which is used as an aid to develop literature search terms and inclusion/exclusion criteria for the systematic review.

Table 1. Population, Exposure, Comparator, and Outcome (PECO) Statement

	Evidence
Population	Humans without restriction based on age, sex, or lifestage at exposure or outcome assessment; experimental animal and *in vitro* systems that are used as models of human health
Exposure	Exposure to mountaintop removal mining activities including residential proximity or occupational exposure, environmental measures (e.g., air, water levels), or experimental exposure to a MTR mining-related mixture (i.e., not studies of individual chemicals)
Comparator	Vehicle-only treatment controls in experimental studies
Outcomes	Any health-related effect or change in physiological or cellular response
Study type	No restrictions

METHODS

Problem Formulation

Nomination History

The NTP Office of Health Assessment and Translation (OHAT) received a nomination in January of 2015 to consider reviewing human health effects of MTR mining. The nomination highlighted several published reports of exposures and health effects associated with MTR mining, the existing health disparities in Appalachia, and relatively little funding for scientific research in these communities. In July of 2015, the directors of the Centers for Disease Control and Prevention (CDC), the National Institutes of Health (NIH), and the U.S. Geological Survey (USGS) received a letter from the West Virginia Department of Health and Human Resources requesting federal expertise in reviewing the research on potential health implications of surface coal mining. A Request for Information on the nomination was released in the Federal Register on October, 7, 2015, and no comments were received at that time. A concept was developed and presented to the NTP Board of Scientific Counselors (BSC) on December 2, 2015 and there were no public comments on the project at that time (http://ntp.niehs.nih.gov/go/780611).

Problem Formulation and Protocol Development

The overall objective, specific aims, and PECO statement included above were developed and refined through a series of problem formulation steps including: (1) review of the topic by NTP staff; (2) consultation with scientists at other Federal agencies represented on the NTP Executive Committee[1]; (3) review of the concept by technical advisors with backgrounds in toxicology and environmental health effects associated with MTR mining; (4) public review of the concept for "Mountaintop Removal Mining: Impacts on Health in the Surrounding Community" by the NTP BSC (http://ntp.niehs.nih.gov/go/9741); (4) guidance outlined in the OHAT Handbook for Conducting a Literature-Based Health Assessment (NTP 2015a).

Consideration of key scientific issues

1. **Limited epidemiological evidence base**

   Based on the literature scoping activities during problem formulation, there are few studies of community health effects of MTR mining and those available are predominantly of limited size and scope. Thus, considering other types of data (i.e., occupational studies, experimental data from animal and in vitro studies) within the comprehensive literature search may inform interpretation of the community health studies. For example, recognizing issues such as differences in exposure, evaluating consistency of exposure-health outcome associations identified in occupational studies with those identified in community studies might assist in assessing plausibility of causation. Experimental studies allow for a controlled exposure to mining-related mixtures to test biological hypotheses generated by human observational

[1] Consumer Product Safety Commission (CPSC), Department of Defense (DoD), Environmental Protection Agency (EPA), Food and Drug Administration (FDA), National Cancer Institute (NCI), National Center for Environmental Health/Agency for Toxic Substances and Disease Registry (NCEH/ATSDR), National Institute of Environmental Health Sciences (NIEHS), National Institute for Occupational Safety and Health (NIOSH), Occupational Safety and Health Administration (OSHA) http://ntp.niehs.nih.gov/go/163

studies. When available, experimental animal and in vitro studies may provide context for interpreting the human studies literature.

2. **Confounding issues**

   Observational epidemiology studies must properly account for all potential sources of confounding. Lower socioeconomic status, smoking, and reduced access to health care are all factors they may be associated with both the observed health effects and living near MTR mining. A systematic review including a transparent evaluation of confounding and other aspects of internal validity (risk of bias) will help document the relative strengths and weaknesses of each study's design and conduct as well as identify areas for improvement in future research in this area. This area may be of particular concern for MTR mining exposures when individual confounding factors and exposure levels have not typically been measured in published reports of community health effects.

3. **Complex Exposure Scenario**

   Unlike relatively contained underground coal mining, MTR mining can expose the surrounding community to hazardous materials, particularly air particulate matter. Such complex mixture exposure scenarios include metals with other components of air particulate matter and selenium with other water pollutants (Palmer et al. 2010). Identifying risks from specific sources and their individual components is extremely difficult in observational epidemiological studies of communities unless the studies cite relevant exposure work in the same communities such that it can be considered jointly by the review. Studies that characterize exposures related to MTR mining (i.e. environmental monitoring) will be identified by the systematic review and summarized to provide context for the observed human health studies. Authoritative reviews of toxicity of individual components will be separately identified and summarized to provide context for interpretation of the studies of health effects.

Search and Select Studies for Inclusion

Literature Search Strategy

A literature search strategy was developed by an informationist familiar with systematic review methodology to identify all relevant published evidence on mountaintop and Appalachian coal mining through: (1) reviewing PubMed's Medical Subject Headings (MeSH) for relevant and appropriate terms, (2) extracting key terminology from relevant reviews and a set of previously identified primary data studies that are known to be relevant to the topic ("test set"), and (3) reviewing search strategies presented in other reviews. Six databases will be searched from the beginning of the database entries: Embase, PsycINFO, PubMed, Scopus, Toxline, and Web of Science. The search strategy was customized for each database because of differences in syntax (see **Appendix 1**). No publication year or language limits will be imposed.

Databases Searched

- Embase
- PsycINFO
- PubMed
- Scopus

- Toxline
- Web of Science

Searching Other Resources

We will use the following methods to find additional studies that were not identified through the electronic searches. Studies will be evaluated using the same inclusion and exclusion criteria as used for screening records retrieved from the electronic search. Relevant studies identified through these steps will be marked as "provided from other sources" in the study selection flow diagram.

- Hand searching the reference lists of relevant reviews, commentaries, government-authored (state and federal) technical reports, or other non-research articles identified during the initial search. Commentaries or letters on specific studies are also reviewed to see if they contain content that should be noted during data extraction or risk of bias assessment of the original report.
- Hand searching the reference lists of included studies after the full text review.
- Reference identified by subject matter experts or the public.

Screening Process

DistillerSR®, a web-based, systematic review software program with structured forms and procedures will be used to screen articles for relevance and eligibility to ensure standardization of process[2]. Initially, results of the literature search are assembled in EndNote software and exact article duplicates removed prior to uploading the references and within the systematic review software program.

Evidence Selection Criteria for (1) health effects and for (2) exposure data only

In order to be eligible for inclusion in the systematic review of health effects, studies must comply with the criteria specified by the PECO statement (**Table 1**) or contain relevant exposure assessment information. In addition to the PECO criteria, the following exclusion criteria will apply: studies that do not contain original data, such as reviews, editorials, or commentaries; and studies that have not been peer-reviewed (e.g., conference abstracts, theses/dissertations, working papers from research groups or committees, and white papers). There are no limitations on the language of the publication.

Exposure Data Only: Studies that only measure exposures will be considered for the MTR Mining Exposure Data Only Study summarization.

Inclusion and exclusion criteria used to screen articles for relevance and eligibility at both the title-and-abstract and full-text screening stages are summarized in **Table 2**.

[2]DistillerSR® (http://systematic-review.net/) is a proprietary project management tool for tracking studies through the screening process and storing data extracted from these studies using user-customized forms.

Table 2. Detailed inclusion and exclusion criteria to determine study eligibility for health effects

Inclusion Criteria	Exclusion Criteria
Participants/Population (human studies or experimental model systems)	
• Humans • Non-human animals, including laboratory animal studies • *In silico* studies or *in vitro* models utilizing organs, tissues, cell lines, or cellular components	• Free living non-human organisms including wildlife, aquatic species, or plants
Exposure	
• Exposure to mountaintop removal mining activities including residential proximity or occupational exposure, environmental measures (e.g., air, water levels) • Exposure to mixtures collected from MTR mining areas in an experimental setting	• Exposure to single chemical components of MTR mining • Studies with unspecified type of mining conducted prior to widespread use of mountaintop removal mining or in geographic areas without mountaintop removal mining (added July 2016[c]) • Exposure to coal samples, dust or leachates *in vitro* (added July 2016[c])
Comparators	
• Vehicle-only treatment controls in experimental studies	• Case series of miners, descriptive without comparator (added July 2016c)
Outcomes	
• Human health-relevant outcomes, including measures of general well-being[a]	• Environmental impacts
Publications (e.g., language restrictions, use of conference abstracts)	
• Study must contain original data and must be peer-reviewed • Studies published in a language other than English will be translated for review	• Articles with no original data, e.g., editorials, reviews[b] • Non-peer reviewed articles: Conference presentations (clarified July 2016[c]) or other studies published in abstract form only, grant awards, and theses/dissertations • Retracted articles

[a] Exposure-only studies will provide critical context for the interpretation of the studies of human-health relevant outcomes and be summarized separately.
[b] Relevant reviews can be used as background and for reference scanning.
[c] Revised exclusion criteria added after screening when identifying studies for data extraction

Title/Abstract Review

Two members of the evaluation design team will independently conduct a title and abstract screen of the search results to determine whether a reference meets the inclusion criteria; studies that are not excluded based on the title and abstract will be screened through a full-text review. Initially, screeners will be trained using project-specific written instructions in a pilot phase undertaken to improve clarity of the inclusion and exclusion instructions and to improve accuracy and consistency among screeners. If changes to the inclusion criteria are made based on the pilot phase, they will be documented in a protocol amendment along with the date modifications were made and the logic for the changes.

Studies are not considered further when the title or abstract clearly indicate that the study does not meet the inclusion criteria. In case of screening conflicts, screeners will independently review their screening results to confirm the inclusion/exclusion decision and, if needed, discuss discrepancies with the other screener(s). Any articles with unresolved screening conflicts at the title and abstract phase will be included in the full text review.

Full-Text Review

After completion of the title/abstract screen, full-text articles are retrieved[3] for those studies that either clearly met the inclusion criteria or where eligibility to meet the inclusion criteria is unclear. Full-text review will be conducted by one member of the review team with a second member of the team confirming any exclusion determination of the first reviewer. True disagreements will be resolved by discussion involving another member(s) of the team or, if necessary, through consultation with technical advisors.

Multiple publications of same data

Multiple publications with overlapping data for the same study (e.g., publications reporting subgroups, additional outcomes or exposures outside the scope of an evaluation, or longer follow-up) are identified by examining author affiliations, study designs, cohort name, enrollment criteria, and enrollment dates. If necessary, study authors will be contacted to clarify any uncertainty about the independence of two or more articles. OHAT will include all publications on the study, select one study to use as the primary, and consider the others as secondary publications with annotation as being related to the primary record during data extraction. The primary study will generally be the publication with the longest follow-up, or for studies with equivalent follow-up periods, the study with the largest number of cases or the most recent publication date. OHAT will include relevant data from all publications of the study, although if the same outcome is reported in more than one report, OHAT will include a single instance of the data (and avoid more than one, i.e. duplicate instances of the data). Although only one study is identified as the primary study, relevant information will be considered from other publications. For example when a study refers to a previous publication for additional details of the methods, those citations will be identified and considered with the primary citation for data extraction and risk of bias evaluation.

Tracking study eligibility and reporting the flow of information

The reason for exclusion at the full-text-review stage will be annotated and reported in a study flow diagram in the final report. Studies will be tracked as eligible for (1) the SR of MTR exposure and health effects or (2) the exposure summary.

To be eligible for the systematic review of health effects the studies if they contain exposure and health effect data. At the full text stage studies, studies will be excluded if: (1) is a review, commentary, or

[3] OHAT will initially attempt to retrieve a full-text copy of the study using an automated program, such as QUOSA, when possible, and NIH library services (NIH subscriptions and interlibrary loans). For publications not available through NIH, OHAT will search the Internet and/or may attempt to contact the corresponding author. Studies not retrieved through these mechanisms are excluded and notated as “not available.”

editorial with no original data; (2) lacks relevant exposure information; (3) lacks relevant health outcome information; (4) is a conference abstract, thesis/dissertation, or (5) full text is "not available".

Studies associating MTR mining with chemical exposures (without health outcomes) will be captured in the literature search, separated and summarized in a critical concurrent step of the evaluation process. These studies are likely to have been identified in the full text screening as containing relevant exposure information, but not including human subject, animal model, or *in vitro* systems.

MTR Mining Exposure Data Only Study Summarization

Residents of communities located near MTR mining activities may be exposed to a heterogeneous mixture of chemicals and particulate matter in air, water, or soil that are attributable to this source. Studies that measure and characterize these mixed exposures are critical to the assessment of potential human health effects. Existing authoritative reviews on individual substances may help prioritize future research efforts of possible impacts of MTR mining on the health of people in nearby communities.

A comprehensive and formal exposure assessment of all the potential health effects of all the components of the mixture associated with MTR mining is beyond the scope of this review. The objective of the section is to succinctly summarize the relevant exposure information. When available, exposure information from secondary sources (reviews) will supplement primary studies that provide key information. The exposure section may consist of subsections on the topics listed below, although the organization may change depending on the available database (NTP 2015b).

Key Topics

Substance identification and properties

- Defines the substance(s) relevant to MTR and provides information on chemical and physical or biological properties

Exposure scenarios and biological indices of exposure

- Provides information on present or past pathways of exposure including levels in air, water, and soil.
- Provides information related to interpreting biological indices.
- Provides data on levels of the substance (or metabolite when relevant) in human tissues or biofluids.
- Provides information on estimated levels from various environmental, occupational, or other sources.

Synthesis of information

- Summarizes what is known regarding the major components of the exposure.
- Discusses whether exposure sources, routes, levels, or patterns have changed over time and geographical space.

Regulations and guidelines

- Lists pertinent regulations from the U.S. regulatory agencies, such as the FDA, Department of Agriculture, Department of Transportation, EPA, Occupational Safety and Health

Administration, Department of Interior (Fish & Wildlife Service and/or Office of Surface Mining), US Army Corps of Engineers, or Mine Safety and Health Administration.
- Lists occupational exposure guidelines (if relevant), such as those published by ACGIH and NIOSH.

Data Extraction

Data Extraction Process and Data Warehousing

Data extraction will be managed with structured forms and stored in a database format using Health Assessment Workspace Collaborative (HAWC, https://hawcproject.org/), an open source, web-based interface.[4] Data extraction elements are listed in appendices for human (**Appendix 2**), experimental animal (**Appendix 3**), and *in vitro* studies (**Appendix 4**). Study information collected during data extraction will be visualized, when appropriate (e.g., when there are data on the same or health effects evaluated across multiple studies), and made publicly available upon publication of the finalized report.

The extracted data will be used to summarize study designs and findings, facilitate assessment of risk of bias, and/or conduct statistical analyses. The content of the data extraction may be revised following the identification of the studies included in the review. Data extraction will be performed by one member of the evaluation team and checked by a second member of the evaluation team for completeness and accuracy. Data extractors from the evaluation team will be trained using project-specific written instructions in an initial pilot phase using a subset of studies. Any discrepancies in data extraction will be resolved by discussion or consultation with a third member of the evaluation team. Information that is inferred, converted, or estimated during data extraction will be annotated (e.g., using brackets [n=10]).

OHAT will attempt to contact authors of included studies to obtain missing data considered important for evaluating key study findings (e.g., level of data required to conduct a meta-analysis). The evaluation report will note that an attempt to contact study authors was unsuccessful if study researchers do not respond to an email or phone request within one month of the attempt to contact.

[4] HAWC (Health Assessment Workspace Collaborative): A Modular Web-based Interface to Facilitate Development of Human Health Assessments of Chemicals (http://hawcproject.org).

Quality Assessment of Individual Studies

Internal validity or risk of bias will be assessed for individual studies using a tool developed by OHAT that outlines a parallel approach to evaluating risk of bias from human and non-human animal studies to facilitate consideration of risk of bias across evidence streams with common terms and categories. The handbook for conducting systematic reviews presents OHAT's tool for assessing study quality, or "risk of bias." The risk-of-bias domains and questions for experimental animal studies are based on established guidance for experimental human studies (randomized clinical trials). Detailed instructions for response are provided in the OHAT tool. Briefly, the risk-of-bias tool is comprised of a common set of 11 questions that are answered based on the specific details of individual studies to develop risk-of-bias ratings (using the four options in **Table 3**) for each question. Study design determines the subset of questions that should be used to assess risk of bias for an individual study (**Table 4**). For example, the subset of risk-of-bias questions applicable to all of the experimental study designs includes a question on randomization of exposure that would not be applicable to observational study designs. Therefore, a similar set of questions are used across experimental study designs (experimental animal and human controlled trials). These categorical ratings facilitate comparison of relative strengths and weaknesses of individual studies' design and conduct and are not intended to replace quantitative consideration of the potential biases.

Table 3. Answers to the Risk-of-Bias Questions Result in One of Four Risk-of-Bias Ratings

++	***Definitely Low*** *risk of bias*: There is direct evidence of low risk-of-bias practices
+	***Probably Low*** *risk of bias*: There is indirect evidence of low risk-of-bias practices **OR** it is deemed that deviations from low risk-of-bias practices for these criteria during the study would not appreciably bias results, including consideration of direction and magnitude of bias
– N	***Probably High*** *risk of bias*: There is indirect evidence of high risk–of-bias practices (indicated with "-") **OR** there is insufficient information provided about relevant risk-of-bias practices (indicated with "NR" for not reported). Both symbols indicate probably high risk of bias.
––	***Definitely High*** *risk of bias*: There is direct evidence of high risk-of-bias practices

Table 4. OHAT Risk of Bias Questions and Applicability by Study Design

Risk of Bias Questions	Experimental Animal*	Human Controlled Trials**	Cohort	Case-Control	Cross-Sectional***	Case Series
1. Was administered dose or exposure level adequately randomized?	X	X				
2. Was allocation to study groups adequately concealed?	X	X				
3. Did selection of study participants result in the appropriate comparison groups?			X	X	X	
4. Did study design or analysis account for important confounding and modifying variables?			X	X	X	X
5. Were experimental conditions identical across study groups?	X					
6. Were research personnel blinded to the study group during the study?	X	X				
7. Were outcome data complete without attrition or exclusion from analysis?	X	X	X	X	X	
8. Can we be confident in the exposure characterization?	X	X	X	X	X	X
9. Can we be confident in the outcome assessment (including blinding of outcome assessors)?	X	X	X	X	X	X
10. Were all measured outcomes reported?	X	X	X	X	X	X
11. Were there no other potential threats to internal validity?	X	X	X	X	X	X

*Experimental animal studies are controlled exposure studies. Non-human animal observational studies can be evaluated using the design features of observational human studies such as cross-sectional study design.

**Human Controlled Trials are studies in humans with controlled exposure (e.g., Randomized Controlled Trials, non-randomized experimental studies)

***Cross-sectional studies include population surveys with individual data (e.g., NHANES) and surveys with aggregate data (i.e., ecological studies).

Studies are independently assessed by two assessors who answer all applicable risk-of-bias questions with one of four options in **Table 3** (Sterne *et al.* 2014) following pre-specified criteria detailed in **Appendix 5**. The criteria describe aspects of study design, conduct, and reporting required to reach risk-of-bias ratings for each question and specify factors that can distinguish among ratings (e.g., what separates "definitely low" from "probably low" risk of bias). The instructions and detailed criteria are tailored to the specific evidence stream and type of human study designs. Risk of bias will be assessed at the outcome level because study design or method specifics may increase the risk of bias for some outcomes and not others within the same study.

Risk-of-Bias Assessment Process

Assessors will be trained using the criteria in **Appendix 5** with an initial pilot phase undertaken to improve clarity of criteria that distinguish between adjacent ratings and to improve consistency among assessors. All team members involved in the risk-of-bias assessment will be trained on the same set of studies and asked to identify potential ambiguities in the criteria used to assign ratings for each question. Any ambiguities and rating conflicts will be discussed relative to opportunities to refine the criteria to more clearly distinguish between adjacent ratings. If major changes to the risk-of-bias criteria are made based on the pilot phase (i.e., those that would likely result in revision of response), they will be documented in a protocol amendment along with the date modifications were made and the logic for the changes. It is also expected that information about confounding, exposure characterization, outcome assessment, and other important issues may be identified during or after data extraction, which can lead to further refinement of the risk-of-bias criteria (Sterne *et al.* 2014).

After assessors have independently made risk-of-bias determinations for a study across all risk-of-bias questions, the two assessors will compare their results to identify discrepancies and attempt to resolve them. Any remaining discrepancies will be considered by the project lead and, if needed, other members of the evaluation design team and/or technical advisors. The final risk-of-bias rating for each question will be recorded along with a statement of the basis for that rating. The risk-of-bias assessment of included studies will be part of the study summaries released in materials for the draft OHAT monograph that will be posted for public comment prior to peer review. Peer review will provide an opportunity for investigators and the public to comment on risk-of-bias.

Missing Information for Risk-of-Bias Assessment

OHAT will attempt to contact authors of included studies by email to obtain missing information considered critical for evaluating risk of bias that cannot be inferred from the study. If additional information or data are received from study authors, risk-of-bias judgments will be modified to reflect the updated study information. If OHAT does not receive a response from the authors by one month of the contact attempt, a risk-of-bias response of "NR" for "not reported; probably high risk of bias" will be used and a note made in the data extraction files that an attempt to contact the authors was unsuccessful.

Organizing and Rating Confidence in Bodies of Evidence

OHAT will consider the collection of studies on the same or closely related outcomes as bodies of evidence and develop overall confidence ratings in these bodies of evidence using a modification of the

GRADE framework (Rooney *et al.* 2014). Considerations for considering quantitative or narrative synthesis and developing confidence ratings for this evaluation are described below.

Considerations for Pursuing a Narrative or Quantitative Evidence Synthesis

Heterogeneity within the available evidence will determine the type of evidence integration that is appropriate: either a quantitative synthesis (meta-analysis) or narrative approach for evidence integration. Where appropriate we will perform a meta-analysis. Summaries of main characteristics for each included study will be compiled and reviewed by two reviewers to determine comparability between studies, identify data transformations necessary to ensure comparability, and determine whether heterogeneity is a concern. The main characteristics considered across all eligible studies include the following:

Human Studies

- Study design (e.g., cross-sectional, cohort)
- Details on how participants were classified into exposure groups (e.g., quartiles of exposure)
- Details on source of exposure data (e.g., questionnaire, area monitoring, biomonitoring)
- Health outcome(s) reported
- Conditioning variables in the analysis (e.g., variables considered confounders)
- Type of data (e.g., continuous or dichotomous), statistics presented in paper
- Variation in degree of risk of bias at individual study level

Animal Studies

- Experimental design (e.g., acute, chronic, multigenerational)
- Animal model used (e.g., species, strain, sex, genetic background)
- Age of animals (e.g., at start of treatment, mating, and/or pregnancy status)
- Developmental stage of animals at treatment and outcome assessment
- Dose levels, frequency of treatment, timing, duration, and exposure route
- Health outcome(s) reported
- Type of data (e.g., continuous or dichotomous), statistics presented in paper
- Variation in degree of risk of bias at individual study level

More detailed guidance on evaluating heterogeneity, transforming or normalizing data to ensure comparability, and the process for determining whether a meta-analysis will be pursued is provided in the OHAT Handbook for Conducting a Literature-Based Health Assessment (see Step 5) (NTP 2015a). We expect to require input from topic-specific experts to help assess whether studies are too heterogeneous for evidence integration steps to be appropriate. Situations where it may not be appropriate to include a study are (1) data on exposure or outcome are too different to be combined, (2) there are concerns about high risk of bias, or (3) other circumstances may indicate that averaging study results would not produce meaningful results. When it is inappropriate or not feasible to proceed with the subsequent steps of the review, OHAT will narratively describe or visually present findings.

Stratified Analyses, Meta-Regression, and Publication Bias

If there is significant study-level heterogeneity, then OHAT may conduct stratified analyses or multivariate meta-regression in an attempt to determine how much heterogeneity can be explained by

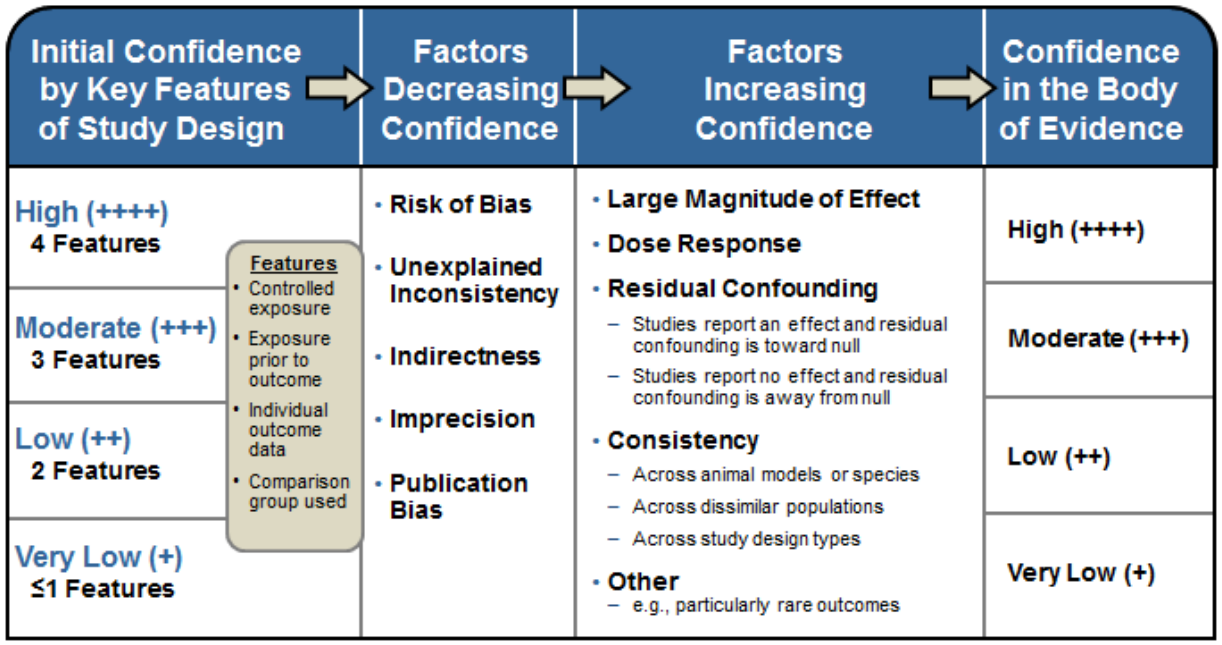
taking into account both within- and between-study variance (Vesterinen *et al.* 2014). Multivariate meta-regression approaches are especially useful for assessing the significance of associations between study design characteristics. These approaches are considered most suitable if there are at least six to ten studies for a continuous variable and at least four studies for a categorical variable (Fu *et al.* 2011). If possible (i.e., if there are enough studies) we will assess potential publication bias by developing funnels and performing Egger regression on the estimates of effect size. In addition, if these methods suggest that publication bias is present, we will use trim and fill methods to predict the impact of the hypothetical "missing" studies (Vesterinen *et al.* 2014).

Confidence Rating: Assessment of Body of Evidence

The quality of evidence for each outcome will be graded using the GRADE system for rating the confidence in the body of evidence (Guyatt *et al.* 2011) as used within the OHAT Approach to Systematic Review and Evidence Integration (Rooney *et al.* 2014). More detailed guidance on reaching confidence ratings in the body of evidence as "high," "moderate," "low," or "very low" is provided in the OHAT Handbook for Conducting a Literature-Based Health Assessment (see Step 5) (NTP 2015a). In brief, available studies on a particular outcome are initially grouped by key study design features, and each grouping of studies is given an initial confidence rating by those features. This initial rating (column 1 of **Figure 2**) is downgraded for factors that decrease confidence in the results (column 2 of **Figure 2** [risk of bias, unexplained inconsistency, indirectness or lack of applicability, imprecision, and publication bias]) and upgraded for factors that increase confidence in the results (column 3 of **Figure 2** [large magnitude of effect, dose response, consistency across study designs/populations/animal models or species, consideration of residual confounding, and other factors that increase our confidence in the association or effect]).

The reasons for downgrading (or upgrading) confidence may not be due to a single domain of the body of evidence. If a decision to downgrade is borderline for two domains, the body of evidence is downgraded once in a single domain to account for both partial concerns based on considering the key drivers of the strengths or weaknesses. Similarly, the body of evidence is not downgraded twice for what is essentially the same limitation (or upgraded twice for the same asset) that could be considered applicable to more than one domain of the body of evidence. Consideration of consistency across study designs, human populations, or animal species is not included in the GRADE guidance (Guyatt *et al.* 2011); however, it is considered in the modified version of GRADE used by OHAT (Rooney *et al.* 2014).

Figure 2. Assessing Confidence in the Body of Evidence



Confidence ratings are independently assessed by federal staff on the evaluation review team, and discrepancies are resolved by consensus and consultation with technical advisors as needed. Confidence ratings are summarized in evidence profile tables.

Relevance of Animal Models to Human Health

- Rats, mice, and other mammalian model systems: No limitations of model systems for mammals have been identified a priori. Thus, studies conducted in mammalian model systems will be assumed to be relevant for humans (i.e., not downgraded for indirectness) unless compelling data to the contrary is identified during the course of the evaluation.
- Birds, reptiles, amphibians, fish, and other non-mammalian vertebrate model systems: use of these model systems to address human health is not as well-established as use of the mammalian model systems (WHO 2012). For this reason, studies conducted in non-mammalian vertebrates will be downgraded one level for indirectness.
- Invertebrate model systems: Due to a large phylogenetic difference, studies conducted in invertebrates will be downgraded two levels for indirectness.

Health Outcomes

For the evaluation of mountaintop removal mining on community health no outcome area has been specified, so no primary and secondary outcomes will be specified a priori. Once relevant literature is collected, we plan to use input from topic-specific experts to help group outcomes and may designate primary and secondary outcomes for biologically related health effects. Primary outcomes or endpoints are those considered to be the most direct indicators of a health effect and secondary outcomes would

be indirect measures or upstream indicators of a health effect and therefore downgraded one level for indirectness.

Exposure

- Human studies: All exposure levels and scenarios encountered in the human studies (e.g., general population, occupational settings, etc.) will be considered direct and not downgraded.
- Dose levels used in animal studies: There will be no downgrade for dose level used in experimental animal studies. We recognize that the level of dose or exposure is an important factor when considering the relevance of animal findings to human health. The relevance of the dose levels will be considered in the report. In addition, in OHAT's process the relevance of the dose or exposure level occurs after hazard identification as part of reaching a "level of concern" conclusion.
- Route of administration in animal studies: All of the most commonly used routes of administration will be considered direct for the purposes of establishing confidence ratings. We recognize that some of these exposure routes may only be relevant for certain human sub-populations. However, in OHAT's process this consideration occurs after hazard identification as part of reaching a "level of concern" conclusion.

    - Oral (no downgrade for indirectness) – Gavage, drinking water, or feeding studies are considered relevant because oral exposure through drinking water is a possible sources of exposure to to MTR mining component in humans.

    - Dermal (no downgrade for indirectness) – Dermal exposure is considered relevant for contact with surface waters, soil, dusts, soil, and direct contact of skin.

    - Inhalation (no downgrade for indirectness) – Inhalation studies are considered relevant because MTR mining components are found in air and relevant to both community and occupational cohorts.

    - Intraperitoneal or subcutaneous injection (one level downgrade for indirectness) – These studies will be downgraded one level because they are not relevant to the nature of human exposure.

Mechanistic Studies

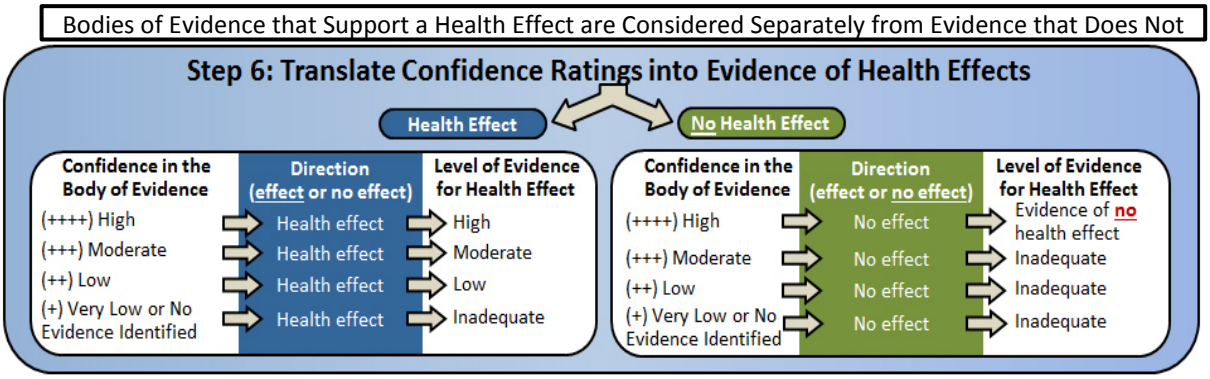
The framework described above only applies to human and animal studies. There is no analogous model to develop confidence ratings for other relevant data such as outcomes from *in vitro*, mechanistic, cellular or genomic studies. Thus our current approach for considering the level of support provided by other relevant data including mechanistic studies is described separately in a later section of this the document when integrating other relevant data (see "Consideration of Mechanistic Data").

Preparation of Draft Level of Evidence Statement

The confidence ratings will be translated into draft level of evidence of health effects for each type of health outcome separately according to one of four statements: (1) High, (2) Moderate, (3) Low, or (4) Inadequate (**Figure 3**). The descriptor "evidence of no health effect" is used to indicate confidence that the substance is not associated with a health effect. Because of the inherent difficulty in proving a

negative, the conclusion "evidence of no health effect" is only reached when there is high confidence in the body of evidence.

Figure 3. Translate Confidence Ratings into Evidence of Health Effect Conclusions



Evidence Descriptors	Definition
High Level of Evidence	There is high confidence in the body of evidence for an association between exposure to MTR mining and the health outcome(s).
Moderate Level of Evidence	There is moderate confidence in the body of evidence for an association between exposure to MTR mining and the health outcome(s).
Low Level of Evidence	There is low confidence in the body of evidence for an association between exposure to MTR mining and the health outcome(s), or no data are available.
Inadequate Evidence	There is insufficient evidence available to assess if exposure to MTR mining is associated with the health outcome(s).
Evidence of No Health Effect	There is high confidence in the body of evidence that exposure to MTR mining is not associated with the health outcome(s).

Integrate Evidence to Develop Hazard Identification Conclusions

Finally, the levels of evidence ratings for human and animal data will be integrated with consideration of mechanistic data to reach one of five possible hazard identification categories: (1) Known, (2) Presumed, (3) Suspected, (4) Not classifiable, or (5) Not identified to be a hazard to humans (**Figure 4**).

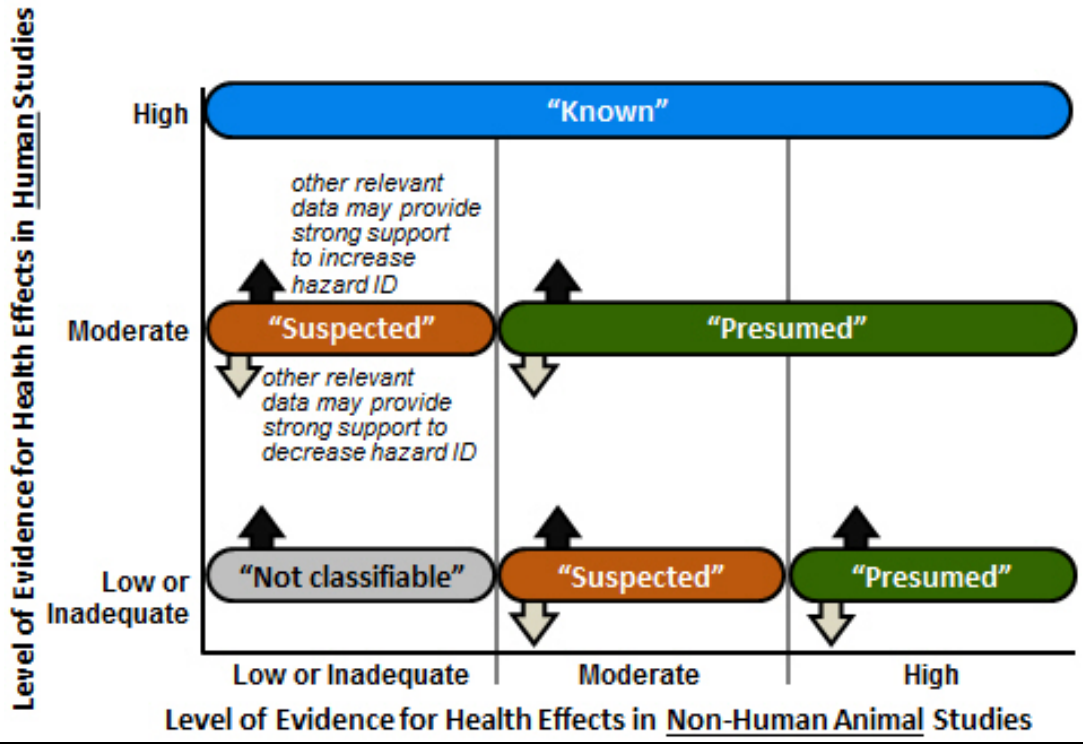
Consideration of Human and Animal Data

Initial hazard identification conclusions will be reached by integrating the highest level-of-evidence conclusion for an effect(s) on an outcome basis for the human and the animal evidence streams. Hazard identification conclusions may be reached on groups of biologically related outcomes as well as more specific endpoints if data are available to make more specific conclusions. If the data support a health

effect, the level-of-evidence conclusion for human data for that health outcome from the previous step will be considered together with the level of evidence for non-human animal data to reach one of four initial hazard identification conclusions: Known, Presumed, Suspected, or Not classifiable. If either the human or animal evidence stream is characterized as "Inadequate Evidence," then conclusions are based on the remaining evidence stream alone (which is equivalent to treating the missing evidence stream as "Low" in **Figure 4**).

If the human level of evidence rating of "Evidence of no health effect" from the previous step is supported by a similar level of evidence rating for animal evidence for no health effect, the hazard identification conclusion would be "Not identified to be a hazard to humans."

Figure 4. Hazard Identification Scheme



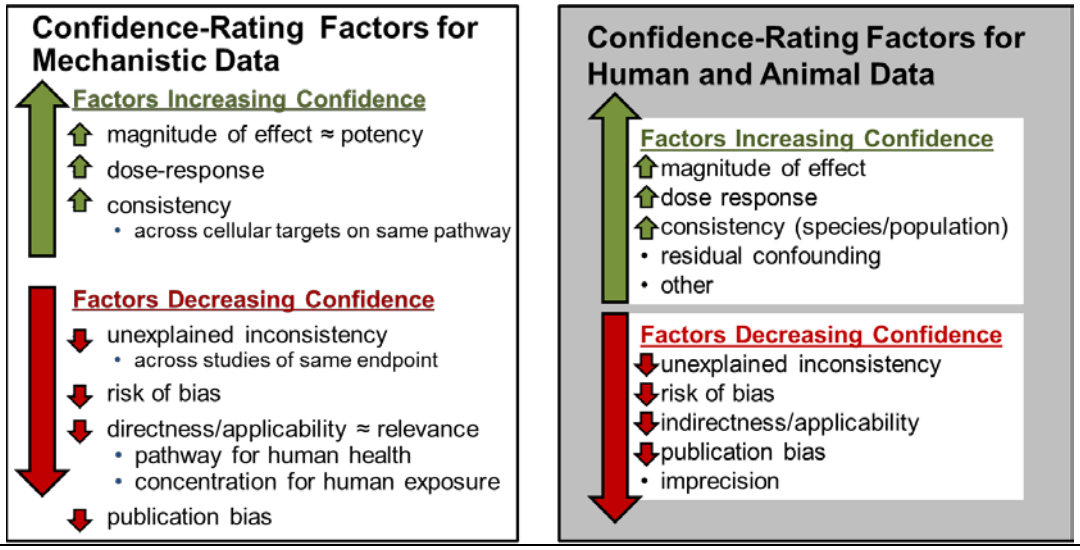
Consideration of Mechanistic Data

The NTP does not require mechanistic or mode-of-action data in order to reach hazard identification conclusions, although when available, this and other relevant supporting types of evidence may be used to raise (or lower) the category of the hazard identification conclusion. Mechanistic data can come from a wide variety of studies that are not intended to identify a disease phenotype. This source of experimental data includes *in vitro* and in vivo laboratory studies directed at cellular, biochemical, and molecular mechanisms that explain how a chemical produces particular adverse effects.

The strength of the support or opposition presented by the other relevant data is evaluated using the guidance presented in **Figure 5**. The factors outlined for increasing or decreasing confidence in that the mechanistic data support biological plausibility are conceptually similar to those used to rate confidence

in bodies of evidence for human or animal in vivo studies. Evaluations of the strength of evidence provided by mechanistic data are made on an outcome-specific basis based on discussion by the evaluation team and consultation with technical advisors as needed.

Figure 5. Factors Considered in Evaluating the Support for Biological Plausibility



- If mechanistic data provide strong support for biological plausibility of the relationship between exposure and the health effect, the hazard identification conclusion may be upgraded (indicated by black "up" arrows in **Figure 4**) the that initially derived by considering the human and non-human animal evidence together.

- If mechanistic data provide strong opposition for biological plausibility of the relationship between exposure and the health effect, the hazard identification conclusion may be downgraded (indicated by gray "down" arrows in **Figure 4**) from that initially derived by considering the human and non-human animal evidence together.

Although it is envisioned that strong evidence for a relevant process from mechanistic data alone could indicate a greater potential that the substance is an hazard to humans, for this evaluation the mechanistic data will only be considered to inform the biological plausibility of observed outcomes from in vivo data.

NTP MONOGRAPH FORMAT

The NTP Monograph on the association between MTR mining exposure and community health effects will include the following information:

Introduction

This section will provide a brief background on the topic.

Methodology

This section will provide a brief overview of the methodologies used in the review process, including:

- the research question
- the search strategy used to identify and retrieve studies
- the process for selecting the included studies
- the methods of data extraction
- the methods used to assess risk of bias of included studies

If conducted, dependent on the extent and nature of the available evidence (i.e., number and similarity of studies):

- the methods used to synthesize the data of included studies
- the methods used to evaluate confidence in the body of evidence
- the methods used to reach hazard identification conclusions

Results

This section will include the results from the systematic review of the evidence for an association between exposure to MTR mining and community health effects. Results will be presented in tables or figures as appropriate using HAWC. The results from the included studies will be discussed by outcome. This will include a description of:

- the number of studies identified that reported the outcome
- the full list of excluded studies, with reasons for exclusion documented for studies excluded at the full text review stage
- the results and risk-of-bias assessment for each included study (including files in downloadable format)
- the summary of MTR mining exposure data only studies and relevant secondary sources (reviews) of health effects of components of the MTR mining mixture

If conducted:

- description of results and ratings for confidence in the bodies of evidence for major outcomes for which there are MTR mining data using the OHAT adaption of GRADE
- evidence profiles for major outcomes for which there are MTR mining data
- presentation of level of evidence and draft hazard identification conclusions for major outcomes for which there are MTR mining data

Discussion

The discussion will provide a summary of the review findings, including:

- Discuss limitations of the systematic review
- Describe limitations of the evidence base
- Identify data gaps and key research needs
- Discuss findings in the context of human exposure scenarios

Conclusion

This will present the conclusion of the review.

REFERENCES

Acton PM, Fox JF, Campbell JE, Jones AL, Rowe H, Martin D, Bryson S. 2011. Role of soil health in maintaining environmental sustainability of surface coal mining. *Environmental science & technology* 45(23): 10265-10272.

Fu R, Gartlehner G, Grant M, Shamliyan T, Sedrakyan A, Wilt TJ, Griffith L, Oremus M, Raina P, Ismaila A, Santaguida P, Lau J, Trikalinos TA. 2011. Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. *Journal of clinical epidemiology* 64(11): 1187-1197.

Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, DeBeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schunemann HJ. 2011. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *Journal of clinical epidemiology* 64(4): 383-394.

Higgins J, Green S. 2011. Cochrane Handbook for Systematic Reviews of Interventions, The Cochrane Collaboration.

Holzman DC. 2011. Mountaintop removal mining: digging into community health concerns. *Environmental health perspectives* 119(11): A476-483.

NTP (National Toxicology Program). 2015a. *Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration - January 9 2015*. Durham, NC: Office of Health Assessment and Translation. Available: http://ntp.niehs.nih.gov/go/38673 [accessed 25 Jan 2015].

NTP (National Toxicology Program). 2015b. *Handbook for Preparing Report on Carcinogens Monographs* Durham, NC: Office of the Report on Carcinogans. Available: http://ntp.niehs.nih.gov/ntp/roc/handbook/roc_handbook_508.pdf [accessed 10 Feb 2016].

Palmer MA, Bernhardt ES, Schlesinger WH, Eshleman KN, Foufoula-Georgiou E, Hendryx MS, Lemly AD, Likens GE, Loucks OL, Power ME, White PS, Wilcock PR. 2010. Science and regulation. Mountaintop mining consequences. *Science* 327(5962): 148-149.

Rooney AA, Boyles AL, Wolfe MS, Bucher JR, Thayer KA. 2014. Systematic Review and Evidence Integration for Literature-Based Environmental Health Science Assessments. *Environmental health perspectives* 122(7): 711-718.

Shamliyan T, Kane RL, Dickinson S. 2010. A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. *J Clin Epidemiol* 63(10): 1061-1070.

Simmons JA, Currie WS, Eshleman KN, Kuers K, Monteleone S, Negley TL, Pohlad BR, Thomas CL. 2008. Forest to reclaimed mine land use change leads to altered ecosystem structure and function. *Ecological applications : a publication of the Ecological Society of America* 18(1): 104-118.

Sterne JAC, Higgins JPT, Reeves BC. 2014. A Cochrane Risk Of Bias Assessment Tool for Non-Randomized Studies of Interventions (ACROBAT-NRSI). ACROBAT-NRSI obotdgf.

Vesterinen HM, Sena ES, Egan KJ, Hirst TC, Churolov L, Currie GL, Antonic A, Howells DW, Macleod MR. 2014. Meta-analysis of data from animal studies: a practical guide. *J Neurosci Methods* 221: 92-102.

ABOUT THE PROTOCOL

Contributors

Evaluation teams are composed of federal staff and contractor staff. Contractor staff members are screened for potential conflicts of interest. Federal staff members should do a self-evaluation. Technical advisors were screened for conflict of interest prior to their service and did not report any conflicts of interest. Epidemiologists and toxicologists on OHAT evaluation teams should have at least three years' experience and/or training in reviewing studies, including summarizing studies and critical review (e.g., assessing study quality and interpreting findings). Team members should have at least a master's degree or equivalent experience in epidemiology, toxicology, environmental health sciences, or a related field.

Federal Staff

Name	Affiliation
Abee L. Boyles	NTP/NIEHS, Project Lead
Frank Bove	ATSDR, CDC
Kacee Deener	US EPA
Lora Werner	ATSDR, CDC

Contract Support Staff: Will assist in screening, data extraction and risk of bias assessment

Name	Affiliation
Robyn Blain	ICF International
Susan Goldhaber	ICF International
Cara Henning	ICF International
Sofie McComb	ICF International
Johanna Rochester	ICF International
Courtney Skuce	ICF International

Technical Advisors

Technical advisors are outside experts retained on an as-needed basis to provide individual advice to the NTP for a specific topic. The technical advisors were selected for their experience with MTR mining toxicology and environmental health.

Name	Affiliation
Michael McCawley, PhD	School of Public Health, West Virginia University
Jerome A. Paulson, MD, FAAP	George Washington University Milken Institute School of Public Health

Sources of Support

National Institute of Environmental Health Sciences/Division of the National Toxicology Program

Protocol History and Revisions

Date	Activity or revision
Feb. 12, 2016:	Draft Protocol reviewed: sent to peer reviewers for comment/review
April 1, 2016:	Protocol posted publically at: http://ntp.niehs.nih.gov/go/780611

Date	Activity or revision
April 1, 2016:	Protocol registered with PROSPERO: http://www.crd.york.ac.uk/PROSPERO/
July 11, 2016:	Revised protocol sent to reviewers with inclusion/exclusion reference lists
July 27, 2016:	Revised protocol posted publically at: http://ntp.niehs.nih.gov/go/780611
July 27, 2016:	List of included and excluded studies posted publically at: http://ntp.niehs.nih.gov/go/780611
July 27, 2016:	Revised protocol updated in PROSPERO: http://www.crd.york.ac.uk/PROSPERO/

APPENDICES

Appendix 1. Mountaintop Removal Mining Search Terms

Database	Mountaintop Mining	Appalachia Coal Mining
Embase	(mountaintop OR 'mountain top'):ab,ti AND (anthracite OR bituminous OR coal OR mine OR mines OR mining OR removal):ab,ti	((Appalachian Region OR Appalachia* OR Kentucky OR Ohio OR Pennsylvania OR Tennessee OR Virginia OR "West Virginia"):ti,ab AND coal mining/exp) OR ((Appalachia*:ti,ab OR Kentucky:ti,ab OR Ohio:ti,ab OR Pennsylvania:ti,ab OR Tennessee:ti,ab OR Virginia:ti,ab OR "West Virginia":ti,ab) AND (anthracite:ti,ab OR bituminous:ti,ab OR coal/exp OR coal:ti,ab) AND (mining/exp OR mine:ti,ab OR mines:ti,ab OR mining:ti,ab))
PsycINFO	(mountaintop OR "mountain top") AND (anthracite OR bituminous OR coal OR mine OR mines OR mining OR removal)	(Appalachia* OR Kentucky OR Ohio OR Pennsylvania OR Tennessee OR Virginia OR "West Virginia") AND (anthracite OR bituminous OR coal) AND (mine OR mines OR mining)
PubMed	((mountaintop OR "mountain top") AND (anthracite OR bituminous OR coal OR mine OR mines OR mining OR removal))	((Appalachian Region[mh] OR Appalachia*[tiab] OR Kentucky[tiab] OR Ohio[tiab] OR Pennsylvania[tiab] OR Tennessee[tiab] OR Virginia[tiab] OR "West Virginia"[tiab]) AND coal mining[mh]) OR ((Appalachian Region[mh] OR Appalachia*[tiab] OR Kentucky[tiab] OR Ohio[tiab] OR Pennsylvania[tiab] OR Tennessee[tiab] OR Virginia[tiab] OR "West Virginia"[tiab]) AND (anthracite[tiab] OR bituminous[tiab] OR coal[mh] OR coal[tiab]) AND (mining[mh] OR mine[tiab] OR mines[tiab] OR mining[tiab]))
Scopus	TITLE-ABS((mountaintop OR "mountain top") W/6 (anthracite OR bituminous OR coal OR mine OR mines OR mining OR removal))	TITLE-ABS((Appalachia* OR Kentucky OR Ohio OR Pennsylvania OR Tennessee OR Virginia OR "West Virginia") AND (anthracite OR bituminous OR coal) AND (mine OR mines OR mining))
Toxline	(mountaintop OR "mountain top") AND (anthracite OR bituminous OR coal OR mine OR mines OR mining OR removal)	((Appalachia* OR Kentucky OR Ohio OR Pennsylvania OR Tennessee OR Virginia OR "West Virginia") AND (anthracite OR bituminous OR coal) AND (mine OR mines OR mining))
Web of Science	TS=((mountaintop OR "mountain top") NEAR/6 (anthracite OR bituminous OR coal OR mine OR mines OR mining OR removal))	TS= ((Appalachia* OR Kentucky OR Ohio OR Pennsylvania OR Tennessee OR Virginia OR "West Virginia") AND (anthracite OR bituminous OR coal) AND (mine OR mines OR mining))

- "Mountaintop" terms identified studies that specified mountaintop removal within the citation. Because MTR mining is currently the predominant form of coal mining in the Appalachian Region of the US, the addition of "Appalachia" terms was necessary to capture studies that did not specify mountaintop mining.
- Terms for other "open" mining practices (e.g. strip mining) were also considered to identify potentially relevant studies, but these results predominantly included coal mining in regions

where MTR mining is not allowed and mining of things other than coal. Due to a significant increase in the number of references retrieved and limited utility of the majority of these studies, these terms are not included in the proposed literature search strategy.

- When the systematic review is conducted, the reference lists of included studies and relevant reviews will be searched for additional relevant publications.
- The list of included (and excluded) studies will also be posted on the OHAT website prior to release of a draft report as an additional strategy to identify potentially relevant studies that may have been missed during the literature search

Appendix 2. Data Extraction Elements for Human Studies

HUMAN	
Funding	Funding source(s)
	Reporting of conflict of interest (COI) by authors (*reporting bias)
Subjects	Study population name/description
	Dates of study and sampling time frame
	Geography (country, region, state, etc.)
	Demographics (sex, race/ethnicity, age or lifestage at exposure and at outcome assessment)
	Number of subjects (target, enrolled, n per group in analysis, and participation/follow-up rates) (*missing data bias)
	Inclusion/exclusion criteria/recruitment strategy (*selection bias)
	Description of reference group (*selection bias)
Methods	Study design (e.g., prospective or retrospective cohort, nested case-control study, cross-sectional, population-based case-control study, intervention, case report, etc.)
	Health outcome, e.g., blood pressure (*reporting bias)
	Diagnostic or methods used to measure health outcome (*information bias)
	Confounders or modifying factors and how considered in analysis (e.g., included in final model, considered for inclusion but determined not needed (*confounding bias)
	Exposure description (treatment, self-report, supplement or fortified food)) (*information bias)
	Methodological details for exposure assessment (e.g., questionnaire used, validation, and definition of variables or assumptions on level e.g. most prenatal vitamins contained X dosage) (*information bias)
	Statistical methods (*information bias)
Results	Description of Exposure levels (e.g., mean, median, measures of variance as presented in paper, such as SD, SEM, 75th/90th/95th percentile, minimum/maximum); range of exposure levels, or number of exposed cases
	Statistical findings (e.g., adjusted β, standardized mean difference, adjusted odds ratio, standardized mortality ratio, relative risk, etc.) or description of qualitative results.
Other	Documentation of author queries, use of digital rulers to estimate data values from figures, exposure unit, and statistical result conversions, etc.

Appendix 3. Data Extraction Elements for Animal Studies

ANIMAL	
Funding	Funding source(s)
	Reporting of COI by authors (*reporting bias)
Animal Model	Sex
	Species
	Strain
	Source of animals
	Age or lifestage at start of dosing and at health outcome assessment
	Diet and husbandry information (e.g., diet name/source)
Treatment	Chemical name and CAS number
	Source of chemical
	Purity of chemical (*information bias)
	Dose levels or concentration (as presented and converted to mg/kg bw/d when possible)
	Other dose-related details, such as whether administered dose level was verified by measurement, information on internal dosimetry (*information bias)
	Vehicle used for exposed animals
	Route of administration (e.g., oral, inhalation, dermal, injection)
	Duration and frequency of dosing (e.g., hours, days, weeks when administration was ended, days per week)
Methods	Study design (e.g., single treatment, acute, subchronic (e.g., 90 days in a rodent), chronic, multigenerational, developmental, other)
	Guideline compliance (i.e., use of EPA, OECD, NTP or another guideline for study design, conducted under GLP guideline conditions, non-GLP but consistent with guideline study, non-guideline peer-reviewed publication)
	Number of animals per group (and dams per group in developmental studies) (*missing data bias)
	Randomization procedure, allocation concealment, blinding during outcome assessment (*selection bias)
	Method to control for litter effects in developmental studies (*information bias)
	Use of negative controls and whether controls were untreated, vehicle-treated, or both
	Report on data from positive controls – was expected response observed? (*information bias)
	Endpoint health category (e.g., reproductive)
	Endpoint (e.g., infertility)
	Diagnostic or method to measure endpoint (*information bias)
	Statistical methods (*information bias)

ANIMAL	
Results	Measures of effect at each dose or concentration level (e.g., mean, median, frequency, and measures of precision or variance) or description of qualitative results. When possible, OHAT will convert measures of effect to a common metric with associated 95% confidence intervals (CI). Most often, measures of effect for continuous data will be expressed as mean difference, standardized mean difference, and percent control response. Categorical data will be expressed as relative risk (RR, also called risk ratio).
	No Observed Effect Level (NOEL), Lowest Observed Effect Level (LOEL), benchmark dose (BMD) analysis, statistical significance of other dose levels, or other estimates of effect presented in paper. **Note:** The NOEL and LOEL are highly influenced by study design, do not give any quantitative information about the relationship between dose and response, and can be subject to author's interpretation (e.g., a statistically significant effect may not be considered biologically important). Also, a NOEL does not necessarily mean zero response. Ideally, the response rate at specific dose levels is used as the primary measure to characterize the response.
	If not presented in the study, statistical power can be assessed during data extraction using an approach that assesses the ability to detect a 10% to 20% change from control group's response for continuous data, or a relative risk or odds ratio of 1.5 to 2 for categorical data, using the outcome frequency in the control group to determine sample size. Recommended sample sizes to achieve 80% power for a given effect size, i.e., 10% or 20% change from control, will be compared to sample sizes used in the study to categorize statistical power as "appears to be adequately powered" (sample size for 80% power met), "somewhat underpowered" (sample size is 75% to < 100% of number required for 80% power), "underpowered" (sample size is 50% to < 75% of number required for 80% power), or "severely underpowered" (sample size is < 50% of number required for 80% power).
	Observations on dose response (e.g., trend analysis, description of whether dose-response shape appears to be monotonic, non-monotonic)
	Data on internal concentration, toxicokinetics, or toxicodynamics (when reported)
Other	Documentation of author queries, use of digital rulers to estimate data values from figures, exposure unit, and statistical result conversions, etc.

Items marked with an asterisk (*) are examples of items that can be used to assess internal validity/risk of bias

Appendix 4. Data Extraction Elements for *In Vitro* Studies

In vitro	
Funding	Funding source(s)
	Reporting of COI by authors (*reporting bias)
Cell/Tissue Model	Cell line, cell type, or tissue
	Source of cells/tissue (and validation of identity)
	Sex of human/animal origin
	Species
	Strain
Treatment	Chemical name and CAS number
	Concentration levels (as presented and converted to μM when possible)
	Source of chemical
	Purity of chemical (*information bias)
	Vehicle used for experimental/control conditions
	Duration and frequency of dosing (e.g., hours, days, weeks when administration was ended, days per week)
Methods	Guideline compliance (i.e., use of EPA, OECD, NTP or another guideline for study design, conducted under GLP guideline conditions, non-GLP but consistent with guideline study, non-guideline peer-reviewed publication)
	Randomization procedure, allocation concealment, blinding during outcome assessment (*selection bias)
	Number of replicates per group (*information bias)
	Percent serum/plasma in medium
	Use of negative controls and whether controls were untreated, vehicle-treated, or both
	Report on data from positive controls – was expected response observed? (*information bias)
	Endpoint health category (e.g., immune)
	Endpoint or assay target (e.g., IL-2 cytokine levels)
	Name and source of assay kit
	Diagnostic or method to measure endpoint (e.g., reporter gene)(*information bias)
	Statistical methods (*information bias)
	Measures of effect at each dose or concentration level (e.g., mean, median, frequency, and measures of precision or variance) or description of qualitative results. When possible, OHAT will convert measures of effect to a common metric with associated 95% confidence intervals (CI). Most often, measures of effect for continuous data will be expressed as mean difference, standardized mean difference, and percent control response. Categorical data will be expressed as relative risk (RR, also called risk ratio).
Results	No Observed Effect Concentration (NOEC), Lowest Observed Effect Concentration (LOEC), statistical significance of other concentration levels, AC50, or other estimates of effect presented in paper. **Note:** The NOEC and LOEC are highly influenced by study design, do not give any quantitative information about the relationship between dose and response, and can be subject to author's interpretation (e.g., a statistically significant effect may not be considered biologically important). Also, a NOEC does not necessarily mean zero response.
	Observations on dose response (e.g., trend analysis, description of whether dose-response shape appears to be monotonic, non-monotonic)
Other	Documentation of author queries, use of digital rulers to estimate data values from figures, exposure unit, and statistical result conversions, etc.

Items marked with an asterisk (*) are examples of items that can be used to assess internal validity/risk of bias

Appendix 5. Risk-of-Bias Criteria

The OHAT risk-of-bias tool for human and animal studies (version date January 2015 and available at http://ntp.niehs.nih.gov/go/38673) reflects OHAT's current best practices and provides the detailed discussion and instructions for the risk-of-bias practices used in this evaluation. The OHAT tool uses a single set of questions (also called "elements" or "domains") to assess risk of bias across various study types to facilitate consideration of conceptually similar potential sources of bias across the human and animal evidence streams with a common terminology. Individual risk-of-bias questions are designated as only applicable to certain study designs (e.g., cohort studies or experimental animal studies), and a subset of the questions apply to each study design (**Table 4**).

The specific criteria used to assess risk of bias for this evaluation are outlined below for Human/observational studies and experimental animal studies. Based on literature searches done for the case study we do not expect any controlled exposure studies in humans (i.e., human controlled trials) and therefore have not included risk-of-bias criteria for that study design. If relevant human controlled trials of MTR Mining are identified, the criteria from the January 2015 OHAT risk–of-bias tool will be used to evaluate risk of bias.

Observational Studies (Human studies)

Cohort studies

1. Was administered dose or exposure level adequately randomized? [NA]

2. Was allocation to study groups adequately concealed? [NA]

3. Did selection of study participants result in the appropriate comparison groups?

Definitely Low Risk of Bias (++)
Direct evidence that subjects (both exposed and non-exposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates, **Note:** A study will be considered low risk of bias if baseline characteristics of groups differed but these differences were considered as potential confounding or stratification variables (see question #4).
Probably Low Risk of Bias (+)
Indirect evidence that subjects (both exposed and non-exposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates, **OR** differences between groups would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
Indirect evidence that subjects (both exposed and non-exposed) were not similar, recruited within very different time frames, or had the very different participation/response rates, **OR** there is insufficient information provided about the comparison group including a different rate of non-response without an explanation (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that subjects (both exposed and non-exposed) were not similar, recruited within very different time frames, or had the very different participation/response rates.

4. Did study design or analysis account for important confounding and modifying variables?

Definitely Low Risk of Bias (++)
• Direct evidence that appropriate adjustments or explicit considerations were made for the variables listed below as potential confounders in the final analyses through the use of statistical models to reduce research-specific bias including standardization, matching, adjustment in multivariate model, stratification, propensity scoring, or other methods that were appropriately justified. Acceptable consideration of appropriate adjustment factors includes cases when the factor is not included in the final adjustment model because the author conducted analyses that indicated it did not need to be included, • **AND** there is direct evidence that primary covariates and confounders were assessed using valid and reliable measurements, • **AND** there is direct evidence that other exposures anticipated to bias results were not present or were appropriately measured and adjusted for. In occupational studies or studies of contaminated sites, other chemical exposures known to be associated with those settings were appropriately considered.
Probably Low Risk of Bias (+)
• Indirect evidence that appropriate adjustments were made, • **OR** it is deemed that not considering or only considering a partial list of covariates or confounders in the final analyses would not appreciably bias results, • **AND** there is evidence (direct or indirect) that covariates and confounders considered were assessed using valid and reliable measurements, • **OR** it is deemed that the measures used would not appreciably bias results (i.e., the authors justified the validity of the measures from previously published research), • **AND** there is evidence (direct or indirect) that other co-exposures anticipated to bias results were not present or were appropriately adjusted for, • **OR** it is deemed that co-exposures present would not appreciably bias results. • **Note:** this includes insufficient information provided on co-exposures in general population studies.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the distribution of important covariates and known confounders differed between the groups and was not appropriately adjusted for in the final analyses, • **OR** there is insufficient information provided about the distribution of known confounders (record “NR” as basis for answer), • **OR** there is indirect evidence that covariates and confounders considered were assessed using measurements of unknown validity, • **OR** there is insufficient information provided about the measurement techniques used to assess covariates and confounders considered (record “NR” as basis for answer), • **OR** there is indirect evidence that there was an unbalanced provision of additional co-exposures across the primary study groups, which were not appropriately adjusted for, • **OR** there is insufficient information provided about co-exposures in occupational studies or studies of contaminated sites where high exposures to other chemical exposures would have been reasonably anticipated (record “NR” as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the distribution of important covariates and known confounders differed between the groups, confounding was demonstrated, and was not appropriately adjusted for in the final analyses, • **OR** there is direct evidence that covariates and confounders considered were assessed using non valid measurements, • **OR** there is direct evidence that there was an unbalanced provision of additional co-exposures across the primary study groups, which were not appropriately adjusted for.

5. Were experimental conditions identical across study groups? [NA]

6. Were the research personnel blinded to the study group during the study? [NA]

7. Were outcome data complete without attrition or exclusion from analysis?

Definitely Low Risk of Bias (++)
Direct evidence that loss of subjects (i.e., incomplete outcome data) was adequately addressed and reasons were documented when human subjects were removed from a study. **Note**: Acceptable handling of subject attrition includes: very little missing outcome data; reasons for missing subjects unlikely to be related to outcome (for survival data, censoring unlikely to be introducing bias); missing outcome data balanced in numbers across study groups, with similar reasons for missing data across groups, **OR** missing data have been imputed using appropriate methods and characteristics of subjects lost to follow up or with unavailable records are described in identical way and are not significantly different from those of the study participants.
Probably Low Risk of Bias (+)
Indirect evidence that loss of subjects (i.e., incomplete outcome data) was adequately addressed and reasons were documented when human subjects were removed from a study, **OR** it is deemed that the proportion lost to follow-up would not appreciably bias results. This would include reports of no statistical differences in characteristics of subjects lost to follow up or with unavailable records from those of the study participants. Generally, the higher the ratio of participants with missing data to participants with events, the greater potential there is for bias. For studies with a long duration of follow-up, some withdrawals for such reasons are inevitable.
Probably High Risk of Bias (-) or (NR)
Indirect evidence that loss of subjects (i.e., incomplete outcome data) was unacceptably large and not adequately addressed, **OR** there is insufficient information provided about numbers of subjects lost to follow-up (record “NR” as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that loss of subjects (i.e., incomplete outcome data) was unacceptably large and not adequately addressed. **Note**: Unacceptable handling of subject attrition includes: reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across study groups; or potentially inappropriate application of imputation.

8. Can we be confident in the exposure characterization?

Definitely Low Risk of Bias (++)
Direct evidence that exposure was consistently assessed (i.e., under the same method and time-frame) using well-established methods that directly measure exposure (e.g., measurement of MTR mining-related components in air or drinking water), **OR** exposure was assessed using less-established methods that directly measure exposure and are validated against well-established methods, **AND** exposure was assessed in a relevant time-window for development of the outcome, **AND** there is sufficient range or variation in exposure measurements, including spatial variation, across groups to potentially identify associations with health outcomes, **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that exposure differences can be distinguished
Probably Low Risk of Bias (+)
Indirect evidence that the exposure was consistently assessed using well-established methods that directly measure exposure), **OR** exposure was assessed using indirect measures (e.g., drinking water levels and residency, questionnaire or occupational exposure assessment by a certified industrial hygienist) that have been validated or empirically shown to be consistent with methods that directly measure exposure (i.e., inter-methods validation: one method vs. another), **AND** exposure was assessed in a relevant time-window for development of the outcome. **AND** there is sufficient range or variation in exposure measurements, including spatial variation, across groups to potentially identify associations with health outcomes (at a minimum from high exposure or ever exposed from low exposure or never exposed), **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay or, if the dataset contains many measurements that are below the limit of quantitation for the assay, exposure groups have been analyzed with statistical methods appropriate for censored datasets.
Probably High Risk of Bias (-) or (NR)
Indirect evidence that the exposure was assessed using poorly validated methods that directly measure exposure **OR** there is evidence that the exposure was assessed using indirect measures that have not been validated or empirically shown to be consistent with methods that directly measure exposure (e.g., questionnaire, job-exposure matrix or self-report without validation) (record "NR" as basis for answer), **OR** there is insufficient information provided about the exposure assessment, including validity and reliability, but no evidence for concern about the method used (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that the exposure was assessed using methods with poor validity, **OR** evidence of exposure misclassification (e.g., differential recall of self-reported exposure).

9. Can we be confident in the outcome assessment?

Definitely Low Risk of Bias (++)
Direct evidence that the outcome was assessed using well-established methods (e.g., gold standard) **AND** subjects had been followed for the same length of time in all study groups, **AND** there is direct evidence that the outcome assessors (including study subjects, if outcomes were self-reported) were adequately blinded to the study group or exposure level, and it is unlikely that they could have broken the blinding prior to reporting outcomes. **NOTE:** Well-established methods will depend on the outcome, but examples of such methods may include: objectively measured antibody or cytokine concentrations with diagnostic methods using commercial kits, commercial laboratories, or standard assays such as ELISAs for IgG with sufficiently low variation and limits of detection to allow discrimination between groups (or evidence that the assay could have detected a difference based on responses to a positive control); doctor diagnosis of asthma or incidence data obtained from medical records; incidence of doctor-diagnosed otitis by trained interviewers; obtained from registries (Shamliyan *et al.* 2010).
Probably Low Risk of Bias (+)
Indirect evidence that the outcome was assessed using acceptable methods (i.e., deemed valid and reliable but not the gold standard), **AND** subjects had been followed for the same length of time in all study groups **OR** it is deemed that the outcome assessment methods used would not appreciably bias results, **AND** there is indirect evidence that the outcome assessors (including study subjects, if outcomes were self-reported) were adequately blinded to the study group, and it is unlikely that they could have broken the blinding prior to reporting outcomes, **OR** it is deemed that lack of adequate blinding of outcome assessors would not appreciably bias results, which is more likely to apply to objective outcome measures, **NOTE:** Acceptable, but not ideal assessment methods will depend on the outcome, but examples of such methods may include proxy reporting of outcomes such as and mining of data collected for other purposes. Proxy reporting (e.g., parental reporting of days sick or doctor-diagnosis) of disease should be considered on a case-by-case basis with consideration of whether or not there is empirical evidence as to the reliability of proxy reporting for that outcome.
Probably High Risk of Bias (-) or (NR)
Indirect evidence that the outcome assessment method is an insensitive instrument (e.g., a questionnaire used to assess outcomes with no information on validation), **OR** the length of follow up differed by study group, **OR** there is indirect evidence that it was possible for outcome assessors (including study subjects if outcomes were self-reported) to infer the study group prior to reporting outcomes, **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that the outcome assessment method is an insensitive instrument, **OR** the length of follow up differed by study group, **OR** there is direct evidence for lack of adequate blinding of outcome assessors (including study subjects if outcomes were self-reported), including no blinding or incomplete blinding.

10. Were all measured outcomes reported?

Definitely Low Risk of Bias (++)
Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction and analyses had been planned in advance.
Probably Low Risk of Bias (+)
Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, **OR** analyses that had not been planned in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).
Probably High Risk of Bias (-) or (NR)
Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported, **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, **OR** there is insufficient information provided about selective outcome reporting (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified or reporting outcomes not pre-specified, or that unplanned analyses were included that would appreciably bias results.

11. Were there no other potential threats to internal validity?

There are no MTR mining-specific additions to the risk-of-bias questions for this evaluation. This question will be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk-of-bias considerations that do not fit under the other questions.

Cross Sectional and Case Series Studies

1. Was administered dose or exposure level adequately randomized? [NA]

2. Was allocation to study groups adequately concealed? [NA]

3. Did selection of study participants result in the appropriate comparison groups?[NA to Case series]

Definitely Low Risk of Bias (++)
Direct evidence that subjects (both exposed and non-exposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates, **Note:** A study will be considered low risk of bias if baseline characteristics of groups differed but these differences were considered as potential confounding or stratification variables (see question #4).
Probably Low Risk of Bias (+)
Indirect evidence that subjects (both exposed and non-exposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates, **OR** differences between groups would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
Indirect evidence that subjects (both exposed and non-exposed) were not similar, recruited within very different time frames, or had the very different participation/response rates, **OR** there is insufficient information provided about the comparison group including a different rate of non-response without an explanation (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that subjects (both exposed and non-exposed) were not similar, recruited within very different time frames, or had the very different participation/response rates.

4. Did study design or analysis account for important confounding and modifying variables?

Definitely Low Risk of Bias (++)
Direct evidence that appropriate adjustments or explicit considerations were made for the variables listed below as potential confounders in the final analyses through the use of statistical models to reduce research-specific bias including standardization, matching, adjustment in multivariate model, stratification, propensity scoring, or other methods that were appropriately justified. Acceptable consideration of appropriate adjustment factors includes cases when the factor is not included in the final adjustment model because the author conducted analyses that indicated it did not need to be included, **AND** there is direct evidence that primary covariates and confounders were assessed using valid and reliable measurements, **AND** there is direct evidence that other exposures anticipated to bias results were not present or were appropriately measured and adjusted for. In occupational studies or studies of contaminated sites, other chemical exposures known to be associated with those settings were appropriately considered.
Probably Low Risk of Bias (+)
Indirect evidence that appropriate adjustments were made, **OR** it is deemed that not considering or only considering a partial list of covariates or confounders in the final analyses would not appreciably bias results, **AND** there is evidence (direct or indirect) that covariates and confounders considered were assessed using valid and reliable measurements, **OR** it is deemed that the measures used would not appreciably bias results (i.e., the authors justified the validity of the measures from previously published research), **AND** there is evidence (direct or indirect) that other co-exposures anticipated to bias results were not present or were appropriately adjusted for, **OR** it is deemed that co-exposures present would not appreciably bias results. **Note:** this includes insufficient information provided on co-exposures in general population studies.
Probably High Risk of Bias (-) or (NR)
Indirect evidence that the distribution of important covariates and known confounders differed between the groups and was not appropriately adjusted for in the final analyses, **OR** there is insufficient information provided about the distribution of known confounders (record “NR” as basis for answer), **OR** there is indirect evidence that covariates and confounders considered were assessed using measurements of unknown validity, **OR** there is insufficient information provided about the measurement techniques used to assess covariates and confounders considered (record “NR” as basis for answer), **OR** there is indirect evidence that there was an unbalanced provision of additional co-exposures across the primary study groups, which were not appropriately adjusted for, **OR** there is insufficient information provided about co-exposures in occupational studies or studies of contaminated sites where high exposures to other chemical exposures would have been reasonably anticipated (record “NR” as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that the distribution of important covariates and known confounders differed between the groups, confounding was demonstrated, and was not appropriately adjusted for in the final analyses, **OR** there is direct evidence that covariates and confounders considered were assessed using non valid measurements, **OR** there is direct evidence that there was an unbalanced provision of additional co-exposures across the primary study groups, which were not appropriately adjusted for.

5. Were experimental conditions identical across study groups? [NA]

6. Were the research personnel blinded to the study group during the study? [NA]

7. Were outcome data complete without attrition or exclusion from analysis?

Definitely Low Risk of Bias (++)
Direct evidence that exclusion of subjects from analyses was adequately addressed and reasons were documented when subjects were removed from the study or excluded from analyses.
Probably Low Risk of Bias (+)
Indirect evidence that exclusion of subjects from analyses was adequately addressed and reasons were documented when subjects were removed from the study or excluded from analyses.
Probably High Risk of Bias (-) or (NR)
Indirect evidence that exclusion of subjects from analyses was not adequately addressed, **OR** there is insufficient information provided about why subjects were removed from the study or excluded from analyses (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that exclusion of subjects from analyses was not adequately addressed. **Note**: Unacceptable handling of subject exclusion from analyses includes: reason for exclusion likely to be related to true outcome, with either imbalance in numbers or reasons for exclusion across study groups.

8. Can we be confident in the exposure characterization?

Definitely Low Risk of Bias (++)
Direct evidence that exposure was consistently assessed (i.e., under the same method and time-frame) using well-established methods that directly measure exposure (e.g., measurement of MTR mining-related components in air or drinking water), **OR** exposure was assessed using less-established methods that directly measure exposure and are validated against well-established methods, **AND** exposure was assessed in a relevant time-window for development of the outcome, **AND** there is sufficient range or variation in exposure measurements, including spatial variation, across groups to potentially identify associations with health outcomes, **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that exposure differences can be distinguished.
Probably Low Risk of Bias (+)
Indirect evidence that the exposure was consistently assessed using well-established methods that directly measure exposure), **OR** exposure was assessed using indirect measures (e.g., drinking water levels and residency, questionnaire or occupational exposure assessment by a certified industrial hygienist) that have been validated or empirically shown to be consistent with methods that directly measure exposure (i.e., inter-methods validation: one method vs. another), **AND** exposure was assessed in a relevant time-window for development of the outcome, **AND** there is sufficient range or variation in exposure measurements, including spatial variation, across groups to potentially identify associations with health outcomes (at a minimum from high exposure or ever exposed from low exposure or never exposed), **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay or, if the dataset contains many measurements that are below the limit of quantitation for the assay, exposure groups have been analyzed with statistical methods appropriate for censored datasets.
Probably High Risk of Bias (-) or (NR)
Indirect evidence that the exposure was assessed using poorly validated methods that directly measure exposure **OR** there is evidence that the exposure was assessed using indirect measures that have not been validated or empirically shown to be consistent with methods that directly measure exposure (e.g., a job-exposure matrix or self-report without validation) (record "NR" as basis for answer), **OR** there is insufficient information provided about the exposure assessment, including validity and reliability, but no evidence for concern about the method used (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that the exposure was assessed using methods with poor validity, **OR** evidence of exposure misclassification (e.g., differential recall of self-reported exposure).

9. Can we be confident in the outcome assessment?

Definitely Low Risk of Bias (++)
Direct evidence that the outcome was assessed using well-established methods (the gold standard), **AND** there is direct evidence that the outcome assessors (including study subjects, if outcomes were self-reported) were adequately blinded to the exposure level, and it is unlikely that they could have broken the blinding prior to reporting outcomes. **NOTE** Well-established assessment methods will depend on the outcome, but examples of such methods may include: objectively measured antibody or cytokine concentrations with diagnostic methods using commercial kits, commercial laboratories, or standard assays such as ELISAs for IgG with sufficiently low variation and limits of detection to allow discrimination between groups (or evidence that the assay could have detected a difference based on responses to a positive control); doctor diagnosis of asthma or incidence data obtained from medical records; obtained from registries (Shamliyan *et al.* 2010).
Probably Low Risk of Bias (+)
Indirect evidence that the outcome was assessed using acceptable methods, **OR** it is deemed that the outcome assessment methods used would not appreciably bias results, **AND** there is indirect evidence that the outcome assessors were adequately blinded to the exposure level, and it is unlikely that they could have broken the blinding prior to reporting outcomes, **OR** it is deemed that lack of adequate blinding of outcome assessors would not appreciably bias results (including that subjects self-reporting outcomes were likely not aware of reported links between the exposure and outcome lack of blinding is unlikely to bias a particular outcome). **NOTE:** Acceptable, but not ideal assessment methods will depend on the outcome, but examples of such methods may include proxy reporting of outcomes such as asthma and mining of data collected for other purposes. Proxy reporting (e.g., parental reporting of days sick or doctor-diagnosis) of disease should be considered on a case-by-case basis with consideration of whether or not there is empirical evidence as to the reliability of proxy reporting for that outcome.
Probably High Risk of Bias (-) or (NR)
Indirect evidence that the outcome assessment method is an insensitive instrument, **OR** there is indirect evidence that it was possible for outcome assessors to infer the exposure level prior to reporting outcomes (including that subjects self-reporting outcomes were likely aware of reported links between the exposure and outcome), **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that the outcome assessment method is an insensitive instrument, **OR** there is direct evidence that outcome assessors were aware of the exposure level prior to reporting outcomes (including that subjects self-reporting outcomes were aware of reported links between the exposure and outcome).

10. Were all measured outcomes reported?

Definitely Low Risk of Bias (++)
Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction and analyses had been planned in advance.
Probably Low Risk of Bias (+)
Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, **OR** analyses that had not been planned in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).
Probably High Risk of Bias (-) or (NR)
Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported, **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, **OR** there is insufficient information provided about selective outcome reporting (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified or reporting outcomes not pre-specified, or that unplanned analyses were included that would appreciably bias results.

11. Were there no other potential threats to internal validity?

There are no MTR mining-specific additions to the risk-of-bias questions for this evaluation. This question will be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk-of-bias considerations that do not fit under the other questions.

Case Control Studies

1. Was administered dose or exposure level adequately randomized? [NA]

2. Was allocation to study groups adequately concealed? [NA]

3. Did selection of study participants result in the appropriate comparison groups?

Definitely Low Risk of Bias (++)
Direct evidence that cases and controls were similar (e.g., recruited from the same eligible population including being of similar age, gender, ethnicity, and eligibility criteria other than outcome of interest as appropriate), recruited within the same time frame, and controls are described as having no history of the outcome, **Note:** A study will be considered low risk of bias if baseline characteristics of groups differed but these differences were considered as potential confounding or stratification variables (see question #4),
Probably Low Risk of Bias (+)
Indirect evidence that cases and controls were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age), recruited within the same time frame, and controls are described as having no history of the outcome, **OR** it is deemed differences between cases and controls would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
Indirect evidence that controls were drawn from a very dissimilar population than cases or recruited within very different time frames, **OR** there is insufficient information provided about the appropriateness of controls including rate of response reported for cases only (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that controls were drawn from a very dissimilar population than cases or recruited within very different time frames.

4. Did study design or analysis account for important confounding and modifying variables?

Definitely Low Risk of Bias (++)
Direct evidence that appropriate adjustments were made for the variables listed below as potential confounders in the final analyses through the use of statistical models to reduce research-specific bias including standardization, matching of cases and controls, adjustment in multivariate model, stratification, propensity scoring, or other methods were appropriately justified, **AND** there is direct evidence that primary covariates and confounders were assessed using valid and reliable measurements, **AND** there is direct evidence that other exposures anticipated to bias results were not present or were appropriately measured and adjusted for.
Probably Low Risk of Bias (+)
Indirect evidence that appropriate adjustments were made, **OR** it is deemed that not considering or only considering a partial list of covariates or confounders in the final analyses would not appreciably bias results, **AND** there is evidence (direct or indirect) that covariates and confounders considered were assessed using valid and reliable measurements, **OR** it is deemed that the measures used would not appreciably bias results (i.e., the authors justified the validity of the measures from previously published research), **AND** there is evidence (direct or indirect) that other co-exposures anticipated to bias results were not present or were appropriately adjusted for, **OR** it is deemed that co-exposures present would not appreciably bias results. **Note:** this includes insufficient information provided on co-exposures in general population studies.
Probably High Risk of Bias (-) or (NR)
Indirect evidence that the distribution of important covariates and known confounders differed between cases and controls and was not investigated further, **OR** there is insufficient information provided about the distribution of known confounders in cases and controls (record "NR" as basis for answer), **OR** there is indirect evidence that covariates and confounders considered were assessed using measurements of unknown validity, **OR** there is insufficient information provided about the measurement techniques used to assess covariates and confounders considered (record "NR" as basis for answer), **OR** there is indirect evidence that there was an unbalanced provision of additional co-exposures across cases and controls, which were not appropriately adjusted for, **OR** there is insufficient information provided about co-exposures in occupational studies or studies of contaminated sites where high exposures to other chemical exposures would have been reasonably anticipated (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that the distribution of important covariates and known confounders differed between cases and controls, confounding was demonstrated, but was not appropriately adjusted for in the final analyses, **OR** there is direct evidence that covariates and confounders considered were assessed using non valid measurements, **OR** there is direct evidence that there was an unbalanced provision of additional co-exposures across cases and controls, which were not appropriately adjusted for.

5. Were experimental conditions identical across study groups? [NA]

6. Were the research personnel blinded to the study group during the study? [NA]

7. Were outcome data complete without attrition or exclusion from analysis?

Definitely Low Risk of Bias (++)
Direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
Probably Low Risk of Bias (+)
Indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
Probably High Risk of Bias (-) or (NR)
Indirect evidence that exclusion of subjects from analyses was not adequately addressed, **OR** there is insufficient information provided about why subjects were removed from the study or excluded from analyses (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that exclusion of subjects from analyses was not adequately addressed. **Note**: Unacceptable handling of subject exclusion from analyses includes: reason for exclusion likely to be related to true outcome, with either imbalance in numbers or reasons for exclusion across study groups.

8. Can we be confident in the exposure characterization?

Definitely Low Risk of Bias (++)
Direct evidence that exposure was consistently assessed (i.e., under the same method and time-frame) using well-established methods that directly measure exposure (e.g., measurement of MTR mining-related components in air or drinking water), **OR** exposure was assessed using less-established methods that directly measure exposure and are validated against well-established methods. **AND** exposure was assessed in a relevant time-window for development of the outcome, **AND** there is sufficient range or variation in exposure measurements, including spatial variation, across groups to potentially identify associations with health outcomes, **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that exposure differences can be distinguished.
Probably Low Risk of Bias (+)
Indirect evidence that the exposure was consistently assessed using well-established methods that directly measure exposure), **OR** exposure was assessed using indirect measures (e.g., drinking water levels and residency, questionnaire or occupational exposure assessment by a certified industrial hygienist) that have been validated or empirically shown to be consistent with methods that directly measure exposure (i.e., inter-methods validation: one method vs. another), **AND** exposure was assessed in a relevant time-window for development of the outcome, **AND** there is sufficient range or variation in exposure measurements, including spatial variation, across groups to potentially identify associations with health outcomes (at a minimum from high exposure or ever exposed from low exposure or never exposed), **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay or, if the dataset contains many measurements that are below the limit of quantitation for the assay, exposure groups have been analyzed with statistical methods appropriate for censored datasets.
Probably High Risk of Bias (-) or (NR)
Indirect evidence that the exposure was assessed using poorly validated methods that directly measure exposure, **OR** there is direct evidence that the exposure was assessed using indirect measures that have not been validated or empirically shown to be consistent with methods that directly measure exposure (e.g., a job-exposure matrix or self-report without validation) (record “NR” as basis for answer), **OR** there is insufficient information provided about the exposure assessment, including validity and reliability, but no evidence for concern about the method used (record “NR” as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that the exposure was assessed using methods with poor validity, **OR** evidence of exposure misclassification (e.g., differential recall of self-reported exposure).

9. Can we be confident in the outcome assessment?

Definitely Low Risk of Bias (++)
Direct evidence that the outcome was assessed in cases (i.e., case definition) and controls using well-established methods (the gold standard), **AND** subjects had been followed for the same length of time in all study groups, **AND** there is direct evidence that the outcome assessors (including study subjects, if outcomes were self-reported) were adequately blinded to the exposure level when outcome was assessed in cases (i.e., case definition) and controls. **NOTE** Well-established methods will depend on the outcome, but examples of such methods may include: doctor diagnosis of asthma or doctor diagnosis obtained from medical records.
Probably Low Risk of Bias (+)
Indirect evidence that the outcome was assessed in cases (i.e., case definition) and controls using acceptable methods), **AND** subjects had been followed for the same length of time in all study groups, **OR** it is deemed that the outcome assessment methods used would not appreciably bias results, **AND** there is indirect evidence that the outcome assessors were adequately blinded to the exposure level when reporting outcomes, **OR** it is deemed that lack of adequate blinding of outcome assessors would not appreciably bias results (including that subjects self-reporting outcomes were likely not aware of reported links between the exposure and outcome or lack of blinding is unlikely to bias a particular outcome). **NOTE** Acceptable, but not ideal assessment methods will depend on the outcome, but examples of such methods may include proxy reporting of outcomes such as asthma and mining of data collected for other purposes. Proxy reporting of disease should be considered on a case-by-case basis with consideration of whether or not there is empirical evidence as to the reliability of proxy reporting for that outcome.
Probably High Risk of Bias (-) or (NR)
Indirect evidence that the outcome was assessed in cases (i.e., case definition) using an insensitive instrument, **OR** there is insufficient information provided about how cases were identified (record "NR" as basis for answer). **OR** there is indirect evidence that it was possible for outcome assessors to infer the exposure level prior to reporting outcomes (including that subjects self-reporting outcomes were likely aware of reported links between the exposure and outcome), **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that the outcome was assessed in cases (i.e., case definition) using an insensitive instrument, **OR** there is direct evidence that outcome assessors were aware of the exposure level prior to reporting outcomes (including that subjects self-reporting outcomes were aware of reported links between the exposure and outcome).

10. Were all measured outcomes reported?

Definitely Low Risk of Bias (++)
Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction and analyses had been planned in advance.
Probably Low Risk of Bias (+)
Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, **OR** analyses that had not been planned in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).
Probably High Risk of Bias (-) or (NR)
Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported, **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, **OR** there is insufficient information provided about selective outcome reporting (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified or reporting outcomes not pre-specified, or that unplanned analyses were included that would appreciably bias results.

11. Were there no other potential threats to internal validity?

There are no MTR mining-specific additions to the risk-of-bias questions for this evaluation. This question will be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk-of-bias considerations that do not fit under the other questions.

Experimental Animal Studies

1. Was administered dose or exposure level adequately randomized?

Definitely Low Risk of Bias (++)
Direct evidence that animals were allocated to any study group including controls using a method with a random component, **AND** there is direct evidence that the study used a concurrent control group as an indication that randomization covered all study groups, **Note:** Acceptable methods of randomization include: referring to a random number table, using a computer random number generator, coin tossing, or shuffling cards (Higgins and Green 2011). **Note:** Restricted randomization (e.g., blocked randomization) to ensure particular allocation ratios will be considered low bias. Similarly, stratified randomization approaches that attempt to minimize imbalance between groups on important prognostic factors (e.g., body weight) will be considered acceptable.
Probably Low Risk of Bias (+)
Indirect evidence that animals were allocated to any study group including controls using a method with a random component (i.e., authors state random allocation, without description of method), **AND** evidence that the study used a concurrent control group as an indication that randomization covered all study groups, **OR** it is deemed that allocation without a clearly random component would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
Indirect evidence that animals were allocated to study groups using a method with a non-random component, **OR** indirect evidence that there was a lack of a concurrent control group, **OR** there is insufficient information provided about how animals were allocated to study groups (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that animals were allocated to study groups using a non-random method including judgment of the investigator, the results of a laboratory test or a series of tests, **OR** direct evidence that there was a lack of a concurrent control group.

2. Was allocation to study groups adequately concealed?

Definitely Low Risk of Bias (++)
Direct evidence that at the time of assigning study groups the research personnel did not know what group animals were allocated to, and it is unlikely that they could have broken the blinding of allocation until after assignment was complete and irrevocable. **Note:** Acceptable methods used to ensure allocation concealment include sequentially numbered treatment containers of identical appearance or equivalent methods.
Probably Low Risk of Bias (+)
Indirect evidence that at the time of assigning study groups the research personnel did not know what group animals were allocated to and it is unlikely that they could have broken the blinding of allocation until after assignment was complete and irrevocable, **OR** it is deemed that lack of adequate allocation concealment would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
Indirect evidence that at the time of assigning study groups it was possible for the research personnel to know what group animals were allocated to, or it is likely that they could have broken the blinding of allocation before assignment was complete and irrevocable, **OR** there is *insufficient* information provided about allocation to study groups (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that at the time of assigning study groups it was possible for the research personnel to know what group animals were allocated to, or it is likely that they could have broken the blinding of allocation before assignment was complete and irrevocable.

3. Did selection of study participants result in the appropriate comparison groups? [NA]

4. Did study design or analysis account for important confounding and modifying variables? [NA]

5. Were experimental conditions identical across study groups?

Definitely Low Risk of Bias (++)
Direct evidence that same vehicle was used in control and experimental animals, **AND** direct evidence that non-treatment-related experimental conditions were identical across study groups (i.e., the study report explicitly provides this level of detail).
Probably Low Risk of Bias (+)
Indirect evidence that the same vehicle was used in control and experimental animals, **OR** it is deemed that the vehicle used would not appreciably bias results, **AND** identical non-treatment-related experimental conditions are assumed if authors did not report differences in housing or husbandry.
Probably High Risk of Bias (-) or (NR)
Indirect evidence that the vehicle differed between control and experimental animals, **OR** authors did not report the vehicle used (record "NR" as basis for answer), **OR** there is indirect evidence that non-treatment-related experimental conditions were not comparable between study groups.
Definitely High Risk of Bias (--)
Direct evidence from the study report that control animals were untreated, or treated with a different vehicle than experimental animals, **OR** there is direct evidence that non-treatment-related experimental conditions were not comparable between study groups.

6. Were the research personnel blinded to the study group during the study?

Definitely Low Risk of Bias (++)
Direct evidence that the research personnel were adequately blinded to study group, and it is unlikely that they could have broken the blinding during the study. Methods used to ensure blinding include central allocation; sequentially numbered treatment containers of identical appearance; sequentially numbered animal cages; or equivalent methods.
Probably Low Risk of Bias (+)
Indirect evidence that the research personnel were adequately blinded to study group, and it is unlikely that they could have broken the blinding during the study, **OR** it is deemed that lack of adequate blinding during the study would not appreciably bias results. This would include cases where blinding was not possible but research personnel took steps to minimize potential bias, such as restricting the knowledge of study group to veterinary or supervisory personnel monitoring for overt toxicity, or randomized husbandry or handling practices (e.g., placement in the animal room, necropsy order, etc.).
Probably High Risk of Bias (-) or (NR)
Indirect evidence that the research personnel were not adequately blinded to study group, **OR** there is insufficient information provided about blinding to study group during the study (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that the research personnel were not adequately blinded to study group.

7. Were outcome data complete without attrition or exclusion from analysis?

Definitely Low Risk of Bias (++)
Direct evidence that loss of animals was adequately addressed and reasons were documented when animals were removed from a study. **Note**: Acceptable handling of attrition includes: very little missing outcome data; reasons for missing animals unlikely to be related to outcome (or for survival data, censoring unlikely to be introducing bias); missing outcome data balanced in numbers across study groups, with similar reasons for missing data across groups; missing outcomes is not enough to impact the effect estimate. **OR** missing data have been imputed using appropriate methods (insuring that characteristics of animals are not significantly different from animals retained in the analysis).
Probably Low Risk of Bias (+)
Indirect evidence that loss of animals was adequately addressed and reasons were documented when animals were removed from a study, **OR** it is deemed that the proportion lost would not appreciably bias results. This would include reports of no statistical differences in characteristics of animals removed from the study from those remaining in the study.
Probably High Risk of Bias (-) or (NR)
Indirect evidence that loss of animals was unacceptably large and not adequately addressed, **OR** there is insufficient information provided about loss of animals (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that loss of animals was unacceptably large and not adequately addressed. **Note**: Unacceptable handling of attrition or exclusion includes: reason for loss is likely to be related to true outcome, with either imbalance in numbers or reasons for loss across study groups.

8. Can we be confident in the exposure characterization?

Definitely Low Risk of Bias (++)
Direct evidence that the exposure to MTR mining mixtures was appropriately characterized **AND** that exposure was consistently administered (i.e., with the same method and time-frame) across treatment groups, **AND** for dietary or drinking water studies that information is provided on consumption or internal dose metrics to confirm expected exposure levels sufficiently to allow discrimination between exposure groups, **AND** if internal dose metrics are available, there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished, **AND** if internal dose metrics are available, the study used spiked samples to confirm assay performance.
Probably Low Risk of Bias (+)
Indirect evidence that the exposure to MTR mining mixtures was appropriately characterized, **AND** that exposure was consistently administered (i.e., with the same method and time-frame) across treatment groups, **AND** for dietary or drinking water studies no information is provided on consumption or internal dose metrics, **AND** if internal dose metrics are available, there is indirect evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished.
Probably High Risk of Bias (-) or (NR)
Indirect evidence that the exposure (including compliance with the treatment, if applicable) was assessed using poorly validated methods, **OR** there is insufficient information provided about the validity of the exposure assessment method, but no evidence for concern (record "NR" as basis for answer), **AND** if internal dose metrics are available, there is indirect evidence that most of the exposure data measurements are below the limit of quantitation for the assay such that different exposure groups cannot be distinguished.
Definitely High Risk of Bias (--)
Direct evidence that the exposure (including compliance with the treatment, if applicable) was assessed using poorly validated methods.

9. Can we be confident in the outcome assessment?

Definitely Low Risk of Bias (++)
Direct evidence that the outcome was assessed using well-established methods (e.g., gold standard) **AND** assessed at the same length of time after initial exposure in all study groups, **AND** there is direct evidence that the outcome assessors were adequately blinded to the study group, and it is unlikely that they could have broken the blinding prior to reporting outcomes. **NOTE** Well-established methods will depend on the outcome, but examples of such methods may include: objectively measured antibody or cytokine concentrations with diagnostic methods using commercial kits, commercial laboratories with experience in the assay, or standard assays such as ELISAs for IgG and with sufficiently low variation and limits of detection to allow discrimination of responses between treatment groups (or direct evidence that the assay could have detected a difference based on responses to a positive control).
Probably Low Risk of Bias (+)
Indirect evidence that the outcome was assessed using acceptable methods (i.e., deemed valid and reliable but not the gold standard), **AND** assessed at the same length of time after initial exposure in all study groups, **OR** it is deemed that the outcome assessment methods used would not appreciably bias results, **AND** there is indirect evidence that the outcome assessors were adequately blinded to the study group, and it is unlikely that they could have broken the blinding prior to reporting outcomes, **OR** it is deemed that lack of adequate blinding of outcome assessors would not appreciably bias results, which is more likely to apply to objective outcome measures. **NOTE** For some outcomes, particularly histopathology assessment, outcome assessors are not blind to study group as they require comparison to the control to appropriately judge the outcome, but additional measures such as multiple levels of independent review by trained pathologists can minimize potential bias. **NOTE** Acceptable assessment methods will depend on the outcome, but examples of such methods may include: objectively measured antibody or cytokine concentrations with diagnostic methods using commercial kits with some variation, but ability to discriminate between the high dose treatment and control group (or indirect evidence that the assay could have detected a difference based on responses to a positive control).
Probably High Risk of Bias (-) or (NR)
Indirect evidence that the outcome assessment method is an insensitive instrument, **OR** the length of time after initial exposure differed by study group, **OR** there is indirect evidence that it was possible for outcome assessors to infer the study group prior to reporting outcomes without sufficient quality control measures, **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
Direct evidence that the outcome assessment method is an insensitive instrument, **OR** the length of time after initial exposure differed by study group, **OR** there is direct evidence for lack of adequate blinding of outcome assessors, including no blinding or incomplete blinding without quality control measures.

10. Were all measured outcomes reported?

Definitely Low Risk of Bias (++)
Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction and analyses had been planned in advance.
Probably Low Risk of Bias (+)
Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, **OR** analyses that had not been planned in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).
Probably High Risk of Bias (-) or (NR)
Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported, **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, **OR** there is insufficient information provided about selective outcome reporting (record "NR" as answer basis).
Definitely High Risk of Bias (--)
Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified or reporting outcomes not pre-specified, or that unplanned analyses were included that would appreciably bias results.

11. Were there no other potential threats to internal validity?

There are no MTR mining-specific additions to the risk-of-bias questions for this evaluation. This question will be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk-of-bias considerations that do not fit under the other questions.