DRAFT

Current Intelligence Bulletin: NIOSH Practices in Occupational Risk Assessment

External Review Draft

June 6, 2018

DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Institute for Occupational Safety and Health

Abstract

Workers are exposed to on-the-job health hazards every day. Unlike safety hazards that may lead to injury, health hazards can lead to various types of illness. For example, exposures to some chemicals used in work processes or for cleaning may cause immediate lung disorders, such as asthma-like symptoms, and in other cases, chemicals may cause cancer in workers that is not observed until years after first exposure. In order to make recommendations for working safely in the presence of chemical health hazards, the National Institute for Occupational Safety and Health (NIOSH) conducts risk assessments. Risk assessment is a way of relating the amount of a hazard, like the concentration of a chemical in the air, to the risk of developing illness because of exposure to that hazard. Risk assessment allows NIOSH to make recommendations for controlling exposures that will keep workers safe and prevent illness.

This document describes the process NIOSH uses to conduct risk assessments. It outlines the logic that NIOSH uses to evaluate the scientific evidence and determine:

- what type of hazard a chemical or other agent might be,
- what scientific evidence is available to help NIOSH determine if the chemical or other agent causes illness or injury,
- the steps NIOSH takes to evaluate the scientific data,
- the mathematical methods that NIOSH uses to determine how much exposure to the chemical or other agent would be harmful to workers, called dose-response assessment,
- the procedures for ensuring that NIOSH carefully considers all the relevant evidence and makes the best, scientifically supported decisions.

A NIOSH risk assessment undergoes scientific peer review before it is published. It may be published in a scientific journal, or become part of a larger NIOSH document that describes the hazard and makes recommendations aimed to improve worker safety and health.

Executive Summary

ES 1.0 INTRODUCTION

Occupational risks are defined as the potential for and severity of adverse effects in workers from their exposure to workplace hazards. These risks can be mitigated by safeguards that are derived via a combination of scientific assessment and best management practices. Risk assessment is an important tool for informed decision-making on workplace safeguards when the hazards and/or health consequences are not fully characterized. Since the 1990s, quantitative risk assessments conducted by the National Institute for Occupational Safety and Health (NIOSH) have buttressed recommendations on limiting chemical exposures and some other workplace hazards, such as ionizing radiation and noise. This document describes the underlying science and general approach used by NIOSH researchers when conducting high quality, scientifically sound quantitative assessments of the risk associated with these workplace hazards. The report focuses on chemical risk assessment practices; however, some of these practices have benefitted NIOSH assessments of other workplace hazards, such as ionizing radiation and noise. This information is intended for NIOSH risk assessors, other scientists, stakeholders, and the public to improve their understanding of the NIOSH risk assessment process. This document is one of many routine exchanges between NIOSH, its stakeholders, and the risk assessment community, both home and abroad, which act to ensure that best practices are followed in risk assessment supporting worker protection.

ES 2.0 RISK ASSESSMENT PROCESS

NIOSH risk assessments are typically carried out by a multidisciplinary team of epidemiologists, toxicologists, biostatisticians, industrial hygienists, other exposure scientists (e.g., health physicists, chemists), and health communications experts, hereafter referred to as 'risk assessors.' NIOSH risk assessments are usually prompted by persons who are at risk (e.g., affected workers), risk managers (e.g., employers, regulators), or risk assessors, alone or in combination, who need information on the probability and severity of potential workplace hazards. In response, NIOSH develops a risk assessment plan containing two key components: 1) a conceptual

model that identifies the hazard (sources, stressors, and pathways), persons potentially at risk, and possible adverse effects; and 2) an analysis plan (work plan) that outlines the analytic components (i.e., data and methods) and interpretative approaches (e.g., risk metrics) to be used [NRC 2009]. Risk assessment planning helps to ensure that the risk assessment applies the best scientific methods, the highest-quality evidence, and addresses the needs of the decision-makers (risk managers).

NIOSH risk assessment is defined as the determination of the relationship between the occupational exposure and adverse effects (e.g., cancer, non-malignant respiratory disease). Data permitting, this determination is preferred to be quantitative; however, qualitative risk assessments are performed on occasion. The quantitative risk assessment comprises three major components that are completed sequentially, namely hazard identification (including exposure assessment), dose-response assessment, and risk characterization. Hazard identification is the systematic process for assessing whether an agent of interest causes an adverse effect in exposed workers. The findings from hazard identification are characteristic descriptions and data on the exposure of interest, any important cofactors (e.g., other risk factors), mode of action, and the adverse effects associated with exposure. These data are prerequisites for conducting the dose-response assessment. In strict terms, ‘dose-response’ refers to the relationship between the amount of an agent administered to, taken up by, or absorbed by an organism, system, or population and the adverse effect developed in that organism, system, or population in reaction to the agent. In practice; however, the terms ‘exposure’ and ‘dose’ have been expressed in many different ways over time and are often used interchangeably. The dose-response assessment provides estimates of the dose-risk relationship for use in the third component of risk assessment, namely risk characterization. Risk characterization is the qualitative and, wherever possible, quantitative determination of the probability of occurrence of known and potential adverse effects in workers under defined conditions of exposure to an agent. It reflects the integration of the science from hazard identification and dose-response assessment with additional information necessary to establish a sound scientific basis for NIOSH recommendations. These recommendations inform decision-makers

who are responsible for managing workplace risk. The components of risk assessment and their relationship with risk management are shown in the figure below.

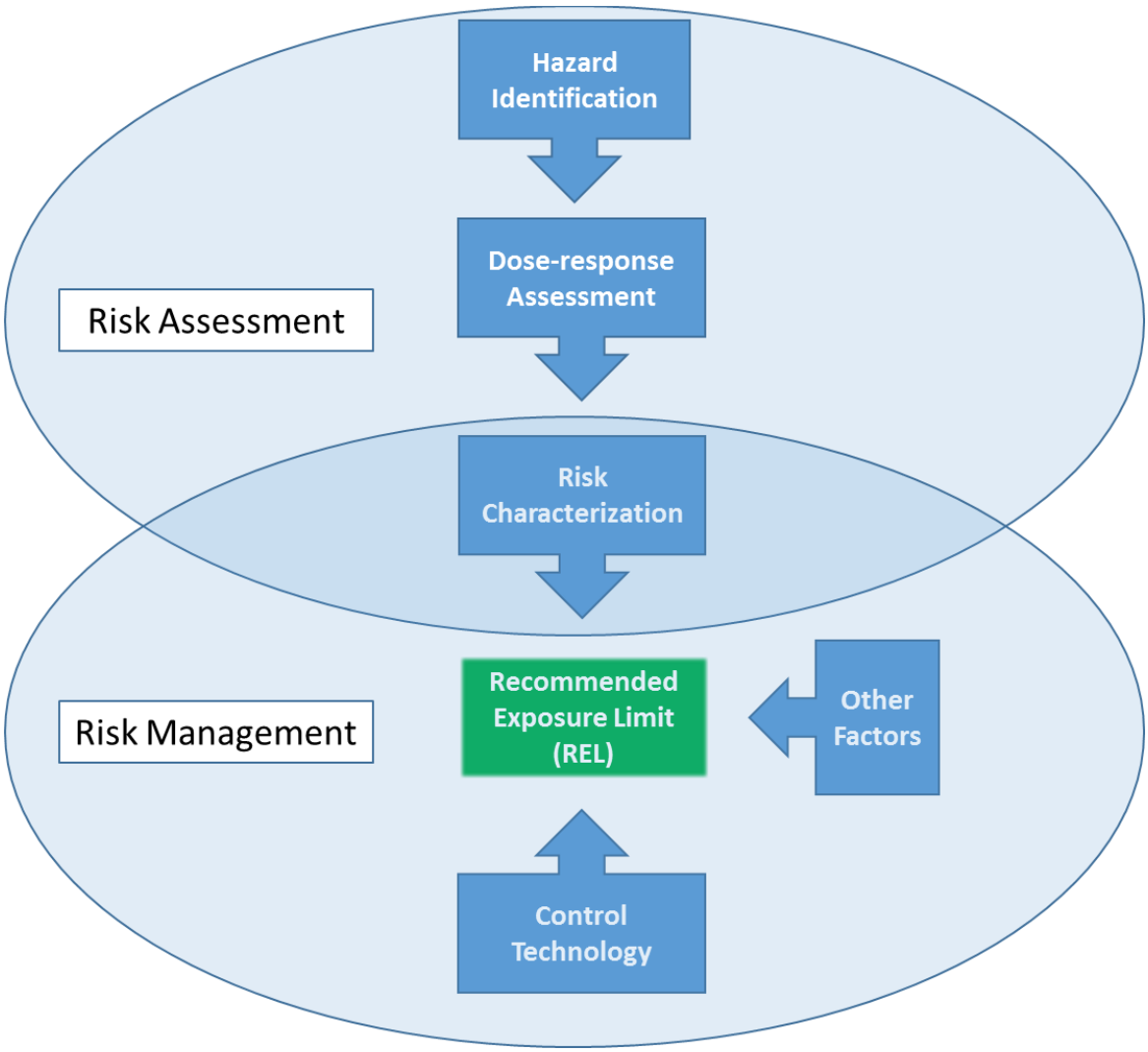

Figure ES-1 NIOSH Risk assessment and risk management processes.

ES 2.1 Hazard Identification

Hazard identification is typically the lengthiest component of the risk assessment process. Identifying hazards requires knowledge of both the agent and the adverse effect. Furthermore, NIOSH risk assessors approach hazard identification in terms of supporting quantification of the dose-risk relationship; therefore, its findings are

intended to define the population at risk, the agent, the adverse effect(s) of interest, and any cofactors (e.g., effect modifiers, confounders, or other sources of uncertainty) in sufficient detail to conduct sound quantitative dose-response analyses. The general framework for gathering and evaluating relevant human and animal study data consists of four basic steps: 1) define the causal questions of interest and develop criteria for study (data) selection; 2) review, identify, and select relevant information; 3) evaluate and integrate evidence across studies; and 4) synthesize and interpret findings [Rhomberg et al. 2013]. The paths to meeting these steps can vary widely with the specific scientific context. In general, risk assessors judge the weight of evidence (WoE) in study evaluation using multiple factors, such as strength of association, consistency, specificity, temporality, biological gradient, plausibility, coherence, experiment, and analogy as first posited by Sir Austin Bradford Hill [1965]. For synthesis and interpretation, risk assessors consider:

- The design and conduct of studies providing data for risk assessment to discern whether study results are generalizable and relevant to the risk assessment problem.
- The characterization of exposure, dose, and adverse effect. What is the utility of the study data for hazard identification? Will these data be suitable for inclusion in the database for the dose-response assessment?
- The degree of data certainty and strength of findings in support of hazard identification. Are results robust under alternative assumptions? How likely are findings due to chance, bias, or residual confounding?

To improve efficiency, NIOSH often utilizes hazard identification by other agencies, such as the U.S. National Toxicology Program (NTP), the U.S. Environmental Protection Agency (EPA), and the International Agency for Research on Cancer (IARC). These agencies have a long history of identifying hazards using sound and transparent methodologies.

Relevant data are primarily derived from epidemiologic and toxicologic studies. Ideally, the direct estimation of risk from human data is always preferred to data from experimental animal studies because: 1) data reflecting actual exposures and responses within the population of interest are intuitively superior for risk

assessment; and 2) the uncertainty in extrapolating data from animal toxicologic studies to predicting human risks can be much larger than that in well-designed epidemiologic studies [Hertz-Picciotto et al. 1995; Smith 1988; Stayner et al. 1999]. Although some epidemiologic data may arise from experimental designs, the vast majority of information pertinent to risk assessment is extracted from observational studies of working populations (e.g., cohort and case-control studies). Although preferred, human data are not without significant limitations; therefore, risk assessments tend to rely on a combination of epidemiologic and toxicologic data for hazard identification and dose-response analyses. It is common to find human data being weighted more than animal data in hazard identification, but be less informative on dose-response. In those instances, human studies provide evidence of an association between exposure and disease, which can guide the choice of agents, exposure routes, and pathological endpoints for examination in toxicological studies that may contribute greatest to quantifying risks.

Environmental risk assessments consider exposure assessment as a separate step for assessing the likelihood of exposure for estimating population risks and/or disease burden. In contrast, NIOSH risk assessments, as described herein, estimate the risk to a hypothetical worker from a known exposure. Although, exposure probabilities are not typically calculated, information on exposure is still needed for dose-response analyses; therefore, NIOSH systematically assesses the availability, magnitude, and validity of exposure data as a part of hazard identification. NIOSH exposure assessments are necessary to identify and characterize exposures to biological, chemical, or physical agents sufficiently to inform analyses of the dose-response association observed in exposed working populations. As such, the exposure assessment focuses on a review of methods used to estimate or measure exposure in informative epidemiologic studies and to synthesize this information for use in dose-response analyses. Specifically, the exposure assessment provides the exposure indices, and attendant uncertainties, that serve as explanatory variables in dose-response regression modeling.

ES 2.2 Dose-response Assessment

The second component of NIOSH risk assessment is the dose-response assessment. The aim of the dose-response assessment is to obtain reliable and valid estimates of the point of departure (PoD) in a cause and effect

relationship for deterministic effects, or the risk per unit dose for stochastic effects. Here, the PoD refers to a point on the dose-response curve that is established from experimental or observational data that corresponds to a level of no (or low) effect without significant extrapolation. These estimates are essential to risk characterization. NIOSH generally obtains dose-response estimates via statistical models constructed to provide the conditional expectation of the dependent variable (the adverse effect) given one or more explanatory variables, but at least including the variable describing the exposure of interest. Model input data are obtained from toxicologic and/or epidemiologic investigations that are identified and assessed in hazard identification. NIOSH risk assessors systematically select modeling data based on their contribution to the weight of evidence of one or more causal associations of interest and their suitability to modeling. As different model specifications can lead to different estimates, a key step in dose-response analysis is model selection. Clearly, it is preferable to base model selection on biologic plausibility, although a strong advantage of one model among several plausible models is rarely evident. Furthermore, data from most studies are imperfect and potentially incomplete; therefore, models may require a number of assumptions based on scientific judgment. Thus, another important part of the dose-response assessment is sensitivity analysis. In a sensitivity analysis, plausible alternative risk assessment strategies, defaults, and assumptions are quantitatively evaluated for their impact on risk estimates. In addition to providing a measure of analysis robustness, sensitivity analyses aid the risk manager by providing a range of plausible estimates of the dose-risk relationship.

ES 2.3 Risk Characterization

The final step in NIOSH risk assessment is risk characterization. It is the translation of information from hazard identification and dose-response assessment into a basis, completely or in part, for recommendations on limiting workplace exposure. For example, a linear dose-response relationship observed between chronic inhalation of methylene chloride and liver and lung tumor incidence in mice may be used to derive a limit on continuous methylene chloride exposure in the workplace that is estimated to result in an increased cancer risk in humans of about one case in 10,000. The process of transporting risks observed in animals in an experimental

study to the risk in workers exposed continuously over the course of their employment is an example of NIOSH risk characterization.

The framework of NIOSH risk characterization centers on a choice between two distinct approaches, based primarily on the evidence supporting the absence or presence of an impairment threshold. Some effects are observed only at doses above a certain level. These effects are sometimes referred to as deterministic. To address deterministic effects, NIOSH typically adjusts the PoD in dose-response analysis using factors that account for natural heterogeneity (e.g., interspecies variability, interindividual variability) to arrive at an estimate of a safe dose. Here the term 'safe' implies that excess risk at this exposure level is absent or negligible. NIOSH used this approach in its risk assessment of nonmalignant pulmonary effects from exposures to carbon nanotubes and nanofibers [NIOSH 2013b]. In contrast, consider a causal agent that is neither necessary nor sufficient to cause disease (e.g., cancer). In this case, cause and effect is best described as a relationship between the probability (but not severity) of disease and the dose level that is absent of a dose threshold. Because of the randomness inherent to cause and effect, these effects are sometimes referred to as stochastic. Cancer from low-dose ionizing radiation is a classic example of a stochastic effect. In a NIOSH risk assessment of radon exposure and lung cancer in uranium miners [NIOSH 1987], a safe level of ionizing radiation exposure was not assured; therefore, residual lung cancer risk under select exposure scenarios were estimated using probabilistic means. When effects appear stochastic, NIOSH obtains quantitative estimates of low-dose risk by model-based extrapolation of the risk at doses below the observed data. For example, probabilistic models have been used by NIOSH to estimate the dose that would cause a lifetime excess cancer risk of 1 in 1000 from occupational exposure to hexavalent chromium [NIOSH 2013a] and titanium dioxide [NIOSH 2011]. There are instances in which the risk characterization approach is less dependent on a determination of whether the process is stochastic or deterministic. For example, the threshold for a deterministic effect may reside far below the observable range in dose-response analyses and may vary widely among exposed individuals. Under this condition, NIOSH may opt for assessing lifetime risks based on model extrapolation. Similarly, an effect that is generally considered stochastic (e.g., cancer) may be

indirectly caused by exposure through a deterministic precursor effect (e.g., inflammation) residing on the causal pathway. If this is the only significant pathway present, then an exposure threshold for cancer is likely.

An important consideration of risk-based characterization is the selection of a target risk, which is a single level of risk that is broadly considered tolerable, given assurances that the risk is managed to an extent that is reasonable and practical. There are multiple methods and principles available for establishing risk acceptance criteria, and the adopted methods and principles will undoubtedly influence the choice of target risk. Thus, risk acceptance (or tolerance) criteria are more likely to be unique to the situation at-hand rather than be pre-defined [Rodrigues et al. 2014; Vanem 2012]. Nevertheless, NIOSH has established a target risk level for non-threshold carcinogens of one excess case per 10,000 workers continuously exposed over a 45-year working lifetime [NIOSH 2017]. This level is intended to be a starting point for initiating a risk management process. The setting of target risk levels for other outcomes is a fundamental component of risk management; therefore, actions are primarily the responsibility of the decision-makers and not the risk assessor. As such, a detailed discussion on the various risk management principles in play for determining these levels is beyond the scope of this report, although discussion is available in several published reports [Aven 2016; HSE 2001; Rodrigues et al. 2014; Tchiehe and Gauthier 2017; Vanem 2012]. Finally, health risk is but one aspect typically needed to derive a target risk level given that risk tolerance can depend on the combination of individual, societal, economic, and environmental impacts. Although employers in managing risks may consider these other factors, NIOSH quantitative risk assessment is solely focused on characterizing health risks.

ES 3.0 CONCLUSIONS

The quantification of occupational risk is paramount to worker protection. NIOSH has a long and rich history of systematically assessing workplace hazards and communicating recommendations aimed to mitigate associated risks. As such, NIOSH is recognized as a leader in risk assessment methods development, and its expertise is often sought by members of the risk assessment community. This report is intended to aid others in their understanding of the NIOSH risk assessment process. To this end, the report describes the NIOSH approach

to addressing hazard identification, dose-response analyses, and risk characterization, including demonstrated examples of NIOSH risk assessments.

Above all, the NIOSH approach stresses careful attention to aims of the risk assessment throughout the risk assessment process. It is important to interrogate key assumptions and provide transparency for both the main analysis and analyses of alternative modeling strategies and defaults. Maintaining mindfulness of the intended audience is of utmost importance; therefore, NIOSH risk assessors endeavor to follow the guiding principles of transparency, clarity, consistency, and reasonableness in risk characterization in conducting risk assessment (Table ES-1).

Table ES-1. Risk Assessment Guiding Principles

Principle	definition	Criteria for risk characterization
Transparency	Explicitness in the risk assessment process.	Use a risk analysis plan Describe assessment approach, assumptions, extrapolations and use of models Describe plausible alternative assumptions Identify data gaps Distinguish science from policy Describe uncertainty Describe relative strength of assessment
Clarity	The assessment itself is free from obscure language and is easy to understand.	Be brief and concise Use plain English (avoid jargon) Avoid technical terms Use simple tables, graphics, and equations
Consistency	The conclusions of the risk assessment are characterized in harmony with other NIOSH actions.	Use this technical report Follow NIOSH policies on technical writing and peer review Place assessment in context with similar risk assessments
Reasonableness	The risk assessment is based on sound judgment.	Use review by peers Use best available scientific information Use good judgment

Adopted from the EPA Risk Characterization Handbook [Fowle and Dearfield 2000]

Risk assessment science is continuously evolving. Methods currently under development may provide additional, powerful tools to assess risks to workers based on very limited data. Validation of these new approaches is a critical need. In efforts to stay abreast of the science, NIOSH will continue to embrace new methodologies, but will do so with appropriate caution and deliberate evaluation of new techniques and approaches.

Table of Contents

Appendices

1.0 BACKGROUND

In their highly cited article, Kaplan and Garrick [1981] first posited that an analysis of 'risk' is an effort to answer three questions:

1. What can happen?
2. How likely is it that it will happen?
3. What are the consequences if it does happen?

Thus, hazards (the risk source) impose risks that are functions of both likelihood and consequence. Risk is omnipresent and diverse in the human experience; therefore, steps are necessary to manage the many different kinds of risks in our daily lives, such as business risk, social risk, political risk, and occupational risk.

Occupational risk is defined as the potential and severity of adverse effects in workers from their exposure to workplace hazards. In this context, the adverse effect of interest is simply a specified unfavorable change in health status of a worker from known exposure. Occupational risks are reduced by safeguards that are carefully derived from scientific assessment and best practices. Risk assessment is an important tool for informed decision-making on workplace safeguards. For example, risk assessment provides the scientific underpinnings to authoritative recommendations, such as occupational exposure limits (OELs). In particular, risk assessment conducted by the National Institute for Occupational Safety and Health (NIOSH) has provided the foundation for Recommended Exposure Limits (RELs) and Risk Management Limits for Carcinogens (RML-CAs) for chemicals and other workplace hazards, such as ionizing radiation and noise.

NIOSH first considered the need to quantify occupational risks when OSHA's standards were challenged in the 1980s, resulting in the well-cited Supreme decision: "Industrial Union Department, AFL-CIO v. American Petroleum Institute", hereafter referred to as the Benzene Decision, 448 U.S. 607 [Industrial Union Department, AFL-CIO v. American Petroleum Institute 1980]. In response, NIOSH developed a "Risk Assessment Team," which was later expanded to a "Risk Assessment Activity." This group of toxicologists, epidemiologists and

statisticians provided quantitative risk assessments for radon; ethylene glycol monomethyl ether, ethylene glycol monoethyl ether, and their acetates; cadmium; 1,3-butadiene [Dankovic et al. 1993]; and coal dust [Kuempel et al. 1997; NIOSH 1991; NIOSH 1987; Stayner et al. 1992a; Stayner et al. 1992b]. The "Activity" was formally organized within NIOSH in 1995. Since that time, NIOSH staff have conducted quantitative risk assessments for a wide variety of agents, including: diesel exhaust [Stayner et al. 1998], 1,3-butadiene [Stayner et al. 2000b], asbestos [Stayner et al. 1997], silica [Park et al. 2002; Rice et al. 2001], noise (with and without co-exposure to carbon monoxide) [NIOSH 1998], titanium dioxide [NIOSH 2011], hexavalent chromium [NIOSH 2013a], manganese in welding fume, carbon nanotubes and nanofibers [NIOSH 2013b], diacetyl and 2,3-pentanedione [NIOSH 2016a] (Table 1-1).

Table 1-1. Examples of NIOSH quantitative risk assessments

Agent	Adverse Effect[1]	Dose-response assessment[2]	Risk Characterization	Reference
1,3-butadiene	leukemia	toxicologic, Weibull time-to-tumor regression model, animal to human extrapolation. epidemiologic and toxicologic, literature review.	extrapolation, excess lifetime risk, target risk unspecified	[Dankovic et al. 1993] [Stayner et al. 2000a]
asbestos	lung cancer, asbestosis	epidemiologic, Poisson regression, additive relative rate function (cancer), power function (asbestosis)	extrapolation, excess lifetime risk, target risk unspecified	[Stayner et al. 1997]
cadmium	lung cancer	epidemiologic, Poisson and Cox PH regression, additive relative rate function	extrapolation, excess lifetime risk, target risk unspecified	[Stayner et al. 1992a; Stayner et al. 1992b]
carbon nanotubes and nanofibers	non-malignant adverse lung effects	toxicologic, NOAEL and BMD assessments	PoD/UF	[NIOSH 2013b]

Agent	Adverse Effect[1]	Dose-response assessment[2]	Risk Characterization	Reference
coal mine dust	coal workers' pneumoconioses, progressive massive fibrosis, pulmonary dysfunction	epidemiologic, logistic and multiple linear regression	extrapolation, excess lifetime risk, target risk unspecified	[Kuempel et al. 1997]
diacetyl and 2,3-pentanedione	pulmonary dysfunction	epidemiologic, linear extrapolation, multiple regression	extrapolation, excess lifetime risk, 10^{-3} target risk	[NIOSH 2016a]
diesel exhaust	lung cancer	toxicologic and epidemiologic (review)	extrapolation, excess lifetime risk, target risk unspecified	[Stayner et al. 1998]
EGME, EGEE, EGMEA, EGEEA	reproduction, developmental, hematotoxic effects	toxicologic, NOAEL and LOAEL assessments	PoD/UF	[NIOSH 1991]
hexavalent chromium	lung cancer	epidemiologic, Poisson regression linear ERR model	extrapolation, excess lifetime risk, 10^{-3} target risk	[NIOSH 2013a; Park et al. 2004]
noise	material hearing impairment	epidemiologic, logistic regression	extrapolation, excess lifetime risk with no target risk level specified	[NIOSH 1998; Prince et al. 2003]
radon	lung cancer	epidemiologic, Cox proportional hazards regression	extrapolation, excess lifetime risk, target risk unspecified	[Hornung and Meinhardt 1987; NIOSH 1987]
silica	lung cancer	epidemiologic, Poisson regression, additive relative rate function	extrapolation, excess lifetime risk, target risk unspecified	[Rice et al. 2001]
silica	non-malignant lung disease	epidemiologic, Poisson regression, additive relative rate function	extrapolation, excess lifetime risk, target risk unspecified	[Park et al. 2002]
titanium dioxide	lung cancer	toxicologic, nonlinear extrapolation, BMD model averaging, quantal endpoint	extrapolation, excess lifetime risk, 10^{-3} target risk	[NIOSH 2011]

1. Analyses may have considered multiple adverse effects. The adverse effect shown in the table was selected as the primary effect in the risk assessment.

2. The dose-response assessment refers to the primary source supporting final models and/or recommendations on risk-based exposure limits.

Abbreviations: BMD, benchmark dose; EGEE, ethylene glycol monoethyl ether; EGEEA, ethylene glycol monoethyl ether acetate; EGME, ethylene glycol monomethyl ether; EGME, ethylene glycol monomethyl ether acetate; ERR, excess relative rate; LOAEL, lowest observable adverse effect level; NOAEL, no observable adverse effect level; PH, proportional hazards; PoD, point of departure; UF, uncertainty factor.

In general, risk assessment is a process in which hazard, exposure, and dose-response information are evaluated to characterize risk. These evaluations determine whether an exposed population is at greater-than-expected risk of adverse effects, such as disease (cancer or non-cancer) or injury. Once the hazard is identified, the magnitude and nature of the increased risk can be explored further, using either qualitative or quantitative approaches. Qualitative risk assessments are typically descriptive and indicate whether a particular adverse effect, such as disease or injury, is likely or unlikely under specified conditions of exposure. Quantitative risk assessments provide a numerical estimation of risk based on mathematical modeling. For example, a quantitative risk assessment may be used to relate conditions of workplace exposure to a value of increased lifetime risk of a disease or injury.

Quantitative risk assessments require: 1) data on exposures relevant to the adverse effect of interest; 2) data on the adverse effect associated with the exposure of interest; and 3) a mathematical model describing that dose-response relationship. Risk assessments based on epidemiologic, population-based studies have real-world relevance to workers, but they generally suffer from a number of limitations inherent to study design and available data. Risk assessments based on experimental animal data provide detailed information on the dose-response relationships; however, there is often concern about the validity of extrapolating animal-based risk assessments to humans who generally have much lower and more variable exposures. The integration of mechanistic, animal, and human data is important for developing a thorough understanding of the risks.

The risk assessment process has become increasingly complex over the past decades. In occupational safety and health regulation, the need to quantify risk became apparent with the Benzene Decision, which established that the Occupational Safety and Health Administration (OSHA) could not issue a standard without demonstrating a significant risk of material health impairment. The ruling allowed (but did not require) for numerical criteria to be used to determine whether a risk is "significant." As a result, risk assessment became standard practice in OSHA rulemaking for health standards, and quantitative risk assessments are now preferred whenever data, modeling techniques, and biological understanding are adequate to support their development.

NIOSH has adopted many of the same risk assessment practices as OSHA in order to keep the analyses relevant and meaningful within OSHA's regulatory context.

1.1 NIOSH Risk Assessment History

Historically, NIOSH employed a variety of methods to establish recommendations intended to prevent adverse effects in workers. NIOSH considered the health effects associated with experimental or observational concentrations and applied a safety factor to ensure that even the most susceptible individual would be generally protected from a hazard. One major exception to that was in addressing issues of carcinogenicity. When evaluating carcinogens, NIOSH typically assumed that no exposure could be considered safe. This led to RELs for carcinogens that were not numerical, but directed employers to keep exposures as low as feasible [Fairchild 1976].

In its first decades, NIOSH was largely uncertain about the utility dose-response modeling, especially for carcinogens. In 1982, NIOSH commented to OSHA on an Advance Notice of Proposed Rulemaking on the Identification, Classification, and Regulation of Potential Occupational Carcinogens [NIOSH 1982]:

> *Because our understanding of the mechanism of carcinogenicity is incomplete, our use of mathematical models to predict its outcome must be employed with extreme caution. To select a model or models from among the many choices and to have them incorporated into Administration policy will not resolve those issues.*

However, just a few years later, NIOSH engaged in quantitative risk assessment. As cited in 1986 NIOSH testimony on OSHA's Proposed Rule on Occupational Exposure to Benzene [NIOSH 1986a], NIOSH drew on the benzene decision [Industrial Union Department, AFL-CIO v. American Petroleum Institute 1980], which focused on significant risk of material impairment of health, and the "lead decision" [United Steel Workers of America, AFL-CIO v. F. Ray Marshall 1980], which discussed acceptable risk of occupational hazards, to conclude that:

These two decisions provided the impetus for the inclusion of a quantitative risk assessment effort in the standards recommending program of NIOSH [NIOSH 1986a].

The first reference to quantitative risk assessment in NIOSH policy statements was in 1986 NIOSH testimony to OSHA recommending 0.1 ppm as a permissible exposure limit (PEL) for benzene, based largely on findings from a NIOSH risk assessment using epidemiologic data [Rinsky et al. 1987]. Risks at 0.1 ppm were determined to be around one excess cancer per 1000 workers over a working lifetime. However, this initial risk-based REL was never incorporated into an updated Criteria Document for benzene and later documentation referred to the limit of quantification (LOQ) of the analytical method (also around 0.1 ppm at the time) as the basis of the 0.1 ppm REL [NIOSH 1988]. Here, the LOQ is the amount or concentration of the analyte at which quantitative results can be reported with a high degree of confidence, which is based on assay-specific acceptance criteria [NIOSH 1995b].

In 1986, although NIOSH did not conduct its own risk assessment for formaldehyde, NIOSH testified that the OSHA risk assessment for formaldehyde was acceptable [NIOSH 1986b]. This risk assessment used animal bioassay data to estimate the human cancer risk of 3.46 cases per 1000 workers exposed over a working lifetime at the proposed PEL of 3 ppm. However, the NIOSH REL of 0.016 ppm as an 8-hour time-weighted average (TWA) was based on the lowest concentration that was considered ‘quantifiable’ at the time.

In 1987, NIOSH published its first Criteria Document to include a quantitative risk assessment: *Criteria for a Recommended Standard for Occupational Exposure to Radon* [NIOSH 1987]. The risk assessment was based on epidemiologic data on excess lung cancer in underground uranium miners exposed to radon [Hornung and Meinhardt 1987]. The risk assessment found that continuous exposure to radon progeny concentrations of one Working Level Month (WLM) annually over a working lifetime corresponded to 5-10 excess lung cancers per 1000 miners. The risk from radon exposure versus the feasibility of controlling exposures was a point of discussion in the document. The REL was ultimately based on the limits of control technology at the time;

however, NIOSH also communicated risks at this level, which supported additional recommendations for continued control technology development.

NIOSH risk assessments during the 1990's largely incorporated data from well-designed occupational epidemiologic studies that had become a mainstay of the Institute's field studies program. Epidemiological risk assessments of lung cancer in humans were conducted for cadmium, chrysotile asbestos, and diesel exhaust [Stayner et al. 1998; Stayner et al. 1997; Stayner et al. 1992a; Steenland et al. 1998]. Cancer as a result of worker exposures to ethylene oxide was also examined [Steenland et al. 2003]. NIOSH also conducted worker-based risk assessments for various lung function measures after coal dust exposure and hearing loss after noise exposure [NIOSH 1995a; Prince et al. 2003]. In addition, although not a complete risk assessment, physiologically based pharmacokinetic (PBPK) modeling was used for dose estimation in a worker study of 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin (TCDD) exposure [Lawson et al. 2004]. Animal-based risk assessments were conducted to predict human risks in the absence of sufficient human data. Toxicologic-based examples include assessments of 1,3-butadiene and cancer at various sites in the mouse and rat, and glycol ethers and reproductive effects in the mouse, rat and rabbit [Dankovic et al. 1993; NIOSH 1991; Stayner et al. 2000b].

In the 2000's and beyond, the need for quantitative risk estimates preferentially based on epidemiologic data resulted in risk assessments becoming increasingly complex. Advances in risk assessment have included innovations in reconstructing past exposures in epidemiological studies, expansion of statistical modeling techniques, increased understanding of the role of particle dosimetry issues in risk assessment, and exploration of dose-response modeling for non-cancer health endpoints. New techniques in statistical modeling methods to account for survivor bias in human studies, incorporating genetics and genomics into risk assessment, and the potential for using quantitative structure activity relationships for risk assessment pose many challenges and opportunities for the future [Buckley et al. 2015; Comber et al. 2003; Weitzel et al. 2011].

2.0 PURPOSE AND SCOPE

Quantitative risk assessment is the foundation of authoritative recommendations. In particular, NIOSH conducts high quality, scientifically sound quantitative assessments of workplace hazards as input to developing its Criteria Documents, including establishing the basis for RELs and alternative forms of authoritative recommendations, such as hazard banding. This document describes the underlying science and general approach used by NIOSH researchers when conducting risk assessments. This information is intended for scientists, stakeholders, and the public to improve their understanding of the NIOSH risk assessment process. It should be understood that every risk assessment is unique; therefore, situations may arise which require steps that are not specifically addressed in this report. Furthermore, discussion on NIOSH risk management and risk communication practices that typically follow the completion of its risk assessments are beyond the scope of this report.

The report is structured to follow the progression of a typical risk assessment, which is defined as the determination of the relationship between the predicted exposure and adverse effects in workers in four major steps: hazard identification, dose–response assessment, exposure assessment, and risk characterization. Tasks within the risk assessment process involve the following three sequential steps (Figure 2-1):

1. Hazard Identification is the identification of the type and nature of adverse effects that an agent has an inherent capacity to cause in an organism, system, or population. An adverse effect is defined as the specified change in the morphology, physiology, growth, development, reproduction, or life span of an organism, system, or population that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress, or an increase in susceptibility to other influences [IPCS 2004]. Hazard identification is the initial stage of the risk assessment. The products of hazard identification are characteristic descriptions and data on the exposure of interest, any important covariates, mode of action, and adverse effects. Preferably, these data are suitable for quantifying the

dose-response relationship. Therefore, hazard identification is the necessary antecedent to dose-response assessment.

2. Dose-response assessment, is an analysis of the relationship between the total amount of an agent administered to, taken up by, or absorbed by the organism, system, individual, or population and the adverse effect developed in that organism, system, individual, or population in reaction to that agent. Desired products of the dose-response assessment are estimates of the risk per unit dose having reasonable statistical properties for use in quantitative risk characterization. In lieu of sufficient data for quantification, the dose-response may be qualitatively described.
3. Risk characterization is the qualitative and, wherever possible, quantitative determination, including attendant uncertainties, of the probability of occurrence of known and potential adverse effects of an agent in workers under defined exposure conditions. It reflects the integration of the sciences from the two preceding steps (i.e., hazard identification and dose-response assessment) with additional information necessary to complete the basis for the REL or other supported recommendation. Some of this information may be based on policy rather than science; therefore, risk characterization is also a component of the risk management process.

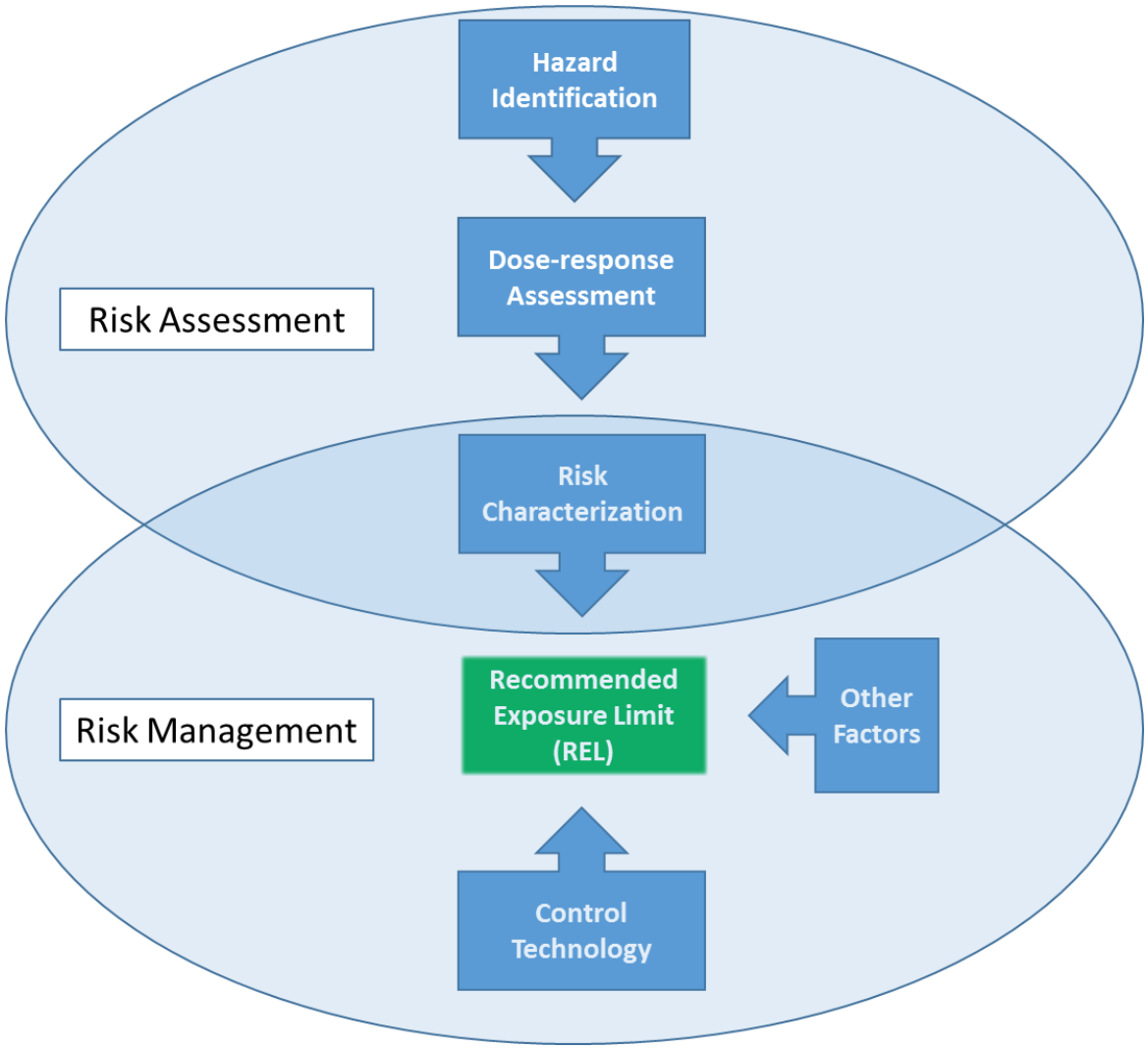

Figure 2-1. NIOSH Risk assessment and risk management processes.

This basic construct is used throughout the risk assessment community; however, some authoritative bodies have described the process using different terms and groupings of steps. Notably omitted from the NIOSH risk assessment process is *exposure assessment*, which follows the dose-response assessment in some risk assessment paradigms (e.g., IPCS, EPA). Exposure assessment is defined as the process of estimating or measuring the magnitude, frequency, and duration of exposure to an agent, along with the number and characteristics of the population exposed. Traditionally, it is defined as providing information on sources, pathways, and routes of exposure necessary to be used in conjunction with dose-response information to project

population risks. This strict definition stems from environmental risk assessment. For example, an exposure assessment may be used in "margin of exposure" analysis, which compares the dose-response assessment with the exposure assessment in the population of interest in order to design appropriate interventions. This type of analysis is not directly applicable to NIOSH risk assessments in which the aim is to estimate human health risks at prescribed exposure levels to support RELs and other occupational exposure limit recommendations. Nevertheless, exposure information is input for dose-response analyses; therefore, the quality of exposure data ultimately used to describe the dose-risk relationship must be assessed as a component of hazard identification. Furthermore, NIOSH exposure assessment, in conjunction with the REL and other risk assessment information, is key for informing risk management decisions.

3.0 RISK ASSESSMENT PLAN (PROBLEM FORMULATION)

In general, a risk assessment has two distinct initiating stages: 1) planning and scoping, and 2) problem formulation (Figure 3-1) [NRC 2009]. Planning and scoping typically involves a dialogue between stakeholders and risk managers (with support from risk assessors) on the hazards and potential risk mitigation strategies, including conceptualizing the need, purpose, structure, and content of a risk assessment to aid in decision-making. Problem formulation occurs from communication between risk managers and risk assessors (with support of stakeholders) on the technical design of the risk assessment, which uses the broad concepts developed in planning and scoping. Although planning and scoping provide input into problem formulation, and therefore is first initiated, activities in both stages will likely progress concurrently. Nonetheless, risk assessors are primarily tasked with problem formulation, which is described in this chapter.

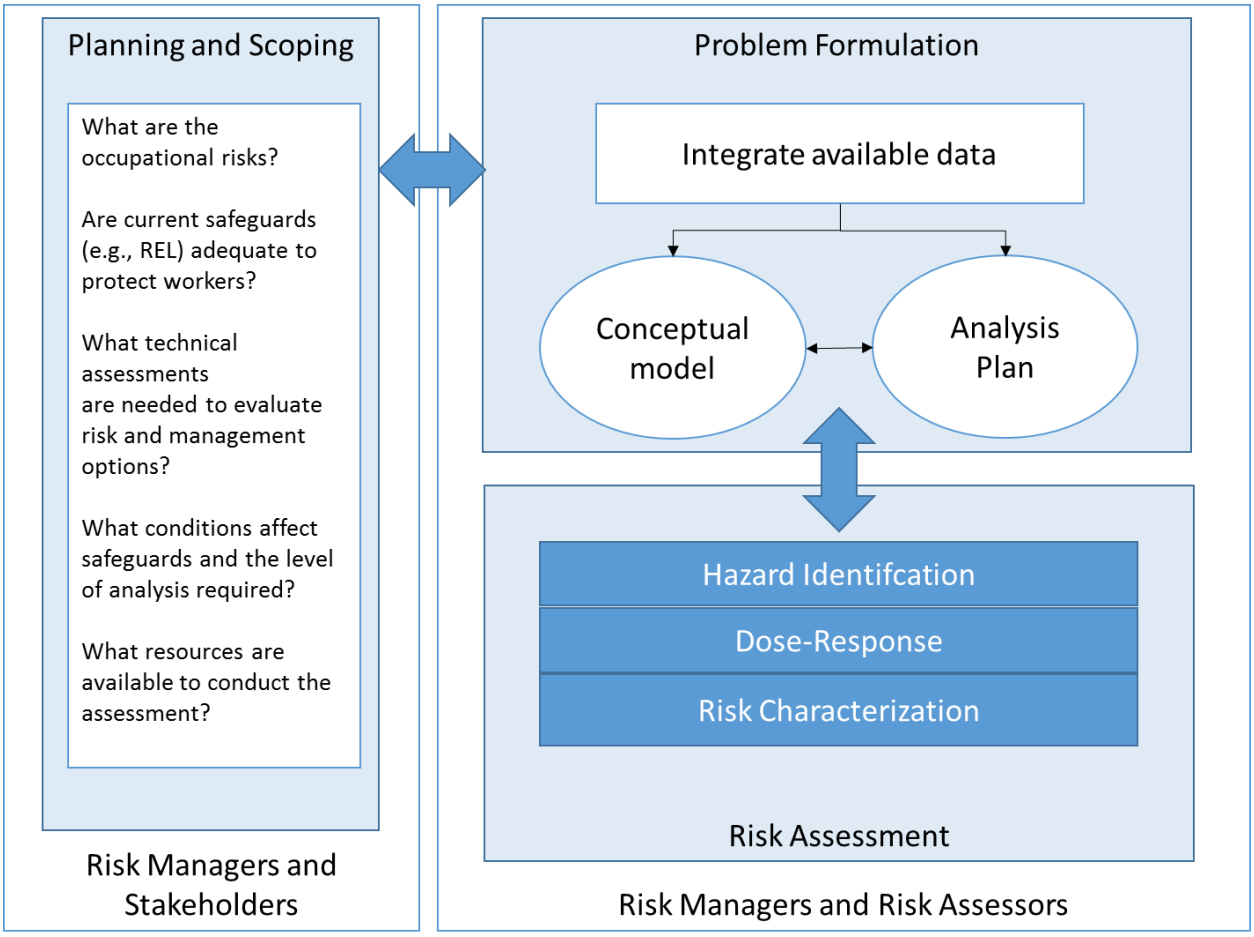

Figure 3-1. The Interrelationship between planning and scoping, problem formulation, and risk assessment.

At the outset of problem formulation, NIOSH investigators develop a risk assessment plan that contains two critical components; 1) a conceptual model that identifies the hazard (sources, stressors, and pathways), persons at risk, and potential adverse effects for analysis and 2) an analysis plan (work plan) that outlines the analytic components (i.e., data and methods) and interpretative approaches (e.g., risk metrics) to be used [NRC 2009]. The conceptual model guides decisions on data needs and the analysis plan matches elements of the conceptual model with a proposed analytic approach. The two overarching principles in developing the plan are: 1) to ensure that the risk assessment uses the best scientific methods and the highest-quality evidence, and 2) to address the needs of the decision-makers (risk managers). Thus, in the problem formulation stage, it is imperative

to be mindful that the risk assessment serves both scientific and communicative needs. This is best accomplished by including input from both scientists and decision-makers in the design of the risk assessment plan, if practical. For complex risk assessments, the analysis plans and conceptual models may benefit from peer review.

One way to formulate the risk assessment plan is to use a series of questions that the risk assessment is intended to address. These could include:

- What agents are involved?
- Who is potentially at risk?
- What are the characteristics of the potential adverse effects caused by the hazard?
- What types of data will be used to support or inform the risk assessment process?
- What dose-response data will be included? For example, what criteria will be used for determining the acceptability of experimental animal data with inhalation as route of exposure?
- How will exposure be expressed (e.g., inhaled dose, absorbed dose, air concentration)? What are the reasonable alternative expressions and how would using those alternatives change the risk assessment?
- If risk quantitation is based on an internal measure of dose, then how will this measure be related to an external exposure supporting a REL?
- How are the health effects defined and measured? Are health effects aggregated (e.g., all cancer)? If aggregated, how would using alternative aggregation strategies alter the risk assessment?
- What basic dose-response relationship should be assumed in analyzing the data? What types of causal mechanisms are likely to be involved, and what tentative inferences should be drawn from those mechanism types to guide initial representations of the data for analysis?
- How should one deal with the background (control) incidence of different effects? Should it be assumed that the processes producing the observed health effects in control animals (or unexposed populations)

interact with processes by which the hazard of interest causes these effects, or should the toxic mechanisms by treated as if they were independent of background processes?

- How will important sources of uncertainty be evaluated and/or accounted for in the risk assessment? Are there any reasonable anticipated adjustments to the exposure or health effect based on mechanism, metabolism, potential confounding factors, other exposures, or other factors that should be considered? What is the anticipated impact on the risk assessment?
- How will the final risks be expressed and, if quantitative analysis, what target risk levels are used? What is the support for those decisions and are there reasonable alternatives? If yes, how would using those alternatives affect the risk assessment?
- What is the timeframe for completing the assessment?

Additional questions may be considered in the context of the risk assessment. As necessary, the risk assessor refers to the plan throughout all aspects of the risk assessment. Appropriate plans include sufficient detail so that another risk assessor could reproduce the analysis. If the risk assessor decides that deviations from the plan are needed, the plan is amended with clear indication that alterations were made, including justification for the alteration.

4.0 HAZARD IDENTIFICATION

The first step in occupational risk assessment is hazard identification, which is the process of determining whether an agent of interest (e.g., benzene) causes an adverse effect (e.g., leukemia) in exposed workers. This process involves characterizing the nature and strength of the evidence of causation [NRC 1983], hereafter referred to as the 'weight-of-evidence' (WoE). The 'evidence' in this case is information on the agent, the adverse effect(s), and their association. Ideally, this evidence will serve as input to the dose-response assessment to support quantitative risk assessment. Evaluating the WoE requires a systematic approach to critically assess and interpret the body of scientific information. This information may stem from epidemiologic studies, animal bioassays, mode-of-action studies, metabolic studies, genetic and epigenetic studies, and *in vitro* studies, which all fall under the general categories of either human or animal data. Human data sources are preferred for quantifying occupational risks; however, hazard identification has at times relied solely on animal data or a combination of human and animal data. When used in combination, either data source may take a supportive role in the risk assessment.

Although methods may differ, the general framework of an acceptable approach to gathering and evaluating relevant human and animal study data consists of four basic steps: 1) define the causal questions of interest and develop criteria for study (data) selection; 2) review, identify, and select relevant information; 3) evaluate and integrate evidence across studies; and 4) synthesize and interpret findings [Rhomberg et al. 2013]. The paths to meeting these steps can vary widely with the specific scientific context of the risk assessment; therefore, precise methods for assessing WoE cannot be prescribed without understanding the individual context, although general guidelines are available [Higgins and Green 2008; NRC 1983; NTP 2015a; NTP 2015b; Rhomberg et al. 2013; WHO 2000]. As such, risk assessors strive to develop and describe their approach in sufficient detail to ensure a transparent and defendable standard of WoE is met for their evaluation [Weed 2005]. Broadly, risk assessors consider:

- The design and conduct of studies providing data for risk assessment. Are study results generalizable and relevant to the risk assessment problem?
- The characterization of exposure, dose, and adverse effect. What is the utility of the study data for hazard identification? Will these data be suitable for inclusion in the database for the dose-response assessment?
- The degree of data certainty and strength of findings in support of hazard identification. Are results robust under alternative assumptions? How likely are findings due to chance, bias, or residual confounding?

In practice, NIOSH risk assessors have sometimes relied on hazard identification by other agencies, such as the U.S. National Toxicology Program (NTP), the U.S. Environmental Protection Agency (EPA), and the International Agency for Research on Cancer (IARC) to inform NIOSH risk assessment. These agencies have a long history of hazard identification using sound and transparent methodologies. The NIOSH Chemical Carcinogen Policy [2017] provides additional information on the use of available cancer hazard assessments. Hazards have also been identified by recent research that has not been reviewed and synthesized by these agencies. This occurs most often in cases where emerging hazards have been identified or when new information on an existing hazard becomes available. In all cases, NIOSH risk assessors evaluate, integrate, and synthesize the existing evidence to characterize the hazard for dose-response analyses and risk characterization. This is accomplished using best practices of the many frameworks established for hazard identification. These practices are discussed in comprehensive reviews [Higgins and Green 2008; NRC 1983; Rhomberg et al. 2013], recent commentaries [Howard et al. 2017; Woodruff and Sutton 2014] and technical reports [NTP 2015a; NTP 2015b; WHO 2000]. In addition, tools for conducting and assessing systematic reviews are available to NIOSH risk assessors, such as A MeaSurement Tool to Assess systematic Reviews (AMSTAR, https://amstar.ca/index.php) and a recent report commissioned by NIOSH as an aid for conducting systematic reviews (http://www.rand.org/pubs/research_reports/RR1463.html). Historically, NIOSH risk assessors have utilized

thorough literature reviews as the foundation for risk assessments. This new report on systematic reviews will serve as an important resource to guide future reviews.

In summary, NIOSH risk assessors are mindful that a 'hazard' is defined as the inherent property of an agent (or situation) having the potential to cause an adverse effect when an organism, system, or population is exposed to that agent. Thus, identifying hazards requires knowledge of both the agent and the adverse effect. Furthermore, NIOSH risk assessors approach hazard identification in terms of supporting the next step in the risk assessment; therefore, data must sufficiently define dimensions of the population at risk, the agent, the adverse effect(s) of interest, and any cofactors (e.g., effect modifiers, confounders, or other sources of uncertainty), which are necessary for conducting sound quantitative dose-response analyses.

4.1 Hill's Views on Causation

Observed associations are typically evaluated by NIOSH against multiple factors to assess WoE. The framework used to make an assessment is likely to be specific to the problem at hand; however, there are a numerous WoE frameworks available to the risk assessor for planning an approach [Rhomberg et al. 2013]. Perhaps the most widely known WoE framework for data integration and evaluation of causation was introduced by Sir Austin Bradford Hill [1965], who proposed nine aspects of association commonly referred to as "Bradford Hill Criteria". These aspects comprise strength of association, consistency, specificity, temporality, biological gradient, plausibility, coherence, experiment, and analogy. It is important to note using these aspects to weight data is but one approach; Hill cautioned against the use of his views as a set of definitive rules and acknowledged that many additional factors may be equally if not more important to WoE. Similar concerns have surfaced in several contemporary critical assessments of Hill's views [Fedak et al. 2015; Hofler 2005; Howick et al. 2009; Ioannidis 2016; Phillips and Goodman 2004; Thygesen et al. 2005]. Thus, the term 'guidelines' is preferred to 'criteria' as posited by Howick et al. [2009]. More information on formulation of Hill's guidelines, including critical assessments of their use in causal inference, is available in the assessments referenced above and in

seminal epidemiologic texts [Checkoway et al. 2004; Rothman et al. 2008]. The guidelines are briefly described below:

- Strength of Association refers to the magnitude and statistical precision of the observed association, whereby a 'strong' association is less likely to be influenced by unmeasured confounders, other sources of bias, or chance alone. Thus, this aspect addresses the feasibility of statistical inference. A strong association is not necessary nor sufficient for a causal relationship. For example, the association between cardiovascular disease and smoking is considered relatively weak; however, it is also considered causal. Conversely, an effect estimate achieving statistical significance provides little evidence of causality without due consideration of other aspects, such as underlying statistical methods, biologic plausibility, and reproducibility of results.

- Consistency refers to the reproducibility of similar effects in different populations (studies). Generally, evidence from a series of studies reporting similar effects is weighted more than findings from a single study. Like strength of association, consistency also addresses the feasibility of statistical inference because increased homogeneity across studies is evidence against poor internal validity. Nevertheless, consistency is neither necessary nor sufficient for a causal relationship.

- Specificity, in Hill's view, is the simple premise that an association is more likely to be causal if it is observed between one cause and one effect. Of course, specificity is reliant on the definitions of the cause (exposure) and effect (disease). In practice, epidemiologic examinations tend to involve complex exposures and multifactorial diseases with similar pathways; therefore, highly specific agent-disease associations are seldom observed. For this reason, many consider specificity to be of less importance for causal inference in most settings.

- Temporality refers to the general acceptance that the cause (exposure) must precede the effect (disease) in time. This is the only criterion that is considered necessary for a causal relation. Thus,

study designs that firmly maintain the temporal progression from cause to effect are far more persuasive in causal inference.

- Biological Gradient refers to the observed presence of a dose-risk relationship (i.e., dose-response). Typically, this is defined as a monotonic trend in disease frequency with increasing levels of exposure. Studies designed to examine dose-response trends are more persuasive for causal inference. Nonetheless, the absence of a monotonic biologic gradient does not preclude the existence of a causal relationship. This aspect of Hill's guidelines is the focus of NIOSH quantitative risk assessment, which is exploited by the dose-response modeling described in Section 5.0.

- Plausibility refers to a measure of biologic reasonableness for explaining the agent-disease association. The guideline is largely a function of the current understanding on toxicity and disease etiology. It is important to synthesize evidence from a wide array of animal and human studies to assess the plausibility of an association between contributing causes and complex diseases. Toxicological data from experimental animal studies can be particularly useful for assessing biological plausibility. For example, agents that cause similar toxicity in animals as that observed in humans is strong evidence of biological plausibility.

- Coherence is related to plausibility; it implies that the interpretation of a causal association is in agreement with known disease etiology. Of course, coherence relies on current knowledge, which is always subject to change. Hill stated that the absence of coherent information should not be considered as evidence against causation. In contrast, the presence of conflicting information is counter to causality. The risk assessor must judge whether the conflict is true (thus potentially negating a cause-and-effect relationship), or false due to study errors or misinterpretation.

- Experiment refers to evidence of a successful intervention; that is, removing (or reducing) the cause results in the disappearance (or attenuation) of the effect. For example, lower lung cancer rates have

followed patterns of decreased smoking. This observation supports the hypothesis that lung cancer is caused by smoking. Hill considered this criterion as "… the strongest support for the causal hypothesis". However, evidence from interventions is rarely available to risk assessors.

- Analogy is related to plausibility; if a causal association is apparent with an agent, then the standard of evidence is lessened for similar agents by analogy. For example, human data on the toxicity of diacetyl are believed informative on risks from exposures to the chemically similar agent 2,3-pentanedione, for which human data are unavailable.

4.2 Laboratory Animal Data

Data from human studies are often inadequate to fulfill hazard identification; therefore, toxicological information from bioassays in animals is used, either alone or in combination with information from human studies. In general, animal studies have been shown to predict human health risks very well [Allen et al. 1988; Crump et al. 1989; Griffin 1986]. In addition, animal models are accepted as valid models for screening potential hazards to humans. However, NIOSH recognizes that some differences exist between species because of the unique inherent physiological, and biochemical mechanisms in each species [Homburger 1987]. Although in general, it is typically assumed that humans are more sensitive to a test compound than experimental animals [Lasagna 1987], there are cases in which this is demonstrably untrue. The case of d-limonene exposure causing kidney tumors in male rats, but not in female rats or either sex of mice, is one example. The male rat kidney tumors have been linked to a metabolite of d-limonene binding to the protein, alpha-2u-globulin, leading to toxicity, cellular regeneration and tumor formation. Humans have no functionally similar protein, therefore this mode of action does not appear to operate in humans, and the male kidney tumors do not indicate a human cancer risk. This case and other examples are summarized in Cohen et al. [2004]. In this paper, the authors describe a framework for evaluating the relevance of chemically induced animal tumors to humans. As part of its hazard identification, NIOSH evaluates mode of action and adverse outcome pathway information in order to determine

whether sufficient evidence exists to establish that the events leading to adverse effects in animals are unlikely to operate in humans.

Animals most often used in the bioassays include rat, mouse, guinea pig, hamster, and rabbit. In addition, monkey, and dog are also used in some bioassays. Codified U.S. EPA guidelines for animal toxicity studies are provided in Title 40, Subchapter R-Toxic Substances Control Act, Part 798 (40 CFR 798), Health Effects Testing Guidelines [1998]. International guidelines are provided by the Organisation for Economic Co-operation and Development (OECD) in: OECD Guidelines for the Testing of Chemicals, Section 4, Health Effects, Test 403, 412, 413, and 452 for inhalation exposure studies [OECD 2009a; OECD 2009b; OECD 2009c; OECD 2009d]; and in the OECD Guidelines for the Testing of Chemicals, Section 4, Health Effects, Test 402, 404, 410, 411, and 429 for dermal toxicity studies [OECD 1987; OECD 2015; OECD 1981a; OECD 1981b; OECD 2010]. These guidelines provide recommendations on physical parameters of test substances and testing conditions; on laboratory animals (e.g., species, number, sex, age, and condition); and on gross and histopathology, and clinical, biochemical, hematological, ophthalmological, and urinary excretion tests to be included in the study. It is generally recommended to conduct toxicity tests for each test compound in at least two species, typically rats and mice [Bingham et al. 2001; Salem and Katz 2014]. In addition to improving consistency between studies, these test guidelines can be a useful for evaluating WoE among studies, where deviations from these guidelines can indicate potential weaknesses.

As in human studies, not all toxicological studies of a particular agent are equally useful. Some studies are limited by virtue of their sample size, experimental design, methods, and the interpretation of the results by authors. It is very important that the toxicity evaluation of a substance be based on information from well-conducted studies. Evaluation of the quality and reliability of individual animal toxicity studies requires consideration of factors associated with a study's hypothesis, design, methods, execution, analysis, and interpretation [Hothorn 2014; Klimisch et al. 1997; Lu and Kacew 2002; NTP 2015a; NTP 2015c; Salem and Katz 2014].

4.2.1 Relevance and Appropriateness of the Animal Model

A relevant and appropriate animal model of human disease is one that includes a living organism in which normative biology or behavior can be studied, or in which a spontaneous or induced pathological process can be investigated, and in which the phenomenon in one or more respects resembles the same phenomenon in humans [NRC 1981]. The term "relevance" refers to the comparability of the observations in animals to those in humans. Clearly, a preferred model is one in which the phenomenon of interest is observed equally in both animals and humans. In practice; however, the degree of direct comparability can be low, which is a limitation in animal studies. Limited comparability does not preclude the use of animal information in human risk assessment. In fact, animal studies may provide the best dose-response information to support human risk assessment. The term "appropriateness" refers to factors that support the choice of animal model for risk assessment, depending on the scientific questions to be addressed. These factors can include animal life-span; genetic homogeneity; specific anatomical, physiological, or behavioral attributes; the frequency of the effect of interest and its background occurrence; availability (supply and cost); and other factors [NRC 1981].

Understanding the mode of action of a chemical helps to establish the best animal model for use in the toxicity testing and risk assessment (mode of action is further described in Section 4.2.2.2). For example, male rats are not a useful model for evaluating the risk of kidney cancer from gasoline exposure in humans [Baetcke et al. 1991]. The mode of action for kidney cancer in male rats from gasoline exposure (as described above also for d-limonene exposure) involves the presence of alpha-2-u globulin protein. This protein combines with the metabolites of gasoline and eventually induces kidney tumors. Humans, female rats, and mice do not have this protein. From everything known to date, the presence of alpha-2-u globulin is necessary for the development of kidney tumors; therefore, no excess kidney cancer has been observed in exposed mice or female rats. For test compounds that depend on a metabolite to produce an adverse effect in the animal, the most appropriate animal species is often the one that has the closest similarity to humans with respect to relevant metabolic processes involved in toxicity of the putative toxicant [Bogaards et al. 2000; Martignoni et al. 2006; Nilsson et al. 2012;

Panchal and Brown 2011]. Sometimes other factors besides metabolism are important in selecting an appropriate animal model. For example, a sufficient number of test animals are needed to ensure that a study has adequate statistical power to detect an adverse effect. Therefore, if rhesus monkeys are most metabolically similar to humans, but only small numbers of these animals were used in the experiment and the toxicologic response was equivocal, then the rhesus monkey may not be the best animal model for the risk assessment.

Ideally, animal studies used in human risk assessment should be performed in appropriate aged animals (adult vs newborn), in both sexes, and with health status (e.g., pregnant vs non-pregnant) that corresponds to human exposure and toxicity. The study should take into consideration the appropriate duration and pattern of exposure (acute versus chronic; single exposure versus repeated administration) that simulate human exposure.

Often, a test compound will have data from several animal studies. The information on test animals should include species, strain, sex, age, and number of animals/group from any individual study. Ideally, an animal model with the most valid biological rationale (e.g., similar pharmacokinetic profiles) should be selected as the animal model most relevant to humans. However, in some cases no such closely relevant model exists. In such cases, the animal model that is most sensitive (i.e., showing a toxic effect at the lowest administered dose) is often used [Barnes and Dourson 1988].

4.2.2 Toxicologic Study Design

To be useful for human risk assessment, NIOSH considers that the design of a toxicologic study should be sufficiently documented to include information on: study aims and hypotheses tested; reasons for selecting the animal model used; species, strain, source, and type of animal used; details of each experiment performed, including its design and number of animals used; exposure including dose, route, schedule and duration; and statistical methods [EPA 1994]. This information may be found in a study protocol or in the methods section of the study report. Superior studies provide sufficient documentation on the methods to replicate findings. Combined with a clear and thorough presentation of findings, the study design information is a valuable resource

for judging WoE. In addition, it is often helpful if individual animal data are made available for additional analysis.

Datasets from studies conducted adhering to good laboratory practices (GLP) [OECD 2005], and according to internationally accepted test guidelines (e.g., OECD, EPA, and EU) are preferred as candidate datasets for risk assessment. In addition, studies with sufficient details on methods, analysis, and results that have been peer-reviewed are often acceptable [Klimisch et al. 1997].

4.2.2.1 Exposure Information

Exposure conditions play a vital role in the experimental design of animal toxicity studies. Determination of the dose that reached the test animal in a study is a complex process. This involves proper methods used for the generation, characterization, and delivery of a test compound [EPA 1994]. The following criteria are used by NIOSH to assess the suitability of toxicological studies for risk assessment purposes.

Ideally, the study should clearly define the physicochemical characteristics of the substance used, such as purity, stability, pH, partition coefficient, particle size and distribution, breathing zone concentration, and vehicle. The concentration of the test compound should be reported as means and variances.

For an inhalation study, the information should include a description of the generation and characterization technology used (e.g., chamber design, type, dimensions, uniformity of distribution, source of air, generating system, air conditioning, and exhaust treatment) [Nelson 1992; Wong 2007]. The number of air changes, air flow rate, oxygen content, temperature, and relative humidity are exposure chamber characteristics that should be monitored and reported as means and variances; The description of the characterization method(s) should also include frequency of measurement, calibration of the measurement instrument, frequency of the calibration, and other quality assurance elements [Barrow 1989; Chen and John 2001; Moss and Cheng 1989; Wong 1999].

The various inhalation exposure techniques include whole-body, head-only, nose-only, intra-tracheal instillation, and laryngeal aspiration [Driscoll et al. 2000; Phalen 1997; Sahu and Casciano 2009; Wong 2007]. Factors such as safe and efficient generation, amount of material, test compound stability, exposure duration, and the measurements desired influence the selection of an exposure technique for a study design. For instance, in chronic inhalation exposure studies whole-body exposure of laboratory animals in cages is the most common method, whereas nose-only exposures are most often used for short duration particle exposures. However, it should be noted that several factors such as heat, stress, and anesthesia could affect the biological patterns of the animal, potentially influencing results [Hughes et al. 1982; Mete et al. 2012; Overmyer et al. 2015; Stratmann et al. 2010; Suvrathan et al. 2010]. Due consideration of these issues should be included in detail in the data analysis to ensure appropriate comparisons.

Test agents may affect lung ventilation, function, clearance mechanism, uptake, and retention of the dose and make animals susceptible to diseases because of long latency on chronic exposure. Particle overloading in the lungs of test animals should also be evaluated [Oberdörster 1997]. Excessive particle exposure may result in an increase in aggregated alveolar macrophages (AM) engorged with phagocytized dust particles. These AMs release an array of mediators resulting in various inflammatory responses and tissue injury [Kanj et al. 2005; Laskin and Pendino 1995]. The issue of overloading mostly occurs for particle exposure and not for gases/vapors exposure. Therefore, careful evaluation is warranted for the applicability of overloaded test compounds to humans [EPA 1994]. A comparison of the particle burden in the lungs of the overloaded test animal to the particle burdens expected in the lungs of occupationally exposed humans might be helpful in determining whether the experimental study in question is relevant to occupational health risk assessment.

The exposure concentration, administration route, exposure schedule, and exposure duration must be described. Consideration should also be given to the concentration, and time of exposure used versus the expected level of human exposure.

Animal studies are conducted using different durations (acute, subchronic, and chronic) and frequency of exposure (single, intermittent, and continuous). All of these studies help to identify the hazard associated with a given test compound but not all of them may be appropriate for quantitative risk assessment, depending on the human exposure of concern.

Appropriate control groups of unexposed (e.g., 'air-only" controls in inhalation studies) and/or vehicle-exposed animals should be included in the study. The control group(s) should be treated similarly to the chemical treated group except that the control group should not receive any of the test compounds [Hayes 2008; Salem and Katz 2014]. In addition, historical control data can also be used to evaluate the differences between control and treated groups. In general, historical control data should be submitted from the same laboratory and should be from animals of the same age and strain generated during the five years preceding the current study [OECD 2009b].

Most often animal bioassays expose animals to higher doses of a chemical than humans are normally exposed to [Klaassen et al. 2013]. Sometimes, the animal doses are comparable to occupational exposures but often, they are significantly higher (Please see Section 6.1.9 for more details). Toxicity observed at high doses may or may not occur at lower doses. Therefore, animal studies should always be evaluated in the context of dose-response relationships. Doses for vapor exposure are usually provided in units of parts per million (ppm) or milligrams per cubic meter (mg/m^3), and doses for airborne particle exposures are usually provided as the mass concentration (mg/m^3) and aerodynamic particle size (e.g., mass median aerodynamic diameter, MMAD). Number concentration is also reported for airborne particles or fibers. The size and shape of particles determines the region in the respiratory system in which particles are deposited (Please see Section 6.1.3 for more details). Care should be taken to ensure that particle morphology under test conditions reflects human exposure patterns. The physicochemical properties of a particle determine whether it will be dissolved in the blood or removed by clearance mechanisms.

In dermal exposure, contact area, absorption, concentration of the chemical, contact frequency, retention time and penetration potential contribute to the dermal toxicity [Marquart et al. 2003; Poet and McDougal 2002; Schuhmacher-Wolz et al. 2003; van Ravenzwaay and Leibold 2004].

4.2.2.2 Consideration of Mode of Action and Adverse Outcome Pathways

Whenever data are available to describe the mechanism of action, the mode of action (MoA) or the adverse outcome pathway (AOP), NIOSH uses this information to evaluate the dose-response information. Mechanism of action is generally thought of as the underlying biochemical interactions that lead to the expression of the adverse effect. Full information on the mechanism of action is rarely available. Mode of action is a more general term referring to the general processes and key events that are involved in the toxicity of a chemical. MoA analysis includes review of physical, chemical, and biological information of the substance [Boobis et al. 2006; Boobis et al. 2008]].

Typically, when conducting risk assessment to inform the development of a REL, NIOSH evaluates health effects that may be experienced by humans or that may be related to health effects experienced by humans, as evidenced by data from human and animal studies. Once the constellation of health effects under consideration has been established, the risk assessor critically evaluates the health effects to determine which effect(s) are of interest. In doing so, the risk assessor clearly explains the rationale for selection of the health effects and their relevance to human health. An understanding of the mode of action of the toxic agent can help define which data are most appropriate for consideration. The EPA has described MoA framework initially for cancer risk assessment in the "Guidelines for Carcinogen Risk Assessment" as a sequence of key events and processes, starting with interaction of an agent with a cell, proceeding through operational and anatomical changes, and resulting in cancer formation" [EPA 2005]. Examples of possible modes of carcinogenic action include "…mutagenicity, mitogenesis, inhibition of cell death, cytotoxicity with reparative cell proliferation, and immunologic suppression". Later the MoA framework concept has been expanded to assess non-cancer endpoints risk as well [Bogdanffy et al. 2001; Julien et al. 2009; Lochner et al. 2005; Seed et al. 2005]. The substance may

induce adverse effects by more than one MoA at a single tissue or at different sites. Therefore, a single MoA for an endpoint may not be expected to apply for all other toxic endpoints unless indicated otherwise.

The expanded framework focuses on the weight of evidence establishing the mode of action in animals, whether the key events identified in animals are plausible in humans, and takes into account kinetic and dynamic factors to determine whether the mode of action is plausible in humans. To use the mode of action framework, the risk assessor asks the following questions (Figure 4-1):

- *Is the weight of evidence sufficient to establish the MoA in animals?* The first step in considering the relevance of MoA information to human health is having sufficient MoA information in the animal species/health effect of interest. The default position is that the health effects observed in animals are relevant to humans. As stated in Seed et al. [2005], "[W]hen data are insufficient to confidently characterize an MoA for test animals, the animal tumor data are presumed to be relevant to humans and a complete risk assessment is necessary."

- *Are key events in the animal MoA plausible in humans?* To evaluate whether the MoA is relevant to humans, there must be sufficient information available regarding the potential for the key events identified in the animal MoA to operate in humans. For example, a key enzymatic pathway observed in the animal should also present in humans or, if the exact enzymatic pathway is not present, there are there pathways that serve a similar or identical function. If there is insufficient information in humans to characterize the relevant pathways, as described above, the animal data are presumed to be relevant to humans and a complete risk assessment is necessary. However, if there is clear evidence that the relevant pathways do not operate in humans, the risk assessor should assume that the observed effects in animals are not relevant for humans for this endpoint. Unless it is known that all the health effects observed in animals derive from a common MoA, this analysis needs to be conducted separately for all health effects under consideration for risk assessment.

- *Taking into account kinetic and dynamic factors, as well as life stages, are key events in the animal MoA plausible in humans?* This step requires quantitative information on the relative kinetic and dynamic factors that would influence risk in humans and animals, as well as consideration of life stages of potential exposure in humans. For example, consider the case in which the MoA has been identified in animals involving toxicant metabolism by a specific enzymatic pathway found in both animals and humans; however, there is a high rate of metabolism by this pathway in the rodent that is not evident in humans. In addition, humans have a competing enzymatic pathway that metabolizes the toxicant much more rapidly. Thorough analysis of the kinetics indicates the potential human toxicity via this MoA is, in fact, very low, and the conclusion may be reached that there is no need to conduct a risk assessment for this endpoint. The same type of analysis could be conducted when considering potential exposures during specific life stages, if that is deemed a critical variable.

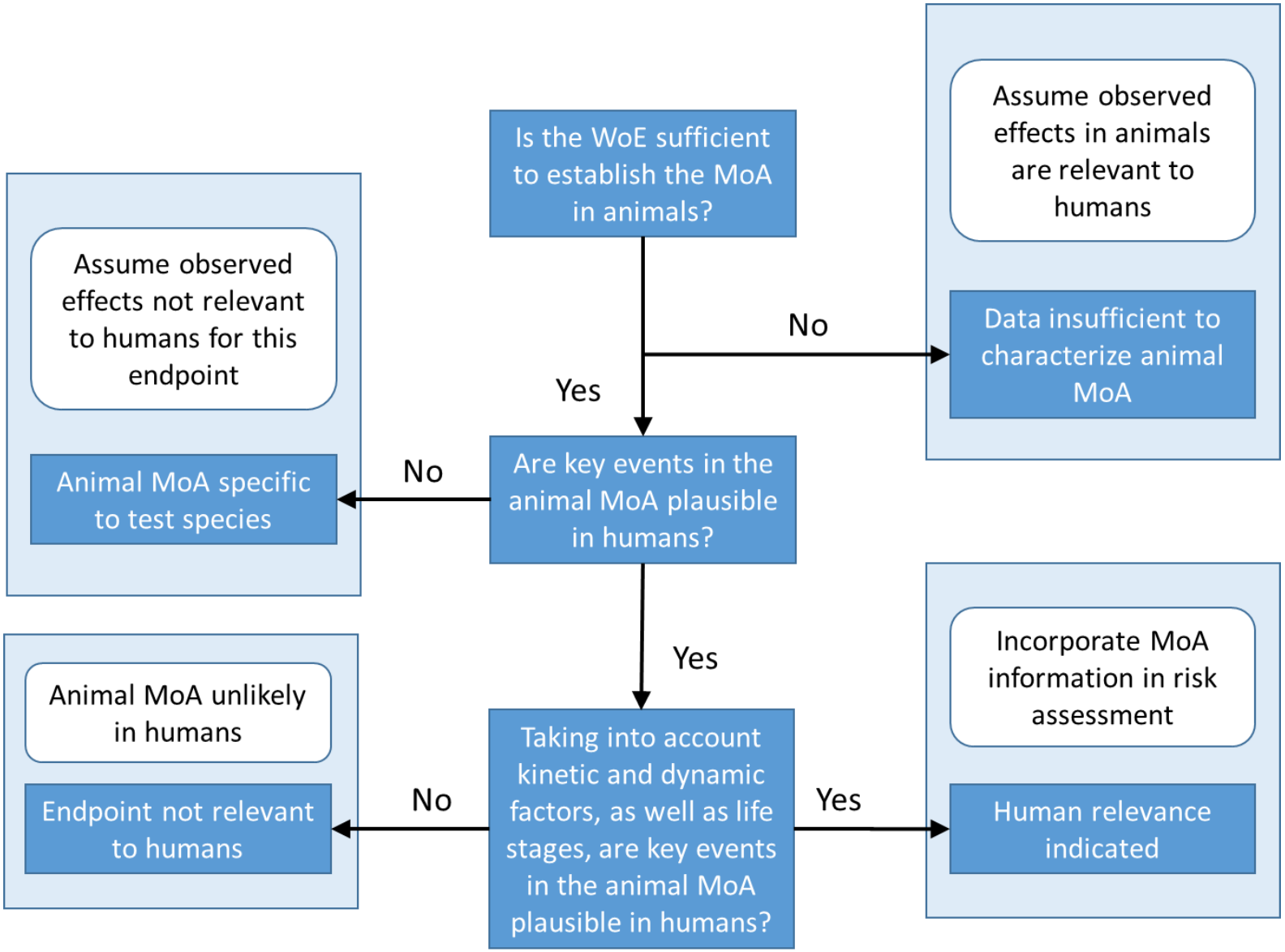

Figure 4-1. Framework for Mode of Action (MoA) assessment (adapted from Seed et al. [2005])

Adverse outcome pathways (AOPs) are structured representations of biological events leading to an adverse effect. AOPs are flexible frameworks that can include linking relationships that are causal, mechanistic, inferential, or correlation based [Ankley et al. 2010]. AOPs have been established for several chemical groups. In AOP, a structured sequential chain of events is constructed using all available scientific information. For example, a chain of events might entail a toxicant exposure causing a molecular initiating event, when then leads to a series of key events causing the adverse outcome of interest.

Data on key events in AOPs, on the events in the biochemical mechanism of action or on the processes involved in the presumed mode of action could give an insight into the toxicity of chemicals, and could help to refine the adverse effect under investigation by providing evidence to support critical biomarkers of effect, using an analysis similar in structure to the MoA framework. It is important to consider the mode of action or adverse outcome pathway during risk assessment to determine if the data offer additional insights as to the shape of the dose-response curve, the key indicators of critical dose or adverse response.

4.2.2.3 Selecting Adverse Response of Interest

After considering mode of action and adverse outcome pathway information, the dose-related changes in observed adverse effects should be biologically and statistically significant. Ideally, a dose-response relationship can be demonstrated. The observed effects should be directly related to the magnitude of exposure to test compounds and not influenced by concurrent exposure to other compounds or already existing health conditions [Lewis et al. 2002]. In general, the recommended list of hematology, clinical biochemistry, and histopathological examinations to be evaluated in the laboratory studies are given in the several guidelines [Crissman et al. 2004; OECD 2009a; OECD 2009b; OECD 2009c; OECD 2009d; Weingand et al. 1996]. NIOSH risk assessors refer to these guidelines, as applicable, for a better understanding and evaluation of the study.

Preferably, the histopathological examinations should be of all tissues for all treated doses and control groups and all tissues from animals dying or killed during the study. However, many published studies, excluding

the NTP studies, have histopathological examination only for the target end points, which although somewhat limited in scope still may be useful for risk assessment.

When multiple toxicity studies are available, the studies should be reviewed with reference to the types of effects observed in different test species. Consistency of response across species, sex, and/or route of exposure increases the weight of evidence that the effect might occur in humans. In contrast, an effect observed in only one species, one sex, may need further evaluation. Results replicated by independent researchers would have increased credibility. Tests conducted using structurally related compounds could also be considered for a comparison of results.

Once the evidence is evaluated, NIOSH assesses the database for completeness. A complete toxicological database includes studies that evaluate carcinogenic, genotoxic, reproductive, developmental, and other organ effects (e.g., immunotoxic, neurotoxic, nephrotoxic, irritation, and sensitization). Ideally, the literature describes the dose-response relationship; concordance across species, strain, sex, exposure routes, or in multiple experiments with respect to adverse effects; effects are biologically plausible and of human relevance; similar effects with structurally related compounds. However, a complete toxicological database is not essential for hazard identification if the observed adverse effects are considered relevant to occupational exposures. If there are only limited data on a specific chemical, then all of the available studies with limited information should be critically evaluated to determine the usefulness of the information for risk assessment. If concordance across species/strain/sex is not observed, additional evaluation is needed and, in the absence of information to the contrary, the more sensitive species/strain/sex is often used.

4.3 Human Data

Ideally, the direct estimation of risk from human data is always preferred by NIOSH. In practice; however, risk assessments usually must rely on a combination of epidemiologic and toxicologic data for hazard identification and dose-response analyses. This is because both data sources are imperfect. On one hand, human data tend to be vulnerable to potential biases from confounding factors. On the other hand, there is large

uncertainty in extrapolating risk in animals to humans. It is common to find human data to be weighted more than animal data for hazard identification, but be less informative on dose-response. In those instances, human studies provide evidence of an association between exposure and disease, which can guide the choice of agents, exposure routes, and pathological endpoints for examination in toxicological studies that contribute greatest to quantifying risks.

4.3.1 Epidemiologic Study Design

A hierarchy of epidemiologic study designs ordered by the potential contribution to WoE is shown in Figure 4-2. Human data for WoE assessment may originate from experimental or observational studies. In regards to the former, study participants are intentionally exposed to an agent under controlled experimental conditions. In this context, 'controlled' refers to design parameters intended to minimize the effects of factors other than the exposure condition on the measured response [NAS 2017]. These studies are sometimes referred to as clinical studies, human challenge studies, or controlled human exposure studies. Adherence to a strict experimental design is a trait of controlled human exposure studies that lessen the potential for significant biases; therefore, these data are well suited to hazard identification and dose-response analyses. Of course, human experimental data on exposures to hazardous agents are sparse for obvious ethical reasons; therefore, observational studies tend to be the most significant information source for directly assessing the dose-risk relationship in humans and are the focus of the discussion on epidemiologic study design.

Observational studies can be further classified as either analytic (e.g., longitudinal and cross-sectional studies) or descriptive (e.g., case reports, case series, and ecologic studies) designs, the latter being the least informative for risk assessment. Detailed descriptions of epidemiologic study designs are available in seminal texts [Breslow and Day 1980; Breslow and Day 1987; Checkoway et al. 2004; Rothman et al. 2008]. Brief descriptions of common observational study designs are provided below.

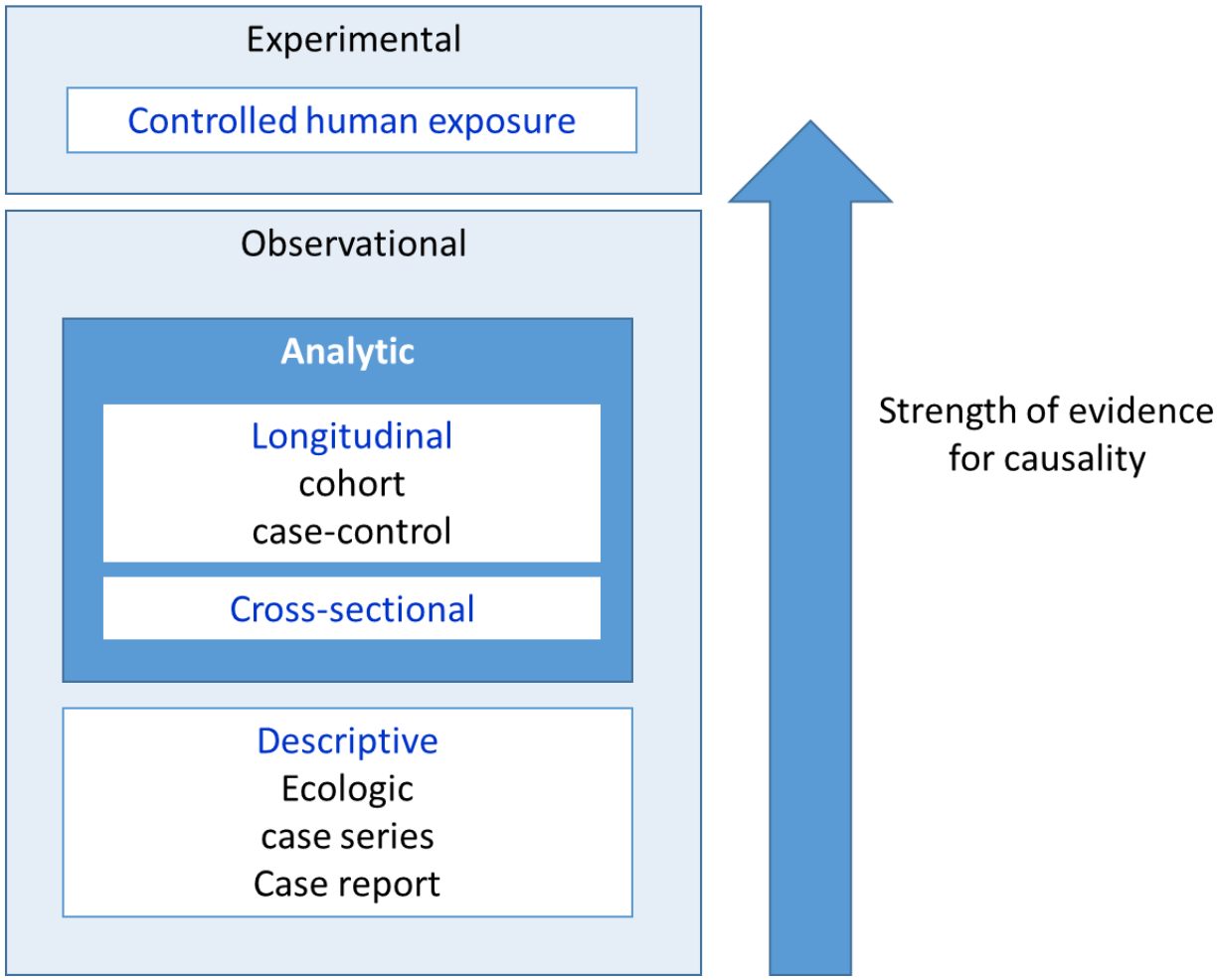

Figure 4-2. Hierarchy of human epidemiologic studies

Of observational research, longitudinal analytic studies (e.g., cohort, panel studies) are most promising with respect to WoE. These studies follow the exposure and health status of each individual in a study sample or population over time. An important strength of this design is its ability to measure temporal changes in exposure and outcome at the individual level. Thus, this study design allows for the direct examination of the dose-response. Cohort studies are the most common source of human data in NIOSH risk assessment. A 'cohort' comprises a group of individuals who share some defining characteristics who are followed in time. Data can be collected prospectively; however, most occupational cohort studies are historical, using data that span a time prior

to initiating the study. A disadvantage of a cohort study is that it may require the recruitment of many participants who must be observed over a long period when examining rare outcomes (e.g., cancers); therefore, a detailed accounting of individual exposures for everyone in the study group may be impractical. Measures of association can vary in cohort studies. If comparisons are made between the study population and an external referent (e.g., U.S. population), common measures of association are the standardized mortality ratio (SMR) or standardized incidence (morbidity) ratio (SIR) [Rothman et al. 2008]. These measures are simply the ratio of the observed number of cases to the number of expected cases, where the expected cases are calculated based on disease rates observed in the referent population that are standardized by characteristics (e.g., age, race, gender, and calendar period) in the study population. Trends in SMRs by categories of exposures can offer crude dose-response information; however, because of indirect standardization methods, comparisons of SMR are vulnerable to bias due to differences (e.g., differences in age, gender, and race) in comparison groups. Internal comparisons (comparisons made within the study population) offer improved dose-response data compared to SMR and SIR analyses. Measures of association from internal comparisons include trends across standardized rates by exposure categories or risk measures from dose-response regression models. Risk measures can be expressed on a relative scale, such as hazard ratios (HRs), rate (or risk) ratio (RR), excess relative risk (ERR), or in terms of risk differences, such as attributable risk or excess absolute risk (EAR).

A case-control (or case-referent) study compares exposure among persons with the outcome of interest (i.e., cases) to exposures among persons preferably drawn from the same population (controls). Thus, the reduction in study size saves time and expense relative to a cohort study. This design is particularly useful when examining rare adverse effects. Cases can be enumerated at a point in time (prevalent cases) or over a period of time (incident cases). An important consideration is the number of matched controls per case. In the absence of a dose-response relationship, reasonable asymptotic relative efficiency is achieved with few controls [Breslow et al. 1983; Goldstein and Langholz 1992; Ury 1975]. However, the actual relative efficiency decreases as the strength of the exposure–response increases and as the skewness of the exposure distribution increases [Bertke et al.

2013]. The standard effect measure of the case-control study is the odds ratio (OR), which approximates the risk ratio (relative risk) if the disease is rare.

A special instance of a case-control study is one that is nested within a specified cohort. This design retains many of the analytic advantages of the large cohort while reducing the number of subjects needing exposure assessment. Thus, a nested case-control study allows for improvements in exposure assessment that can lead to better information on dose-response. As in cohort studies, a nested-design also allows for the precise treatment of the timescale; therefore, measures of association related to events per unit person-time can be estimated using dose-response regression modeling (e.g., HR, RR, and ERR).

A cross-sectional study (e.g., survey) examines the frequency or level of a particular attribute, (e.g., exposure and/or adverse effect) in a defined population at a particular point in time. This design is often used to examine the prevalence of nonfatal diseases or symptoms that typically do not rapidly lead to employment termination (e.g., mild decreases in lung function, blood pressure, pre-clinical biomarkers of early effect, and skin irritation). This design is a poor choice for examining diseases that are rare or periodic. Because cross-sectional studies are based on prevalent cases, this design has limited value for examining etiologic relationships. Other disadvantages are the lack of information on the temporal sequence between cause and effect and bias when health-related employment termination was present prior to ascertainment.

Epidemiologic studies that are conducted with observation at the group level (e.g., plant, cities, counties, and nations) instead of the individual level are called ecologic or aggregate studies [Rothman et al. 2008]. These studies can involve a single cross-sectional survey or repeat measures (i.e., time-trend design). Ecologic studies may be a practical alternative to individual level studies when exposures and disease are relatively homogeneous within a population but differ between populations, or when individual exposure estimates are not possible. Within-group heterogeneity is a likely condition; therefore, extrapolation to the individual level is not feasible. Thus, an association at the group level does not imply the same association at the individual level. This significant limitation is known as the ecologic fallacy. Another major limitation is that these studies lack the ability for

adequate control of confounding. For these reasons, ecologic studies have limited value in assessing causal associations. Nevertheless, ecologic studies may provide descriptive information on differences in population that may be suggestive of a potential cause-and-effect relationship. This information can be used to support findings from analytic studies.

Other descriptive studies, such as case reports and case series provide information on symptomology, disease history, diagnostic features, and outcomes for one or more subjects under observation. A 'case report' refers to a description of a person with a disease, while 'case series' refers to a series of related case reports that were typically collected at a specific practice, clinic, or hospital over a defined time period. Similar to cross-sectional studies, routine data studies are of limited value for examining etiologic relationships, but may initiate larger investigations that are better designed to inform on causation. In addition, these studies are purposed for identifying emerging trends in adverse health, high-risk populations, and unrecognized hazards; therefore, they can be an important data source for hazard identification in risk assessment. In particular, case reports can be very informative on rare diseases when exposures are well defined. For example, 4,4´-methylenebis(2-chloroaniline) (MBOCA) is considered carcinogenic in humans (IARC Group 1) because of sufficient evidence in experimental animals and strong mechanistic evidence, however, evidence in humans was deemed inadequate [IARC 2012]. Liu et al. [2005] reported on a 52-year-old non-smoking male bladder cancer patient who was significantly exposed to MBOCA while employed as a chemical worker for 14 years. Reconstruction of his past exposures yielded no other exposures to bladder carcinogens. Thus, this case study provided direct evidence of MBOCA potentially acting as a human bladder carcinogen.

When assessing human studies for WoE, NIOSH risk assessors often use available checklists of study design and analyses criteria that have been widely vetted. For example, clinical research has greatly improved under the Consolidated Standards of Reporting Trials (CONSORT) statement, which includes a 22- item checklist and flow diagram [Moher et al. 2001]. Similar checklists are available for meta-analyses of clinical trials and observational studies [Moher et al. 1999; Stroup et al. 2000]. The checklists developed under the Strengthening

the Reporting of Observational Studies in Epidemiology (STROBE) initiative are useful tools for assessing the strengths and weaknesses of standalone observational studies [von Elm et al. 2007]. The STROBE website (http://www.strobe-statement.org/index.php?id=strobe-home) provides separate checklists for cohort, case-control, and cross-sectional studies. Also for longitudinal studies, risk assessors are encouraged to use the checklist offered by Tooth et al. [2005]. This checklist consists of 33 questions on study design and analysis criteria. These available checklists may augments the systemic approach to assessing WoE; however, they should be viewed only as potentially useful tools among many. Given that each risk assessment is unique, reliance solely on published checklists should be avoided.

In summary, the majority of human data used in NIOSH occupational risk assessments are drawn from observational studies and preferably from longitudinal studies of working populations. These data include information on disease status from registries, death certificates, medical records, diagnostic exams, or self-report (i.e., questionnaires). Exposure data results from personal or ambient measurements, modeling, constructed proxies (e.g., job-exposure matrix), self-report or any combination of these sources (See Section 4.3.3). Risk assessors must fully understand the nature and limitations of the data in studies(s) selected for risk assessment.

4.3.2 Adverse Effects

Adverse effects in workers, sometimes referred to as the ‘adverse health effect’, ‘outcome of interest’, or simply the ‘response’, must be clearly defined in the population at risk and comparative populations, such as control groups or the general population. NIOSH risk assessors affirm that case definitions and ascertainment methods used in candidate epidemiologic studies are sufficient for risk assessment. When considering cancer and non-cancer adverse effects or biomarkers for those effects, it is important to understand the progression of disease and select a measurable adverse effect as early in the process or with the least severity of effect as possible. Ideally, an empirically observable endpoint that is clearly a key event or precursor to the adverse effect of interest should be targeted for risk assessment, but the strength of association between the exposure and outcome and the potential for confounding are important to consider. Furthermore, an agent may involve multiple adverse effects.

Frequently, risk assessors have limited evaluations to the most sensitive effect by examining multiple effects separately and then choosing the effect offering the greatest risk per unit exposure. However, recent efforts have shifted toward a more holistic approach of estimating aggregate risks from the combined effects of exposures to one or more agents.

4.3.2.1 Cancer

'Cancer' is a term used to describe over 100 different diseases in which abnormal cells divide without control and can invade nearby tissues. Given this broad definition, there are many possible characterizations of cancer as an adverse effect used in epidemiologic studies, ranging from all cancers combined to a precise classification of a primary malignant tumor. In dose-response analyses, studies reporting specific adverse effects are superior to those reporting the effects of all cancers combined given varying etiology among types of cancer; however, specificity of the adverse effect may come at a cost of statistical imprecision given the rarity of most individual cancers. Moreover, most occupational studies have examined mortality data from death certificates, which often lack desired cancer specificity. Therefore, human studies have infrequently examined specific malignancies (e.g., lung adenocarcinoma). Instead, cause-specific cancer endpoints are typically constructed by grouping multiple tumors that share common traits (e.g., lung cancer, respiratory cancers, and solid tumors). It is important to consider the potential effects of a heterogeneous grouping on the dose-response. For example, ionizing radiation is a known leukemogen; however, research suggests that chronic lymphocytic leukemia (CLL) is nonradiogenic. If nonradiogenic, combining CLL with radiogenic leukemias will act to attenuate the dose-response between the grouped outcome and ionizing radiation exposure; therefore, most studies have excluded CLL from the leukemia grouping.[1] Group definitions can vary between studies and even within a study due to

[1] The research on CLL radiogenicity is inconsistent; however, in 2011 after a review of the literature and consultation with subject matter experts, NIOSH recommended that CLL be considered radiogenic for purposes of compensating workers covered under the Energy Employee Occupational Illness Compensation Program Act (EEOICPA).

changing diagnostic criteria over time. When combining data for hazard identification and subsequent dose-response analyses, the risk assessor considers the compatibility of the adverse effect definition between studies.

In human studies, cancer cases are typically ascertained from causes of death listed on death certificates, or from diagnostic information found in cancer registries, medical records, or patients (or proxy) self-report. Cancer incidence data are generally considered superior to mortality data, because of improved diagnostic information in registries and medical charts and greater ascertainment of highly survivable cancers. In addition, incidence data are less susceptible to survival effects (e.g., competing risks and healthy worker survivor effects) in dose-response analyses. However, incidence data are more susceptible to screening bias, which occurs when cancer screening differs among comparison groups. In addition, U.S. cancer registry data are relatively incomplete prior to the mid-1990s and are managed differently among individual states; therefore, studies using these data have limited but complicated followup. Most studies have preferred cancer registries to medical charts for incidence data because registries are less affected by losses due to death or followup. However, ascertainment from registries can be quite poor for some cancers that are underreported, such as melanoma treated in private clinics [Cockburn et al. 2008], or not reported (e.g., basal and squamous cell carcinomas, excluding genital sites). Self-reported incidence data tend to be the least preferred because they are more susceptible to bias from incomplete followup and patient recall. However, results from comparisons made between self-report and registries vary by cancer site, with reasonable agreement indicated for some outcomes, such as lung, breast, prostate, and uterine cancers [Bergmann et al. 1998; Desai et al. 2001].

Combining self-reported data with medical follow-back is a viable option to using registry data, especially, is situations where searching multiple state registries is infeasible or if the date range coverage is insufficient. In addition, these methods provide opportunity for gathering information on important covariates. An example of this design is the recent NIOSH study of breast cancer incidence in a cohort of U.S. flight attendants [Schubauer-Berigan et al. 2015]. For this study, cases that were first identified by self (or proxy) report were then

verified by contacting the physician, hospital, or other health care organization in which the cancer diagnosis was made and obtaining supporting documentation of the diagnosis.

4.3.2.1.1 Carcinogenesis

The mechanisms of carcinogenesis are rapidly becoming an important aspect of hazard identification. It is now known that human carcinogenesis is a multistage process that can involve numerous mechanisms causing various biological changes leading to tumorigenesis. These mechanisms can vary widely by agent; therefore, a systematic approach to assessing available mechanistic data is needed to appropriately characterize the dose-risk relation and assess the overall carcinogenic hazard of an agent. For example, IARC has identified 10 characteristics of human carcinogens to be used in a systematic strategy of assessing mechanistic data for hazard identification. These characteristics are the abilities of an agent to: 1) act as an electrophile either directly or after metabolic activation; 2) be genotoxic; 3) alter DNA repair or cause genomic instability; 4) induce epigenetic alterations; 5) induce oxidative stress; 6) induce chronic inflammation; 7) be immunosuppressive; 8) modulate receptor-mediated effects; 9) cause immortalization; and 10) alter cell proliferation, cell death, or nutrient supply [Smith et al. 2016]. Most carcinogens demonstrate more than one of these traits. IARC recommends that the key characteristic be used in three steps: identify relevant information, 2) screen and organize mechanistic data, and 3) synthesize mechanistic information (e.g., develop adverse-outcome pathways). In this way, the IARC approach provides a foundation for carcinogen classification (i.e., hazard identification); however, mechanistic data can also inform choices made in risk characterization, such as estimating the response expected at low doses.

4.3.2.2 Non-cancer

For non-cancer risk assessment, it is important to evaluate issues of severity, reversibility, progression to more serious conditions and other pertinent issues. NIOSH has typically conducted quantitative risk assessment on non-cancer adverse effects assuming chronic exposure (see Table 1-1). However, there may be instances where the effects after short-term or intermediate-length exposure are determined to be critically important. In those

cases, the risk assessor must evaluate and document the impact of the exposure in the context of a shorter-term exposure duration and any longer term sequelae.

A key question is whether or not the observed pathophysiologic change used to define the adverse effect indicates the development of irreversible health effects or whether the impairment is temporary and reversible with cessation of exposure. Unfortunately, data are often insufficient to answer this question. Two examples of NIOSH risk assessments best illustrate this issue. First, NIOSH examined Parkinson disease-like symptoms and manganism resulting from manganese exposure in welders [Park et al. 2009]. The NIOSH risk assessment quantified the relationship between manganese exposures in confined-space welding and cognitive deficits such as for working memory or verbal IQ. However, there was insufficient information to conclude whether or not the risk of exposure-related neurobehavioral deficits persisted after cessation exposure. Without the data, NIOSH conservatively assumed that excess risk accrued with exposure and persisted afterwards (i.e., similar to exposure related cancer). Similarly, a risk assessment of diacetyl exposure and the development of bronchiolitis obliterans (BO) used multiple definitions of pulmonary dysfunction, as a case-surrogate for the onset of BO [NIOSH 2016a]. Again, risk accumulation and persistence was assumed in lieu of contrary information. In both risk assessments, the assumption of irreversible adverse effects had significant impact on estimates of lifetime risk per unit exposure.

4.3.2.3 Human Data Sources

Mortality is a common endpoint in epidemiologic studies. Cause of death information stems primarily from death certificates, which can provide information on multiple causes of death. Typically, the underlying cause of death (UCOD) is preferred given longstanding and well-accepted use in public health, although some studies examined multiple causes to increase ascertainment information, especially for rare outcomes. The UCOD is defined as the disease or injury, which initiated the train of events leading directly to death, or the circumstances of the accident or violence, which produced the fatal injury [WHO 1977]. Although the unidimensional UCOD is conceptually easy to understand, it does not take into account other important

contributors to death that may also be listed on the death certificate. This is especially true for complex chronic illnesses, which are typically characterized by multiple contributing causes. To make better use of available information, multiple causes of death (MCOD) data have become an appealing alternative in some studies [Chamblee and Evans 1982; Israel et al. 1986; Redelings et al. 2007].

For study purposes, death causes are usually translated to codes from the International Classification of Diseases (ICD). Coding death certificates is a highly specialized and interpretive process that is nearly always conducted only by a qualified nosologist. Agreement between nosologists tends to be high, but some disagreement and errors in coding are unavoidable as are inaccuracies in the death certificates themselves. For example, the underlying cause of death recorded on death certificates have differed upwards of 20-40 percent when compared to autopsy [Cameron and McGoogan 1981; Engel et al. 1980; Maudsley and Williams 1996; Sehdev and Hutchins 2001]. Coding sequence errors in translating information from the death certificate are also likely. In both cases, the effects of these errors may be offset in analyses using MCOD data.

Morbidity data are typically abstracted from disease registries. Obtaining morbidity information is generally more difficult compared to mortality data given there are few reportable diseases (e.g., cancer) and most U.S. disease registries have originated relatively recently compared to mortality databases. Cancer registries are perhaps the most informative given that nearly all states have registries acquiring data since the early 1990s. Nevertheless, the U.S. lacks of a national cancer incidence database, therefore studies of U.S. workers require matching to multiple state cancer registries, the number of which depend on the potential for workers to leave the covered area during the observation period. Many workers employed in northern states retire to southern states after employment termination; therefore, their cancers will likely be unobserved unless ascertainment includes the retirement state. Acquiring data from multiple cancer registries can be onerous, costly, and time-consuming. Thus, cancer ascertainment has primarily relied on mortality data in worker studies. In general, morbidity data from disease registries are preferable to death certificates; therefore, registries are generally considered the gold standard for use in human epidemiologic studies. Nevertheless, U.S. disease registries may have limitations due to

a number of factors, such as poor coverage, underreporting of some diseases (e.g., melanoma), no reporting of others (e.g., basil cell carcinoma), and duplicate reporting among multiple database (e.g., neighboring state cancer registries), varying data acquisition procedures across databases, and relatively short time span since registry inception.

Data on the adverse effects may be self-reported or stem from expert diagnosis. For example, study data may originate from members (or proxies) of the population at risk who have reported specific symptoms or diagnosed conditions that may be related to the exposure of interest. These data are vulnerable to errors from inaccurate recall [Atkinson et al. 2016; Howell et al. 2015; Wallace and Kohatsu 2008], which may be attenuated somewhat using medical follow-back to confirm a reported diagnosis (e.g., Schubauer-Berigan et al., [2015]). Expert assessment of signs or diagnoses requires the uniform application of an adverse effect definition by knowledgeable evaluators. In addition, it should be ascertained whether evaluators were kept uninformed of study subjects' exposure status when assessing health effects. If information exists regarding the validity of expert assessment of health effects, this information may inform a determination of background rates in general or comparison populations, and allow for the development of levels of excess cases of the health effect relevant to occupational populations for purposes of risk assessment. A recent example of using expert assessment in a NIOSH risk assessment is found in the criteria document supporting the REL for diacetyl and 2,3-pentanedione [NIOSH 2016a]. In that study, researchers quantitatively assessed the effects of diacetyl and 2,3-pentanedione exposures on pulmonary function, using spirometry data and defined case definitions based on expert assessments of forced expiratory volume (FEV) and forced vital capacity (FVC). These case definitions were used in models quantifying the dose-response relationship between diacetyl exposures and changes in pulmonary function. The models also included data gathered using a medical questionnaire to collect self-report information on respiratory health, dermal symptoms, allergies, smoking habits, coexposures, and protective equipment used.

4.3.3 Exposure Assessment

The purpose of exposure assessment for NIOSH risk assessment is to sufficiently identify and characterize exposures to biological, chemical, or physical agents that occur, or are anticipated to occur, in a working population. The exposure assessment is conducted as part of the hazard identification process to support WoE evaluations and to provide necessary input to dose-response analyses (e.g., explanatory variable(s) in dose-response regression models). As such, the exposure assessment focuses on review of methods used to estimate or measure exposure in existing epidemiologic studies, and to synthesize exposure information from selected studies for use in dose-response analyses.

Ideally, the dose-response relationship between an adverse effect and an agent is quantified using complete exposure histories on each subject in the affected population. Of course, ideal conditions are rarely present in observational studies of working populations; therefore, many challenges are faced by exposure assessors, such as:

- The reliance on data from previous studies or employer information that are suboptimal for risk assessment purposes.
- A lack of sensitive, specific, precise, and accurate measurements of worker exposures. Exposure values are often derived indirectly from employment information (job titles and employment), and other proxy sources (e.g., other research, industrial hygiene data, process records, and institutional knowledge).
- Incomplete information on exposure or other risk factors that could influence effect measures. For example, exposures that occur while a worker was employed elsewhere (i.e., outside of studied facilities) are rarely known.
- Temporal and spatial variation in occupation characteristics (e.g., tasks, chemical inventories, and engineering controls) can result in wide-ranging inter and intra-individual variation in exposure. These differences add to the uncertainty in exposure indices.

- Industry settings that involve complex exposures to combinations of hazardous agents rather than a single agent of interest. Health effects from exposures to an agent may be entangled with effects from other agents. Furthermore, the combined effects of a mixture of agents (i.e., cumulative risk) may differ from the additive effect of separate exposures to these agents.

Due to the evolution of workplace hazard controls, present day exposures to hazardous agents tend to be lower compared to earlier times, resulting in less evident exposure-related adverse effects. Thus, there is increased need for the most informative exposure estimates for quantifying a correspondingly smaller attributable risk. As such, the field of exposure science is rapidly progressing to meet the demand for improved methods for estimating exposures. Many of these methods are summarized in several works on occupational epidemiology and risk assessment [Checkoway et al. 2004; EPA 1992; Nieuwenhuijsen 2010; NRC 1983; White et al. 2008]. This section defines useful terms and presents methods of exposure assessment that are commonly used to construct exposure indices used in risk assessment.

The quality of exposure information in observational studies is often limited by data availability. When considering data for the purpose of risk assessment, the risk assessor generally weights available information by the hierarchy shown in Table 4-1. This order supports a general preference of individual exposure estimates over group estimates and quantitative values over exposure classes. Thus, exposure assessment information ranges from individual exposure estimates derived from personal monitoring as best, to exposure status that is dichotomously assigned as a least favorable approach.

Table 4-1. Types of exposure data in occupational epidemiologic studies (adapted from Checkoway et al. 2004)

Type of Data	Dose Approximation
Quantified personal measurements on all workers	Most Precise
Quantified area- or job-specific measurement data	
Ordinal ranking jobs and tasks (e.g., SEGs, JEMs using exposure categories)	

Duration of employment in the industry	
Ever/never employed in the industry	Least Precise

Abbreviations: JEM, job-exposure matrix, SEG, similar exposure group

4.3.3.1 Exposure Indices

The output from exposure assessment is typically quantified directly or indirectly in terms of either exposure or dose. The metrics derived are referred to as exposure indices. The meanings of terms 'exposure' and 'dose' have been expressed in many different ways over time and are often used interchangeably; however, a distinction between these terms is generally recognized, where exposure refers to contact between an agent and a target at an exposure surface and dose is the amount of the agent that crosses the exposure surface and enters the target [IPCS 2004]. Strict adherence to this distinction relies on the choice of target and exposure surface, and dose estimation may require a complete accounting of various physiologic and metabolic systems that modify the amount deposited into the chosen human target. In practice, the choice of exposure or dose metrics depends on the aims of the candidate study, which may or may not align with the needs of the risk assessment. Therefore, NIOSH risk assessors consider how the choice affects the WoE provided by the study during hazard identification, and if data are selected for dose-response analyses, what additional steps (if any) are needed to convert the quantity used in the dose-response analysis to the quantity needed for suitable REL.

Exposure indices can be expressed in many ways using information on three basic dimensions: intensity (e.g., concentration, mass), duration (e.g., hours days), and frequency (e.g., times per day). Indices may include each dimension separately or in combination, such as assessing ionizing radiation exposure as a time-integrated dose [e.g., lifetime dose equivalent measured in sievert (Sv)] or a time-averaged dose (dose equivalent rate measured in Sv per hour). In epidemiology, the choice is largely based on the expected effect. For example, exposure indices used to examine acute toxicity effects are typically based on short-term or instantaneous intensity (e.g., peak airborne concentration), whereas cumulative dose (i.e., the time integral of exposure intensity) is generally preferred for dose-response analysis involving chronic effects in which biologic damage

appears proportional to the delivered dose quantity (e.g., silica and silicosis or ionizing radiation and cancer). As another example, an adverse effect may be reversible by elimination of toxic agents from the body over time. In this instance, a measure of the amount of the hazardous agent residing in the body (i.e., body burden) may be more directly related to an adverse response than another metric. Body burdens are a metric of internal exposure and are typically determined via biomonitoring methods that measure the hazardous agent, its metabolites, or other reaction products in a biologic matrix (e.g., human tissues, saliva, blood, or excreta) [Needham et al. 2007]. Pharmacokinetic models can also be used to estimate body burdens. Thus, an understanding of the underlying biological mechanism related to the outcome of interest is important for index selection. If an understanding of the expected response is lacking, then it may be necessary to assess the dose-response by examining multiple indices [Blair and Stewart 1992]. In this case, the choice of the best index is based on its validity and reliability, and its utility in subsequent dose-response analyses.

Summary (aggregate) exposure metrics (e.g., average, geometric mean, or peak exposures) are often used to assign group level exposure indices in the absence of individual data. Clearly, the choice of summary metric can have a marked effect on results from dose response analyses. More often than not; however, the choice is limited to published results that may not be best suited for risk characterization. For example, exposure distributions of most occupational agents tend to be right-skewed and geometric mean values are used in many studies as a measure of central tendency. Although these measures may be appropriate for the intended purpose, the choice was likely made without consideration of a subsequent use in describing population risk. It has been shown that the appropriate group assignments for risk characterization is largely dependent on the expected shape of the dose response, regardless of the underlying exposure distribution [Crump 1998; Seixas et al. 1988]. In most situations, the anticipated response is increasing with dose; therefore, the arithmetic mean (i.e., average) provides a better approximation for assessing population risk [Crump 1998].

4.3.3.2 Direct Assessment Methods

Direct methods of exposure assessment refer to measurements of the agent of interest that are obtained at the individual or group level. Exposure information that is directly obtained from personal measurements is likely to provide the most accurate estimate of individual exposure. Personal exposure monitoring can be conducted in the environment of the worker (e.g., direct reading dosimeters and breathing zone air samples), *in vivo* (e.g., whole body radiation counter), or can involve biomarkers of the agent of interest, its metabolites, or its effects (e.g., chromosome aberrations from ionizing radiation exposure) in biologic media, such as blood, hair, excreta, sputum, sweat, or exhaled breath. Ideally, personal monitoring for every agent of interest is best performed during a series of tasks that are representative of the occupation of each worker and over a sufficient period to inform on the exposure distribution. For example, many workers who have been employed in the nuclear industry have worn personal radiation dosimeters in radiation areas throughout their careers beginning as early as the late 1940's. However, personal monitoring of ionizing radiation exposure is the exception; limitations in logistics, costs, and technology, have excluded widespread use of personal monitoring in most other industries.

A more common approach is the use of group level measurements. Group level measurements pertain to either: 1) measurements from personal monitoring of a worker or a sample of workers who represent a similar exposure group; or ambient measurements in work areas occupied by the similar exposure group. A summary measure of exposure is assigned to each member of the similar exposure group; thus, estimates rely on the underlying assumption of similar exposure level and variation among persons within the similar exposure group. In both instances, the use of group level exposure measurements to represent all members of their similar exposure groups should be evaluated in the risk assessment whenever practical.

Ambient (stationary or area) measurements are further limited by the required translation to individual exposure. For example, measurements from a fixed air sampler placed between the exposure source and the exposed person may tend to overestimate the individual exposure. Furthermore, sampling plans are often designed to describe maximum exposures for regulatory compliance purposes. Hence, exposure estimates from such

sampling plans would be susceptible to overestimation. Risk assessors consider the potential for exposure misclassification resulting from the design and conduct of ambient exposure measurements that are subsequently used in risk assessment. With regard to group assignments based on personal monitoring, exposure estimates are strengthened by increased sample sizes and the use of repeat measures that enable an assessment of between and within-worker variability. NIOSH risk assessors consider sample size and the availability of repeat measures when assessing the validity of group assigned exposures from personal monitoring data.

4.3.3.3 Indirect Assessment Methods

There is a paucity of historical industrial hygiene monitoring data available for most hazards; therefore, indirect methods of exposure assessment are commonplace in occupational studies. Exposure estimates can be derived indirectly from proxy measures, questionnaires, expert judgement, job-exposure matrices (JEMs), statistical models, or any combination of these sources. There are several comprehensive reviews on data sources, assessment methods, uncertainties, and validation techniques available to risk assessors [Kauppinen 1994; Seixas and Checkoway 1995; Stewart et al. 1996; Teschke et al. 2002].

4.3.3.3.1 Self-Report or Proxy Respondent Data

When data are obtained directly from individual study participants or indirectly from proxy responses to a study interview or questionnaire, assessments are subject to bias from recall that has been influenced by case status (i.e., recall bias). The literature is abundant with reports examining the validity and reliability of exposure assessment methods using self or proxy reported data [Ahlborg Jr 1990; Baumgarten et al. 1983; Benke et al. 2001; Bond et al. 1988; Bourbonnais et al. 1988; Fritschi et al. 1996; Joffe 1992; Nieuwenhuijsen 2010; Stewart et al. 1987; Teschke et al. 1994].

4.3.3.3.2 Job Exposure Matrix

Job exposure matrices (JEM) are widely used by NIOSH for estimating exposure indices, whereby a ‘job’, which is defined by relevant employment information (e.g., job title, task, department, and plant), is

systematically linked to an exposure level. Noteworthy early JEM examples involve assessments of exposures to silica, asbestos, and solvents [Dement et al. 1983; Eisen et al. 1984; Gardner et al. 1986; Rice et al. 1984; Rinsky et al. 1987; Seixas et al. 1997; Stewart et al. 1986]. JEMs have been used to identify similar exposure groups, provide individual qualitative and quantitative exposure estimates, and in conjunction with algorithms and statistical models, to fill in gaps in exposure information during time periods when monitoring data were unavailable [Coughlin and Chiazze 1990; Dement et al. 1983; Eisen et al. 1984; Hallock et al. 1994; Hornung et al. 1994; Seixas et al. 1997; Woskie et al. 1988].

In its simplest form, the JEM is a table with rows and columns characterizing occupation and exposure, respectively. Thus, each cell represents an estimate of the exposure for individuals linked to an occupation. Strata for occupation and exposure are optimized to increase estimate precision while reflecting the exposure gradient, which is necessary for dose-response analyses. Of course, there is still a large potential for exposure misclassification in a two-dimensional JEM. This misclassification can be reduced by adding dimensions. Contemporary JEMs typically describe exposures along four axes, comprising strata for: the agent, job or task, time, and location.

NIOSH risk assessors evaluate the quality of source data and methods used in the JEM to reduce exposure misclassification. Given the JEM's reliance on employer-provided information, the completeness, accuracy, and scale of these data are typically scrutinized. These data fall into two categories corresponding to the primary dimensions: 1) individual employment information used to establish task, time, and location of the worker, and 2) process information and plant industrial hygiene data used to assess the exposure potential (i.e., agent). Worker data often stem from personnel records, medical histories, or questionnaires. Exposure data often include job descriptions, chemical inventories, monitoring data, and incident and accident reports. Information on exposure modifiers is likely found in plant records on engineering controls, administrative policies, and personal protective equipment use. Employer-provided information is rarely complete; therefore, JEMs are sometimes augmented by data from other sources (e.g., new measurement data, statistical models, and other JEMs). For example, industrial

hygiene data from routine sampling that began in the 1980's may have supported exposure estimates for previous decades. In this case, the risk assessor assesses the methods used to extend estimates to unmonitored periods.

In some instances, data are available from other sources (e.g., measurement data, statistical models, and other JEMs) that can be used to assess the quality of the JEM and/or quantify the magnitude of potential misclassification. For example, consider a cohort study that estimated exposure using personnel records and ambient air measurements but had personal monitoring data available on a subset of the study population. These monitoring data could then be used as a standard for comparison to study estimates and be a means to calibrate the JEM.

The exposure assessment information in the epidemiologic report is likely to be brief and have limited use for assessing data completeness. Fortunately, superior JEMs are often documented in separate detailed reports that are available in the published literature or can be found in study records. For example, an exposure assessment for a cohort mortality study of beryllium processing workers in multiple plants [Schubauer-Berigan et al. 2011] relied on data from three separately published JEMs [Chen 2001; Couch et al. 2011; Sanderson et al. 2001] for dose-response analyses that were subsequently used in quantitative risk assessment by for developing permissible exposure levels [Schubauer-Berigan et al. 2017]. When JEM data are not published, risk assessors are encouraged to contact investigators for additional documentation needed to assess the quality of the exposure assessment supporting study findings.

4.3.3.3.3 Expert Assessment

Employment information and/or self-reported data are often used in tandem with expert judgement by industrial hygienists, chemists, engineers, and other professionals to estimate exposure [Nieuwenhuijsen 2010; Teschke et al. 2002]. It is generally thought that experts, having a better understanding of exposure mechanisms because of their training, can more accurately estimate exposures compared to the workers themselves. Furthermore, if the experts are kept blind to case status, the potential for information bias is reduced. Nevertheless, it may be impractical for experts to become suitably familiar with all exposure conditions over the

period of interest without detailed information from employer records and the affected workforce. Another disadvantage of expert judgment is an inherent inconsistency among experts given relatively unstructured opinions about the exposure that have developed from varying levels of training and experience. In preferred studies using expert judgement, reliability is typically assessed by comparing estimates from two or more experts [Benke et al. 1997; Kromhout et al. 1987; Ramachandran et al. 2003; T Mannetje et al. 2003; Van Wendel De Joode et al. 2005a]. When available, comparisons with measurement data are preferred for assessing validity [Benke et al. 1997; Tielemans et al. 1999; Van Wendel De Joode et al. 2005b].

4.3.4 Factors Compromising Validity

Bias is defined as a deviation of the results or inferences from the truth, or processes leading to that deviation [Gail and Benichou 2000]. Study designs are typically evaluated by risk assessors to ensure that candidate epidemiologic studies are absent of major systematic errors. Sources of systematic errors can be classified into four general forms:

- Selection bias, resulting from procedures used to select participants into or out of the study or that inherently occurs as part of the normal occupational setting (e.g. healthy worker survivor bias effects described below).
- Information bias, resulting from misclassification of the study participants' disease or exposure status.
- Confounding, which is a mixing of the effects from the exposure of interest with the effects of other measured or unmeasured factors (confounders) on the risk of the adverse effect. Insufficient accounting for confounding factors can lead to significantly biased risk estimates.
- Healthy worker effects, which are a combination of selection and confounding biases resulting from relationships between heath status, employment, and exposure. This source of potential bias is restricted to observational studies of working populations.

Based on these general forms, risk assessors must answer the following questions:

1. Is there any evidence suggesting a potential for a strong selection bias?
2. Is there any evidence supporting the potential for a strong information bias?
3. Could the findings of the study be attributed to confounding by other risk factors either because of inadequate control (residual confounding) or because of lack of control?
4. In lieu of significant bias, how likely are the study findings to have resulted from chance?

Ideally, the risk assessment relies on study results that are not attributable to bias or chance; therefore, a positive response to any of the questions above is grounds for removal of the study from further consideration. The term "strong" is used to describe an unacceptable bias by its degree; therefore, it implies that a small distortion of the effect relative to its reported size (i.e., a potential weak bias) may not be disqualifying for risk assessment purposes. Appendix B of this report provides specific information on potential biases as an aid the risk assessor in responding to these questions.

Bias can occur in any stage of the research, including the literature review, study design, data collection, analysis, interpretation of results, and publication. When reviewing studies for validity, risk assessors avoid using a "guilty until proven innocent approach" whereby one assumes that the study design and analysis are inadequate unless sufficient information to the contrary is provided by the study authors. Instead, the risk assessor evaluates the potential impact of study limitations and omissions on findings for determining WoE [Zaccai 2004]. While some discussion on specific types of biases that may be encountered is provided in Appendix B, risk assessors may consult several highly cited articles for additional information [Arrighi and Hertz-Picciotto 1994; Delgado-Rodríguez and Llorca 2004; Greenland et al. 1999; Grimes and Schulz 2002; Sackett 1979] and available epidemiologic texts [Breslow and Day 1980; Breslow and Day 1987; Checkoway et al. 2004; Gail and Benichou 2000; Rothman et al. 2008]. .

5.0 DOSE RESPONSE ASSESSMENT

5.1 Introduction

The dose-response assessment is the second step of NIOSH risk assessment. Here, 'dose-response' refers to the relationship between the amount of an agent administered to, taken up by, or absorbed by an organism, system, or population (i.e., the "dose") and the adverse effect developed in that organism, system, or population in reaction to the agent (i.e., the 'response") [IPCS 2004]. The aim of the dose-response assessment is to obtain reliable and valid estimates of the point of departure in a cause and effect relationship or the risk per unit dose that can be used in risk characterization. For example, the dose-response information for an occupational carcinogen can be used to estimate a chemical concentration corresponding to a risk level for invoking risk management decisions [NIOSH 2017]. Alternatively, a benchmark dose can be estimated, uncertainty factors applied and a "safe" concentration determined for non-cancer effects such as reproductive or developmental effects.

The "dose" represents a quantitative metric, *d*, usually derived from some external exposure, and believed predictive of an adverse effect. Whenever the relationship between the biologically effective dose and another dose metric, e.g., absorbed dose, inhaled dose or exposure concentration, is well-described by a constant ratio, i.e., a directly proportional relationship over the range of doses under consideration, then these doses are interchangeable and their dose-responses are equivalent. For example, if the inhalation rate during exposure is constant over the range of concentrations then the inhaled dose rate is proportional to the concentration; but if the inhalation rate is not constant then the dose-response based on inhaled dose requires adjustment to obtain the corresponding dose-response based on exposure concentration. As another common example, consider a non-linear rate of metabolic activation. The rate of metabolic activation of a toxicant may be best approximated as linear at low exposure concentrations, but may become non-linear at high concentrations. If the biologically effective dose is the amount that is metabolically activated, then the dose-response analysis is usually based on the amount activated rather than the exposure concentration, if that has been quantitatively described and validated.

5.2 Dose-Response Modeling

Dose-response regression modeling provides a basis to estimate the expected response as a function of dose *d* and possibly other risk factors, $X_1, X_2, \ldots, X_c$, together with assumptions about the variability of responses. In animal toxicology studies, the dose-response is often simplified to expected response = $f(d)$ since the other risk factors are controlled by design or by the random assignment to dose levels. As an illustration, consider the outline of animal study data in Table 5-1. At each dose level *d*, there are *n* animals exposed, and the corresponding response, *Y*, is number of animals presenting with the adverse effect of interest. The expected proportion of animals with the adverse effect is related to each dose *d* and is equivalent to the probability of the adverse effect, $f(d)$; the function, $f(d)$, is generally modeled by a continuous function that can be evaluated at any dose between the background response, $f(0)$, when $d = 0$ and the maximum observed dose.

Table 5-1. Illustration of data from an animal bioassay for a dichotomous response[1]

Dose (d_i)	Number of exposed animals (n_i)	Number observed with the response (Y_i)	Observed proportion ($\boldsymbol{Y_i/n_i}$)
d_0[2]	n_0	Y_0	Y_0/n_0
d_1	n_1	Y_1	Y_1/n_1
…	…	…	…
d_D	n_D	Y_D	Y_D/n_D

1. For example, cancer of a target organ or tissue.
2. Typically, an unexposed group of controls are used and $d_0=0$.

Clearly, it is preferable to base model selection on biologic plausibility. In practice; however, model specification with a clear advantage based on biology is seldom observed. Instead, a suite of plausible models is usually fit to the data. When multiple models of a response adequately describe the data, the model selected for a risk assessment is generally chosen using criteria that is defined *a priori*. For example, the Akaike information criterion (AIC) is a criterion frequently used to select the preferred model, which is a measure of model fit with a penalty for model complexity, defined as the number of estimated parameters in the model [Akaike 1974]. The

strategy for selecting a model is generally specified before the modeling results are examined. Preferably, these methods are clearly described in the risk assessment. As different model-selection criteria can lead to different model choices, model selection is often an area explored in sensitivity analysis. Multiple (alternative) estimates are then reported with a description of how each estimate was derived.

5.2.1 Parametric Dose-Response Modeling

The function $f(d_i; \boldsymbol{\theta})$, that describes the relationship between dose and the expected response for observation *i*, is often assumed to have a known form that depends on a vector of parameters $\boldsymbol{\theta}$ whose unknown values are estimated. This assumption places strong constraints on the shape of the dose-response curve and the data are used to estimate $\boldsymbol{\theta}$. Unknown quantities of critical interest such as risk associated with a given dose or the dose associated with a given risk are estimated based on the fitted dose-response $f(d_i; \widehat{\boldsymbol{\theta}})$. Within the form adopted for $f(\cdot)$ multiple ways to describe the effect of dose may be available, e.g., $\theta_1 B_1(d) + \theta_2 B_2(d) + \cdots + \theta_K B_K(d)$, where the $B_k(d)$ are known functions of dose d. As an illustration, if the effect of dose is to be described by $\theta_1 d + \theta_2 d^2$ then $B_k(d) = d^k$ and $\theta_3 = \theta_4 = \cdots = \theta_K = 0$. This suggests that a hierarchy of increasing flexibility may be examined, e.g., $\theta_1 d$ followed by $\theta_1 d + \theta_2 d^2$, etc., but this should be done carefully since the inclusion of unnecessary terms degrades the statistical precision of estimation and this degradation can be substantial. However, the omission of a necessary term is likely to introduce a statistical bias into the estimation. Thus, there is trade-off between a potential for bias associated with an overly constrained model of the dose-response versus a degradation of precision, i.e., increased variance, from an unnecessarily flexible model; this relationship between potential bias vs increased variance holds in general and is referred to as a bias-vs-variance trade-off. The *a priori* specification of model selection criteria such as AIC can be helpful with assessing it.

5.2.1.1 Dichotomous Response Data Modeling

Many different parametric models have been proposed for toxicology or epidemiologic data. For example, in the toxicology setting, the following model specifications are commonly used in dichotomous dose-response modeling:

Logistic
$$f(d) = \frac{1}{1+\exp\left[-\left(\alpha+\beta d\right)\right]} \quad \text{(eqn. 5-1)}$$

Log-logistic
$$f(d) = \gamma + \frac{(1-\gamma)}{1+\exp\left[-\left(\alpha+\beta \ln(d)\right)\right]}, \; 0 \le \gamma < 1, \; \beta \ge 1 \quad \text{(eqn. 5-2)}$$

Gamma
$$f(d_i) = \gamma + (1-\gamma)\frac{1}{\Gamma(\alpha)}\int_0^{\beta d} t^{\alpha-1}e^{-t}dt, \; 0 \le \gamma < 1, \; \alpha \ge 1, \; \beta \ge 0 \quad \text{(eqn. 5-3)}$$

Multistage (degree=2)
$$f(d) = \gamma + (1-\gamma)\left[1-\exp(-\theta_1 d - \theta_2 d^2)\right], \; 0 \le \gamma < 1, \; \theta 1 \ge 0, \; \theta 2 \ge 0 \quad \text{(eqn. 5-4)}$$

Probit
$$f(d) = \Phi\left(a+\beta d\right) \quad \text{(eqn. 5-5)}$$

Log-probit
$$f(d) = \gamma + (1-\gamma)\Phi\left[a+\beta \ln d\right], \; 0 \le \gamma < 1, \; \beta \ge 0.5 \quad \text{(eqn. 5-6)}$$

Quantal-linear
$$f(d) = \gamma + (1-\gamma)\left[1-\exp(-\beta d)\right], \; 0 \le \gamma < 1 \quad \text{(eqn. 5-7)}$$

Quantal-quadratic
$$f(d) = \gamma + (1-\gamma)\left[1-\exp(-\beta d^2)\right], \; 0 \le \gamma < 1 \quad \text{(eqn. 5-8)}$$

Weibull
$$f(d) = \gamma + (1-\gamma)\left[1-\exp(-\beta d^{\alpha})\right], \; 0 \le \gamma < 1, \; \alpha \ge 0.5, \; \beta \ge 0 \quad \text{(eqn. 5-9)}$$

where $f(d)$ represents the probability of adverse response given dose *d*, Φ(x) is the cumulative distribution function of a standard normal random variable at x (i.e., the integral of a N(0,1) density from –∞ to x), and *f(0)* = γ when *d*=0 for models (5-2 to 5-4) and (5-6 to 5-9). Some bounds in the above models are arbitrarily set to prevent extreme properties and attendant computational problems although nonlinear dose-response patterns remain available. Hence, modification of these constraints may be necessary when consideration of either pattern is unwarranted. Furthermore, although the model forms (eqn. 5-1through 5-9) above encompass a wide variety of curves to represent the dose-response and have readily available software for their implementation, other parametric forms could be considered if necessary.

Dichotomous outcomes from animal bioassays are often modeled under an assumption that the sampling variation of the underlying experimental process is well represented by a binomial distribution. In some cases, especially where the data are pooled from multiple studies or substantial genetic variations of the animals are present, this assumption is not appropriate and extra binomial variability, or "over-dispersion" is observed. In these instances, beta-binomial, quasi-likelihood methods, or more fully defined models incorporating random effects are preferred.

5.2.1.2 Continuous Response Data Modeling

Continuous data arise when response values come from a continuous distribution, for example, precisely measured liver weights or pulmonary function tests. In these situations, a variety of parametric models exists to predict the mean response $f(d) = \mu(d)$. For example, the following five parametric models are often considered when modeling continuous data:

Linear $$f(d) = \beta_0 + \beta_1 d \qquad \text{(eqn. 5-10a)}$$

Quadratic $$f(d) = \beta_0 + \beta_1 d + \beta_2 d^2 \qquad \text{(eqn. 5-10)}$$

Power $$f(d) = \beta_0 + \beta_1 d^{\beta_3} \qquad \text{(eqn. 5-11)}$$

Hill $$f(d) = \beta_0 + \beta_1 \frac{d^{\beta_3}}{\beta_2^{\beta_3} + d^{\beta_3}}$$ (eqn. 5-12)

Exponential $$f(d) = \beta_0 \left\{\beta_2 - (\beta_2 - 1)\exp\left[\left(-\beta_1 d\right)^{\beta_3}\right]\right\}$$ (eqn. 5-13)

The parameters β_0, β_1, β_2, and β_3 are specific to the given model. Unlike the previous set of dichotomous models, these models are unbounded.

The above models describe the mean response; however, additional modeling assumptions on the variance of the response data are generally needed. For example, in many situations, the variance may be a function of the mean, e.g., it may be a positive constant, or it may be proportional to the mean or power of the mean such as its square. Model fit is examined to assess whether the response mean and variance structures are supported.

5.2.1.3 Parametric Dose-Response Modeling including Other Predictors

The extension to include other predictors into the function $f(d_i, X_{1i}, X_{2i}, \cdots, X_{ci}; \boldsymbol{\theta})$ to describe the expected response shares much with the dose-only modeling described above in that the data are used to estimate $\boldsymbol{\theta}$ to obtain $f\left(d_i, X_{1i}, X_{2i}, \cdots, X_{ci}; \widehat{\boldsymbol{\theta}}\right)$. However, the models of the data are more complex and the vector of parameters $\boldsymbol{\theta}$ contains coefficients that govern the effect of dose and coefficients for the effects of the other variables. In epidemiological studies, d_i, is an exposure metric constructed from possibly complex employment histories, and identifying the optimum construct for d_i may, itself, be an important component of the modeling procedure. For human observational studies, predictors in the function $f(d_i, X_{1i}, X_{2i}, \cdots, X_{ci}; \boldsymbol{\theta})$ usually include age, sex and other demographic variables and may include interactions or effect modifiers, e.g., dose-rate effects or other effect modifiers that allow for the effect of dose to depend on the other predictors. These additional factors can make model selection using human data more complex because confounders, effect modifiers, and complex selection processes can be present and there are often many ways they can enter the model. In addition,

the differences between the study population and the target population should be considered for the risk estimation. If interactions are present then the estimates of interest may be made conditional on fixed values of $X_{1i}, X_{2i}, \cdots, X_{ci}$ or averaged over the appropriate marginal distribution of $X_{1i}, X_{2i}, \cdots, X_{ci}$; or, a combination of fixing some values and averaging over the others.

5.2.2 Model Uncertainty and Model Averaging

Model averaging takes into account model uncertainty by incorporating results from all models into the estimation process through a weighted average of the model-specific excess risk estimates. This technique has been applied in a general modeling context by Raftery [1995], who suggested the use of the posterior model probabilities as weights derived from a Bayesian analysis of all models considered. As a full Bayesian analysis is frequently computationally burdensome, Buckland et al. [1997] proposed simpler methods, where weights are based upon the penalized likelihood functions formed from the AIC and Bayesian Information Criteria (BIC) [Schwarz 1978]. The AIC and the BIC are defined on likelihood functions where the AIC = -2 log L + 2p and the BIC = -2 log (L) + p log (n), and p is the number of parameters in the model, L is the maximum likelihood value, and n is the sample size.

The NIOSH approach is to use a model-averaged fit to synthesize risk estimates across multiple fitted parametric models. An estimate of the dose-response function $\hat{f}_{\text{MA}}(d)$ is calculated as a weighted average of K model-specific dose-response estimates $f_k(\hat{\boldsymbol{\theta}}_k, d)$ for k=1, …, K. Formally this is represented as $\hat{f}_{\text{MA}}(d) = \sum_{k=1}^{K} f_k(\hat{\boldsymbol{\theta}}_k, d) \cdot w_k$, where $f_k(\hat{\boldsymbol{\theta}}_k, d)$ represents the adverse effect given the dose d using the k^{th} model, $\hat{\boldsymbol{\theta}}_k$ is the estimated parameter vector for the k^{th} model, and w_k represents the corresponding weight for the k^{th} model (e.g. $\hat{\boldsymbol{\theta}}_k = \lfloor \hat{a} \quad \hat{\beta} \rfloor$ for model 5). Given the model M_k in the model space that includes K models, the weight w_k is:

$$w_k = \frac{\exp(-I_k / 2)}{\sum_{i=1}^{K} \exp(-I_i / 2)},$$

where I_i represents the penalized information criterion described above (e.g. AIC or BIC). Other weighting mechanisms exist; for more information on these different strategies, see Morales et al. [2006] and Moon et al. [2005].

5.2.3 Semiparametric or Nonparametric Modeling

The use of parametric models to describe the dose-response relationship may not be necessary. Instead, a nonparametric curve can be used that allows for a more flexible approach of fitting data to a dose response curve. The methodologies available to achieve this vary and often make the mild assumption of monotonicity with a possible smoothness constraint.

Wheeler and Bailer [2012] describe a Bayesian semiparametric method that uses a flexible spline construction for dose-response analyses. This method was shown to be superior to the model averaging method of Wheeler and Bailer [2007] in terms of its statistical properties. The method is fully Bayesian, which requires attention to the specification of prior distributions but it allows one to include prior information on such things as the incidence of the response in historical animal controls or in human reference populations. Even though this method is free of the model selection issues encountered in benchmark dose modeling, informed choices must still be addressed with this method. Its use requires the choice of spline basis functions located at specific knot locations, which should be selected before modeling begins. Ultimately, flexibility in the choice of these models comes at the expense of statistical and computational challenges in fitting such models.

Other fully semiparametric/nonparametric modeling methodologies have been developed for dichotomous and continuous data [Guha et al. 2013; Lin et al. 2015; Piegorsch et al. 2013; Piegorsch et al. 2012; Wheeler et al. 2015], some of which overcome the known selection problems of Wheeler and Bailer [2007]. These methods are

fully nonparametric in that they assume no prior form of the dose-response curve except monotonicity. Lin et al., [2015] showed that their continuous method would converge to the true underlying dose-response curve for large samples. The method of Wheeler et al. [2015] accounts for uncertainty in the specified response distribution for continuous outcomes and the dose-response. This method was shown through simulation to produce accurate estimates of excess risk provided studies had sufficient numbers of observations. Like model averaging, these methods allow for a flexible representation of the dose-response curve and are often preferable to a single parametric model fit.

5.3 Point of Departure

The point of departure (PoD) is defined as the point on the dose-response curve that is established from experimental or observational data generally corresponding to an estimated level of no effect or a low effect level that is without significant extrapolation to lower doses. The PoD is used in conjunction with uncertainty factors to predict a safe level of exposure or to mark the beginning of model-based low-dose extrapolation to dose points associated with a target risk level. This is necessary when there is instability in model-based estimation at very low doses. These PoD concepts have their origins, and continue to be widely used, in animal toxicologic studies; therefore, much of the discussion on PoD metrics is in context of methods using animal bioassay data. Nevertheless, some of these concepts have been adapted to epidemiologic data from observational studies [Bailer et al. 1997; Budtz-Jørgensen et al. 2001; Noble et al. 2009]. With respect to these metrics, three definitions of the PoD are commonly used in NIOSH risk assessments. These are: 1) the No Observed Adverse Effect Level (NOAEL), 2) the Lowest Observed Adverse Effect Level (LOAEL), and 3) the Benchmark Dose (BMD). These are described in the following sections.

5.3.1 NOAEL/ LOAEL-Based Assessments

The NOAEL is defined as the highest dose level at which there are no significant increase in the frequency or severity of adverse effects between the exposed population and its appropriate control; some effects

may be produced at this dose level, but they are not considered adverse or precursors of adverse effects observed [EPA 2012b]. For example, given a rank order series of exposure groups in a toxicologic study, the NOAEL is the administered dose level in the exposure group that immediately precedes the first exposure group in which the frequency of the observed adverse effect significantly differs from that in the control (no exposure) group. Similarly, the LOAEL is *lowest* dose or concentration at which there are significant increases in frequency or severity of adverse effects between the exposed population and its appropriate control group [EPA 2012b].

Usually, statistical hypothesis tests with a significance level set to 5% is used to identify NOAELs or LOAELs. As such, problems arise in studies with few subjects observed at low exposure levels due to insufficient signal to noise ratios and statistical power. Nevertheless, it has been shown in most animal studies that the highest exposure group qualifying as a NOAEL is, on average, equivalent to model-based BMD using a BMR of 10%, which is empirical evidence that many NOAELs were associated with an increased response that did not meet the standard significance level of 0.05 [Wignall et al. 2014]. Other limitations in using a NOAEL/LOAEL approach include: a) it ignores the shape of the dose-response curve which would inform estimation at lower levels, b) it is constrained to be one of the levels of exposure selected in the experiment, c) the spacing of exposures in an experiment can result in only high doses having sufficient power to detect statistically significant differences from the background condition [Crump 1984] even though biologically significant effects at lower doses may have been missed due to limited statistical power or sampling error; hence, basing an interpretation of a NOAEL as representing a threshold below which effects are null is generally unfounded. Despite these limitations, the NOAEL/LOAEL approach may be the only alternative for determining a PoD for application of uncertainty factors when data are insufficient to model the dose-response adequately. NIOSH has used a NOAEL/LOAEL approach in assessing risks of occupational exposures to ethylene glycol ethyl ethers and associated chemicals [NIOSH 1991].

5.3.2 The Benchmark Dose Approach

Given the limitations of the NOAEL/LOAEL approach, the risk assessment community has widely adopted an approach originally proposed by Crump [1984] for determining a benchmark dose (BMD) as the PoD when observed data are adequate to model the dose-response (see Section 5.5). The BMD is defined as the dose or concentration that produces a predetermined change in the response rate of an adverse effect relative to the background response rate of this effect. This predetermined change is called a "benchmark response" or BMR. The BMR is usually in the range of 5-10% for toxicologic data, which is the limit of responses typically observed in well-conducted animal experiments [EPA 2012b]. Given a BMR value that is selected *a priori*, the risk assessor fits various dose-response models to the observed data. This approach is applicable to dichotomous, ordinal, or continuous response data and categorical or continuous exposure data [Chen and Chen 2014; Crump 1995; Crump 1984]. For continuous response data, the BMR is usually based on a central measure of the biological effect (e.g., mean organ weight), a measure of its variability (e.g., standard error), and the number of observations at each dose level [Davis et al. 2011]. Regression models are fit to dose-response data that should include at least two dose groups above the control and in the low-dose range of interest (e.g., in the range of the BMR). The resulting curve(s) is used to calculate the *BMD* and its one-sided lower 95% confidence limit (*BMDL*). The BMDL is typically used to define the PoD. This process accounts for the variability and uncertainty in the experimental results (but not uncertainty in model selection) [Davis et al. 2011]. The general BMD approach is illustrated in Figure 5-1 below.

This BMD approach is preferred by the EPA [2012b], who has developed benchmark dose software (BMDS) that is readily accessible to risk assessors worldwide. Specifically, the EPA software allows for the examination of a suite of dose-response functions for selection of the best single dose-response model. Model fitting is achieved by maximum likelihood. The adequacy of models is judged by likelihood goodness of fit (i.e., testing for lack of fit) typically using a critical value of 0.1 as a threshold for acceptance. Selecting the 'best fit' model from a set of adequately fitting nested models can be accomplished, in part, by likelihood ratio tests.

Similarly, selecting a model from a set on unrelated models can be achieved by comparing AIC or BIC values. NIOSH risk assessors examine the variability in BMD or BMDL estimates across adequately fitting models. Reasonable agreement in estimates among a set of models suggests little model dependence; therefore, selection based on the lowest AIC value is well supported. Conversely, divergence in model estimates is indicative of model dependence. It is important to note that model dependence results from extrapolation; therefore, it is more likely to occur when the BMR value is below the observable range. It is prudent to examine adequately fitted models closely to determine if the variability is attributable to anomalies in the data. In particular, risk assessors may reject models that do not adequately describe the low-dose portion of the dose-response relationship, as determined by examining residuals and model plots. When the group of adequately fitted models is divergent and in lieu of other evidence supporting model rejection, a health-protective approach is to select the model that provides the lowest BMDL estimate [EPA 2012b]. Other options include summary estimates from multiple models, such as in model averaging analyses (Section 5.2.2) or the use of semiparametric on nonparametric models (Section 5.2.3).

Ideally, the dose-response and its associated uncertainty at the target risk level can be directly estimated from the data. Still, it is frequently the case that the PoD is determined at a higher response rate than a response of interest; therefore, extrapolation toward the origin of the dose-response curve may be required. For example, a PoD based on a BMR of 10% excess risk of cancer is likely to require extrapolation to a much lower dose-risk region of interest to support a suitable estimate of lifetime risk. The common practice of 1) setting the BMR at 10% extra risk, 2) using the BMDL as the PoD, and 3) linearly extrapolating to the risk level of interest, is well supported by studies showing that the BMD is often in the range of the NOAEL [Sand et al. 2011; Wignall et al. 2014]. For example, NIOSH typically uses linear extrapolation for cancer risk assessments unless mechanistic or mode of action data support a different approach (Section 7.2.1). In cases in which data support a nonlinear dose-response, low-dose extrapolation is accomplished via the selected parametric dose-response curve or by model averaging, semiparametric, or nonparametric methods.

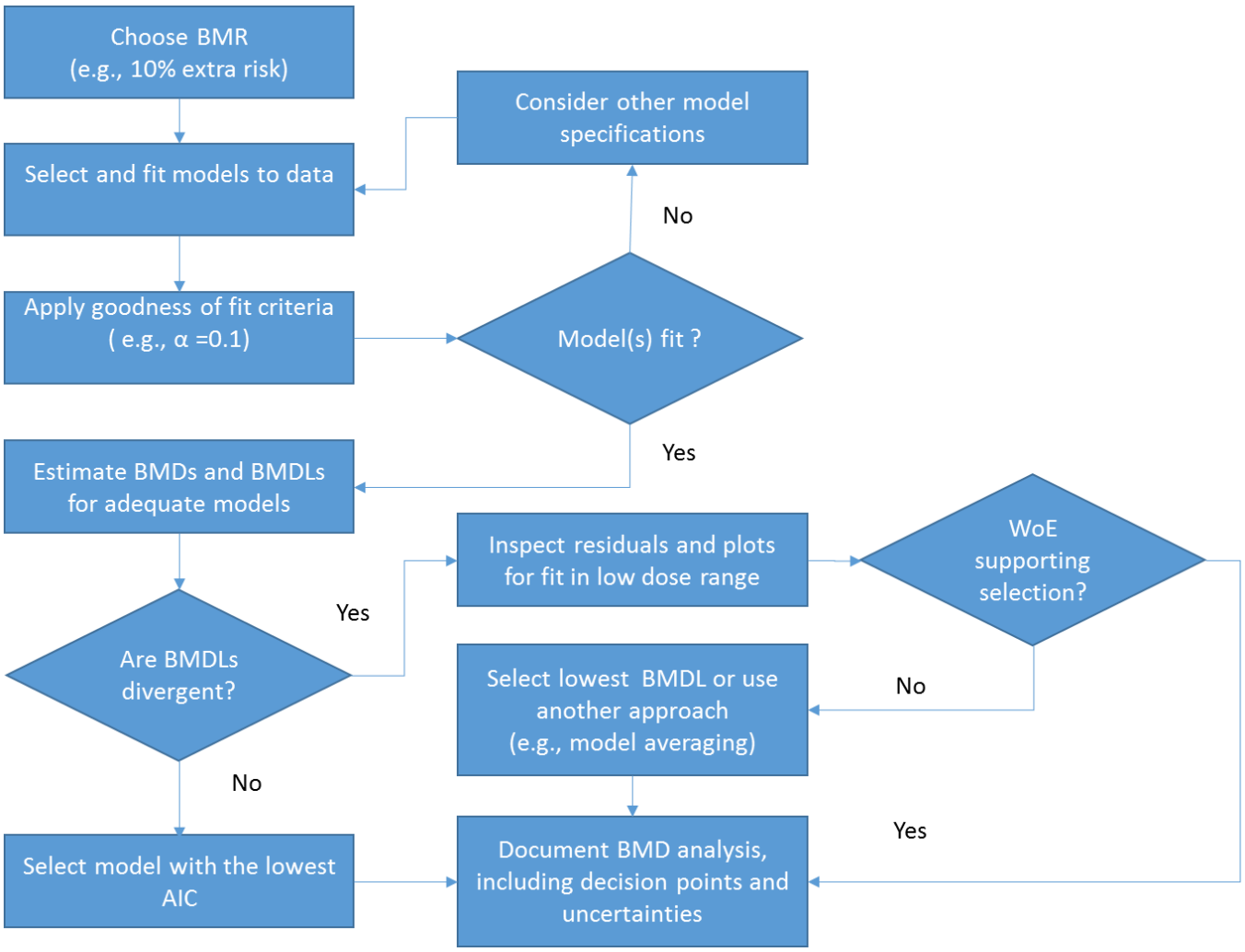

Figure 5-1. The Benchmark dose method, selecting a single parametric model (adapted from Davis et al. [2011])

NIOSH used the BMD approach in its risk assessment of occupational exposures to carbon nanotubes and nanofibers [NIOSH 2013b]. The dataset was abstracted from short-term and subchronic studies of nonmalignant pulmonary responses in exposed rats and mice. Both quantal and continuous response data were examined. The BMR was set at 10% added risk of early stage adverse lung effects. The one-sided 95% BMDL was selected as the PoD. Modeling was conducted using the EPA benchmark modeling software. Although several models were specified, only a multistage (polynomial degree 2) model adequately fit quantal response data used in this risk

assessment. The continuous dose-response data were fit with a second order polynomial model for all data with three or more dose groups, and a linear model for data with two groups. As a departure from traditional methods, model goodness-of-fit was considered adequate at $P > 0.05$. The authors, explained the choice as "… a trade-off in the type I or type II error." Nevertheless, all selected models (i.e., NIOSH 2013b, Tables A-3 to A-5) met $P \geq 0.10$. NIOSH also applied BMD concepts to epidemiologic data in its assessment of the risk associated with occupation diacetyl exposure [NIOSH 2016a]. Multiple BMRs describing pulmonary impairment were derived from continuous data on pulmonary function.

5.3.2.1 Determining the PoD using Model Averaging and Semiparametric Methods

Linear extrapolation below the PoD is unnecessary when using model averaging, semiparametric or nonparametric approaches [Wheeler and Bailer 2012; Wheeler and Bailer 2007] because the estimation of exposures corresponding to small excess risks is model-based. Actually, the model-based extrapolation may result in a value similar to a linear extrapolation from a PoD unless substantial evidence against the latter is present in the data. For example, when applied to actual data and investigated in simulation studies, these model-averaging and semiparametric approaches have adequately described both the model and statistical uncertainties at excess risk levels well below the 5 or 10% level. Wheeler and Bailer [2013] found that for dose-responses that were low-dose linear these approaches yielded estimates that differed negligibly from a linear extrapolation from the 10% level for target risks as low as 0.001%. For non-linear dose-response relationships, these methodologies were observed to provide superior estimates (i.e., BMDLs that maintained nominal coverage but were closer to the point estimate) than the PoD linear extrapolation while still accurately describing the risk. These two methodologies were also observed to be internally consistent producing similar estimates, usually within a factor of three, across all excess risk levels examined. Parametric extrapolations to excess risks as low as 0.001 often produce BMD estimates that differ by multiple orders of magnitude between different models, which is why BMRs are often set to 10% when parametric models are used.

In addition to examining single parametric models, NIOSH used a model averaging method to summarize risk estimates from linear-quadratic, Weibull, and log-probit models in its risk assessment of lung cancer and titanium dioxide (TiO_2) exposure [NIOSH 2011]. This approach was chosen because the dose-response relationship appeared nonlinear, and the specific models used in the three-model average procedure did not impose (although allowed) low-dose linearity for risk extrapolation. In this model, Weibull and log-probit models were weighted more heavily than the linear-quadratic, which supported a dose-response that was sublinear at low doses (Figure 5-2).

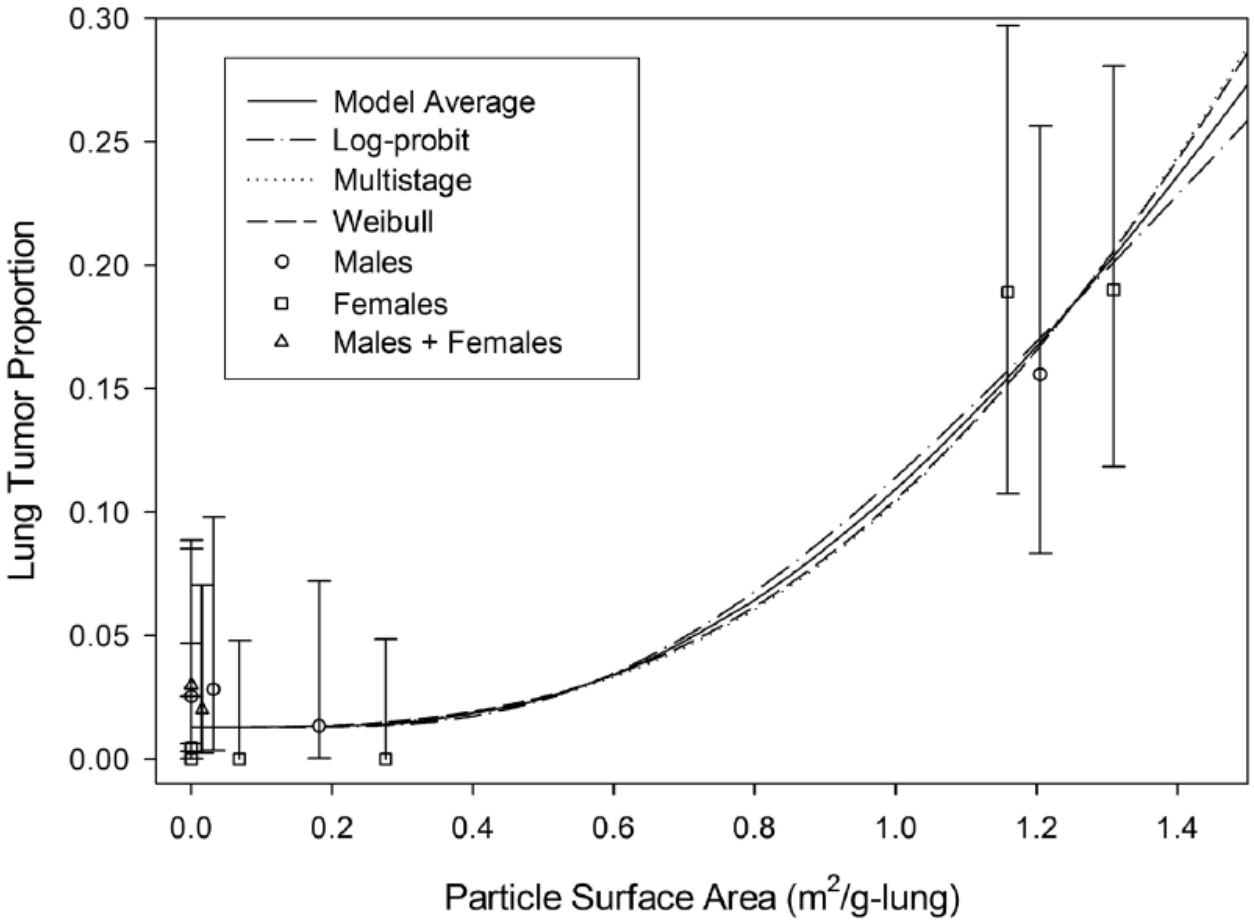

Figure 5-2. BMD models and three-model average fit to the lung tumor data (without squamous cell keratinizing cysts) in male and female rats chronically exposed to fine or ultrafine TiO_2 [NIOSH 2011]

5.4 Selecting a Dose-response Modeling Method

The estimated dose response curves from multiple biologically plausible models can differ substantially over a range of doses that can include the low dose region. Thus in addition to biologic plausibility, the strength of the data and the statistical methodologies must be assessed to inform the choice on approaches to estimating risks and quantify relevant uncertainties. Non-parametric, semi-parametric, and model averaging modeling techniques have been shown to be both robust and flexible; therefore, NIOSH generally prefers these methods to stand-alone models for assessing the dose-response. Nevertheless, every dose-response analysis is unique and requires careful consideration of the approach used. For example, possible exceptions to using non-parametric, semi-parametric, and model averaging techniques are: 1) compelling mechanistic or statistical evidence supporting a specific dose-response function or 2) data limitations require a simpler approach or a more parsimonious model.

5.5 Laboratory Animal Data

The adequacy of the database to support dose-response analysis based on animal studies is an important consideration in occupational risk assessment. Animal studies are evaluated in context of the risk assessment question under investigation. Studies are identified that may shed light on the research question. For example, a single dose, acute toxicity study may not have much relevance for assessing chronic exposures to a chemical, but may be useful for setting an immediately dangerous to life and health (IDLH) value or a short-term exposure limit (STEL), depending on the specifics of the study. Ideally, all studies that may contain relevant information should be considered for dose-response assessment. Each study is evaluated for adequacy of study design and conduct (duration of exposure, dosing regimen, species, numbers and sexes of animals, description of experimental conditions), health endpoints observed, statistical analyses conducted, and how the data supports the conclusions of the study. Preferably, the rationale for including/excluding studies, dose-groups, or health endpoints from analysis is clearly documented in the risk assessment.

For dose-response modeling of dichotomous response data, both the number of animals showing the effect in the group and the total number of subjects in the group are necessary at a minimum. For modeling of continuous response data, individual animal data are strongly recommended although information on central tendency and variability may be sufficient; typically, data on the number of subjects, mean of the response, and variability measure (e.g., standard deviation, standard error, or variance) for each group are adequate to perform the analysis. For dose-response modeling of categorical responses, individually recorded data are strongly recommended although the number individuals examined and the counts for each response category of each dose group are generally sufficient [EPA 2012b; Hertzberg 1989].

Once the relevant studies are gathered, each study is assessed using a WoE approach. Endpoints with a statistically and/or biologically significant dose-response association relevant to the risk assessment question are considered for assessment. Further endpoint selection may be based on factors such as the relevance of the endpoint to human health, severity of the health endpoint, and the sensitivity of the health endpoint.

When the dose-response analysis is conducted using benchmark dose analysis and/or model averaging, there are specific requirements for the data sets. In general, toxicology animal studies with more than one dose group are required for dose-response analysis. Ideally, there are responses in more than one dose group that are different from background and different from the maximal response. Multiple intermediate responses of this type increase confidence that the study contains adequate information on the dose-response curve and does not represent only background or only maximal responses. It may be possible to calculate a BMD and BMDL with only a single dose showing a response near the BMR [Kavlock et al. 1996]. However, if the studies show responses in more than one dose group, but all the responses are at the background level, near the maximal response level, or appear as a very steep rise of the dose response curve over a small range of doses, the data may not be adequate for regression modeling. Thus, it is preferable to have toxicological studies with observed responses sufficient to provide a unique solution to the optimizing procedure. For the dichotomous data models (eq. 5-1 thru 5-9) above this usually requires at least two dose groups with responses intermediate between

background and maximal. An additional advantage accrues from having at least one dose group near the BMR, to yield a better estimate of the BMD [EPA 2012b].

Overall, the specific type of toxicity information required is dependent upon the question to be addressed and the interplay with human data. In general, for occupational risk assessment, the availability of a well-conducted chronic bioassay data, preferably in more than one species (typically one in rats and one in mice); a two-generation reproductive study; and a developmental study in mammalian species would provide a reasonable database to reduce the uncertainty and increase the confidence in the risk estimates. A well-conducted subchronic study that evaluated a comprehensive array of endpoints could also be useful, especially in the absence of chronic bioassays. In most cases, NIOSH is concerned about chronic exposures to hazards, but in some cases, acute or intermediate-duration hazards may be of concern. In other cases, data needs are endpoint-specific. For example, if acute or subchronic data demonstrate neurotoxic, immunotoxic, or cardiotoxic effects, a neurotoxicity, immunotoxicity, or cardiotoxicity battery of tests could satisfy the data requirements [EPA 1994; EPA 2002]. When the typical animal bioassays are not available, data from alternative testing systems such as high throughput molecular toxicity assays and QSAR models could be used to inform the risk assessment and to fill the data gaps. NIOSH has not had extensive experience in using these types of data, so each use would be on a case-by-case basis.

5.5.1 Parallelogram Approach

First introduced by Sobels [1977], the "parallelogram approach" (Figure 5-3) is an argument by analogy for inferring missing data when you have closely related data, especially useful for cross-species extrapolation. It has been used in genotoxicity studies to predict human germ cell mutations from measured mouse germ cell mutations, mouse somatic cell mutations and human somatic cell mutations [Anderson et al. 1994]. It has been used in physiologically-based pharmacokinetic studies to predict human *in vivo* metabolic parameters from measured mouse *in vitro* parameters, mouse *in vivo* parameters and human *in vitro* parameters [Kienhuis et al. 2009]. NIOSH, in part, used this technique to assess comparative potency of closely related chemicals diacetyl

and 2,3-pentanedione, when data on human toxicity and animal toxicity data were available for diacetyl but only data on animal toxicity was available for 2,3-pentanedione [NIOSH 2016a].

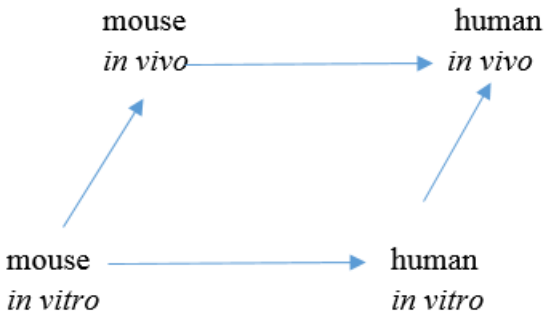

Figure 5-3. An example of the parallelogram approach.

The parallelogram approach is conceptually very simple but requires explicit assumptions. For example, if one would like to estimate the metabolic constants for a substance to use in a physiologically based pharmacokinetic model, it must be assumed that:

- There is a constant and knowable relationship between metabolic constants measured in vitro and metabolic constants measured in vivo within a species.
- The relationship between in vivo and in vitro metabolic constants is the same, regardless of species.

Therefore, once the ratio between mouse *in vitro* and *in vivo* metabolic constants has been measured and the human *in vitro* metabolic constant is known, the mouse ratio can be applied to the human *in vitro* constant to estimate the human *in vivo* metabolic constant.

Similarly, for genotoxicity studies, it must be assumed that:

- There is a constant and knowable ratio between somatic mutations and germ cell mutations within a species.
- The relationship between somatic mutations and germ cell mutations is the same regardless of species.

Therefore, once the ratio between mouse somatic mutations and mouse germ cell mutations has been measured and the human somatic mutations have been measured, the mouse ratio can be applied to the human somatic mutations to estimate the human germ cell mutations.

For the comparative potency example, NIOSH had mouse toxicity data and human epidemiology data on diacetyl. NIOSH was also interested in a closely related (1-carbon different) flavoring chemical, 2,3-pentanedione. However, there was no human data on 2,3-pentanedione toxicity. In this case, NIOSH assumed:

- There is a constant and knowable relationship between the lung toxicity in mice and the lung toxicity in humans for a chemical.
- Di-alpha-ketones such as diacetyl and 2,3-pentanedione are closely enough related that they share toxic modes of action.
- The relationship between lung toxicity and hazardous exposure in mice and humans is constant for these closely related chemicals.

Although NIOSH did not follow this logic through to predict human risk estimates for 2,3-pentanedione, the same logical structure applies. NIOSH stopped with an assessment that 2,3-pentanedione was in a similar range of toxicity as diacetyl and used the diacetyl risk assessment to set a recommended exposure limit for 2,3-pentanedione. In this case, the uncertainties in the method and the sparseness of the data argued for cautious application [NIOSH 2016a].

Using a parallelogram approach requires that the measured values used to construct the ratios reflect the same or very closely allied methods and data sources. This will not work if the technique or type of tissue used for

mouse *in vitro* assays is substantially different from the human assays. The assumptions regarding which values in the parallelogram are similar should be carefully examined. The uncertainty in the parallelogram approach is lessened with cross-species validation. For example, if the ratios of *in vitro* to *in vivo* metabolic parameters in mice, rats and hamsters are measured and they all have a similar ratio or a predictable relationship, it strengthens the argument that it is reasonable to extrapolate to humans. Depending on the data available, this step is not always possible. Finally, the parallelogram approach is a useful tool to consider when key data are unavailable, but it requires strong assumptions that must be closely examined and carefully justified.

5.6 Dose-Response Modeling with Epidemiologic Data

NIOSH prefers the direct estimation of occupational risks using epidemiologic data from studies of working populations, whenever data permit, for two main reasons: 1) data reflecting actual exposures and responses within the population of interest are intuitively superior for risk assessment; and 2) the uncertainty in extrapolating data from animal toxicologic studies to predicting human risks can be much larger than that in well-designed epidemiologic studies [Hertz-Picciotto et al. 1995; Smith 1988; Stayner et al. 1999]. Of the NIOSH risk assessments listed in Table 1-1, nine (70%) quantitatively examined the dose-response relationship by statistical models using epidemiologic data. In contrast, epidemiologic data have been used in less than 10% of Integrated Risk Information System (IRIS) assessments conducted by the EPA [Persad and Cooper 2008].

Many of the concepts discussed previously concerning animal data are also applicable to human data, especially for experimental designs or when modeling binary outcome data from observational studies without time-dependent variables (e.g., using logistic regression). Although methods of analyses may be identical, one must acknowledge that the majority of human data for risk assessment stems from observational studies by necessity, which have less control of extraneous factors, and thus are more prone to error compared to experimental data (see Section 4.3.4). The design of epidemiologic studies contributing to risk assessment can vary between studies, as can study aims, which also may not fully align with risk assessment goals. Thus, dose-

response information from human studies is often considered less precise than that obtained from experimental studies.

NIOSH risk assessors first decide, in a systematic way, if human data are suitable for quantitative dose-response analyses, and if suitable, whether the data will serve as: 1) the primary basis for risk extrapolation or 2) supporting information for a toxicologic-based risk assessment. The evaluation may be made concurrently with the WoE assessment in hazard identification, although data supporting hazard identification may lack the rigor necessary for dose-response analyses. In any event, it is desired that all decisions on data suitability be fully described in the risk assessment documentation. As a starting point, risk assessors have applied the framework first described by Hertz-Picciotto [1995], who suggested judging the suitability of epidemiologic data for quantifying dose-response using five criteria. These criteria, slightly modified for NIOSH risk assessment purposes, are as follows:

1. The data consistently indicate a stable positive statistical association between the agent and adverse effect.
2. The data are abstracted from studies that are of high overall quality.
3. There is no substantial potential for confounding or other source of major bias.
4. There is a quantitative assessment of exposure that is deemed sufficient for dose-response analyses.
5. There is evidence of a monotonic dose-response.

Hertz-Picciotto [1995] suggested that compliance with Criteria 1-4 provides a minimum basis for risk extrapolation using human data. Compliance with two of Criteria 1-3 is considered suitable for quantifying risks as a plausibility check with toxicologic based assessments. Thus, it is clear that more weight is to be placed on Criteria 1-3. Criterion 1 is directly related to Hill's guidelines on strength of association and consistency (see Section 4.1). This criterion differs from that originally specified, which included only *moderate to strong* positive associations [Hertz-Picciotto et al. 1995]. This modification was made in recognition that excellent studies

reporting weakly positive but consistent associations can inform on the dose-response. For example, Park et al. [2004] conducted a quantitative risk assessment of lung cancer from exposure to hexavalent chromium that used data from Gibbs et al. [2000], who reported modestly elevated lung cancer risk (e.g., SMR <2) in chromium production workers compared to the general population (SMR=1.80; 95% CI: 1.49-2.14). This risk assessment helped form the basis for the NIOSH REL on hexavalent chromium exposure [NIOSH 2013a]. Criteria 2 and 3 are related to one another; both preferring study designs that reduce the potential for an inaccurate estimated effect. Criterion 4 was modified to recognize that quantitative exposure data *at the individual level*, as originally recommended [Hertz-Picciotto et al. 1995], is sparse in epidemiologic studies. The lack of individual exposure data should not disqualify study data from quantitative risk assessment; however, its presence is clearly preferred to aggregate exposure measures. Criterion 5 coincides with Hill's guideline on a biologic gradient, which is not necessary in either case but certainly supports data use. It is important to note that there are many explanations for a lack of observed monotonicity in dose-response data, such as measurement error, biologic saturation, and depletion of a susceptible population. Lastly, there may be exceptional circumstances in which other criteria may better apply or in which modification to existing criteria is prudent. These exceptions are preferably described in the risk assessment document. For example, a potential for substantive confounding may exist (Criterion 3); however, data may allow for an examination and/or adjustment of its effect on dose-response estimates. In this example, the risk assessor should fully describe the potential for bias, the alternative analyses for examining the effects on dose-response estimates, and any consequent actions in the risk assessment document.

NIOSH risk assessors strive to make the best use of epidemiologic data that are available in dose-response modeling strategies, given that these data provide the important advantage of directly assessing human risk. When epidemiologic data are available and appear suitable for quantifying exposure-related effects, the risk assessor generally adopts a statistical modeling approach that includes an evaluation of potential sources of biases that may exist. NIOSH risk assessors endeavor to select statistical methods that best account for identified sources of uncertainty and therefore improve the reliability and validity of dose-response estimates. As discussed previously,

exposure estimation in human studies is fraught with limitations. Risk assessors consider ways to account for exposure uncertainty in developing a risk modeling approach. Given the relative uniqueness of most epidemiologic datasets, it is not feasible to describe all possible modeling strategies in this report. However, there are some overall modeling approaches using human data that can be discussed, and this section provides some information in that regard. In particular, methods that are unique to aggregate data from published reports and time-to-event data from longitudinal studies are presented.

5.6.1 Limited Data

Although dose-response analyses using individual exposure and outcome data are preferred, the lack of these data does not preclude examining the dose-risk relationship using summary estimates from human data. In fact, risk assessments have used limited data comprising only an aggregate exposure measure (e.g., average cumulative exposure) and a measure of relative risk (e.g., SMR, SIR, and OR). For example, a simple dose-response model can be specified using data from a study reporting only a cohort SMR and average exposure by assuming a linear relationship exists between the SMR (or any measure of relative risk) and exposure (x): $SMR=1+x\beta$, where the dose response slope, β, represents the change in relative risk (e.g., the SMR) per unit dose [Smith 1988]. This relationship is plotted for a hypothetical cohort in Figure 5-4.

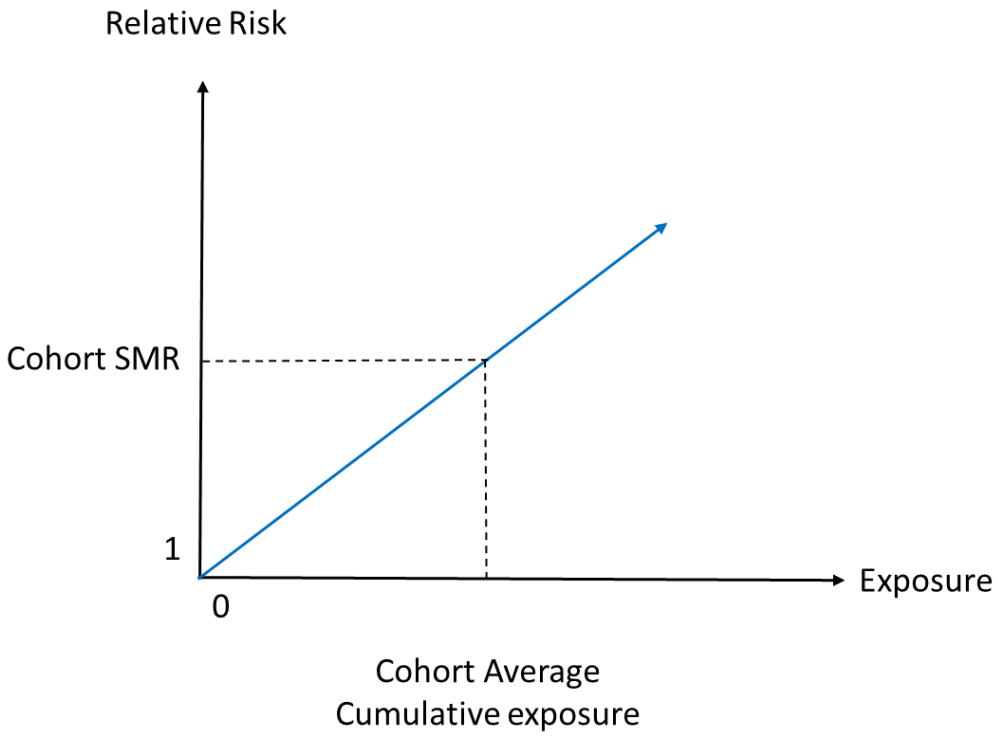

Figure 5-4. Linear dose response slope estimates using average cumulative exposure and reported standardized mortality ratio (SMR) from an epidemiologic study (adapted from Smith et al. [1988]).

An SMR from an occupational study may be negatively biased from a healthy worker hire effect (See Appendix B for more information). This effect can be countered using an adjusted SMR, which is derived based on the study type and outcome [Park et al. 1991]. Similarly, information on other potential sources of bias can be included as model covariates. As another example, if SMRs are reported at different levels of exposure, then weighted least-squares regression or maximum likelihood estimation methods can be used. Examples of these techniques have been described in several reports [Breslow and Day 1987; Crump and Allen 1985; Hanley and Liddell 1985; Rothman et al. 2008; Smith 1988; Smith et al. 1994; Steenland and Savitz 1997].

There are some important limitations in the methods described above. First, we assume the true dose-response is linear given that: it is biologically plausible, generally appears conservative in the low dose range

compared to alternative models, and tends to fit epidemiologic data. Of course, other dose-response functions could be used, which may result in marked differences in estimates in the range of interest. In practice; however, sufficient data to support an alternative assumption are unlikely. Second, pooling SMRs in dose-response analysis is less than ideal given that multiple SMRs (stratum-specific) may not be comparable due to indirect standardization. Pooling SMRs could bias estimates due to differences in age, race, sex or some other confounder across exposure groups. Fortunately, strong statistical confounding from stratum heterogeneity is not typically observed in most cases [Breslow and Day 1987]. Nonetheless, risk assessors must address the potential for bias from heterogeneous comparison groups. Third, using weighted least squares to regress multiple responses at different exposure levels (e.g., SMRs or log-SMRs) does not account for correlations between response measures induced by sharing a common reference group. Methods have been developed to account for these correlations in trend estimation in both single study and meta-analytic (meta-regression) designs [Greenland and Longnecker 1992; Hamling et al. 2008; Orsini et al. 2012].

A NIOSH example of a limited data approach is not available; however, examples are available in the literature [Chovil et al. 1981; Crump and Allen 1985; Hanley and Liddell 1985; Smith 1988; Steenland and Savitz 1997]. For example, Steenland and Savitz [1997] used a simple linear model to examine the dose response between airborne nickel levels and lung cancer mortality. The dataset was abstracted from a previous epidemiologic study of Ontario nickel refinery workers (n=495) followed from 1963 to 1978 [Chovil et al. 1981]. The relative risk per unit exposure was estimated by the slope parameter from a weighted least squares linear regression of the SMRs at specified cumulative dose levels. The expected numbers of lung cancer deaths were used as weights and the model forced the intercept at unity. A simple estimate of lifetime excess cancer was estimated by $R_x = R_0(x\beta)$, where R_x is the added lifetime risk from exposure x, R_0 is the background lifetime risk of lung cancer death, and β is the upper 95% confidence limit on the slope parameter.

5.6.2 Longitudinal Data

In longitudinal studies, data on observation time, demographics (e.g., age, race, and gender), and time-varying predictors (e.g., exposures) are available. Approaches to modeling must be consistent with the data although more than one approach may be available. For example, data from a cohort study of cause-specific mortality can be expressed as the amount of observation time and an observed count of adverse responses cross-classified based on the other predictors in order to estimate the incidence rate.

In general, previous risk assessments have applied a tiered approach, whereby categorical analyses and splines are first used to evaluate the shape of the dose-response curve, which aids in defining a set of parametric models that are most appropriate for risk assessment [Steenland and Deddens 2004]. The risk assessor may then select a preferred model from the set of models based on prior knowledge of the expected response (biologic plausibility) and model fit. In any event, the choice of the 'best' model should not rest solely on statistical grounds [Breslow 1990]. This is because competing statistical models can often yield roughly equivalent fits to the data in the observable effect dose range, yet extrapolation below the observable range (i.e., in the range of interest) can result in estimates that are orders of magnitude apart [Brown and Koziol 1983]. Methods within the framework of this approach can vary and an exhaustive discussion of all modeling possibilities is beyond the scope of this report. More information is available in many important epidemiologic texts [Breslow and Day 1987; Rothman et al. 2008; Woodward 2013].

Dose-response modeling of longitudinal data has been approached using a wide array of methods, but is generally conducted by regression of survival data (i.e., failure-time data) or person-years data. Survival regression models can be fully parametric models of the distribution of failure times (e.g., Weibull models) or semi-parametric (i.e., Cox proportional hazards model). Poisson regression modeling is an example of a modeling approach for data on response counts and person-years of observation. Most epidemiologic studies have examined dose-response relationships from longitudinal study data using general relative risk models with maximum

likelihood estimates obtained from Cox proportional hazards or Poisson regression techniques for cohort data and conditional logistic regression for nested case-control designs.

5.6.2.1 Poisson Regression

Followup data can be recorded as counts of responses, i.e., the number of events (e.g., deaths) and the number of person-years in strata of other variables (e.g., categories of age). Furthermore, dose-response curves can be fitted to the count data Y_i and person-years τ_i based on Poisson regression modeling, i.e., $Y_i \sim Poisson[f(d_i, X_{1i}, X_{2i}, \cdots, X_{ci}; \boldsymbol{\theta}) \cdot \tau_i]$ where Y_i represents the count observed during an accumulation of person-time τ_i; if each record of the data is constructed from one person then $Y_i \sim Binomial[f(d_i, X_{1i}, X_{2i}, \cdots, X_{ci}; \boldsymbol{\theta}) \cdot \tau_i | n_i = 1]$ may be substituted. If a reference population is available that provides information on the rate associated with age, sex and other demographic variables then it can be incorporated to improve estimate precision. However, the assumption of Poisson or binomial variations is a strong one and it may be necessary to accommodate response variation that exceeds those predicted by the model (i.e., over-dispersion), such as using quasi-likelihood methods or models that incorporate random effects. Another approach to the estimation of the dose response relationship is to model the distribution of age or time of the response using methods appropriate for failure-time data that are incomplete for those individuals who were alive when follow-up was ended or when their times to the cause-specific response event were censored by competing risks; usually, the hazard function of the distribution is the focus of the model [Moeschberger et al. 2007]. Although applications are generally amenable to either approach, i.e., Poisson regression and failure-time modeling, the latter may be advantageous for modeling varying susceptibility or "frailty" over the individuals by a continuous distribution of random effects. However, if the cohort is the union of a sensitive subpopulation who has a homogeneously higher susceptibility and its complement who has a homogeneously lower susceptibility then a finite mixture model where the two subpopulations are mixed together may be available to analyze either form of response data, i.e., counts or failure-times. Furthermore, these frailty models can account for an attenuation of the dose-response curve initially observed using a $\sqrt{dose}$ model and should avoid basing estimation

of risk quantities on a dose-response curve whose slope becomes unbounded as dose approaches zero as would be expected under a simple $\sqrt{dose}$ model of the dose effect.

Examples of Poisson regression modeling in NIOSH risk assessment include dose-response models of the relationships between lung cancer and hexavalent chromium [NIOSH 2013a; Park et al. 2004] and asbestos [Stayner et al. 1997].

5.6.2.2 Cox Proportional Hazards Regression

In survival analyses, the hazard function (or hazard) is the rate of failure at an instant in time, *t*, given that the individual survives up to *t*. In other words, it is the instantaneous risk that the event (e.g., death, cancer diagnoses) will occur at *t*. In most longitudinal studies, the time scale of interest is age. The hazard ratio (HR), is the hazard of one individual (e.g., the exposed) divided by another individual (e.g., the unexposed), typically holding all other predictors constant, thus it is a measure of the relative risk. Since its introduction in 1972, the Cox Proportional Hazards (PH) regression model has become the most widely used approach to quantifying conditional hazards [Cox 1972]. A general form of the PH model for the hazard, *h*, cumulative dose, *D*, and attained age, *t*, is:

$$h\big(t\big|D(t),Z(t)\big) = h_0(t)\, f(D(t);\, \beta) \exp\big[\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{Z}\big],$$

where: h_0 is the baseline hazard, $\boldsymbol{Z}$ represents a vector of model covariates, model parameters β and $\boldsymbol{\gamma}$ are to be estimated, and $f(D(t);\, \beta)$ is the relative rate as a function of cumulative dose at attained age. This model is semi-parametric because the baseline hazard is an unspecified function, but a parametric form is assumed for the effect of predictors on the hazard. Several options for specifying a dose rate function are available; the most common is an exponential form, i.e.,$(D(t);\, \beta) = \exp(\beta D(t))$, which is sometimes referred to as a loglinear dose-response model. In this form, the PH model is a simple additive model of the log of the hazard. Another common form is a linear response function $f(D(t);\, \beta)=1+\beta D(t)$, thus β is the excess relative rate per unit dose in the exposed individual relative to the unexposed. Validity of this model relies on a rather strong assumption that the hazards in

the group of interest are proportional to the hazards in the referent group, and this proportionality is constant over time when $D(t)$ is constant. A significant interaction between $D(t)$ and t would be evidence against such proportional hazards. Additional statistical methods (e.g., stratification, fully parametric or piecewise proportional models) may be necessary in the event of strong modification of the dose effect on the hazard over time [Allison 2010].

Examples of Cox PH regression methods in NIOSH risk assessment include dose-response modeling of the relationships between lung cancer and exposures to radon [Hornung and Meinhardt 1987] and cadmium [Stayner et al. 1992a; Stayner et al. 1992b].

5.6.2.3 Conditional Logistic Regression

Case-control designs are typically analyzed with logistic regression, as previously described. The fitting of matched or stratified logistic regression models is sometimes referred to as conditional logistic regression [Breslow and Day 1980; Rothman et al. 2008]. When time-dependent predictors are present, case-control studies often rely on conditional logistic regression. For matched case-control studies with one case per matched set (i.e., 1: *n* matching), the form of the likelihood function for conditional logistic regression reduces to that of the Cox PH model for the continuous time scale. In both cases, the data are organized into risk sets (sometimes referred to as a matched set in conditional logistic regression), whereby a risk set is the collection of individuals at risk for the event at each time point in which a failure is observed. For example, a nested case control study may specify a conditional logistic regression model that uses controls drawn from risk sets of individuals matched to cases on attained age. In this instance, the controls are selected using incidence density sampling methods [Beaumont et al. 1989]. Computational limitations may restrict full risk set analyses (i.e., Cox PH model); therefore, a nested case-control study using conditional logistic regression is an appealing alternative to full cohort modeling.

5.6.2.4 Additional Considerations

In epidemiological studies, d_i, is an exposure metric constructed from possibly complex employment histories, and identifying the optimum construct for d_i may, itself, be an important component of the modeling procedure (see Section 4.3.3.1). Moreover, predictors in the function $f(d_i, X_{1i}, X_{2i}, \cdots, X_{ci}; \boldsymbol{\theta})$ usually include age, gender and other demographic variables and may confound or modify the effect of dose. Thus, the risk assessor must consider effects on estimates from the selected model or set of models that are due to the exposure metric construct and other predictors (See Appendix B for discussion on possible study biases).

Attenuation of the dose-risk relationship at higher doses is a common observation in occupational epidemiologic studies [Stayner et al. 2003; Steenland et al. 2015]. This effect typically presents as a monotonically increasing slope at low exposure levels that diminishes or becomes negative at high exposure levels. Among possible explanations are: a depletion of the susceptible population, healthy worker survivorship, a natural limit on the relative risk for diseases with a high background rate, errors in measurement of the exposure that are proportional to the exposure level, influence of unknown risk factors that may vary by the level of exposure, adaptive responses, and biologic saturation. Regardless of cause, the risk assessor must consider the possible effects of high-dose attenuation when estimating responses at low doses, given that a 'best-fit' model may actually be a poorer choice for risk assessment. For example, a linear excess relative risk model that is best fit to the full range of exposures may underestimate the low-dose response as a result of risk attenuation (artifactual or otherwise) at high doses. Conversely, using a logarithmic transformation of exposure (i.e., a power model) may improve the model fit; however, this model is prone to overestimation of the response at low doses [Ginevan and Watkins 2010; Steenland and Deddens 2004; Steenland et al. 2011]. The potential for high-dose attenuation can be explored using categorical models, transformations of the exposure metric such as square-root or logarithmic and the use of splines; however, the response in the low dose-region of the dose-response curve can widely vary between these approaches [Steenland and Deddens 2004; Steenland et al. 2011]. When selecting preferred models for risk assessment, the risk assessor must evaluate the low-dose behavior of the models with

respect to the potential effects of attenuation. Modification or replacement of the best-fit model may be required in order to avoid unrealistic estimates of effects in the range of dose that is most meaningful to the protection of workers. For example, simple piecewise linear models that allow for different slopes between high and low-dose regions (e.g., a two-piece linear spline) may be appealing, given they account for high-dose attenuation but allow for interpretation of risk at low dose that is suitable for risk assessment purposes [Steenland and Deddens 2004; Steenland et al. 2011].

Dose-rate effects are of interest because they can have a substantial impact on low-dose extrapolation common in risk assessment. A dose-rate effect occurs when the dose-response relationship is best described by exposure intensity (e.g., airborne chemical concentration) rather than accrued dose (e.g., the product of concentration and exposure duration). A positive dose rate effect (i.e., higher exposure intensities having a greater effect) suggest that transient or peak exposures may have an important role in disease induction [Checkoway and Rice 1992; Esmen 1984; Rappaport 1991]. For example, peak exposure is obviously most important for evaluating acute toxicity. Peak exposure can also be the primary index when the agent is rapidly eliminated from the body or when nonlinear rates of biologic damage occur during periods of intense exposure [Esmen 1984; Rappaport 1991]. Nonlinearity may result from exposure-related responses that are reversible and/or have a threshold for the onset of biologic damage. Dose-rate effects that act to attenuate the response at higher exposure intensities are sometimes referred to as protraction enhancement or inverse dose rate effects. These effects were evident in studies of the dose-response association between lung cancer and radon exposures in underground uranium miners [Lubin et al. 1995]. The actual mechanisms involved in the radon inverse dose rate effect are unknown; however, plausible explanation are: nonlinear cellular responses, such as a bystander effect (i.e., a dose effect observed in non-irradiated cells) at very low dose rates [Brenner et al. 2001], or physical differences in the particle size distribution of radon progeny at different airborne concentrations [Leonard 2007]. Addressing dose-rate effects are challenging in dose-response modeling. Interpretation will be largely dependent on the dose index used; however, given the complex and largely unknown biology associated with these nonlinear effects,

mechanistic data are likely insufficient to inform on a modeling strategy. Nevertheless, complete understanding of the underlying cause and effect relationship may not be necessary if a dose-response relationship between the chosen exposure metric and the adverse effect can be quantified.

5.7 Sensitivity Analysis

The choice of modeling approach can markedly influence risk estimates. Moreover, limitations in available data often require scientific judgment in order to fill gaps in model specifications. NIOSH risk assessors generally conduct additional analyses to test plausible alternative hypotheses, examine the robustness of main analyses, and improve transparency in the risk assessment process. These alternative analyses are also known as sensitivity analyses. Sensitivity analysis is defined as a study of the uncertainty in estimates from the mathematical model that can be apportioned to uncertainties in its inputs. In other words, it is a study of the robustness of the modeling results in the presence of uncertainty. In a sensitivity analysis, plausible alternative risk assessment strategies, defaults, and assumptions are quantitatively evaluated for their impact on risk estimates. As stated in *Science and Decisions* [NRC 2009], ". . . [S]ensitivity analysis could be performed when risk estimates for alternative hypotheses that are sufficiently supported by evidence are reported . . . The goal is not to present the multitude of possible risk estimates exhaustively but to present a small number of exemplar, plausible cases to provide the risk manager a context for understanding additional uncertainty contributed by considering assumptions other than the default." This means a targeted, hypothesis-driven strategy for conducting sensitivity analysis is preferred.

In large part, sensitivity analyses are examinations of risk estimates over a range of plausible values for uncertain data that are used in the risk assessment. Largely divergent estimates (or large uncertainties) suggest a high degree of model dependence while reasonable agreement in findings suggests estimate robustness. Sensitivity analysis is also useful for identifying factors that have the most influence on worker risks, which could then be targeted as priorities in risk management. Finally, sensitivity analyses can be a useful tool for model development and refinement [Frey and Patil 2002].

Whether sensitivity analysis is conducted as part of the main analysis or after the main analysis is completed, it should be part of the initial risk assessment plan. Analysis planning generally includes a description of any iterative methods intended for model development and refinement. A sensitivity analysis can be structured into the main analysis so that a variety of risk estimates is produced and the decision path to the final risk estimate is well supported and transparent. Risk assessors are cautioned against *post hoc* analyses as a substitution for planned sensitivity analyses. Nevertheless, there are instances when these analyses are appropriate or even expected. For example, subsequent analyses may occur in response to review comments.

It is not practical to list all possible sensitivity analyses that may be found in risk assessments; however, Table 5-2 lists some examples of sensitivity analyses that appear most often in the literature. Some areas of sensitivity analyses are discussed in subsequent sections.

Table 5-2. Example scenarios for sensitivity analyses

Source of Uncertainty	Research question	Possible sensitivity analysis
Response variable	Are there alternative definitions of the adverse effect? If so, how do these definitions affect dose-response estimates?	Alternative models using different specifications of the response variable.
	Is more than one adverse effect (not on the same causal pathway) associated with the hazardous exposure? If so, how do risks differ?	Alternative models using array of plausible responses.
Explanatory variables	How does measurement error in the primary exposure affect risk estimates?	Alternative models using array of plausible estimates of exposure based on uncertainty.
	Are there alternative exposure metrics? If so, how do risk estimates differ across metrics?	Alternative multiple models using array of exposure definitions.
	If exposure is categorical, how does the choice of category cutpoints affect risk estimates	Alternative models with varying exposure cutpoints.
	Is there a potential for unmeasured confounding (e.g., smoking data unavailable in analysis of cancer) by one or more sources? Can these effects be estimated?	Alternative models using array of plausible estimates of the confounder.
Model specification	How does model choice of dose-response function affect risk estimates?	Alternative models using array of plausible dose-response functions

	How does the choice of confounding control (e.g., stratification versus covariate control) affect	Alternative models using array of methods for confounding control.
Animal to human risk transport	How do assumptions on the animal-human relationship for metabolism, distribution, and toxicity affect risk estimates?	Alternative models using array of plausible assumptions.

5.7.1 Choice of Adverse Effect

There may be more than one adverse effect available for dose-response analysis. Decisions about which adverse effect to analyze rely on consideration of the site of the effect and its relevance to the human toxicity of concern, the severity of effect, reversibility of effect, mode of action, the sensitivity of the test species (or human subpopulation), and consistency of effects across sex/species (or population groups). Sensitivity analyses should include plausible alternative adverse effects. In epidemiological analyses, different adverse effects could include different measures of lung function (for example, FEV1, FEV1/FVC), self-report of symptoms, and/or diagnosed respiratory effects. For cancer studies, a variety of tumor sites could be analyzed. In animal toxicology studies, adverse effects could include analysis of both cancer and non-cancer effects, a selection of tumor sites, and more. The rationale for selecting the adverse effects in the main analysis and the sensitivity analysis should be thoroughly explained.

5.7.2 Sensitivity Analyses in Dose Response Modeling

A standard practice in NIOSH dose-response modeling is to first specify models of interest *a priori* and then test the specification by examining alternative models. Alternative models should be plausible and parsimonious. It is preferred that the sensitivity analysis approach used, including the suite of alternative models to be examined, be specified *a priori*; however, model output information has been used in *post hoc* specification of alternative models in some analyses. This is likely to occur when results from main models point to the need for further development or if new information is found during analysis or in review afterward.

The validity of modeling defaults should be examined in a sensitivity analysis when additional chemical-specific information is available that challenges those values. In addition, when there are alternative plausible assumptions on explanatory variables used in an analysis, it is reasonable to explore the impact of these assumptions in sensitivity analysis. Typical examples of key assumptions in epidemiologic studies include exposure lag times, homogenous dose-response among grouped outcomes (e.g., all-cancers) and irreversible effects of chronic exposure, especially for non-cancer adverse effects. When examining alternative assumptions or default vales, it is important to use credible values that reflect the available data.

In some cases, the potential effects of measurement error or unmeasured confounding can be examined by sensitivity analysis [Chu et al. 2006; Greenland 1996; Groenwold et al. 2010]. For example, consider a study reporting a positive dose-response association between lung cancer and exposure to chemical X. Smoking data are unavailable. One could assume a range of plausible smoking behaviors (and their effects) that vary by degree of correlation with chemical X as a means to estimate the potential for residual confounding by smoking in main analyses. If a significant effect is not observed under plausible scenarios, then it is unlikely that smoking patterns explain the dose-response observed. The complexity of these sensitivity analyses can vary widely from a simple examination of a single binary variable to complex computer simulations for examining joint effects of multiple factors. Some examples of sensitivity analyses over this range are readily available in a number of highly cited articles [Frey and Patil 2002; Greenland 1996; Greenland et al. 2005; Lash and Fink 2003; Lin et al. 1998]. Regardless of the analyses design, it should be evident that the reasonableness of these analyses hinges on the range and values examined; therefore, risk assessors must carefully consider the choice of plausible values.

In all modeling efforts, including sensitivity analyses of alternative models, NIOSH risk assessors must clearly describe the approach used in sufficient detail such that results can be replicated. Special attention should be given to providing a sound basis for any *post hoc* analyses, if conducted. Risk assessors should be aware that model specifications made using *post hoc* information, say from a stepwise regression approach, can introduce bias from a lack of accounting for the informed choices made [Harrell 2015]. For example, consider a dose-

response model that includes an assumption of an exposure lag period to account for disease latency. If the choice is made *a priori*, then the uncertainty in model parameter estimates attributable to the estimated lag period is taken into account. However, if the exposure lag period used in the main model is selected from a suite of models that vary the exposure lag to find the best fitting model, then the data-based selection of the exposure lag should be compensated for in the estimation process [Richardson et al. 2011]. Without this adjustment, modeling uncertainty is likely underestimated, which could bias risk estimates. As an alternative, the sensitivity analysis comprising the models using alternative lags can inform on the appropriateness of the exposure lag selected assumed *a priori* in the main analysis.

5.7.3 Extrapolation Methods

Model extrapolation occurs when inferences are made beyond the calibration or validation of the model [Frey and Patil 2002]. This can occur when model inputs used to predict risk are beyond the dataset used to develop the model. For example, the application of toxicologic data to assess human risk is a common extrapolation. When extrapolating risk from animals to humans, depending on the metabolism, distribution and toxicity of the chemical, there may be a choice of extrapolation methods. In this case, it is reasonable to explore the impact of plausible extrapolation methods in the sensitivity analysis. An example of this can be found in the diacetyl risk assessment, in which the $BW^{3/4}$ extrapolation method was compared to the EPA RDGR method for reactive gases/vapors [NIOSH 2016a].

6.0 DOSIMETRY ADJUSTMENTS FOR HUMAN EQUIVALENT CONCENTRATIONS

6.1 Particle Exposure

6.1.1 Overview

Understanding the relationship between the exposure to a substance, internal dose, and the biological response provides the information needed to develop occupational risk assessments. Dosimetry models are used by NIOSH to estimate the internal dose of a hazardous substance given exposure, with focus on the target tissue for an adverse effect [EPA 1994; Kuempel et al. 2015].

A critical dose in an animal study (i.e., a dose associated with an adverse effect) is extrapolated to humans using dosimetry modeling to estimate the human-equivalent dose. Such models account for interspecies differences in the factors that determine the deposition, clearance, retention, or clearance of particles (spherical or nonspherical) from the respiratory tract. Dose estimation is one of the major sources of uncertainty in a risk assessment (e.g., as discussed for carbon nanotubes in NIOSH [2013b], Section A.6.3). Use of validated dosimetry models reduces the uncertainty in extrapolating animal data to humans.

To estimate a human-equivalent internal dose or exposure concentration of particles using animal data, the main dosimetry method options include: 1) application of uncertainty factors (see UF section); 2) general/categorical adjustments (e.g., EPA "Regional Deposited Dose Ratio" in respiratory tract); and 3) substance-specific PBPK models (e.g., to account for particle dissolution and translocation beyond the respiratory tract). Application of UFs is simpler and requires less data, but is also associated with greater uncertainty. Other methods generally require more detailed data and rigorous analysis, but may provide more accurate dose estimates for the risk assessment. Information needed for dosimetry and risk assessment of inhaled particles is shown in Table 6-1.

.

1

Table 6-1. Basic parameters needed for dose estimation and risk assessment of inhaled particles [Kuempel et al. 2012].

Parameter or Issue	Purpose
Particle size, shape, density	Estimate inhalation and lung region-specific deposition fraction
Physicochemical properties, including particle surface area, reactivity, solubility	Evaluate mode of action and effects
Multiple exposure or dose groups	Describe dose-response relationship, and estimate benchmark dose
Biological significance of response	Evaluate severity and relevance to humans
Body and lung weights; target lung region	Normalize dose from animals to humans

6.1.2 Respiratory Tract Regions

The respiratory tract in both humans and laboratory experimental animals are divided into three regions based on their structure, size, and function: the extrathoracic region (ET) extends from nose to larynx, the tracheobronchial region (TB) extends from trachea to the terminal bronchioles and the pulmonary region (PU) that includes the respiratory bronchioles, alveolar sacs, alveolar ducts and alveoli (Figure 6-1). The pulmonary region is where gas-exchange occurs (i.e., uptake of oxygen and release of carbon dioxide. Diseases of the respiratory tract have been associated with exposure to substances that deposit in each of these regions. These regions also correspond to the inhalable, thoracic, and respirable particle size fractions for airborne sampling [ACGIH 2015].

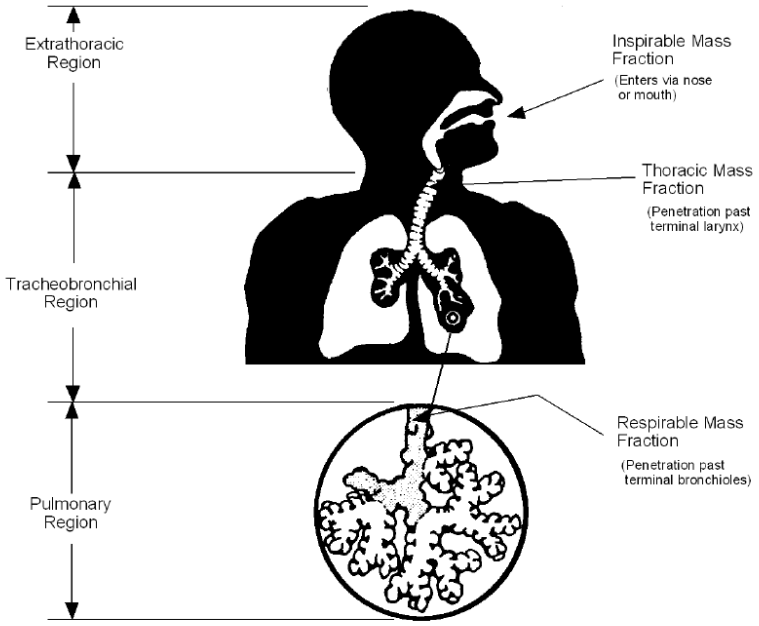

Figure 6-1. Human respiratory tract regions [EPA 1994].

6.1.3 Deposition Mechanisms

Particle size is a key factor in estimating the deposited doses in the respiratory tract region. Standard definitions of airborne particle size fractions include inhalable, thoracic, and respirable [ACGIH 2015]. "Inhalable" particles are those capable of entering the nose or mouth and depositing anywhere in the respiratory tract. For example, particles with aerodynamic diameter of 100 µm have an approximately 50% probability of being inhaled and deposited in the respiratory tract. The extrathoracic fraction is the mass fraction of inhaled particles with low probability of penetrating beyond the larynx. The thoracic fraction refers to particles capable of reaching beyond the larynx into the thoracic region and depositing in the lung airways. The respirable fraction is the mass fraction of inhaled particles that is capable of reaching and depositing in the gas exchange region of the lungs [Brown et al. 2013].

Aerodynamic equivalent diameter is defined as the diameter of a standard-density of one gram per cubic centimeter (1 g/cm^3) sphere having the same terminal velocity when settling under gravity as the particle under

consideration [Hinds 1999]. Diffusion equivalent diameter is defined as diameter of a sphere with the same thermal or Brownian diffusivity as the particle under consideration [Hinds 1999]. For nonspherical particles such as fibers, shape and orientation are additional factors that can influence deposition [Sturm and Hofmann 2009]. Particle dosimetry models generally use particle mass as the dose metric.

Respiratory tract deposition models can take into account these particle properties to predict the deposited dose in each region. In addition to the particle properties, the lung morphology can influence particle deposition. Differences in airway structure, lung volume, and breathing patterns (e.g., nasal only or oronasal) have been observed among individuals and are also related to age, gender, and race [Schulz et al. 2000]. Some deposition models account for inter-individual variability in lung morphology [ARA 2009; ICRP 1994]. Activity level (e.g., resting or exercising) influences the ventilation rate and thereby particle deposition in the respiratory tract.

At a minimum, data are generally available to estimate the deposited dose of particles in a respiratory tract region of humans or animals, given the exposure concentration, duration, and airborne particle size estimates. Examples of these basic methods and information sources are discussed below.

6.1.4 Ventilation Rates and Activity Levels

6.1.4.1 Humans

NIOSH generally uses the ICRP standard reference value in workers for the total air intake (volume inhaled), which is 9.6 m^3 in an 8-hour workday [ICRP 1994]. This total air intake is equivalent to an average minute ventilation rate of 20 liters of air per minute (L/min) [i.e., 9.6 m^3 = 20 L/min x 480 min x 0.001 m^3/L]. These reference values are based on adult males, assuming 5.5 hours of light exercise and 2.5 hours of rest/sitting. The adult male minute ventilation rates are 25 L/min for light exercise and 9 L/min for resting (sitting).

Thus, the total air intake in an 8-hour workday in men is calculated as follows:

$$9.6\ \mathrm{m^3} = [(5.5\ \mathrm{hours} \times 60\ \mathrm{minutes\ per\ hour}) \times 25\ \mathrm{L/min}] + [(2.5\ \mathrm{hours\ x\ 60\ minutes\ per\ hour}) \times 9\ \mathrm{L/min}] \div [1{,}000\ \mathrm{L/m^3}]$$

(eqn. 6-1)

Minute ventilation (V_E, in L/min) is calculated as the product of the tidal volume (V_T, in L) and the breathing frequency (*f*, in min^{-1}):

$$V_E = V_T \times f$$

(eqn. 6-2)

These respiratory values (tidal volume and breathing frequency) vary by age, gender, and activity level [ICRP 1994]. For example, V_E of 25 L/min (as used in eqn. 6-1) is calculated from V_T of 1.25 and *f* of 20 (as shown in Table 8 of ICRP [1994]). For adult female workers, the average air intake is 8.2 m^3 in an 8-hour workday, assuming the same activity levels and using the gender-specific values for V_T and *f* in ICRP [1994].

For dosimetry modeling, these respiratory values are used to estimate deposited dose given the exposure scenario. In the MPPD model [ARA 2015], the breathing frequency and tidal volume are required input values. Tidal volume is the volume of air inspired or expired in each respiratory cycle [EPA 1994]. The default values (resting) are V_T of 625 milliliters (ml) and *f* of 12 [ARA 2015], which correspond to adult male values reported in ICRP [1994]. For workers, NIOSH [2011] used the values of 1,143 ml for V_T and 17.5 for *f*, which are weighted averages of the respiratory values that correspond to the average male worker reference values of 20 L/min (V_E) and 9.6 m^3 (total volume inhaled) in an 8-hour workday, as described above.

6.1.4.2 Animals

Ventilation rates by species are required to estimate the deposited dose of airborne particles in the respiratory tract of animals. When experimental ventilation rates are not available, species-specific average ventilation rates can be calculated using the following allometric scaling equation:

$$\ln(V_E) = b_0 + b_1 \times \ln(BW)$$

(eqn. 6-3)

where V_E is the minute ventilation (L/min); BW is body weight in kilograms (kg); and b_0 and b_1 are the species-specific parameters; for the rat, estimates of $b_0 + b_1$ are -0.578 and 0.821, respectively (in Table 4-6 of EPA [1994]).

For example, the default value for minute ventilation in the MPPD rat model [ARA 2015] is 0.21 L/min, based on the default values of 2.1 ml (V_T) and 102 min^{-1} (*f*):

$$0.21(\mathrm{L/min}) = 2.1(\mathrm{ml}) \times 102(\mathrm{min}^{-1}) \times 0.001(\mathrm{L/ml}) \quad \text{(eqn. 6-4)}$$

This minute ventilation corresponds to a 300 g rat, based on eqn. 6-3:

$$0.21\,\mathrm{L/min} = \exp[-0.578 + 0.821 \times \ln(0.3)] \quad \text{(eqn. 6-5)}$$

6.1.5 Deposited Dose Calculation

The deposited dose of inhaled particles in the respiratory tract region is a biologically relevant estimate of equivalent dose in humans or animals. Equivalent dose metrics are needed to extrapolate dose-response relationships and risk estimates from animals to assess human risk.

The deposited lung dose can be estimated as follows:

$$\begin{aligned}\text{Deposited lung dose (mg)} &= \text{exposure concentration (mg / m}^3\text{)} \times \text{duration in hours (hours per day} \times \text{days per week} \\ &\times \text{weeks exposed)} \times \text{ventilation (L / min)} \times 0.001\ \text{m}^3\text{ / L} \times \text{60 minutes per hour} \\ &\times \text{regional deposition fraction}\end{aligned} \quad \text{(eqn. 6-6)}$$

Exposure concentration and duration would be as reported in the animal or human study, or for the exposure scenario of interest. Minute ventilation is calculated as shown in Section 6.1.4. The regional deposition fraction of interest is estimated for the respiratory tract region associated with the adverse effect in the risk assessment. The regional deposition fraction is estimated from the airborne particle diameter, and these values have been measured in various particle sizes, including in a study of several small laboratory animals [Raabe et al. 1988]. The deposition fraction can also be estimates in MPPD, v. 3.04 [ARA 2015] for several species (human, rat, mouse, rhesus monkey, pig, or rabbit). Airborne particle size and density are required input values in MPPD.

For example, to estimate the deposited lung dose in a rat subchronic (13-week) inhalation study at a pulmonary effect level of 5 mg/m^3, the exposure concentration and duration are as reported in the study. The minute ventilation can be calculated as shown in eqn. 6-5, and the pulmonary deposition fraction can be estimated in MPPD. Typically, the particle size data would also be reported in the study. For simplicity, assuming particle mass median aerodynamic diameter (MMAD) of 1 µm, monodisperse (geometric standard deviation of 1), and unit density (1 g/cm^3), a rat pulmonary deposition fraction of 0.06 is estimated in MPPD v. 3.04 [ARA 2015] (using default values for the other model parameters). The total deposited dose in rats in this example would be calculated as follows:

$$1.4 \text{ mg} = 5\ mg/m^3 \times (6 \text{ hours per day x } 5 \text{ days per week } \times 13 \text{ weeks}) \times (0.21\ L/min \text{ x } 0.001\ m^3/L \times 60 \text{ minutes per hour}) \times 0.06 \quad \text{(eqn. 6-7)}$$

If lung doses were not reported in a rodent study, the deposited dose can be estimated using this method. The worker-equivalent airborne concentration can then be estimated by "back-calculating" to determine the airborne concentration that would result in the equivalent pulmonary-deposited dose in humans (Figure 6-2). More biologically relevant dose estimates may also take account of the clearance of particles by respiratory tract region to estimate the retained dose over time, as discussed below.

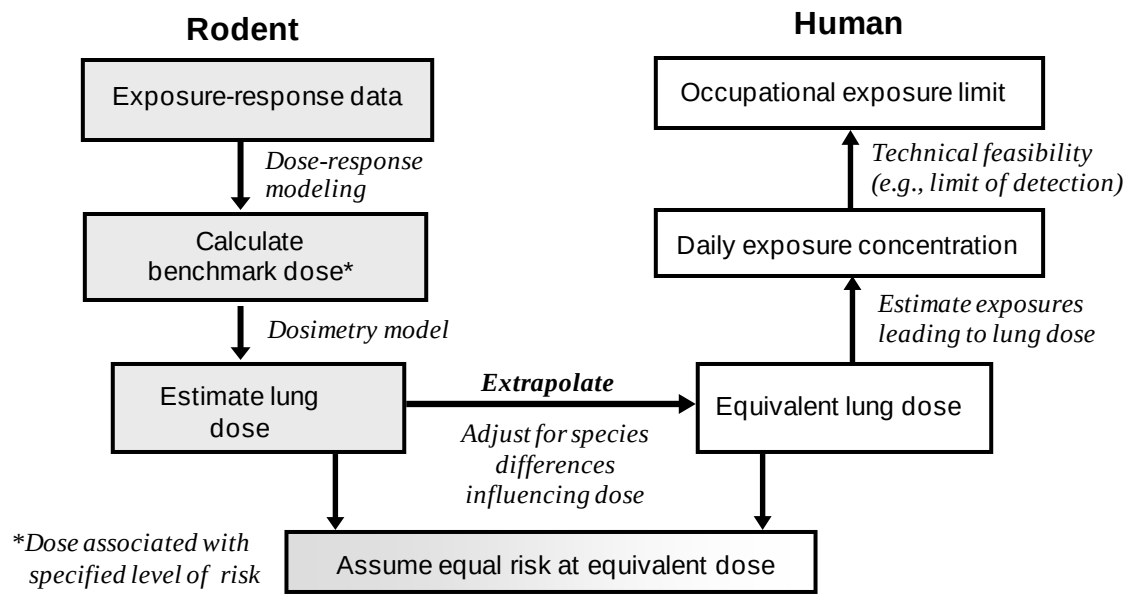

Figure 6-2. Dosimetry and risk assessment steps used by NIOSH to develop occupational exposure limits for airborne particles extrapolated from animal data (Kuempel [2011]; adapted from Oberdörster [1989]).

6.1.6 Biokinetic Mechanisms and Models of Inhaled Particles

6.1.6.1 Clearance, Retention, and Translocation

The biological mechanisms of particle clearance depend on the respiratory tract region in which the particles deposit and on the physicochemical properties of the particles. Particles that deposit in the bronchial region are cleared mainly by the mucociliary pathway, which carries particles or other exogenous materials towards the mouth where they are swallowed or expectorated. Particles that deposit in the pulmonary region are cleared primarily by alveolar macrophages that phagocytose (engulf) particles, where they are dissolved or transported to the tracheobronchial region for mucociliary clearance [Schlesinger 1985]. Poorly soluble particle clearance can differ across species due to differences in the rates of mucociliary transport in the conducting airways and macrophage-mediated clearance from the alveolar region [Miller 2000; Snipes 1989]. Pulmonary clearance is approximately 10 times slower in humans than in rats, based on first-order clearance assumptions [Snipes 1989].

Retention is described as the temporal distribution of uncleared particles in the respiratory tract [Lioy et al. 1984]. In humans, two distinct phases of particle retention have been observed. The first phase is thought to represent mucociliary clearance of particles depositing in the tracheobronchial region and is complete within approximately 24 hours, although a particle size-dependent slow clearance fraction has also been demonstrated [ICRP 1994; Stahlhofen et al. 1989]. The second phase, which involves retention half-times from approximately 30 to several hundred days, may represent particle clearance within the alveoli (air sacs) and interstitium (connective tissue separating the alveoli) of the pulmonary region.

Particles or fibers that are not cleared from the lungs can move into the lung interstitial tissue (either alone or inside macrophages). Particle retention in the interstitium increases the risk of fibrosis for poorly soluble particles. Translocation of particles from the lungs to the lung-associated tissues and systemic organs has also been reported, for particles from coal dust to carbon nanotubes [LeFevre et al. 1982; Mercer et al. 2013].

The physicochemical properties that influence the clearance or retention of particles from the respiratory tract include the chemical composition, size, surface properties, solubility, and shape [Kreyling et al. 2013].

Dosimetry models have been developed to describe both deposition and clearance of poorly-soluble spherical particles from the respiratory tract in animals and humans, for example, in MPPD 3.04 [ARA 2015], which is freely available and widely used. Recent updates to the human long-term clearance model are discussed in Section 6.1.6.2. Regional respiratory tract deposition fractions for fibers can also be estimated in MPPD 3.04, but fiber-specific clearance is not included. Dissolution of particles and extrapulmonary translocation are also not currently included in MPPD.

6.1.6.2 Models of Long-term Particle Retention in Humans

Studies in workers have suggested that the long-term retention of respirable particles involves the sequestration of some portion of the dust in the lungs, even at low exposures that would be below overloading in rats [Gregoratto et al. 2010; Kuempel et al. 2001]. These independent studies include workers exposed to particles

from relatively low (radioactive cobalt) to high (coal dust) mass concentrations. The human pulmonary clearance and retention models that include an interstitial sequestration compartment have been shown to provide better prediction of long-term retained lung burdens in humans with either low or high dust exposures compared to models with either simple first-order clearance model or dose-dependent overloading (first-order clearance until reaching a critical dose associated with decreasing clearance rate) [Gregoratto et al. 2010; Kuempel and Tran 2002; Kuempel 2000; Kuempel et al. 2001; Tran and Buchanan 2000]. Consistent with these findings, a study comparing rat and human particle retention patterns in the lungs showed that coal miners retained a greater proportion of particles in the alveolar interstitial tissue, while rats retained a greater proportion of particles in the alveolar spaces [Nikula et al. 2000].

The ICRP [1994] model includes three first-order pulmonary (alveolar-interstitial) clearance compartments. A fixed proportion of respirable particles deposition in the alveolar region is assigned to each compartment (i.e., 30, 60, and 10% for AI_1 AI_2, and AI_3, respectively). The first-order clearance rate coefficients are 0.02, 0.001, and 0.0001 day^{-1}, corresponding to retention half-times of 34, 693, and 6,930 days, respectively.

The MPPD human clearance and retention model (including v. 1.0 to current v. 3.04) [ARA 2015; Price et al. 2002] uses the ICRP model to predict clearance and retention in humans. Higher worker lung burdens were estimated in the interstitial sequestration model [Kuempel et al. 2001] than in the MPPD model v. 1.0 [Price et al. 2002], as reported in Dankovic et al. [2007]. Thus, the current widely used dosimetry model may under predict the average long-term particle retention in humans, and therefore may underestimate the risk of adverse effects associate with retained particle dose in the lungs.

6.1.7 PBPK models to Estimate Dose

When the relationship between the external exposure and internal dose is nonlinear, PBPK models are preferred for temporal extrapolation. This is because PBPK models can account for capacity-limited processes in the absorption/uptake, distribution, metabolism, and/or excretion of a toxicant. Capacity limitation may occur due

to saturation of a key process, e.g., involving a receptor or enzyme. The overloading of pulmonary clearance of particles or fibers in rodents via alveolar macrophage dysfunction results in a dose-dependent increase in the particle retained dose [Bellmann et al. 1991; Bolton et al. 1983; Morrow 1988; Stöber et al. 1990]. As a result, the biological mechanisms and pathways operating at lower, non-overloading doses can differ from those operating at higher doses when defenses of cells or organism are overwhelmed [McClellan 1997; Oberdörster et al. 2005]. In this case, a default dosimetry adjustment to a higher shorter-term dose may provide a poor estimate of the response at an equivalent dose delivered over a longer term at a lower rate, which does not impair clearance capacity [Kuempel et al. 2015]. PBPK models, which are also known as dosimetry models, have been used in several NIOSH risk assessments of inhaled particles, including titanium dioxide, carbon nanotubes, and silver nanoparticles [NIOSH 2011; NIOSH 2013b; NIOSH 2016b].

6.1.8 Overloading Considerations in Rodent Model and Dose Estimation

The effects of particle overloading of lung clearance in rats and mice involves a sequence of events including persistent pulmonary inflammation in both rodent species, fibrosis primarily in rats, and cancer in rats [Baan 2007; Elder et al. 2005; Oberdörster 1995]. Rats have been shown to be better predictors of lung cancer from inhaled particles that are carcinogenic to humans (i.e., classified by IARC as having limited or sufficient evidence) compared to mice or hamsters, which give false negative results more often [Mauderly 1997].

This well-studied rodent phenomenon of particle overloading of pulmonary clearance is the basis for the risk assessment approach of identifying the non-overloading dose in rats as the NOAEL to extrapolate to humans [Morrow et al. 1991; Pauluhn 2010]. While this concept seems reasonable based on the rat data, it may not be adequate to estimate chronic responses in humans due to differences in the clearance and retention kinetics in humans (as discussed in Section 6.1.9).

The dose metrics associated with overloading of lung clearance include particle mass (unit density particles), particle volume (particles with density other than 1 g/cm^3), or particle surface area (nanoparticles)

[Bellmann et al. 1991; Morrow 1988; Tran et al. 2000]. In contrast to microscale particles, nanoscale particles or highly toxic particles have been shown to cause impaired pulmonary clearance at a lower mass or volumetric particle dose than for microscale poorly soluble low toxicity particles (PSLT) [Bellmann et al. 1991; Oberdörster et al. 1994]. Particle surface area has been shown to better describe the decreased clearance and pulmonary responses to nanoscale compared to microscale particles [Tran et al. 2000]. Since particle dosimetry models are generally based on the particle mass, dose conversion may be necessary between the estimation of effect level in the rodent study (e.g., surface area dose associated with adverse effect) and the estimation of the equivalent dose in humans.

6.1.9 Interspecies Dose Estimation in Risk Assessment

Scientific models are generally available to estimate the human-equivalent lung doses of inhaled particles to those in rodents [ARA 2015; Paquet et al. 2015]. Less well understood are the human and rat biological responses to equivalent mass, surface area, or volumetric particle lung doses. For example, the biological mode of action for the development of lung tumors in rats exposed to PSLTs by chronic inhalation appears to involve secondary genotoxicity resulting from chronic inflammation and cell proliferation [IARC 2010; ILSI 2000; NIOSH 2011]. Thus, at low lung doses in rats (i.e., below lung overload), where inflammation and cell proliferation are not present, lung cancer would not be anticipated [Greim et al. 2001]. Mice also showed overloading of lung clearance but had lower inflammatory response than rats in a subchronic inhalation study of a PSLT (carbon black); hamsters did not show overloading or lung inflammation in that study [Elder et al. 2005]].

The interpretation and use of rat dose-response data of inhaled particles in human hazard and risk assessment and OEL development has been discussed and debated for many years [Cherrie et al. 2013; IARC 2010; ILSI 2000; Kuempel et al. 2014; Morfeld et al. 2015; Oberdörster 1995; Pauluhn 2014; Warheit et al. 2016; Yu 1996]. Yet, the rat chronic bioassay data have been shown to give fewer false negatives for inhaled particles classified by IARC as human carcinogens than have the mouse and hamster data [Mauderly 1997]. Moreover,

human particle lung doses in workers in dusty jobs such as coal mining have been shown to be equivalent to the mass overloading doses in rats [IARC 2010; Kuempel et al. 2014; NIOSH 2011].

In general, the rat model is considered as a useful model for human non-neoplastic lung responses to PSLT, and in the absence of mechanistic data to the contrary, the rat model is relevant to identifying potential carcinogenic hazards in humans [ILSI 2000]. Rat chronic inhalation data of PSLT were used by IARC [2010] in its evaluation of the carcinogenicity of inhaled PSLT (titanium dioxide and carbon black) and by NIOSH [2011] in its hazard classification and REL for nanoscale and microscale titanium dioxide.

Scientific questions on rat lung overload that still need to be resolved were discussed by Borm et al. [2015], who cite two papers that contribute to the debate [Morfeld et al. 2015; Pauluhn 2014]. To date there is no clear resolution of this issue in the scientific community. Therefore, interpretations of the rat dose-response data for risk assessment have differed widely for inhaled PSLT including for nanoscale titanium dioxide, using the same basic data [NIOSH 2011; Relier et al. 2017; Warheit et al. 2016]. Although the scientific debate may continue, dosimetric adjustments to account for differences in PSLT aerosol particle size and respiratory tract disposition and/or clearance between rodents and workers have been used to account for toxicokinetic differences, and uncertainty factors can be applied to account for toxicodynamic differences [EPA 1994; ICRP 1994; Jarabek et al. 2005; Kuempel et al. 2015; Oller and Oberdörster 2016]. Animal inhalation studies used in risk assessment should include sufficient doses to characterize the dose-response relationship, including low doses to overloading doses [ILSI 2000; Kuempel et al. 2014; Oberdörster 1997; Pauluhn 2011].

Despite the differences in particle clearance and retention kinetics, the overloaded rat model may be relevant to predicting risk to workers exposed to inhaled particles. Overloading doses of microscale PSLT in rats have been observed as low as 0.5 mg/g lungs, with complete cessation of clearance at dose >10 mg/g lungs [Muhle et al. 1990; Oberdörster 1995]. By comparison, workers in dusty jobs historically have had average retained particle mass doses >10 mg/g lungs [Douglas et al. 1986; Freedman and Robinson 1988; Stöber et al. 1965]. Thus, only at overloading doses does the particle lung burden in rats reach the higher levels that have been

reported in coal miners. These findings suggest that the rat is an appropriate model for human health risk assessment of respirable particles.

Studies in mice and hamsters are not as predictive of the human particle-associated lung responses, and were negative for some particles that have been classified as known human carcinogens [Mauderly 1997]. In a quantitative comparison of lung cancer risk estimates in rats and humans associated with chronic exposure to various types of respirable PSLT (coal mine dust, carbon black, titanium dioxide, or crystalline silica), the rat- and human-based estimates were statistically consistent given the level of imprecision in the animal and human data [Kuempel et al. 2009; NIOSH 2011]. These studies suggest that the rat may be the most reasonable and sensitive rodent model to estimate the risk of chronic exposure to respirable particles, despite the species differences in the clearance and retention kinetics, which can be adjusted for by using dosimetry modeling.

6.1.10 Tools/Models (deposition and/or clearance)

The most widely used dosimetry models for inhaled particles and fibers for more than a decade are found in the MPPD suite of models [ARA 2015; Price et al. 2002]. These models have largely replaced the U.S. EPA Regional Deposited Dose Ratio (RDDR) model, which allowed estimation of the equivalent deposited dose in the respiratory tract across species, but did not include clearance [EPA 1994].

Several deposition and clearance models are included in MPPD, as described in the model overview and details in the software (MPPD v. 3.04). The MPPD has been developed over a decade or more with funding by various U.S. governmental (including EPA, Navy, and NIOSH) and nongovernmental sources. It is available at https://www.ara.com/products/multiple-path-particle-dosimetry-model-mppd-v-211].

NIOSH-funded revisions to earlier versions include: batch capability in running the deposition and clearance models in humans and rats (in MPPD v. 2.1 [ARA 2009]); addition of oronasal deposition in animals and humans including olfactory deposition of nanoscale particles [Garcia and Kimbell 2009; Garcia et al. 2015]; extension of the spherical particle model to include nonspherical and fibrous particles based on aerosol

characterization and measurement of deposition efficiency in human respiratory tract replicas [Su and Cheng 2015; Su and Cheng 2014].

The MPPD model is for poorly soluble particles (spherical or nonspherical), but does not account for particle dissolution. It is also limited to the respiratory tract, and does not include translocation to other organs. Other dosimetry/PBPK models are needed to estimate internal dose of soluble particles in the lungs or other organs.

In general, data available for PBPK modeling are limited. When validated models are available, they are preferred to application of uncertainty factors to estimate human-equivalent dose because they account for material- and species-specific factors influencing the dose to target tissues. Many individual PBPK (dosimetry) models have been developed for inhaled particles and fibers; their use would need to be evaluated on a case-by-case basis. Some useful tools and references associated with dosimetry modeling are listed in Table 6-2.

In literature searching for information on lung dosimetry models of aerosols, it should be noted that multiple databases might need to be used, such as PubMed, Web of Science or Scopus, Toxline, and/or Embase. Although PubMed is a major research database, and perhaps the most widely used, it does not provide citations for some of the journals in which aerosol research is published (e.g., the Journal of Aerosol Science). Past practices have demonstrated that broader search strategies may be needed to identify relevant articles in this area.

Table 6-2. Examples of available tools and resources for dosimetry modeling (adapted from Kuempel et al. [2015]).

Name of Tool or Resource	Description	Source and Availability
Multiple-path particle dosimetry model (MPPD)	Deposition, clearance, and retention estimation of inhaled particles in the respiratory tract of the human, rat, and mouse	Freely available at: http://www.ara.com/products/mppd.htm Based on several models including: Anjilvel and Asgharian [1995], Asgharian et al. [2001; 2014], and ICRP models [ICRP 1994; Paquet et al. 2015].
Respiratory tract region deposited dose equations	Deposited dose estimation of inhaled particles or vapors Interspecies dosimetric adjustments. Derivation of reference concentrations	U.S. EPA [2012a; 1994] http://www.epa.gov http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=71993 http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=244650 (freely available)
Human respiratory tract model	Deposition, clearance, and retention estimation of inhaled particles (including non-radioactive) in the human respiratory tract	ICRP Publication 66 [ICRP 1994] http://www.icrp.org/ http://www.sciencedirect.com/science/journal/01466453/24/1-3 (freely available)
Physiologically-Based Pharmacokinetic (PBPK) modeling guidance	Guidance on principles of characterizing and applying PBPK models in risk assessment	U.S. EPA 2006 http://www.epa.gov http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=157668; IPCS 2010 http://www.who.int/ipcs/en/ http://www.inchem.org/documents/harmproj/harmproj/harmproj9.pdf ; (freely available) Loizou et al. [2008] and McLanahan et al. [2012]
Human reference values	Anatomical and physiological parameters (reference values) in humans Inter-individual variability by age and gender Parameters for PBPK models	ICRP Publication 89 [Valentin 2002] http://www.icrp.org/ http://www.sciencedirect.com/science/journal/01466453/32/3-4 (freely available)
Interspecies reference values	Physiological parameters for dose normalization or PBPK modeling Application to Biological Exposure Indices	Brown et al. [1997], Davies and Morris [1993], Mercer et al. [1994], Stone et al. [1992]; Boxenbaum [1982], and Fiserova-Bergerova [1990].
Particle size definitions	Criteria for airborne sampling of particle size fractions by probability	ACGIH 2014 (moderate fee for purchase); ACGIH 1984 and Lioy et al. [1984]

	of deposition in human respiratory tract regions

6.2 Gas and Vapor Exposures

It is often useful to estimate the dose delivered to the target sites in laboratory animals and humans. Occupational risk assessment based on the biologically effective dose that mediates the adverse response would be ideal but is generally unavailable. However, internal dose metrics are often available and may be advantageous for assessing risks, particularly when they are not linearly related to the air concentration of the chemical. Therefore, understanding how gases and vapors are absorbed throughout the respiratory tract is important to determine an accurate dose estimate in the test species and in humans. Dosimetry models can reduce the uncertainty associated with extrapolating risks from test species to humans. This section describes the methods used for calculating human equivalent concentrations (HEC) for gas or vapor exposures.

As previously described in Section 6.1.2, the major respiratory tract regions include extrathoracic (nasal, pharyngeal, laryngeal), tracheobronchial (airways), and pulmonary (alveolar) regions (Figure 6-1 and Table 6-3). In general, the major factors influencing the internal dose from gas or vapor inhalation are anatomy (ventilation rate), physiology (diffusion, dissolution, blood flow, metabolism, and elimination rates), physicochemical properties (e.g., gas or vapor solubility, reactivity) of the chemical [Bogdanffy and Jarabek 1995; Hanna et al. 2001; Jarabek 1995; Kuempel et al. 2015].

Table 6-3. Respiratory tract regions

Region		Anatomic Structures	Other Terminology
Extrathoracic (ET)		Nose Mouth Nasopharynx Oropharynx Laryngopharynx Larynx	Head airways region Nasopharynx (NP) Upper respiratory tract (URT) Upper airways
Thoracic (TH)	Tracheobronchial (TB)	Trachea Bronchi Bronchioles (to terminal bronchioles)	Conducting airways
	Pulmonary (PU)	Respiratory bronchioles (not found in rodents) Alveolar ducts and sacs Alveoli	Gas exchange region Alveolar region Parenchyma

Adapted from EPA [1994] and Phalen et al. [1988]

The components of the inhalation dosimetry adjustment for gases are:

1. Conversion of units from ppm to mg/m^3: The concentration in the inhalation toxicity studies on gases are usually reported in units of ppm or mg/m^3. For exposure levels reported as ppm, this should be converted to the standard units of mg/m^3 by using the following formula.

$$\text{mg/m}^3 = \frac{\text{ppm x MW}}{24.45}$$

Where MW is the molecular weight in grams and 22.45 is the volume occupied by 1 g-mol of any compound in the gaseous state at 25 °C and 760 mm Hg.

2. Duration adjustment: Many inhalation toxicity studies in laboratory animals are conducted with discontinuous exposure, often with exposure frequencies of 6 to 8 hours per day and 5 days per week. Occupational risk estimates are derived with the intention to protect workers against the exposure of 8

hours per day for 5 days per week totaling 40 hours per workweek. Therefore, duration-adjusted exposure level is:

$$Adjusted\ concentration\ (mg/m^3) = E(mg/m^3) \times D/8 \times W/5$$

Where E is the experimental exposure concentration, D is the work day adjustment of the number of hours exposed in 8-hour daily increments, and W is the workweek adjustment of the number of days of exposure in 5-day workweek increments.

3. Human Inhalation Rate: The human inhalation rate for light exertion while doing work of 9.6 m^3/8-hours should be included in the risk estimate.

4. Human Equivalent Concentration (HEC): The HEC is the concentration of a substance in humans that is believed to produce an equal effect by a dose in experimental animals adjusted for exposure duration and physiological parameters, such as breathing rate.

6.2.1 NIOSH Practice:

The current practice of calculating HEC in NIOSH is as follows:

1. Experimental animal dose in ppm is converted to daily mg/kg inhaled dose.
2. In the absence of chemical-specific information on metabolism or dosimetry, this dose is extrapolated to humans, assuming dose equivalence in units of mg/kg·day scaled according to body weight to the 0.75 power.
3. The human mg/kg-day dose is then converted to ppm for an 8-hour workday.

The following example is taken from the NIOSH 1-bromopropane criteria document. Experimental animal dose in ppm is converted to daily mg/kg inhaled dose.

$$\text{Mouse BMDL} = 0.64\ \text{ppm} \times 5.031\ \frac{\text{mg/m}^3}{\text{ppm}} \times 0.060\ \text{m}^3/\text{day} \times 6\ \text{hour}/24\ \text{hour} / 0.0353\ \text{kg}$$
$$= 1.3682\ \text{mg/kg} \cdot \text{day}$$

where, 0.64 ppm is the dose, 5.031 mg/m^3 per ppm is the molecular weight (122.99 g/mol) divided by 24.45, 0.060 m^3/day is the reference inhalation rate for female B6C3F1 mouse, and 0.0353 kg is the reference body weight for female B6C3F1 mouse in chronic study.

This dose is extrapolated to humans, assuming dose equivalence in units of mg/kg·day scaled according to body weight to the 0.75 power:

$$\text{Human BMDL} = \text{Mouse BMDL}\ (1.3682\ \text{mg/kg} \cdot \text{day}) \times (0.0353\ \text{kg}/70\ \text{kg})^{0.25} = 0.205\ \text{mg/kg} \cdot \text{day}$$

Where 70 kg is the reference human body weight. Here the exponent value of 0.25 reflects body weight (BW) scaling to the ¾ power (i.e., BW$^{3/4}$) in units of mg/kg-day (rather than mg/day) such that BW$^{1/1}$ / BW$^{3/4}$ = BW$^{1/4}$.

The human mg/kg·day dose is then converted to ppm for an 8-hour workday:

$$\text{Human BMCL} = 0.205\ \text{mg/kg} \cdot \text{day} \times \frac{70\text{kg}}{9.6\ \text{m}^3/8-\text{hour workday}} \times \frac{1\text{ppm}}{5.031\ \text{mg/m}^3}$$
$$= 0.297\ \text{ppm, or about } 0.3\ \text{ppm}$$

where 9.6 m^3 per 8-hourwork day is the reference inhalation rate for humans.

6.2.2 U.S. EPA Practice

The HEC is calculated from a PoD (e.g., NOAEL, LOAEL, or BMDL) by using the following formula:

$$\text{PoD}_{[\text{HEC}]}(\text{mg/m}^3) = \text{PoD}_{[\text{ADJ}]}(\text{mg/m}^3) \times \text{RGDR}$$

where, RGDR is the regional gas deposited ratio, which is the ratio of regional gas dose in laboratory animal species to that of humans for the target region.

The EPA categorized the gas based on the physicochemical properties and the regions of the effect in the respiratory tract into categories 1, 2 and 3 (please see Table 6-4 for the details of each categories). Category selection for a given chemical should be based on the properties of the chemical and its target effects in the body, as described in Table 6-4.

Table 6-4. Gas categories and characteristics

	Characteristics				
Category	Water solubility	Reactivity	Accumulation in blood	Site of Toxicity	Examples
1	High	Rapidly irreversibly reactive	Not significant	Portal of entry	Hydrogen fluoride, chlorine, formaldehyde, volatile organic acids and esters
2	Moderate	Rapidly reversibly reactive, or moderately to slowly irreversibly metabolized in respiratory tract tissue	Potential	Portal of entry, maybe systemic	Ozone, sulfur dioxide, xylene, propanol, isoamyl alcohol
3	Low	Unreactive in surface liquid and tissue	Yes	Systemic toxicity	Styrene

Adapted from EPA [2012a]

6.2.3 Category 1 Gases

Category 1 gases are highly water soluble or reactive and thus produce an effect mostly in the respiratory tract itself. Because of the high-level deposition along with high reactivity, local tissue damage is expected from these gas exposures. Only a small fraction of these gases could penetrate deeper than the ET region under normal circumstances. However, during heavy exercise, fires, explosions etc., these gases could penetrate deeper, leading to tissue damage in the distal respiratory tract. The following equations are used to calculate RGDR gas ratio for different regions of Category 1 gas.

1. $RGDR_{ET}$ for Category 1 Gas:

$$RGDR_{ET} = \frac{(V_E/SA_{ET})_A}{(V_E/SA_{ET})_H}$$

Where V_E is the minute volume (cm^3/min), SA_{ET} is the surface area of the extrathoracic region (cm^2), and terms A, H represent laboratory animal and human, respectively.

2. $RGDR_{TB}$ for Category 1 Gas:

$$RGDR_{TB} = \frac{(V_E/SA_{TB})_A}{(V_E/SA_{TB})_H}$$

Where V_E is the minute volume (cm^3/min), SA_{TB} is the surface area of the tracheobronchial region (cm^2), and terms A, H represent laboratory animal and human, respectively.

3. $RGDR_{PU}$ for Category 1 Gas:

$$RGDR_{PU} = \frac{(Qalv/SA_{PU})_A}{(Qalv/SA_{PU})_H}$$

Where Qalv is the alveolar ventilation rate (mL/min) and is equal to 0.6 x V_E.

6.2.4 Category 2 Gases

Category 2 gases are moderately water-soluble and have the potential to penetrate into bronchi and thereby to the blood. Therefore, both local and systemic effects could be observed following exposure to these gases. HECs for respiratory tract effects are calculated using the equations for a Category 1 gas, whereas HECs for extra-respiratory effects are calculated using the Category 3 equations. In cases where respiratory tract effects

caused by systemic distribution of the chemical, such as chloroform and naphthalene, the HEC should be calculated as a Category 3 gas. Therefore, the mode of action determines the category.

6.2.5 Category 3 Gases

Category 3 gases have very low water solubility and limited reactivity with respiratory epithelium. These gases readily penetrate to the pulmonary region and are absorbed into the systemic circulation. Most of the effects are observed distal to the respiratory system except in cases where metabolism in the upper-respiratory tract leads to local effects. The following equation is used to calculate the RGDR for Category 3 gases.

$$\mathrm{RGDR} = \frac{\left(\mathrm{H_{b/g}}\right)_{\mathrm{A}}}{\left(\mathrm{H_{b/g}}\right)_{\mathrm{H}}}$$

Where the value $\frac{\left(\mathrm{H_{b/g}}\right)_{\mathrm{A}}}{\left(\mathrm{H_{b/g}}\right)_{\mathrm{H}}}$ is the ratio of the blood: gas (air) partition coefficient of the chemical for the laboratory animal species to the human value. A value of 1.0 is used for the ratio of $(Hb/g)_A > (Hb/g)_H$. A value of 1.0 is used as the default when one or both of the partition coefficients are not available. Blood: air partition coefficients for a number of chemicals are available from Gargas et al. [1989]

6.2.6 PBPK and Computational Fluid Dynamics Approaches

PBPK modeling is used to derive target tissue dose estimates in various species. The construction and development of a PBPK model for an individual chemical involves a significant amount of data and understanding of the absorption, metabolism, distribution, and elimination of the chemical in the test species and in humans. One of the examples of PBPK modeling used in occupational risk assessment is use of the methylene chloride PBPK model to derive target tissue dose estimates for lung tumors in mice [OSHA 1997].

Computational fluid dynamics-PBPK models are designed to model the fluxes of vapor between tissue phases (eg. Between epithelial and submucosal tissues) and also allow for a differential blood flow and coupling

the respiratory tract to the whole body. This type of model has been used to evaluate the dosimetry of many compounds including diacetyl, styrene [Gloede et al. 2011; Sarangapani et al. 2002].

Guidance on developing PBPK and/or computational fluid dynamics models are beyond the scope of this document; however, additional information is available in: "Approaches for the Application of Physiologically Based Pharmacokinetic (PBPK) Models and Supporting Data in Risk Assessment" [EPA 2006a].

7.0 RISK CHARACTERIZATION

Risk characterization is the third and final step in the NIOSH risk assessment process. It is the qualitative and, wherever possible, quantitative determination, including attendant uncertainties, of the probability of occurrence of known and potential adverse effects of an agent in a given organism, system, or population, under defined exposure conditions. In environmental risk assessment, risk characterization is meant to describe the likelihood and severity of exposure-related adverse effects using information on the degree of potential exposure to the hazard within a population and its dose-response relationship with the adverse effect. Risk characterization in NIOSH risk assessments is restricted to health risks and focuses on the translation of information on the risk of workplace exposures into a basis for recommendations on limiting exposure. For example, a dose-response relationship observed between chronic inhalation of methylene chloride and liver and lung tumor incidence in mice may be used to derive a limit on methylene chloride exposure over a working lifetime that corresponds to an increase risk in humans on the order 10^{-4}. The process of transporting risks observed in animals in an experimental study to workers exposed over the course of their employment is an example of NIOSH risk characterization.

Risk characterization is the culmination of all of the information gathered for the risk assessment to meet its intended purpose of informing risk management decisions. As such, the risk characterization is formally documented in NIOSH Criteria Documents or Current Intelligence Bulletins containing RELs, RML-CAs and alternative authoritative recommendations, such as hazard banding. Risk characterization is meant to synthesize and communicate the risk assessment science to a broad audience. Thus to the maximum extent practicable, NIOSH risk assessors follow the guiding principles of transparency, clarity, consistency, and reasonableness in risk characterization, as first described by the EPA [Fowle and Dearfield 2000] (see Table 7-1).

Table 7-1. Risk Characterization Guiding Principles[1]

Principle	definition	Criteria for risk characterization
Transparency	Explicitness in the risk assessment process.	Use a risk analysis plan Describe assessment approach, assumptions, extrapolations and use of models Describe plausible alternative assumptions Identify data gaps Distinguish science from policy Describe uncertainty Describe relative strength of assessment
Clarity	The assessment itself is free from obscure language and is easy to understand.	Be brief and concise[2] Use plain English (avoid jargon)[2] Avoid technical terms Use simple tables, graphics, and equations
Consistency	The conclusions of the risk assessment are characterized in harmony with other NIOSH actions.	Follow NIOSH policies on technical writing and peer review Place assessment in context with similar risk assessments
Reasonableness	The risk assessment is based on sound judgment.	Use review by peers Use best available scientific information Use good judgment

1. Adopted from the EPA Risk Characterization Handbook [Fowle and Dearfield 2000]

2. Brevity may be at odds with clarity, as complex analyses may require detailed explanations for understandability. Similarly, the avoidance of jargon and technical terms may not be feasible for some complex analyses.

7.1 Risk Definitions

For NIOSH purposes, risk can be described as the probability of the adverse effect, although it is better defined as the incidence of the adverse effect (e.g., disease onset) occurring in subject(s) over a specified period of time given that the subject(s) were disease free at the beginning of time (i.e. cumulative incidence rate). Absolute risk describes the total risk within a working population independent of cause, while the attributable risk (sometimes referred to as added risk) refers to the component of absolute risk that is related to the hazardous agent of interest. Quantitative risk assessment relies on estimates of excess risk per unit dose that are derived from information obtained in the dose-response assessment. Here, excess risk is broadly defined as the increased

incidence of the adverse effect above a control level or background that is attributable to the exposure. However, excess risk can be expressed in multiple ways; such as:

- Added Risk: The difference in risk (or in the probability of a response) between subjects exposed and those not exposed to a hazard. For example, it is the increment by which the probability of adverse effect exceeds background probability, calculated as P(*d*)–P(*0*), where P(*d*) is the probability of response at dose, *d*, and P(*0*) is the probability of response at zero dose (i.e., background risk). Added risk is sometimes referred to as attributable risk.
- Extra Risk: The measure of the proportional increase in risk of an adverse effect adjusted for the background incidence of the same effect. In other words, extra risk is the added risk relative to the proportion of the population not responding to the background risk, calculated as [P(*d*)–P(*0*)]/ [1–P(*0*)]. For example, dose-response analyses of quantal experimental response data tend to use a BMR of 10% extra risk. Extra risk approaches added risk with decreasing contributions from background.
- Relative Risk: Typically reserved for human studies, the relative risk is the ratio of the risk of the adverse effect among people who are exposed to the hazardous agent, to the risk among those who are unexposed (or exposed to a lesser degree). Relative risk is synonymous with risk ratio. Relative risk has also been described as the ratio of the cumulative incidence rates among those exposed, to those unexposed. Rate ratios (i.e., the ratio of the instantaneous rate of disease in the exposed to those unexposed), hazard ratios, odds ratios and SMRs are frequently used as approximations of relative risk. Observational studies may also refer to excess relative risk (ERR), which is relative risk (or rate) -1.

7.2 Risk Characterization Framework

The direct measurement of exposure-related risk in the region of interest is not practical in most cases given that acceptable risks dwell below the observational level in toxicologic and epidemiologic research. Instead,

risk characterization relies on the extension of dose-response information using one of two general approaches to using dose-response data:

- PoD/UF Approach (Health-based): Divide the estimated PoD by factors (see Section 7.3.4) that account for identified sources of uncertainty to arrive at an estimate of safe dose. Here the term 'safe' implies an exposure level in which the associated risk is absent or negligible. In this approach, risk is not explicitly quantified (i.e., the dose is implicitly risk-free); however, probabilistic means may be used to quantify risk from exposure above the safe level.
- Extrapolation Approach (Risk-based): Obtain quantitative estimates of low-dose risk by model-based extrapolation of the risk at doses below the observable data. For example, a linear non-threshold (LNT) model would support extrapolation by extending a line from the origin of the dose response curve (i.e., the point of no exposure and no excess risk) to the human equivalent PoD in the observable range. This approach assumes that a safe level of exposure cannot be assured; therefore, residual risks are typically reported under one or more exposure scenarios using probabilistic means (e.g., the dose estimated to cause a lifetime excess risk of 1 in 10,000). In addition to an array of risks, a target risk level may be specified by risk managers to estimate an exposure limit (see Section 7.3.1).

The PoD/UF approach is generally applied when there is evidence or an assumption of a toxicity threshold in the dose-response curve at low doses. The concept that toxic effects have exposure thresholds is fundamental to toxicology [Aldridge 1986; Klaassen et al. 2013; Rhomberg et al. 2011; Rodricks et al. 2007]. As such, chemical risk assessments related to occupational diseases, excluding cancer, have mostly used a PoD/UF approach. In contrast, early risk assessments of cancer from ionizing radiation exposure recognized that induced mutagenesis exhibited effects that were proportional to dose and absent of a dose-response threshold [NRC 1956; Sievert and Failla 1959]. Continued research into low-dose radiation effects have led to the generally accepted notion of the LNT dose-response for radiocarcinogenesis [NRC 2006; UNSCEAR 2015]. Assuming the LNT

dose-response was also applicable to chemical carcinogenesis, the EPA adopted LNT extrapolation in its risk assessments of carcinogens beginning in the late 1970s [Albert et al. 1977].

Refinements in risk assessment methods since the 1970s have placed more emphasis on MoA evaluations given that some carcinogens exhibit nonlinearity at low doses. In fact, many of the factors contributing to nonlinearity in the dose-response curve at low doses for noncarcinogenic agents (e.g., clearance pathways, cellular defenses, and repair processes) may also support nonlinearity at low doses for some carcinogens. Conversely, some noncancer endpoints may be better suited to risk extrapolation, with an allowance for a dose-response that appears LNT at low doses. For example, large interindividual variability in the low-dose threshold of a noncancer endpoint (i.e., widely varying susceptibility) can result in a dose-response that approaches linearity at low dose. Exceptions to the existing cancer/noncancer dichotomy have prompted calls for the harmonization of risk characterization methods [Barton et al. 1998; Crump 2011; Crump et al. 1997; NRC 2009; Rhomberg et al. 2011; White et al. 2009]. In response, some researchers have suggested a unified approach to risk characterization that is either extrapolation [NRC 2009; White et al. 2009], PoD/UF-based [Crump 2011; Crump et al. 1997; Gaylor et al. 1999] or some combination of the two [Baird et al. 1996; Chen et al. 2007; Chiu and Slob 2015]. Still, others have suggested that harmony is best achieved by a framework allowing a choice of either approach [Barton et al. 1998; Rhomberg et al. 2011]. For example, Rhomberg et al. [2011] suggested that the cancer/noncancer paradigm is valid in most cases, yet acknowledged that exceptions may occur; therefore, the choice between approaches should be based on the degree of compatibility of the method on a case-by-case basis.

The approach used for risk characterization can have significant impact on its findings; therefore, its selection is a critical decision point in NIOSH quantitative risk assessment. Unfortunately, science is generally incapable of resolving the correct approach given typically sparse MoA data, an inability to observe the true shape of the dose-response curve at very low doses, and other limitations such as measurement error and interindividual variability in the dose-risk relationship. Therefore, methods can appear interchangeable and a preference for one over the other can be perceived as less than objective. To avoid inconsistency among risk assessments and to ease

transparency, NIOSH has developed a risk characterization framework that incorporates decision logic for systematically selecting a strategy for using extrapolation and PoD/UF approaches in conjunction with current science and NIOSH policy (Figure 7-1). NIOSH risk assessors are encouraged to follow this logic for planning and conducting risk characterization. NIOSH realizes that exceptions to the framework are possible given nuances in every risk assessment; therefore, risk assessors are discouraged from forcing a fit. Above all, a WoE approach for evaluating and applying MoA must be the foundation of any method selected for risk characterization. For example, extrapolation is the preferred risk characterization approach for the general class of chemical carcinogens. However, it is plausible that a non-genotoxic or non-DNA reactive carcinogen may have sufficient MoA information to support a practical response threshold at low doses. In this instance, it may be more appropriate to use PoD/UF rather than extrapolation. In addition, it should be understood that data availability is an important factor for deciding on a risk characterization approach. For example, a PoD can be determined from a NOAEL or LOAEL even if data are insufficient to quantify the dose-response relationship. Furthermore, this framework is applicable only to NIOSH risk assessments; factors used in its development may be unrelated to, or may weigh differently on, risk characterization conducted elsewhere.

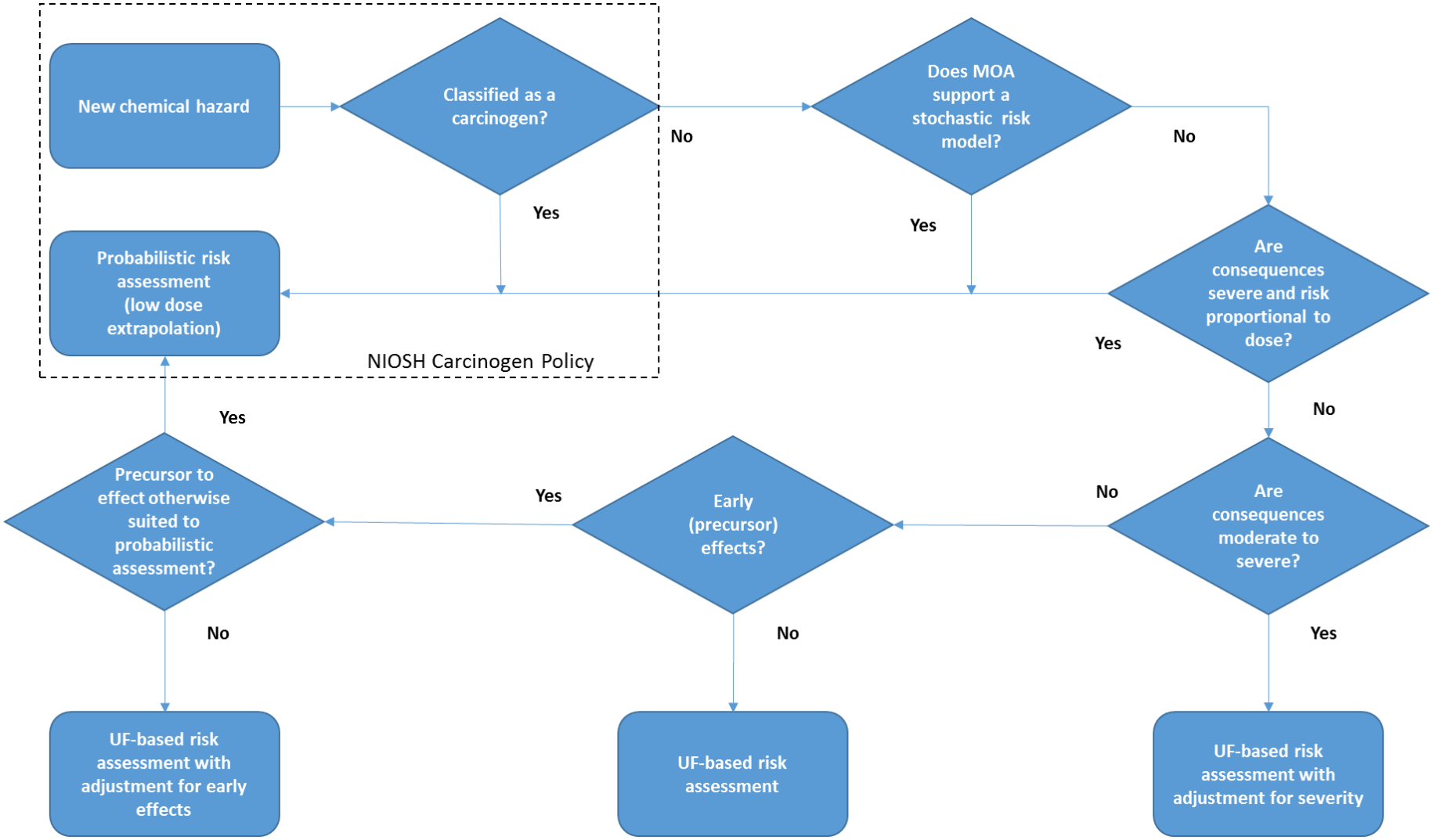

Figure 7-1 Framework for risk characterization of a potential chemical hazard.

7.2.1 Carcinogens

NIOSH has separately published its policy on the classification and risk characterization of chemical carcinogens [NIOSH 2017]. The policy is partly founded on the premise that, for most chemical carcinogens, there is no *known* safe level of exposure; therefore, an extrapolation approach is generally required for characterizing carcinogenic risk at low doses. That being said, there is emerging scientific evidence supporting that some carcinogens may have sufficient MoA information to conclude that the dose-response is nonlinear at low doses. In these situations, simple linear extrapolation may significantly overestimate cancer risk. Thus, the policy allows for nonlinear extrapolation for chemical carcinogens with sufficient MoA evidence supporting nonlinear dose-response relationships at low doses. Specifically, NIOSH recognizes three general types of carcinogens based on the weight of MoA evidence for carcinogenesis (adapted from Streffer et al. [2004]):

- Genotoxic carcinogens consistent with LNT: All mutagens and most direct-acting (DNA-reactive) genotoxic carcinogens separated into two subgroups: 1) those in which the WoE supports LNT (e.g., ionizing radiation and vinyl chloride); and 2) those in which mechanisms are uncertain or generally unsupportive of a threshold at low doses (e.g., acrylamide, acrylonitrile). For the latter, a default assumption of LNT is used as a health-protective measure. For example, acrylamide is clearly genotoxic at the chromosome level and is metabolized through cytochrome P450 CYP2E1 pathway to a potentially reactive metabolite; therefore, it has generally been treated as a direct-acting mutagen [Streffer et al. 2004]. There is a growing body of evidence of nonlinearity in the slope of the response for acrylamide; however, underlying genotoxic mechanisms are still poorly understood [Maier et al. 2012; Shipp et al. 2006]. Until the WoE is supportive of an alternative approach, risk characterization for acrylamide would likely prefer LNT extrapolation as a primary risk characterization approach.

- Genotoxic carcinogens inconsistent with LNT: This category is characterized by genotoxic carcinogens that have sufficient evidence of underlying mechanisms suggesting nonlinearity in the response at low doses. These carcinogens are primarily non-DNA reactive substances in which the interaction is with

proteins or protein systems at the chromosome level (e.g., aneugenicity, or clastogenicity). These substances have a weak potency for direct mutagenicity relative to secondary mechanisms. This group also includes those substances in which carcinogenesis is associated with repetitive local tissue damage and cell proliferation (e.g., chloroform, formaldehyde, and vinyl acetate). For example, existing evidence suggests that chloroform is a substance in which carcinogenicity is achieved through cytolethality and regenerative cell proliferation. As such, the EPA considers chloroform to be a probable human carcinogen that is not likely to cause cancer in humans without exposure conditions that cause cell death and regrowth (i.e., a practical threshold exists) [EPA 2001].

- Non-genotoxic carcinogens that act solely through secondary mechanisms (e.g., endocrine-modification, tumor-promotion, immunosuppression, and inflammation). Non-genotoxic carcinogens have widely varied MoA and tissue specificity, but generally act through perturbation of cellular structures that can result in genomic instability. These processes are deterministic and complex; therefore, non-genotoxic carcinogens are generally thought to be best described by sublinear or threshold responses at low doses [Hernández et al. 2009]. For example, TiO_2 is not directly genotoxic; however, a plausible mechanism for carcinogenesis is a nonchemical interaction of inhaled particles with the cells in the lung that is characterized by persistent inflammation and mediated by secondary genotoxic processes. This complex mechanism may explain the sublinear carcinogenic response observed at low doses, as described in the NIOSH risk assessment [NIOSH 2011].

Genotoxic carcinogens consistent with LNT extrapolation are most commonly observed in risk assessment. The other types of carcinogens form a much smaller subset that are either non-genotoxic or their genotoxicity plays a limited role compared to other mechanisms (e.g., cell proliferation); therefore, nonlinear extrapolation may be preferred in risk characterization. For example, Bevan and Harrison [2017] identified a small number of genotoxic substances that have recommended health-based OELs founded on MoA evidence of

practical thresholds (Table 7-2). Similarly, Hernández et al. [2009] estimated that non-genotoxic carcinogens comprise about 12% of substances listed in IARC Groups 1, 2A and 2B.

A logic diagram to aid in choosing an extrapolation approach is illustrated in Figure 7-2. This diagram is a slight modification of concepts adopted by Scientific Committee on Occupational Exposure Limits [Bolt and Huici-Montagud 2008]. As in the risk characterization framework, this diagram is a generalization that may not accurately depict the specific situation encountered in an actual risk assessment.

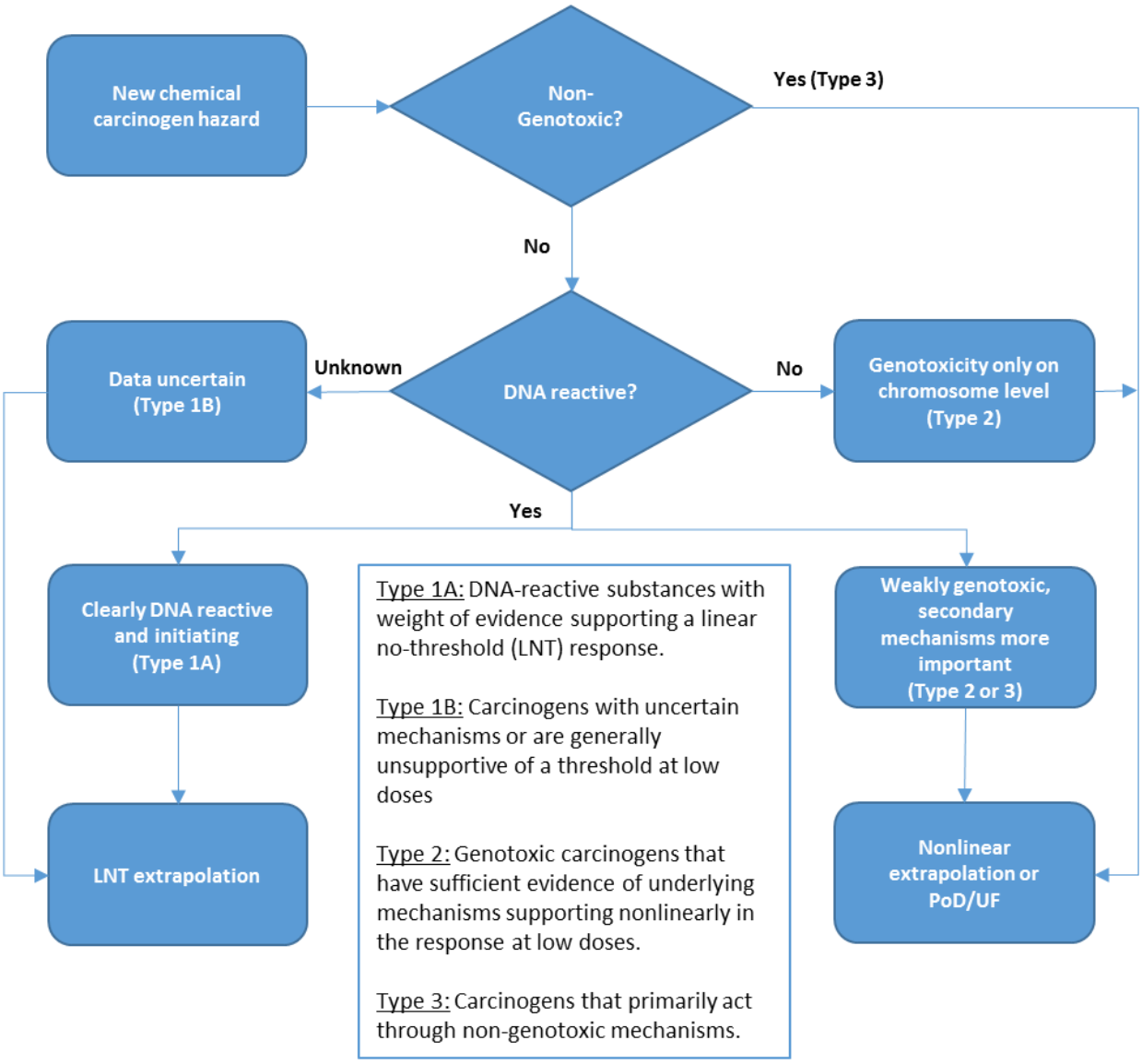

Figure 7-2. Risk characterization of chemical carcinogens using weight of evidence (adapted from Bolt and Huici-Montagaud [2008].

Table 7-2. Examples of genotoxic carcinogens with evidence against LNT response (adapted from Bevan and Harrison [2017]).

Substance	Primary Cancers	Mechanism and OEL Supporting Document[1]
Cadmium (and cadmium compounds)	Lung, kidney, and prostate.	Indirect genotoxic mode of action characterized by different and non-mutually exclusive mechanisms, including: oxidative DNA damage, induction of oxidative stress, inhibition of DNA repair, and deregulation of cell proliferation [SCOEL 2010].
Formaldehyde	Nasopharynx	Genotoxic amplification (at low exposures) by chronic proliferative processes caused by the cytotoxic effects [SCOEL 2008]
Nickel compounds (water soluble)	Lung, nasal cavity, and paranasal sinuses.	Indirect genotoxic mode of action characterized by interference with DNA repair systems and DNA methylation patterns, which lead to clastogenicity and an increased genomic instability [SCOEL 2011]

3. All OEL recommendations made by SCOEL, who advises the European Commission.

Abbreviations: DNA, Deoxyribonucleic acid; OEL, Occupational Exposure Limit, SCOEL, Scientific Committee on Occupational Exposure Limits.

Methods for carcinogenic risk characterization have varied within the risk assessment community. In assessments supporting regulation in the U.S., most have estimated the carcinogenic risk at low doses using LNT, with the exception of the EPA's assessment of chloroform [EPA 2001]. In contrast, some European countries have used a PoD/UF approach to derive OELs for non-genotoxic and some genotoxic carcinogens [Seeley et al. 2001]. NIOSH carcinogenic risk assessments have exclusively used extrapolation by mathematical models to quantify risks at low doses. Of agents assessed, only TiO_2 demonstrated a nonlinear response, which was accounted for in the dose-response modeling. The lack of evidence of a threshold at low doses for any carcinogen does not prove the absence of an exposure level at which cellular homeostasis is maintained and risk is negligible or zero. Similarly, strong evidence of a threshold may still be insufficient to estimate a numerical value for exposure that is considered risk-free, given statistical limitations, inter-individual variability, and other sources of uncertainty. Thus, PoD/UF methods and threshold-based mathematical models have not been used in previous

NIOSH risk assessments of occupational carcinogens, but they may be viable alternatives to linear and nonlinear extrapolation in future assessments.

Exposure effects of carcinogens are generally assumed cumulative and irreversible; therefore, lifetime risks are typically estimated. In most cases, significant background cancer risk is expected in a population (i.e., due to factors other than the occupational exposure); therefore, a competing risk model is preferred. Risks are generally estimated for an array of exposure scenarios. A target risk level is used to recommend a limit on exposure to carcinogens, hereafter known as a Risk Management Limit for Carcinogens (RML-CA). The target risk level for cancer, as stated in the NIOSH Chemical Carcinogen Policy [NIOSH 2017], is one excess cancer case in 10,000 workers exposed in a 45-year working lifetime (i.e., 10^{-4}). In the absence of opposing evidence, exposure-related cancer risk is believed to be persistent; therefore, the excess risk that accrues over a 45-year working lifetime persists afterwards up to the age at death. The age at death used in NIOSH risk assessments has varied over time; however, recent assessments have projected risks to age 85 years, based on the availability of stable population rates.

Recent examples of NIOSH risk assessments include occupational carcinogens such as hexavalent chromium [NIOSH 2013a] and titanium dioxide [NIOSH 2011]. The MoA evidence for these materials differ; supporting low-dose linear response modeling for hexavalent chromium and nonlinear dose-response modeling for titanium dioxide. Risk characterization for these carcinogens follow the framework discussed, except that a target risk level of 10^{-3} was used for both materials according to previous NIOSH policy.

7.2.2 Non-Carcinogens

The NIOSH risk characterization framework generally considers a nonmalignant disease to have a toxicity threshold unless there is evidence to the contrary. Therefore, the PoD/UF approach is preferred for many noncancer endpoints. However, there are two notable exceptions. First, MoA information suggesting the lack of a toxicity threshold (e.g., genotoxicity) and response that is proportional to cumulative dose (i.e., a stochastic

disease process) should prompt an extrapolation approach to risk characterization. Second, extrapolation is preferred in the absence of MoA evidence when the adverse effect is severe and risk appears proportional to dose. Here, the term 'severe' refers to a noncancer adverse effect that resembles cancer (without treatment) with respect to disability, survivorship, progression, and risk persistence. The intent of this exception is to recognize that a chronic illness, if left untreated, may take a course from inception at exposure to eventual resolution that is similar in health consequence relative to cancer, and therefore may merit an analogous approach to risk characterization.

Ideally, sufficient MoA information would be available to support the decision on risk characterization without equivocation; however, in practice this is rarely the case. Instead, the decision usually requires careful consideration of the nature and severity of the adverse effect and its observed association with the agent of interest. For example, occupational pneumoconioses (e.g., silicosis, asbestosis, and coal worker's pneumoconiosis) are severe apical health effects in terms of disability, survivorship, and risk persistence. As such, lung diseases are among the most common noncancer endpoint investigated in NIOSH quantitative risk assessments and most have invoked an extrapolation approach to risk characterization when quantitative dose-response data were available. As an example, NIOSH recently completed its assessment of the risks of obliterative bronchiolitis from diacetyl exposure in the workplace. Diacetyl, and some related chemicals such as 2,3-pentanedione, is used in the manufacture of food flavorings. Obliterative bronchiolitis is a rare, fibroproliferative, incurable, and potentially fatal disease of the small airways of the lung that has been linked to diacetyl exposure in some epidemiologic studies of flavoring workers. However, data from these studies were insufficient for direct quantification of the excess risk of obliterative bronchiolitis (i.e., the apical adverse effect) from diacetyl exposure. Instead, NIOSH assessed data on changes in lung function in exposed workers, which was presumed to be a precursor effect related to obliterative bronchiolitis given that respiratory obstruction is a common presentation of the disease. Pulmonary dysfunction observed among exposed diacetyl workers appeared irreversible. The natural history of obliterative bronchiolitis is highly variable and there is a paucity of information on its pathology related to initiation by toxic exposure; therefore, a practical threshold for diacetyl

toxicity (although perhaps present on an individual basis) is not known. As with cancer risk assessments, airborne concentrations corresponding to a variety of target risk levels assuming a 45-year working lifetime were estimated. In this case, the risk-based REL was derived using a target excess risk of one case in 1000 [NIOSH 2016a].

7.3 Using Risk Assessment as a Basis for RELs or RML-CAs

NIOSH risk assessments provide the quantitative scientific basis for NIOSH recommendations including RELs for noncancer agents and RML-CAs for carcinogens. Although the ultimate decision on a REL or RML-CA is a risk management decision and outside the scope of this report, it is important for risk assessors to understand the issues that contribute to those decisions in order to provide well-supported advice for the risk manager.

Although NIOSH may develop RELs to protect against occupational exposures of any duration, and, in fact the bases of many RELs in the *NIOSH Pocket Guide* are adverse effects due to acute exposures, RELs (and RML-CAs) based on quantitative risk assessment usually focus on longer duration exposures. In other words, NIOSH has typically conducted quantitative risk assessments for serious, chronic adverse effects such as cancer, pneumoconioses, neurological disorders, reproductive outcomes, and other exposure-related cumulative health effects. In part, this is in response to the NIOSH mandate in the Occupational Safety and Health Act [Occupational Safety and Health Act of 1970 1970] to:

> *". . . develop criteria dealing with toxic materials and harmful physical agents and substances which will describe exposure levels that are safe for various periods of employment, including but not limited to the exposure levels at which no employee will suffer impaired health or functional capacities or diminished life expectancy as a result of his work experience." [29 USC 669 (a) (3)]*

And also the codified directive to OSHA to assure:

". . . on the basis of the best available evidence, that no employee will suffer material impairment of health or functional capacity even if such employee has regular exposure to the hazard . . . for the period of his working life." [29 USC 655 (b) (5)]

Because data describing health effects to workers exposed over a working lifetime are extremely rare, risks are estimated using the guidelines described in this document. This includes integration of hazard identification (including exposure assessment) and dose-response analysis. In applying these procedures, some assumptions and defaults are generally necessary to synthesize the information into risk estimates. The following sections describe the targets, defaults, and assumptions that are used in the risk assessment process to provide an integrated picture of risks to workers.

7.3.1 Target Risk Levels

As previously discussed, the foundation of model-based extrapolation in quantitative risk assessment is that any level of exposure to the agent, no matter how small, has an associated health risk. Complete removal of the agent, albeit ideally preferred, is not practical in many industrial settings; therefore, a continuum of exposure-related risk exists that must be managed. This continuum represents a gradient of occupational health risks ranging from high levels that are clearly unacceptable, to extremely low levels in which efforts further reducing exposure result in a negligible reduction in risk. The upper and lower boundaries of this gradient define the *unacceptable* and *broadly acceptable* regions, respectively (Figure 7-3). Between these regions lies the *tolerable* region, which is characterized by a general willingness to tolerate risks in the region given assurances that the risk is managed to an extent that is reasonable and practical [HSE 2001; Tchiehe and Gauthier 2017]. Here the terms "reasonable" and "practical" refers to avoidance of risk mitigation that is disproportionate to the magnitude of the risk involved rather than engineering plausibility [Jones-Lee and Aven 2011]. This is a commonly used risk-reduction principle sometimes referred to as the As Low as Practicable (ALARP) principle. Historically, NIOSH used this principle in its recommendations for exposure to carcinogens. Instead of a numerical REL, the

carcinogen was given a "Ca" designation, which indicated that employers should implement substitution, engineering, work practice, and personal protective equipment strategies to reduce exposures as low as feasible.

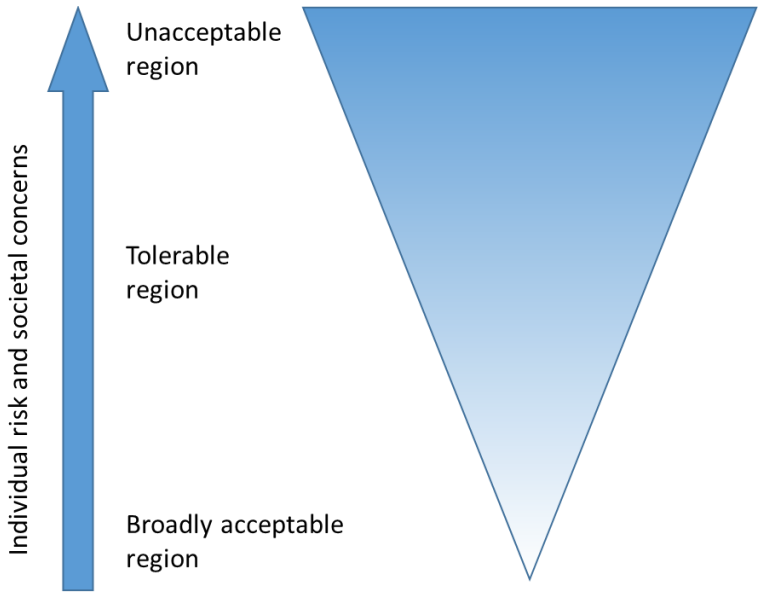

Figure 7-3. Framework for the tolerability of risk (adapted from HSE [2001])

Quantitative risk assessment will often estimate an array of risk levels for risk management purposes. For example, NIOSH uses quantitative risk assessment to estimate chemical exposures corresponding to risks ranging from one excess cancer case in 100 workers (10^{-2}) to 10^{-6} assuming continuous workday exposure over a 45-year working lifetime [NIOSH 2017]. In addition, for a particular hazard, NIOSH typically estimates the airborne concentration at a single level within the tolerable region on which its recommendations are based. This level is called a target risk level. There are multiple methods and principles available for establishing risk acceptance criteria, and the adopted methods and principles will undoubtedly influence the choice of criteria [Vanem 2012]. Thus, risk acceptance (or tolerance) criteria are more likely to be unique to the situation at-hand rather than be pre-defined [Rodrigues et al. 2014; Vanem 2012]. Nevertheless, some examples of target risk levels are available for major hazards [HSE 2001; NIOSH 2017]. For example, the British Health and Safety Executive (HSE)

established the tolerable region for work-related fatality lies between an individual risk 10^{-3} and 10^{-6} per annum [HSE 2001]. The HSE recommends using ALARP to manage risks within the tolerable range but toward the lower bound. As another example, NIOSH has recently established a target risk level for non-threshold carcinogens of one excess case per 10,000 workers continuously exposed over a 45-year working lifetime [NIOSH 2017]. Prior to this policy, assessments have used a target risk of 10^{-3} lifetime catastrophic disease risk from occupational exposures [NIOSH 2013a; NIOSH 2016a; NIOSH 2011]. As in the HSE, NIOSH target risk values have established reasonable starting places for risk mitigation strategies for chemical carcinogen exposure. A simple framework for determining target risk levels based on the relationship between the severity of the adverse effect and its probability of occurrence is shown in Figure 7-4. For example, given a relationship between a catastrophic adverse effect and some hazardous exposure (e.g., leukemia from benzene exposure) the chart reveals a target level for a remote excess working lifetime risk that lies between 10^{-3} and 10^{-5}.

The setting of target risk levels is a fundamental component of risk management; therefore, actions are primarily the responsibility of the decision-makers and not the risk assessor. As such, a detailed discussion on the various principles in play for determining these levels is beyond the scope of this report, although discussion is available in several published reports [Aven 2016; HSE 2001; Rodrigues et al. 2014; Tchiehe and Gauthier 2017; Vanem 2012]. Finally, it should be clear that health risk is but one aspect typically needed to derive a target risk level given that risk tolerance can depend on the combination of individual, societal, economic, and environmental impacts. Although these other factors may be considered by employers in managing risks, NIOSH quantitative risk assessment focuses solely on characterizing health risks; therefore, criteria for establishing NIOSH target risk levels do not consider costs and benefit.

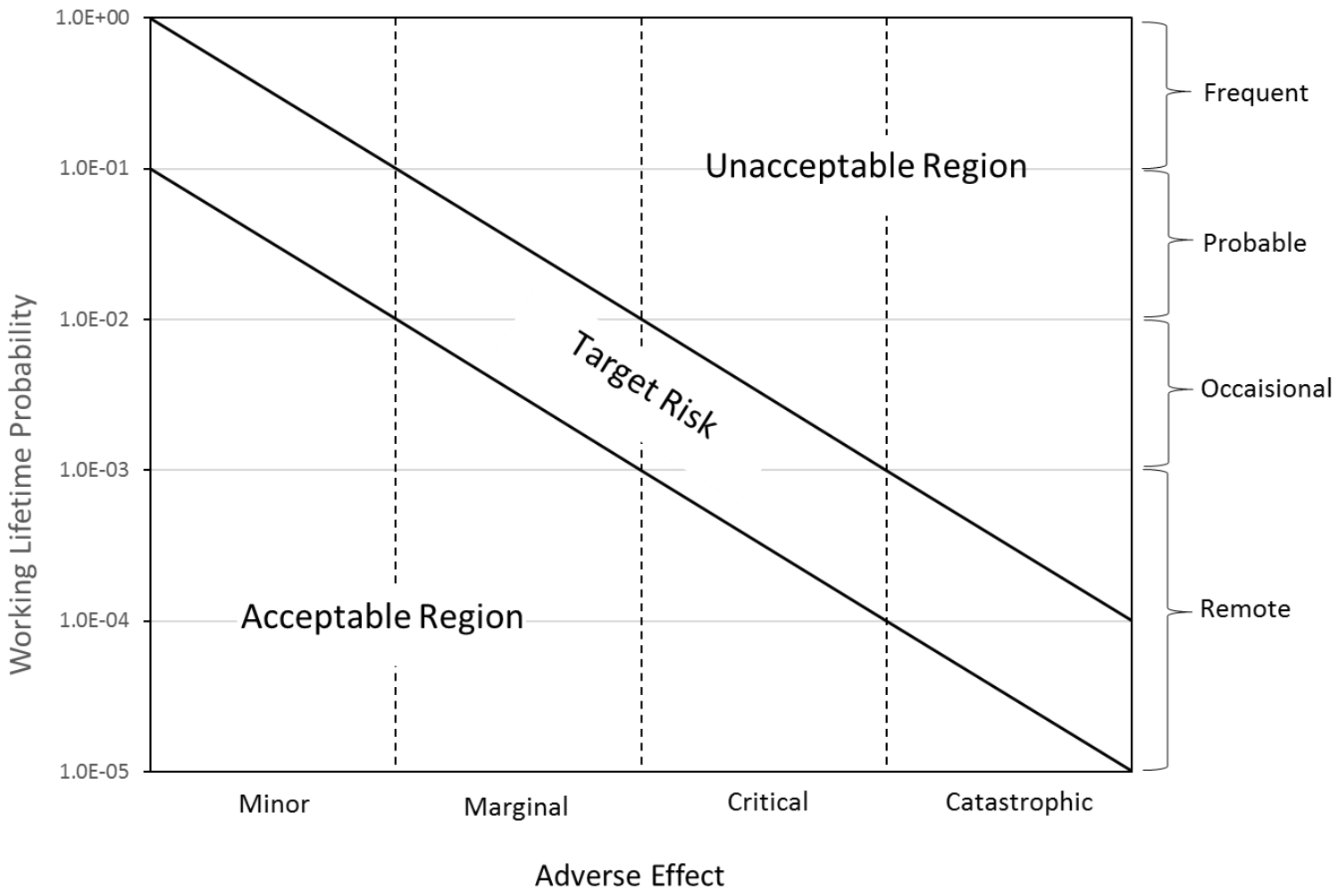

Figure 7-4. Example of target risk based on adverse effect severity and probability of occurrence.

7.3.2 Working Lifetime and Persistent Risk

A common definition of the duration of chronic cumulative exposures is needed in order to maintain comparability among quantitative risk assessments. To that end, NIOSH defines a working lifetime as exposure to a chemical for an 8-hour shift, 5-days a week, 50-weeks a year for 45-years of exposure (i.e., from age 20 to age 65). This represents the maximum amount of exposure anticipated for a worker. However, because the adverse effects of interest are typically chronic effects, the distribution of exposure over a week (or a year) does not usually affect the risk estimate. Therefore, whether a worker is exposed 4-days a week for 10-hours a day or 5-days a week for 8-hours a day does not usually make a difference in the final working lifetime risk estimate or the resulting 8-hour TWA REL or RML-CA. For chronic, cumulative hazards, it is presumed that if exposures were less than working lifetime, risks would be lower than estimated. For risks that do not accumulate across a lifetime (for example, short-duration hazards or adverse effects with an exposure threshold), the 45-year working lifetime is not a relevant measure.

The exposure-related biologic insult may be irreversible for some toxicants and the initiated toxicity pathway may progress throughout life after exposure has ended. For example, significant excess solid cancer risk is still observed in the Japanese atomic-bomb survivors 60 years after their acute exposure to ionizing radiation [Ozasa et al. 2012]. To account for risk persistence, the added or extra risk used in developing the OEL is projected to end of life. Different values for terminal age have been used in risk assessments over the years. For consistency, recent NIOSH risk assessments have assumed a terminal age of 85 years. This value takes into account the limitations on data describing background rates of chronic illnesses at older ages. Examples of NIOSH risk assessments projecting persistent lifetime excess risk are available for: diacetyl exposure and pulmonary impairment [NIOSH 2016a]; nonmalignant respiratory disease and silica exposure [Park et al. 2002]; lung cancer and exposure to asbestos [Stayner et al. 1997], hexavalent chromium [NIOSH 2013a; Park et al. 2004], silica [Rice et al. 2001], and cadmium [Stayner et al. 1992a; Stayner et al. 1992b]

7.3.3 Competing Risks in Projecting Lifetime Risk

Many exposure-related chronic illnesses present very late in life and are not exclusively caused by occupational exposures. Therefore, the risk assessor should account for the competing risks of mortality and background disease rates when projecting lifetime risks of some adverse effects. Among a number of available approaches, competing risk models have most commonly been accomplished using actuarial methods (life-table analysis) that account for age-specific death rates and background disease incidence, under the common assumption that the relative risk, conditional on exposure, is independent of age [Cornfield 1957; Goldberg et al. 1956; NRC 1988; Zdeb 1977]. A lifetable provides a systematic record of the rate at which members of a hypothetical cohort (say 10,000 workers who are 'risk-free' at beginning of working age) withdraw during followup by either death or the illness based on reference mortality and incidence rates that vary by age. This record is used to project risks within age intervals that are conditional on survival to each age interval for intervals specified over the working lifetime period. The summation of the conditional probabilities of diagnoses (or death) in each interval using baseline disease rates provides an estimate of the lifetime risk in the unexposed (R_0). Likewise, summing the conditional probabilities calculated using rates adjusted for exposure provides a corresponding risk measure, R_x, in the exposed. These measures can then be used to determine the excess lifetime risk [e.g., lifetime additive risk = $(R_x - R_0)$ or lifetime extra risk = $(R_x - R_0)/(1 - R_0)$]. These excess lifetime risks are then used to determine the health-basis for the REL or RML-CA.

7.3.4 Application of Uncertainty Factors

For non-cancer adverse effects, NIOSH has conducted risk assessments that assume a threshold or level below which there is no significant risk. For these types of risk assessments, factors that account for sources of uncertainty are applied to estimates from experimental or observational data. The adjusted estimate represents a "safe" level of exposure, which is essentially the PoD (e.g., NOAEL, LOAEL, or BMDL) for the critical effect divided by the series of UFs. For example, NIOSH based its REL for occupational exposure to glycol ether using

a PoD (i.e., NOAEL and LOAEL) and application of uncertainty factors [NIOSH 1991]. NIOSH determined a NOAEL from animal studies of reproductive and developmental toxicity (the most sensitive adverse effect). The NOAEL was adjusted for the animal inhalation rate, body weight, and fraction of the day exposed and converted to an equivalent exposure for humans. Two UFs were applied: a factor of 10 for interspecies variability and another factor of 10 for intraspecies variability (i.e., a divisor of 100). The mg/kg value was then converted to ppm. The resulting concentration was adopted as the REL; however, no attempt to quantify risk was made.

The application of UFs is intended to derive a level of exposure that is protective for workers against all adverse effects related to a substance of concern [Dankovic et al. 2015]. Uncertainty factors are typically applied to non-cancer effects that are assumed to have a threshold of toxicity. The dose-response relationship is usually analyzed using benchmark dose analysis or with consideration of the NOAEL/LOAEL.

During evaluation of a substance for its potential adverse effects, all the available data on that substance must be thoroughly reviewed, including the information on substance analogues in cases where sufficient information is not available for a substance itself. When there are not enough data available to derive the substance-specific or analogue-specific adjustment factors known as chemical specific adjustment factors (CSAFs), uncertainty factors should be applied. CSAFs are used when appropriate chemical-specific data are available; for example suitable quantitative data defining the interspecies differences or human variability in toxicokinetics or toxicodynamics [Meek et al. 2002; WHO 2005]. There will always be some level of uncertainty, especially if the PoD is derived from an experimental animal study instead of in humans. The scientific bases for uncertainty factors have been previously described [Dourson et al. 1996; Dourson and Stara 1983; Naumann and Weideman 1995].

7.3.4.1 Animal-to-Human Uncertainty Factor (UF_A)

UF_A accounts for the uncertainty in extrapolating laboratory animal data to average healthy workers. When human data are used for hazard characterization and deriving OELs, no inter-species extrapolation is necessary and therefore UF_A would be unity. However, when data from laboratory animal studies are used to

assess the risks to workers, the application of UF_A is used to address the differences in sensitivity between animal and humans. By applying UF_A, it is assumed that humans are more sensitive to substances than animals. It may be that humans are equally or less sensitive than animals for specific exposures, but unless it is demonstrated with experimental data, a UF_A should be applied.

The UF_A can be further subdivided into factors that account for toxicokinetic and toxicodynamic differences between species. Toxicokinetic differences arise because of differences in body size and metabolic rate. One way to address the toxicokinetic difference is by using an allometric scaling approach. Allometric scaling is based on the assumption that toxicological effects are driven by physiological parameters and basal metabolic rate, in the absence of cross-species data on chemical-specific metabolism. As discussed previously in Section 6.2.1, an allometric scaling factor, or species-specific dosimetric adjustment factor (DAF), is calculated by:

$$\mathrm{DAF} = (\mathrm{BW_a}/\mathrm{BW_h})^{0.25}$$

for body weights (BW) of the animal (a) and human (h). Thus, different allometric scaling factors would be applied for different species. Allometric scaling is generally applicable in most cases except when the substances cause toxicity only at the portal of entry, such as can occur for the skin, respiratory tract, or gastrointestinal tract (i.e., not dependent on absorption or metabolic rate), and for the acute lethal effects [EPA 2006b]. Allometric adjustments replace the toxicokinetic portion of the UF_A.

Other replacements for the toxicokinetic portion of the UF_A are: 1) a dosimetric adjustment factor that is applied when information is available describing a more proximal (and presumably more relevant) dose; and 2) when compound-related metabolic information is available on humans and animals in the form of physiologically-based pharmacokinetic (PBPK) model, provided that the model is validated and applicable to the specific agent.

Different agencies apply different default uncertainty factors for UF_A. For instance, WHO applies a sub-factor of 4 for toxicokinetics and 2.5 for toxicodynamics [WHO 1994; WHO 2005] while US EPA typically uses equal sub-factors of √10, or approximately 3 [EPA 2002]. The UF_A of 1-10 should be applied based on the available data on toxicokinetics and toxicodynamics. NIOSH risk assessors use the WHO values of 4 for toxicokinetics and 2.5 for toxicodynamics.

7.3.4.2 Interindividual (Human) Variability Uncertainty Factor (UF_H)

UF_H accounts for the variation in sensitivity among the members of worker population. Like UF_A, UF_H is a result of toxicokinetic and toxicodynamic differences between the average and the most sensitive worker population. NIOSH considers the overall UF_H to be a factor of 10 with the sub-factors for toxicodynamics and toxicokinetics to each account for √10 of the variability (often rounded to 3). Chemical-specific toxicokinetic information can be adjusted, for example if a subset of the population has a genetic polymorphism in a metabolic pathway that increases or decreases susceptibility. In addition, for chemicals that cause respiratory irritation upon inhalation, the UF_H is typically adjusted to 3. This is the result of a toxicokinetic sub-factor of 1 (because irritation is typically considered a direct acting effect and not the result of metabolism) and a toxicodynamic sub-factor of 3 for differences in sensitivity among workers.

Some organizations consider a working population to be less heterogeneous than the general population and use a UF_H of less than 10. For example, the European Chemicals Agency recommends a UF_H of 5 to address interindividual variability in workers and a UF_H of 10 for the general population when establishing derived no-effect levels [ECHA 2008]. However, it should be noted that working populations might also include sensitive individuals like asthmatics, pregnant women, older workers, and others who may be more susceptible. Therefore, NIOSH typically uses a factor of 10 for the overall UF_H unless chemical-specific information is available to the contrary. Whenever a factor other than 10 is used, the rationale must be fully explained.

7.3.4.3 LOAEL-to-NOAEL Uncertainty Factor (UF_L)

UF_L is intended to account for the uncertainty in extrapolating from LOAELs to NOAELs. When the starting point for the exposure level of concern calculation is a LOAEL, an additional UF_L between 3 (for minimal toxicological severity, such as fatty liver) and 10 (for a severe effects, such as hepatic necrosis) should be applied to estimate a dose where no adverse effect would occur [Dourson et al. 1996; Dourson and Stara 1983; Naumann and Weideman 1995].

When the starting point is a NOAEL or a BMDL, no additional UF_L is required and the UF_L value should be unity. However, an additional UF may be applied when the PoD is a NOAEL in certain cases, such as with: 1) a poor quality study where very few animals and doses are used; or 2) very severe effects occur at the slightly higher next dose, which is the LOAEL. The justification for this UF should be clearly explained.

7.3.4.4 Shorter-Term-to-Longer-Term Uncertainty Factor (UF_S)

Ideally, data from long-term (chronic) toxicology studies are available to estimate lifetime excess risks of chronic disease in humans. In practice, however, data may be limited to shorter than lifetime bioassays (e.g., two years for mice and rats). In these cases, a Shorter-Term-to-Longer-Term Uncertainty Factor (UF_S) may be necessary to adjust for differences in duration of exposure. The UF_S (also known as a subchronic to chronic factor) assumes that an effect observed at subchronic exposure levels will be seen at lower levels of chronic exposure[Dourson et al. 1996]. An exception would be evidence that risk (i.e., the incidence or severity) is unrelated to exposure duration or is fully characterized by the shorter-term study. For example, some effects like sensory irritation of the skin/respiratory tract and effects caused by a reactive metabolite may not increase with duration. In these circumstances, additional correction using $UF_S > 1$ may be unwarranted [Dankovic et al. 2015].

Typically, a factor from 3 to 10 is applied for sub-chronic to chronic extrapolation. This factor can be modified by chemical-specific experimental data. For study durations longer than 3 months, but less than a full lifetime, lower UFs could be applied. Extrapolation to chronic exposure from sub-acute (28-days) or acute (<24

hours) studies is not generally recommended, although prediction of chronic effects from short-term studies is an active area of research.

7.3.4.5 Database Inadequacy (incomplete data) Uncertainty Factor (UF_D)

UF_D is intended to account for the inability of the available toxicity database to address all likely adverse effects in humans. Evaluation of the total toxicological database should address whether the derived exposure level of concern is protective enough against all potential adverse outcomes for the substance.

Based on the available information for the substance in evaluation, additional studies may be warranted and in those cases, an additional UF_D (1-10) may be applied. If the data are available from only one species, or if developmental and reproductive toxicity studies are not available, then an $UF_D > 1$ may be applied to account for potential toxicity in the unstudied toxic endpoints. In addition, if the preliminary data indicate some evidence for neurotoxic/immunotoxic effects, a detailed study to evaluate neurotoxicity and immunotoxicity may be needed to assess risk adequately. An additional UF_D in this case would represent the concern for the endpoint and the uncertainty in existing data. NIOSH has not typically applied the uncertainty factor for database inadequacy in its risk assessments to date.

7.3.4.6 The Composite Uncertainty Factor (UF_C)

Once the individual uncertainty factors are assigned, they are multiplied to yield the overall composite factor (UF_C). Typically, when an UF of $\sqrt{10}$ is used, convention says that $\sqrt{10}$ is approximately 3.16, which is rounded to 3. When multiplying UFs of 3 and 3, the result used should be 10 and not 9. If a risk estimate incorporates UFs of all five types, the database should be carefully evaluated for its sufficiency to derive an exposure level of concern with enough confidence. UFs may not be independent; therefore, multiplying several factors can result in over-conservatism. As a rule of thumb, a maximum UF_C should be limited to no more than 3,000 in the derivation of an exposure level of concern.

7.4 Special Considerations When Developing a Short-Term Exposure Limit (STEL)

In some cases, the available health effects data may elicit a concern for short-term exposure limits (STELs). For example, if peak exposures increase CNS symptoms, asthma attacks, or other acute-onset health effects, the data may be informative for developing short-term exposure limits. Evidence that peak exposures cause specific health effects should be considered in evaluating the need for a STEL. For example, if the 8-hour TWA REL is 1 ppm and peak exposures at 25 ppm cause nasal and eye irritation, it is prudent to consider that data in developing a STEL. Without a STEL, in this case, it is theoretically possible to be exposed to 32 ppm for 15-minutes with zero exposure for the remainder of the day and still be under the REL. If there is good dose-response data at levels of concern for acute adverse effects, it is possible to conduct a quantitative risk assessment to support a numerical STEL. More often, quantitative data on the effects of peak exposures are not available. In those cases, a concern for acute exposures may be supported by data or a plausible concern may exist based on mode of action, analogous chemicals, or other considerations. A simple way of testing this is to multiply the TWA REL or RML-CA by 32. If the resulting exposure would elicit concern for short-term effects, a STEL is needed. For example, if a REL based on quantitative risk assessment is calculated to be 10 ppm, 10 ppm X 32 = 320 ppm. If there is evidence or a concern for short-term effects after exposure to 320 ppm for 15-minutes, a STEL is needed. If there is quantitative data showing health effects from short-term exposures (for example, respiratory irritation after exposure to 200 ppm for 10 minutes), that should be used to inform or establish the STEL. Alternatively, if there is a concern for short-term exposures, but no data for quantitative assessment, a STEL may be determined based on industrial hygiene practice (for example, STEL = 5 X 8-hour TWA REL). This provides a maximum peak exposure that serves to both reduce peak exposures and to reduce overall TWA exposure.

7.5 More on NIOSH RELs and RML-CAs

Adequate control of causative agents of occupational illness and disability is fundamental to the health and safety of the American workforce. To that end, NIOSH synthesizes relevant information on occupational hazards to formulate hazard mitigation strategies, including publication of RELs and RML-CAs as discussed in

this report. Preferably, these OELs stem from a quantitative assessment of the occupational risk associated with the hazard, although analytic and technical feasibility are also considered. For example, the NIOSH REL for occupational diacetyl inhalation exposure is 5 ppb and was based primarily on the findings from a quantitative risk assessment using epidemiologic data [NIOSH 2016a]. From these data, NIOSH predicted that the risk of significant lung impairment was in the range of a target risk of 10^{-3} excess lifetime risk for workers exposed to 5 ppb over a 45-year working lifetime.

As we have shown, the occupational risk to workers from exposure to a hazard is best characterized by a probability distribution rather than a point estimate, given unavoidable variability in exposure and response. In addition, NIOSH typically integrated both risk science and health policy, such as the feasibility of analytic methods and the achievability of control technology, into deriving RELs and RML-CAs, which introduces further uncertainty (Figure 7-5). There is often considerable uncertainty and generous professional judgment in OEL development. Therefore, the REL and RML-CA are not intended as a 'bright line' between safe and unsafe exposure for all workers. Instead, these OEL are better described as levels on the dose-risk continuum prompting evaluation and control (i.e., risk management), given the extent of current scientific knowledge.

NIOSH recommends that decision-makers consider NIOSH recommendations on exposure levels, including its basis, and magnitude of occupational risk and attendant uncertainties, as well as competing risks from substitution or hazard controls in adopting a risk management strategy. Fundamental to this strategy, employers should consider continued improvement in controls until exposure levels below the REL or RML-CA be confidently attained. However, actual exposure situations can vary widely; therefore, information used by risk managers in assessing reasonableness and/or practicality of implementing hazard control strategies can differ by situation. Thus, underlying any risk management strategy is an assurance that risk mitigation efforts are not disproportionate to the magnitude of the risk involved. In the example case of diacetyl, it is important to understand that the REL protects workers from a chronic hazard from lifetime exposure in which excess risk is

kept relatively low compared to that faced in everyday life. At these low risk levels, situations can arise in which further reduction can be extremely difficult if not impractical to achieve.

The REL and RML-CA are recommendations based on the best available science; neither are enforceable or binding limits on exposure. Application is solely at the discretion of the end-user. However, NIOSH RELs and RML-CAs are derived to be protective for most workers and in most occupational settings; therefore, in the absence of situational risk management, the recommended level can serve as an appropriate control level. As such, many industries have voluntarily adopted NIOSH RELs and RML-CAs as a part of their risk management practices. Moreover, some regulatory agencies have incorporated NIOSH recommendations into enforceable limits on exposure. That said, NIOSH RELs and RML-CAs are determined exclusively for worker protection; therefore, these recommendations are not directly applicable to the protection of consumers or members of the public. However, the science behind these recommendations is likely to be useful for deriving similar public health standards.

In forming its recommendations on exposure, NIOSH identifies uses and manufacturing operations for the given hazard to recognize effective control strategies and describe engineering achievability. NIOSH may also indicate when the nature of job activities presents a challenge to meeting the REL or RML-CA. Although routine attainment of exposures below the NIOSH recommended limits may not occur in all work settings initially, it does represent a reasonable objective that employers can work to achieve through modification of work or the introduction or improvement of engineering controls. In this way, the REL and RML-CA encourages technology improvements to limit exposures. For some operations, additional protective measures, such as administrative controls and personal protective equipment may be necessary to achieve risk mitigation goals.

Finally, NIOSH RELs and RML-CAs facilitate hazard communication, as NIOSH urges employers to disseminate related information to workers and customers, and encourages manufacturers to convey this information to downstream users. NIOSH also requests that professional and trade associations and labor organizations inform their members about workplace hazards. This communication should include a description

of NIOSH recommendations and the risk associated with exposures at control levels. In communicating these risks, it may be helpful to include context, such as risks from other sources encountered in the human experience.

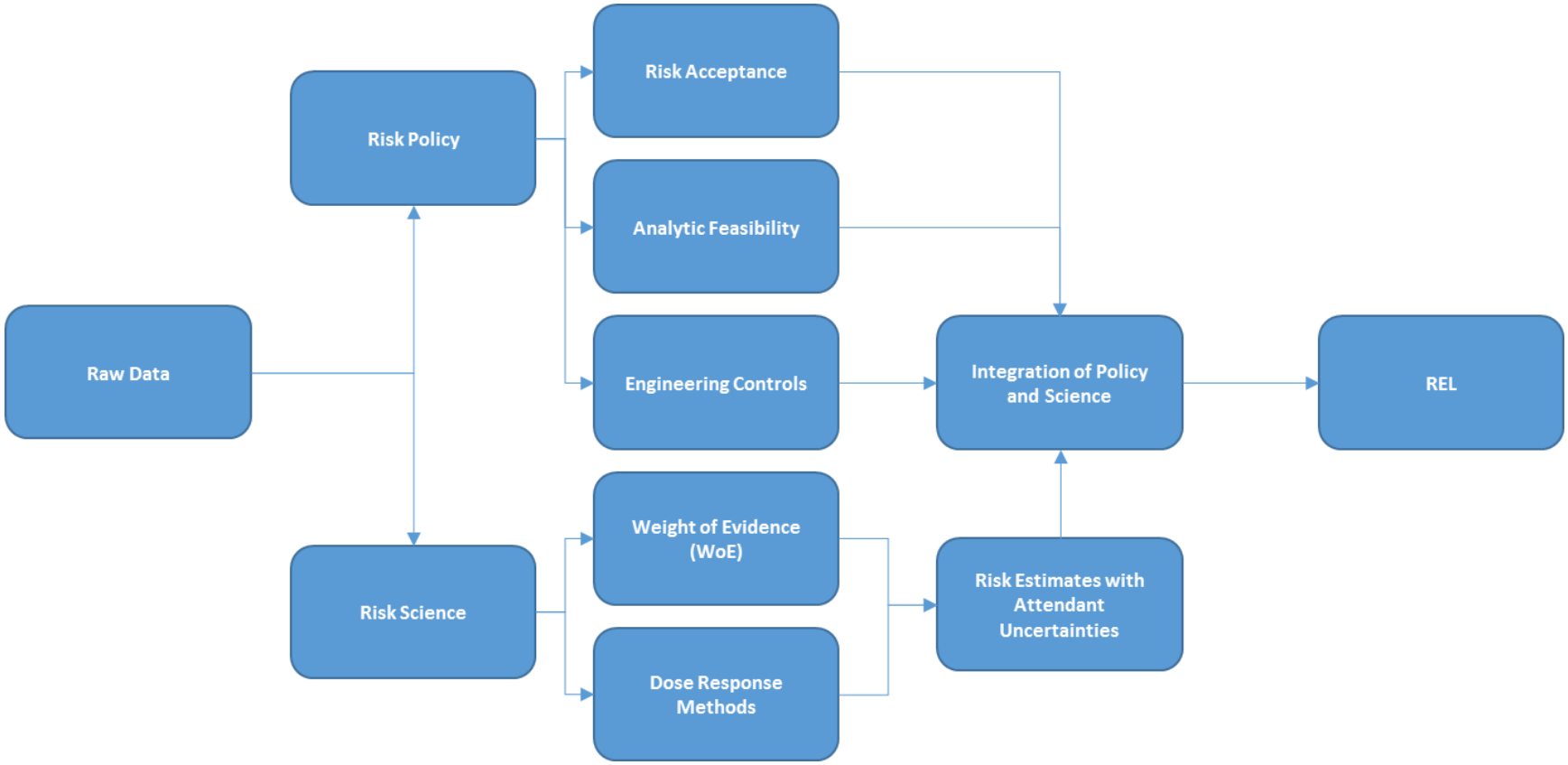

Figure 7-5. Sources of uncertainty in the derivation of the Recommended Exposure Level (REL). Adapted from Waters et al. [2015].

8.0 CONCLUSIONS

Pulling all the pieces of a risk assessment together requires careful attention to the purpose of the risk assessment. It is important to interrogate key assumptions and provide transparency for both the main analysis and analyses of alternative modeling strategies and defaults. If innovative or unusual modeling or analytical strategies are used, it is critical that these be presented in a clear manner, drawing the reader's attention to departures from past practice. Ideally, novel or unusual methods would be published in the peer reviewed scientific literature before they are used in a NIOSH numbered publication, although this may not always be possible. One set of questions from the risk assessment plan at the beginning of this document deserves particular attention.

- How will risks be expressed and, if in quantitative analysis, what are the target risk levels used? What is the support for those decisions and are there reasonable alternatives? If yes, how would using those influence the risk assessment?

Among completed NIOSH quantitative risk assessments, most have examined the risk from occupational carcinogens. The risk assessment assumptions regarding cumulative exposure and chronic expression of cancer are well supported and have numerous NIOSH precedents. Non-cancer risk assessments, on the other hand, have diverse health impacts and exposure profiles. These require thoughtful discussion of the assumptions in the dose-response analysis. Is a cumulative dose appropriate? Is there clear evidence or expectation of irreversibility of health effects? While harmonization of cancer and non-cancer risk assessments is a desirable goal, it is critical to keep in mind the differences in mode of action and natural history of disease when using the risk assessment information to derive a REL for non-carcinogens.

In this era of shrinking data on individual chemicals and increasing complexity of statistical methods, NIOSH risk assessors must take care that the resulting risk assessments are a rational and clear portrayal of the available data. When there is uncertainty in the risk assessment, NIOSH risk assessors may err on the side of protecting workers; however, the risk assessor must keep in mind that decisions based on an overestimation of

risk can actually be detrimental to occupational health. Thus, the potential for bias in either direction must be clearly identified to inform risk managers. Above all, the risk assessment must balance the protection of workers with the strength of the data to ensure that all of the NIOSH recommendations are well supported by sound science.

Risk assessment science is continuously evolving given a wide-array of uncharacterized hazards and a large community of risk assessment practitioners in academia, industry, and government. NIOSH risk assessments, although purposed for worker protection, can have relevance outside of the workplace. Similarly, activities intended for characterizing risks in other populations can also inform on worker risks. Given these conditions, overlapping activities are anticipated among multiple agencies or risk assessment programs. For example, a recent review by the United States Government Accountability Office (GAO) examined overlap among federal and state chemical toxicity assessment programs [GAO 2014]. The GAO findings suggest there was ample room for improvement in risk assessment through shared resources. Thus, it is clear that routine exchange between NIOSH and the risk assessment community, both home and abroad, is paramount to ensuring best practices are followed, including improved efficiency and effectiveness by reducing duplication of effort. For these reasons, NIOSH maintains active collaborations within the risk assessment community and coordinates its risk assessment activities with stakeholders and the public.

Methods currently under development provide additional, powerful tools to assess risks to workers based on very limited data. Validation of these approaches is a critical need. Occupational risk assessment needs to move forward and embrace new methodologies, but with caution and deliberate evaluation of new techniques and approaches. Examples of methods currently under development are contained in Appendix C. As these methods are validated and demonstrate their utility for occupational risk assessment, it is hoped that they will be adopted by NIOSH risk assessors and will be moved to the main body of this document.

9.0 REFERENCES

40 CFR 798 [07/01/1998]. Code of Federal Regulations.Washington, DC: U.S. Government Printing Office, Office of the Federal Register.

ACGIH [2015]. Threshold limit values for chemical substances and physical agents and biological exposure indices. Cincinnati, OH: American Conference of Governmental Industrial Hygienists, pp.

Ahlborg Jr GA [1990]. "Validity of exposure data obtained by questionnaire. Two examples from occupational reproductive studies." Scand J Work Environ Health *16*(4): 284-288.

Akaike H [1974]. "A new look at the statistical model identification." IEEE Trans Automatic Control *19*: 716–723.

Albert RE, Train RE, Anderson E [1977]. "Rationale developed by the Environmental Protection Agency for the assessment of carcinogenic risks." J Natl Cancer Inst *58*(5): 1537-1541.

Aldridge WN [1986]. "The biological basis and measurement of thresholds." Annu Rev Pharmacol Toxicol *VOL.* *26*: 39-58.

Allen BC, Crump KS, Shipp AM [1988]. "Correlation between carcinogenic potency of chemicals in animals and humans." Risk Anal *8*(4): 531-544.

Allison PD [2010]. Survival analysis using SAS: a practical guide. Vol.: Sas Institute, pp.

Anderson D, Sorsa M, Waters MD [1994]. "The parallelogram approach in studies of genotoxic effects." Mutat Res *313*(2-3): 101-115.

Anjilvel S, Asgharian B [1995]. "A multiple-path model of particle deposition in the rat lung." Fundam Appl Toxicol *28*(1): 41-50.

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, Serrrano JA, Tietge JE, Villeneuve DL [2010]. "Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment." Environ Toxicol Chem *29*(3): 730-741.

ARA [2009]. Multiple-path particle dosimetry (MPPD 2.1) model. Raleigh, NC: Applied Research Associates, Inc., pp.

ARA [2015]. Multiple-Path Particle Dosimetry (MPPD) model, v. 3.04. Raleigh, NC: Applied Research Associates, Inc., pp.

Arrighi HM, Hertz-Picciotto I [1994]. "The evolving concept of the healthy worker survivor effect." Epidemiology *5*(2): 189-196.

Asgharian B, Hofmann W, Bergmann R [2001]. "Particle deposition in a multiple-path model of the human lung." Aerosol Sci Technol *34*(4): 332-339.

Asgharian B, Price OT, Yurteri CU, Dickens C, McAughey J [2014]. "Component-specific, cigarette particle deposition modeling in the human respiratory tract." Inhal Toxicol *26*(1): 36-47.

Atkinson TM, Ryan SJ, Bennett AV, Stover AM, Saracino RM, Rogak LJ, Jewell ST, Matsoukas K, Li Y, Basch E [2016]. "The association between clinician-based common terminology criteria for adverse events (CTCAE) and patient-reported outcomes (PRO): a systematic review." Support Care Cancer *24*(8): 3669-3676.

Aven T [2016]. "Risk assessment and risk management: Review of recent advances on their foundation." Eur J Oper Res *253*(1): 1-13.

Baan RA [2007]. "Carcinogenic hazards from inhaled carbon black, titanium dioxide, and talc not containing asbestos or asbestiform fibers: recent evaluations by an IARC Monographs Working Group." Inhal Toxicol *19 Suppl 1*: 213-228.

Baetcke K, Hard G, Rodgers I, McGaughy R, Tahan L [1991]. Alpha 2u-globulin: Association with chemically induced renal toxicity and neoplasia in the male rat. US Environmental Protection Agency, Risk Assessment Forum, Washington, DC.

Bailer AJ, Stayner LT, Smith RJ, Kuempel ED, Prince MM [1997]. "Estimating benchmark concentrations and other noncancer endpoints in epidemiology studies." Risk Anal *17*(6): 771-780.

Baird SJ, Cohen JT, Graham JD, Shlyakhter AI, Evans JS [1996]. "Noncancer risk assessment: A probabilistic alternative to current practice." Hum Ecol Risk Assess *2*(1): 79-102.

Barnes DG, Dourson M [1988]. "Reference dose (RfD): description and use in health risk assessments." Regul Toxicol Pharmacol *8*(4): 471-486.

Barrow C [1989]. Generation and characterization of gases and vapors. Concepts in Inhalation Toxicology. R. F. Henderson, R. McClellam and R. Henderson. New York, NY

Hemisphere Publishing Corp**:** 63-84.

Barton HA, Andersen ME, Clewell Iii HJ [1998]. "Harmonization: Developing consistent guidelines for applying mode of action and dosimetry information to cancer and noncancer risk assessment." Hum Ecol Risk Assess *4*(1): 75-115.

Baumgarten M, Siemiatycki J, Gibbs GW [1983]. "Validity of work histories obtained by interview for epidemiologic purposes." Am J Epidemiol *118*(4): 583-591.

Beaumont JJ, Steenland K, Minton A, Meyer S [1989]. "A computer program for incidence density sampling of controls in case-control studies nested within occupational cohort studies." Am J Epidemiol *129*(1): 212-219.

Bellmann B, Muhle H, Creutzenberg O, Dasenbrock C, Kilpper R, MacKenzie JC, Morrow P, Mermelstein R [1991]. "Lung clearance and retention of toner, utilizing a tracer technique, during chronic inhalation exposure in rats." Fundam Appl Toxicol *17*(2): 300-313.

Benke G, Sim M, Forbes A, Salzberg M [1997]. "Retrospective assessment of occupational exposure to chemicals in community-based studies: Validity and repeatability of industrial hygiene panel ratings." Int J Epidemiol *26*(3): 635-642.

Benke G, Sim M, Fritschi L, Aldred G, Forbes A, Kauppinen T [2001]. "Comparison of occupational exposure using three different methods: Hygiene panel, job, exposure matrix (JEM), and self reports." Appl Occup Environ Hyg *16*(1): 84-91.

Bergmann MM, Calle EE, Mervis CA, Miracle-McMahill HL, Thun MJ, Heath CW [1998]. "Validity of self-reported cancers in a prospective cohort study in comparison with data from state cancer registries." Am J Epidemiol *147*(6): 556-562.

Bertke S, Hein M, Schubauer-Berigan M, Deddens J [2013]. "A Simulation Study of Relative Efficiency and Bias in the Nested Case-Control Study Design." Epidemiol Method *2*(1): 85-93.

Bevan RJ, Harrison PTC [2017]. "Threshold and non-threshold chemical carcinogens: A survey of the present regulatory landscape." Regul Toxicol Pharmacol.

Bingham E, Cohrssen B, Powell CH [2001]. Patty's toxicology. Volume 1: introduction to the field of toxicology. Vol. 1: John Wiley and Sons, pp.

Blair A, Stewart PA [1992]. "Do quantitative exposure assessments improve risk estimates in occupational studies of cancer?" Am J Ind Med *21*(1): 53-63.

Bogaards JJ, Bertrand M, Jackson P, Oudshoorn MJ, Weaver RJ, van Bladeren PJ, Walther B [2000]. "Determining the best animal model for human cytochrome P450 activities: a comparison of mouse, rat, rabbit, dog, micropig, monkey and man." Xenobiotica *30*(12): 1131-1152.

Bogdanffy MS, Daston G, Faustman EM, Kimmel CA, Kimmel GL, Seed J, Vu V [2001]. "Harmonization of cancer and noncancer risk assessment: proceedings of a consensus-building workshop." Toxicol Sci *61*(1): 18-31.

Bogdanffy MS, Jarabek AM [1995]. "Understanding mechanisms of inhaled toxicants: implications for replacing default factors with chemical-specific data." Toxicol Lett *82-83*: 919-932.

Bolt HM, Huici-Montagud A [2008]. "Strategy of the scientific committee on occupational exposure limits (SCOEL) in the derivation of occupational exposure limits for carcinogens and mutagens." Arch Toxicol *82*(1): 61-64.

Bolton RE, Vincent JH, Jones AD, Addison J, Beckett ST [1983]. "An overload hypothesis for pulmonary clearance of UICC amosite fibres inhaled by rats." Br J Ind Med *40*(3): 264-272.

Bond GG, Bodner KM, Sobel W, Shellenberger RJ, Flores GH [1988]. "Validation of work histories obtained from interviews." Am J Epidemiol *128*(2): 343-351.

Boobis AR, Cohen SM, Dellarco V, McGregor D, Meek ME, Vickers C, Willcocks D, Farland W [2006]. "IPCS framework for analyzing the relevance of a cancer mode of action for humans." Crit Rev Toxicol *36*(10): 781-792.

Boobis AR, Doe JE, Heinrich-Hirsch B, Meek ME, Munn S, Ruchirawat M, Schlatter J, Seed J, Vickers C [2008]. "IPCS framework for analyzing the relevance of a noncancer mode of action for humans." Crit Rev Toxicol *38*(2): 87-96.

Borm P, Cassee FR, Oberdörster G [2015]. "Lung particle overload: old school –new insights?" Particle and Fibre Toxicology *12*(1): 10.

Bourbonnais R, Meyer F, Theriault G [1988]. "Validity of self reported work history." Br J Ind Med *45*(1): 29-32.

Boxenbaum H [1982]. "Interspecies scaling, allometry, physiological time, and the ground plan of pharmacokinetics." J Pharmacokinet Biopharm *10*(2): 201-227.

Brenner DJ, Little JB, Sachs RK [2001]. "The bystander effect in radiation oncogenesis: II. A quantitative model." Radiation Research *155*(3): 402-408.

Breslow N [1990]. "Biostatistics and Bayes." Stat Sci *5*(3): 269-284.

Breslow NE, Day NE [1980]. "Statistical methods in cancer research. Volume I - The analysis of case-control studies." IARC Sci Publ(32): 5-338.

Breslow NE, Day NE [1987]. "Statistical methods in cancer research. Volume II--The design and analysis of cohort studies." IARC Sci Publ(82): 1-406.

Breslow NE, Lubin JH, Marek P, Langholz B [1983]. "Multiplicative models and cohort analysis." J Am Stat Assoc *78*(381): 1-12.

Brown CC, Koziol JA [1983]. "Statistical aspects of the estimation of human risk from suspected environmental carcinogens." SIAM Review *25*(2): 151-181.

Brown JS, Gordon T, Price O, Asgharian B [2013]. "Thoracic and respirable particle definitions for human health risk assessment." Particle and Fibre Toxicology *10*: 12-12.

Brown RP, Delp MD, Lindstedt SL, Rhomberg LR, Beliles RP [1997]. "Physiological parameter values for physiologically based pharmacokinetic models." Toxicol Ind Health *13*(4): 407-484.

Buckland ST, Burnham KP, Augustin NH [1997]. "Model selection: an integral part of inference." Biometrics *53*: 603–618.

Buckley JP, Keil AP, McGrath LJ, Edwards JK [2015]. "Evolving methods for inference in the presence of healthy worker survivor bias." Epidemiology *26*(2): 204-212.

Budtz-Jørgensen E, Keiding N, Grandjean P [2001]. "Benchmark dose calculation from epidemiological data." Biometrics *57*(3): 698-706.

Cameron HM, McGoogan E [1981]. "A prospective study of 1152 hospital autopsies: I. Inaccuracies in death certification." J Pathol *133*(4): 273-283.

Chamblee RF, Evans MC [1982]. "New dimensions in cause of death statistics." American journal of public health *72*(11): 1265-1270.

Checkoway H, Pearce N, Kriebel D [2004]. Research methods in occupational epidemiology. Vol. New York: Oxford University Press, pp.

Checkoway H, Rice CH [1992]. "Time-weighted averages, peaks, and other indices of exposure in occupational epidemiology." Am J Ind Med *21*(1): 25-33.

Chen B, John W [2001]. Instrument calibration. Aerosol measurement. P. A. Baron and K. Willeke. New York, NY, John Wiley & Sons**:** 627-666.

Chen CC, Chen JJ [2014]. "Benchmark dose calculation for ordered categorical responses." Risk Anal *34*(8): 1435-1447.

Chen JJ, Moon H, Kodell RL [2007]. "A probabilistic framework for non-cancer risk assessment." Regul Toxicol Pharmacol *48*(1): 45-50.

Chen MJ [2001]. Development of beryllium exposure matrices for workers in a former beryllium manufacturing plant, University of Cincinnati.

Cherrie JW, Brosseau LM, Hay A, Donaldson K [2013]. "Low-toxicity dusts: current exposure guidelines are not sufficiently protective." Annals of occupational hygiene *57*(6): 685-691.

Chiu WA, Slob W [2015]. "A unified probabilistic framework for dose–response assessment of human health effects." Environ Health Perspect *123*(12): 1241-1254.

Chovil A, Sutherland RB, Halliday M [1981]. "Respiratory cancer in a cohort of nickel sinter plant workers." Br J Ind Med *38*(4): 327-333.

Chu H, Wang Z, Cole SR, Greenland S [2006]. "Sensitivity Analysis of Misclassification: A Graphical and a Bayesian Approach." Ann Epidemiol *16*(11): 834-841.

Cockburn M, Swetter SM, Peng D, Keegan THM, Deapen D, Clarke CA [2008]. "Melanoma underreporting: Why does it happen, how big is the problem, and how do we fix it?" J Am Acad Dermatol *59*(6): 1081-1085.

Cohen SM, Klaunig J, Meek ME, Hill RN, Pastoor T, Lehman-McKeeman L, Bucher J, Longfellow DG, Seed J, Dellarco V, Fenner-Crisp P, Patton D [2004]. "Evaluating the human relevance of chemically induced animal tumors." Toxicol Sci *78*(2): 181-186.

Comber MHI, Walker JD, Watts C, Hermens J [2003]. "Quantitative structure-activity relationships for predicting potential ecological hazard of organic chemicals for use in regulatory risk assessments." Environ Toxicol Chem *22*(8): 1822-1828.

Cornfield J [1957]. "Estimation of the probability of developing a disease in the presence of competing risks." American Journal of Public Health and the Nations Health *47*(5): 601-607.

Couch JR, Petersen M, Rice C, Schubauer-Berigan MK [2011]. "Development of retrospective quantitative and qualitative job-exposure matrices for exposures at a beryllium processing facility." Occupational and Environmental Medicine *68*(5): 361-365.

Coughlin SS, Chiazze L, Jr. [1990]. "Job-exposure matrices in epidemiologic research and medical surveillance." Occup Med *5*(3): 633-646.

Cox DR [1972]. "Regression Models and Life-Tables." J R Stat Soc Series B Stat Methodol *34*(2): 187-220.

Crissman JW, Goodman DG, Hildebrandt PK, Maronpot RR, Prater DA, Riley JH, Seaman WJ, Thake DC [2004]. "Best practices guideline: toxicologic histopathology." Toxicol Pathol *32*(1): 126-131.

Crump K, Allen B, Shipp A [1989]. "Choice of dose measure for extrapolating carcinogenic risk from animals to humans: an empirical investigation of 23 chemicals." Health Phys *57 Suppl 1*: 387-393.

Crump KS [1995]. "Calculation of Benchmark Doses from Continuous Data." Risk Anal *15*(1): 79-89.

Crump KS [1984]. "A new method for determining allowable daily intakes." Fundam Appl Toxicol *4*(5): 854–871.

Crump KS [1998]. "On summarizing group exposures in risk assessment: is an arithmetic mean or a geometric mean more appropriate?" Risk Anal *18*(3): 293-297.

Crump KS [2011]. "Use of threshold and mode of action in risk assessment." Crit Rev Toxicol *41*(8): 637-650.

Crump KS, Allen BC [1985]. "Methods for quantitative risk assessment using occupational studies." American Statistician *39*(4): 442-450.

Crump KS, Clewell HJ, Andersen ME [1997]. "Cancer and non-cancer risk assessment should be harmonized." Hum Ecol Risk Assess *3*(4): 495-499.

Dankovic D, Kuempel E, Wheeler M [2007]. "An approach to risk assessment for TiO2." Inhal Toxicol *19 Suppl 1*: 205-212.

Dankovic DA, Naumann BD, Maier A, Dourson ML, Levy LS [2015]. "The Scientific Basis of Uncertainty Factors Used in Setting Occupational Exposure Limits." J Occup Environ Hyg *12 Suppl 1*: S55-68.

Dankovic DA, Smith RJ, Stayner LT, Bailer AJ [1993]. "Time-to-tumour risk assessment for 1,3-butadiene based on exposure of mice to low doses by inhalation." IARC Sci Publ(127): 335-344.

Davies B, Morris T [1993]. "Physiological parameters in laboratory animals and humans." Pharm Res *10*(7): 1093-1095.

Davis JA, Gift JS, Zhao QJ [2011]. "Introduction to benchmark dose methods and U.S. EPA's benchmark dose software (BMDS) version 2.1.1." Toxicol Appl Pharmacol *254*(2): 181-191.

Delgado-Rodríguez M, Llorca J [2004]. "Bias." J Epidemiol Community Health *58*(8): 635-641.

Dement JM, Harris RL, Jr., Symons MJ, Shy CM [1983]. "Exposures and mortality among chrysotile asbestos workers. Part I: exposure estimates." Am J Ind Med *4*(3): 399-419.

Desai MM, Bruce ML, Desai RA, Druss BG [2001]. "Validity of self-reported cancer history: A comparison of health interview data and cancer registry records." Am J Epidemiol *153*(3): 299-306.

Douglas AN, Robertson A, Chapman JS, Ruckley VA [1986]. "Dust exposure, dust recovered from the lung, and associated pathology in a group of British coalminers." Br J Ind Med *43*(12): 795-801.

Dourson ML, Felter SP, Robinson D [1996]. "Evolution of science-based uncertainty factors in noncancer risk assessment." Regul Toxicol Pharmacol *24*(2 Pt 1): 108-120.

Dourson ML, Stara JF [1983]. "Regulatory history and experimental support of uncertainty (safety) factors." Regul Toxicol Pharmacol *3*(3): 224-238.

Driscoll KE, Costa DL, Hatch G, Henderson R, Oberdorster G, Salem H, Schlesinger RB [2000]. "Intratracheal instillation as an exposure technique for the evaluation of respiratory tract toxicity: uses and limitations." Toxicol Sci *55*(1): 24-35.

ECHA [2008]. Guidance on information requirements and chemical safety assessment, pp.

Eisen EA, Smith TJ, Wegman DH, Louis TA, Froines J [1984]. "Estimation of long term dust exposures in the Vermont granite sheds." Am Ind Hyg Assoc J *45*(2): 89-94.

Elder A, Gelein R, Finkelstein JN, Driscoll KE, Harkema J, Oberdörster G [2005]. "Effects of subchronically inhaled carbon black in three species. I. Retention Kinetics, lung inflammation, and histopathology." Toxicol Sci *88*.

Engel LW, Strauchen JA, Chiazze L, Heid M [1980]. "Accuracy of death certification in an autopsied population with specific attention to malignant neoplasms and vascular diseases." Am J Epidemiol *111*(1): 99-112.

EPA [2012a]. Advances in Inhalation Gas Dosimetry for Derivation of a Reference Concentration (RfC) and Use in Risk Assessment. Washington, DC: U.S. Environmental Protection Agency, EPA/600/R-12/044, pp.

EPA [2006a]. Approaches for the Application of Physiologically Based Pharmacokinetic (PBPK) Models and Supporting Data in Risk Assessment. Washington, DC: U.S. Environmental Protection Agency, Office of Research and Development, EPA/600/R-05/043F, pp. 1-123.

EPA. (2012b). "Benchmark Dose Technical Guidance." Retrieved December 31, 2013, from http://www.epa.gov/raf/publications/pdfs/benchmark_dose_guidance.pdf.

EPA [2005]. Guidelines for carcinogen risk assessment. Washington, DC: Risk Assessment Forum, US Environmental Protection Agency, EPA/630/P-03/001F, pp.

EPA [1992]. Guidelines for Exposure Assessment. Washington, DC: U.S. Environmental Protection Agency (US EPA), EPA/600/Z-92/001, pp. 139.

EPA [2006b]. Harmonisation in interspecies extrapolation: use of BW3/4 as default method in derivation of the oral RfD. EXTERNAL REVIEW DRAFT. Washington, DC: U.S. Environmental Protection Agency, Office of Research and Development, EPA/630/R-06/001, pp. 1-34.

EPA [1994]. Methods for derivation of inhalation reference concentrations and application of inhalation dosimetry. Research Triangle Park, NC: Environmental Criteria and Assessment Office, Office of Health and Environmental Assessment, Office of Research and Development,, pp. 409.

EPA [2002]. A review of the reference dose and reference concentration processes. Washington, D.C.: Risk Assessment Forum, U.S. Environmental Protection Agency, EPA/630/P-02/002F, pp.

EPA [2001]. Toxicological review of chloroform in support of summary information on the Integrated Risk Information System (IRIS). Washington, DC: U.S. Environmental Protection Agency, EPA/635/R-01/001, pp. 112.

Esmen N [1984]. On estimation of occupational health risks. Occupational and Industrial Hygiene: Concepts and Methods. A symposium in honor of Theordore F. Hatch. N. A. Esmen and M. A. Mehlman. Princeton, NJ, Princeton Scientific Publishers, Inc. VIII**:** 45-84.

Fairchild EJ, II [1976]. "Guidelines for a NIOSH policy on occupational carcinogenesis." Ann N Y Acad Sci *271*: 200-207.

Fedak KM, Bernal A, Capshaw ZA, Gross S [2015]. "Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology." Emerg Themes Epidemiol *12*: 14.

Fiserova-Bergerova V [1990]. "Application of toxicokinetic models to establish biological exposure indicators." Ann Occup Hyg *34*(6): 639-651.

Fowle JR, Dearfield KL [2000]. Risk characterization handbook. Vol.: Science Policy Council, US Environmental Protection Agency, pp.

Freedman A, Robinson S [1988]. Noninvasive magnetopneumographic studies of lung dust retention and clearance in coal miners, University Park, PA: Penn State University Press**:** 181-186.

Frey HC, Patil SR [2002]. "Identification and review of sensitivity analysis methods." Risk Anal *22*(3): 553-578.

Fritschi L, Siemiatycki J, Richardson L [1996]. "Self-assessed versus expert-assessed occupational exposures." Am J Epidemiol *144*(5): 521-527.

Gail MH, Benichou J [2000]. Encyclopedia of epidemiologic methods. Vol. Chichester ; New York: Wiley, pp.

GAO [2014]. Chemical Assessments: Agencies Coordinate Activities, but Additional Action Could Enhance Efforts. Report to the Chairman, Subcommittee on Oversight, Committee on Science, Space, and Technology, House of Representatives. Washington, DC: United States Government Accountability Office, pp. 64.

Garcia GJ, Kimbell JS [2009]. "Deposition of inhaled nanoparticles in the rat nasal passages: dose to the olfactory region." Inhal Toxicol *21*(14): 1165-1175.

Garcia GJ, Schroeter JD, Kimbell JS [2015]. "Olfactory deposition of inhaled nanoparticles in humans." Inhal Toxicol *27*(8): 394-403.

Gardner MJ, Winter PD, Pannett B, Powell CA [1986]. "Follow up study of workers manufacturing chrysotile asbestos cement products." Br J Ind Med *43*(11): 726-732.

Gargas ML, Burgess RJ, Voisard DE, Cason GH, Andersen ME [1989]. "Partition coefficients of low-molecular-weight volatile chemicals in various liquids and tissues." Toxicol Appl Pharmacol *98*(1): 87-99.

Gaylor DW, Kodell RL, Chen JJ, Krewski D [1999]. "A unified approach to risk characterization." Inhal Toxicol *11*(6-7): 575-578.

Gibb HJ, Lees PSJ, Pinsky PF, Rooney BC [2000]. "Lung cancer among workers in chromium chemical production." Am J Ind Med *38*(2): 115-126.

Ginevan ME, Watkins DK [2010]. "Logarithmic dose transformation in epidemiologic dose-response analysis: Use with caution." Regul Toxicol Pharmacol *58*(2): 336-340.

Gloede E, Cichocki JA, Baldino JB, Morris JB [2011]. "A validated hybrid computational fluid dynamics-physiologically based pharmacokinetic model for respiratory tract vapor absorption in the human and rat and its application to inhalation dosimetry of diacetyl." Toxicol Sci *123*(1): 231-246.

Goldberg ID, Levin ML, Gerhardt PR, Handy VH, Cashman RE [1956]. "The probability of developing cancer." J Natl Cancer Inst *17*(2): 155-173.

Goldstein L, Langholz B [1992]. "Asymptotic Theory for Nested Case-Control Sampling in the Cox Regression Model." 1903-1928.

Greenland S [1996]. "Basic methods for sensitivity analysis of biases." Int J Epidemiol *25*(6): 1107-1116.

Greenland S, Copas J, Jones DR, Spiegelhalter D, Rice K, Armstrong B, Senn S, Carpenter J, Kenward M, De Stavola B, Nitsch D, Nitsch D, De Stavola B, Muirhead CR, Hodges J, Longford NT, Gelman A, Draper D, Gustafson P, McCandless L, Rubin DB [2005]. "Multiple-bias modelling for analysis of observational data." J R Stat Soc Ser A Stat Soc *168*(2): 267-306.

Greenland S, Longnecker MP [1992]. "Methods for trend estimation from summarized dose-response data, with applications to meta-analysis." Am J Epidemiol *135*(11): 1301-1309.

Greenland S, Robins JM, Pearl J [1999]. "Confounding and Collapsibility in Causal Inference." Stat Sci *14*(1): 29-46.

Gregoratto D, Bailey MR, Marsh JW [2010]. "Modelling particle retention in the alveolar-interstitial region of the human lungs." J Radiol Prot *30*(3): 491-512.

Greim H, Borm P, Schins R, Donaldson K, Driscoll K, Hartwig A [2001]. "Toxicity of fibers and particles - report of the workshop held in Munich, Germany, 26–27 October 2000." Inhal Toxicol *13*.

Griffin PJ [1986]. Predictive value of animal toxicity studies. Long-term animal studies: their predictive value for man. S. R. Walker and A. D. Dayan. Lancaster, PA, Springer Netherlands**:** 156.

Grimes DA, Schulz KF [2002]. "Bias and causal associations in observational research." Lancet *359*(9302): 248-252.

Groenwold RHH, Nelson David B DB, Nichol Kristin L KL, Hoes AW, Hak E [2010]. "Sensitivity analyses to estimate the potential impact of unmeasured confounding in causal research." Int J Epidemiol *39*(1): 107-117.

Guha N, Roy A, Kopylev L, Fox J, Spassova M, White P [2013]. "Nonparametric Bayesian methods for benchmark dose estimation." Risk Anal *33*(9): 1608–1619.

Hallock MF, Smith TJ, Woskie SR, Hammond SK [1994]. "Estimation of historical exposures to machining fluids in the automotive industry." Am J Ind Med *26*(5): 621-634.

Hamling J, Lee P, Weitkunat R, Ambuhl M [2008]. "Facilitating meta-analyses by deriving relative effect and precision estimates for alternative comparisons from a set of estimates presented by exposure level or disease category." Stat Med *27*(7): 954-970.

Hanley J, Liddell D [1985]. "Fitting relationships between exposure and standardized mortality ratios." J Occup Environ Med *27*(8): 555-560.

Hanna LM, Lou SR, Su S, Jarabek AM [2001]. "Mass transport analysis: inhalation rfc methods framework for interspecies dosimetric adjustment." Inhal Toxicol *13*(5): 437-463.

Harrell F [2015]. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Vol.: Springer, pp.

Hayes AW [2008]. Principles and methods of toxicology. Vol. Boca Raton, FL: CRC Press, pp.

Hernández LG, van Steeg H, Luijten M, van Benthem J [2009]. "Mechanisms of non-genotoxic carcinogens and importance of a weight of evidence approach." Mutat Res *682*(2-3): 94-109.

Hertz-Picciotto I, Wartenberg D, Simon R [1995]. "Epidemiology and quantitative risk assessment: A bridge from science to policy." Am J Public Health *85*(4): 484-493.

Hertzberg RC [1989]. "Fitting a model to categorical response data with application to species extrapolation of toxicity." Health Phys *57 Suppl 1*: 405-409.

Higgins JP, Green S [2008]. Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series. Vol., pp.

Hill AB [1965]. "The Environment and Disease: Association or Causation?" Proc R Soc Med *58*: 295-300.

Hinds WC [1999]. Aerosol technology: properties, behavior, and measurement of airborne particles. Vol. New York, NY: John Wiley & Sons, pp. 233-259.

Hofler M [2005]. "The Bradford Hill considerations on causality: a counterfactual perspective." Emerg Themes Epidemiol *2*: 11.

Homburger F [1987]. "The necessity of animal studies in routine toxicology." Comments on Toxicology *1*(5): 245-255.

Hornung RW, Greife AL, Stayner LT, Steenland NK, Herrick RF, Elliott LJ, Ringenburg VL, Morawetz J [1994]. "Statistical model for prediction of retrospective exposure to ethylene oxide in an occupational mortality study." Am J Ind Med *25*(6): 825-836.

Hornung RW, Meinhardt TJ [1987]. "Quantitative risk assessment of lung cancer in U.S. Uranium miners." Health Phys *52*(4): 417-430.

Hothorn LA [2014]. "Statistical evaluation of toxicological bioassays–a review." Toxicology Research *3*(6): 418-432.

Howard J, Piacentino J, MacMahon K, Schulte P [2017]. "Using systematic review in occupational safety and health." Am J Ind Med *60*(11): 921-929.

Howell D, Molloy S, Wilkinson K, Green E, Orchard K, Wang K, Liberty J [2015]. "Patient-reported outcomes in routine cancer clinical practice: a scoping review of use, impact on health outcomes, and implementation factors." Ann Oncol *26*(9): 1846-1858.

Howick J, Glasziou P, Aronson JK [2009]. "The evolution of evidence hierarchies: what can Bradford Hill's 'guidelines for causation' contribute?" J R Soc Med *102*(5): 186-194.

HSE [2001]. Reducing risks, protecting people. HSE's decision-making process. Vol. Sudbury, UK: Health and Safety Executive, pp.

Hughes EW, Martin-Body RL, Sarelius IH, Sinclair JD [1982]. "Effects of urethane-chloralose anaesthesia on respiration in the rat." Clin Exp Pharmacol Physiol *9*(2): 119-127.

IARC [2010]. IARC monographs on the evaluation of carcinogenic risks to humans: Carbon black, titanium dioxide, and talc. Lyon, France: International Agency for Research on Cancer, pp.

IARC [2012]. IARC monographs on the evaluation of carcinogenic risks to humans: chemical agents and related occupations. Lyon, France: International Agency for Research on Cancer, pp.

ICRP [1994]. "Human respiratory tract model for radiological protection. A report of a Task Group of the International Commission on Radiological Protection." Ann ICRP *24*(1-3): 1-482.

ILSI [2000]. "Risk Science Institute Workshop Participants, the relevance of the rat lung response to particle overload for human risk assessment: a workshop consensus report." Inhal Toxicol *12*.

448 U.S. 607 [1980]. Indus. Union Dept. v. Amer. Petroleum Inst., U.S. Supreme Court. Industrial Union Department, AFL-CIO v. American Petroleum Institute.

Ioannidis JP [2016]. "Exposure-wide epidemiology: revisiting Bradford Hill." Stat Med *35*(11): 1749-1762.

IPCS [2004]. IPCS risk assessment terminology. Vol. Geneve: OMS, pp.

Israel RA, Rosenberg HM, Curtin LR [1986]. "Analytical potential for multiple cause-of-death data." American Journal of Epidemiology *124*(2): 161-179.

Jarabek AM [1995]. "The application of dosimetry models to identify key processes and parameters for default dose-response assessment approaches." Toxicol Lett *79*(1-3): 171-184.

Jarabek AM, Asgharian B, Miller FJ [2005]. "Dosimetric adjustments for interspecies extrapolation of inhaled poorly soluble particles (PSP)." Inhalation toxicology *17*(7-8): 317-334.

Joffe M [1992]. "Validity of exposure data derived from a structured questionnaire." Am J Epidemiol *135*(5): 564-570.

Jones-Lee M, Aven T [2011]. "ALARP - What does it really mean?" Reliab Eng Sys Saf *96*(8): 877-882.

Julien E, Boobis AR, Olin SS, Ilsi Research Foundation Threshold Working G [2009]. "The Key Events Dose-Response Framework: a cross-disciplinary mode-of-action based approach to examining dose-response and thresholds." Crit Rev Food Sci Nutr *49*(8): 682-689.

Kanj RS, Kang JL, Castranova V [2005]. "Measurement of the release of inflammatory mediators from rat alveolar macrophages and alveolar type II cells following lipopolysaccharide or silica exposure: a comparative study." J Toxicol Environ Health A *68*(3): 185-207.

Kaplan S, Garrick BJ [1981]. "On The Quantitative Definition of Risk." Risk Anal *1*(1): 11-27.

Kauppinen TP [1994]. "Assessment of exposure in occupational epidemiology." Scand J Work Environ Health *20 Spec No*: 19-29.

Kavlock RJ, Schmid JE, Setzer RW, Jr. [1996]. "A simulation study of the influence of study design on the estimation of benchmark doses for developmental toxicity." Risk Anal *16*(3): 399-410.

Kienhuis AS, Van de Poll MCG, Wortelboer H, Van Herwijnen M, Gottschalk R, Dejong CHC, Boorsma A, Paules RS, Kleinjans JCS, Stierum RH, Van Delft JHM [2009]. "Parallelogram approach using rat-human In vitro and rat in vivo toxicogenomics predicts acetaminophen-induced hepatotoxicity in humans." Toxicol Sci *107*(2): 544-552.

Klaassen C, Amdur M, Doull J [2013]. Casarett & Doull's Toxicology: The basic science of poisons. Vol. New York NY: McGraw-Hill Education, pp.

Klimisch HJ, Andreae M, Tillmann U [1997]. "A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data." Regul Toxicol Pharmacol *25*(1): 1-5.

Kreyling WG, Semmler-Behnke M, Takenaka S, Moller W [2013]. "Differences in the biokinetics of inhaled nano- versus micrometer-sized particles." Acc Chem Res *46*(3): 714-722.

Kromhout H, Oostendorp Y, Heederik D, Boleij JS [1987]. "Agreement between qualitative exposure estimates and quantitative exposure measurements." Am J Ind Med *12*(5): 551-562.

Kuempel E, Smith R, Dankovic D, Stayner L [2009]. Rat-and human-based risk estimates of lung cancer from occupational exposure to poorly-soluble particles: a quantitative evaluation. Journal of Physics: Conference Series, IOP Publishing.

Kuempel E, Tran C [2002]. "Comparison of human lung dosimetry models: implications for risk assessment." Ann Occup Hyg *46*(suppl 1): 337-341.

Kuempel ED [2011]. "Carbon nanotube risk assessment: implications for exposure and medical monitoring." J Occup Environ Med *53*(6 Suppl): S91-97.

Kuempel ED [2000]. "Comparison of human and rodent lung dosimetry models for particle clearance and retention." Drug Chem Toxicol *23*(1): 203-222.

Kuempel ED, Attfield MD, Stayner LT, Castranova V [2014]. "Human and animal evidence supports lower occupational exposure limits for poorly-soluble respirable particles: letter to the editor re:'Low-toxicity dusts: current exposure guidelines are not sufficiently protective'by cherrie, brosseau, Hay and Donaldson." Annals of occupational hygiene *58*(9): 1205-1208.

Kuempel ED, Castranova V, Geraci CL, Schulte PA [2012]. "Development of risk-based nanomaterial groups for occupational exposure control." J Nanopart Res *14*: 1029.

Kuempel ED, O'Flaherty EJ, Stayner LT, Smith RJ, Green FH, Vallyathan V [2001]. "A biomathematical model of particle clearance and retention in the lungs of coal miners." Regul Toxicol Pharmacol *34*(1): 69-87.

Kuempel ED, Smith RJ, Attfield MD, Stayner LT [1997]. "Risks of occupational respiratory diseases among U.S. Coal miners." Appl Occup Environ Hyg *12*(12): 823-831.

Kuempel ED, Sweeney LM, Morris JB, Jarabek AM [2015]. "Advances in Inhalation Dosimetry Models and Methods for Occupational Risk Assessment and Exposure Limit Derivation." J Occup Environ Hyg *12 Suppl 1*: S18-40.

Lasagna L [1987]. "Predicting human drug safety from animal studies: Current issues." The Journal of Toxicological Sciences *12*(4): 439-450.

Lash TL, Fink AK [2003]. "Semi-automated sensitivity analysis to assess systematic errors in observational data." Epidemiology *14*(4): 451-458.

Laskin DL, Pendino KJ [1995]. "Macrophages and inflammatory mediators in tissue injury." Annu Rev Pharmacol Toxicol *35*: 655-677.

Lawson CC, Schnorr TM, Whelan EA, Deddens JA, Dankovic DA, Piacitelli LA, Sweeney MH, Connally LB [2004]. "Paternal occupational exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin and birth outcomes of offspring: birth weight, preterm delivery, and birth defects." Environ Health Perspect *112*(14): 1403-1408.

LeFevre ME, Green FH, Joel DD, Laqueur W [1982]. "Frequency of black pigment in livers and spleens of coal workers: correlation with pulmonary pathology and occupational information." Hum Pathol *13*(12): 1121-1126.

Leonard BE [2007]. "Examination of underground miner data for radon progeny size reduction as cause of high radon "inverse" dose rate effect." Health Phys *93*(2): 133-150.

Lewis RW, Billington R, Debryune E, Gamer A, Lang B, Carpanini F [2002]. "Recognition of adverse and nonadverse effects in toxicity studies." Toxicol Pathol *30*(1): 66-74.

Lin DY, Psaty BM, Kronmal RA [1998]. "Assessing the sensitivity of regression results to unmeasured confounders in observational studies." Biometrics *54*(3): 948-963.

Lin L, Piegorsch WW, Bhattacharya R [2015]. "Nonparametric benchmark dose estimation with continuous dose-response data." Scand Stat Theory Appl *42*(3): 713-731.

Lioy P, Lippmann M, Phalen R [1984]. Rationale for particle size-selective air sampling. Ann Am Conf Ind Hyg.

Liu CS, Liou SH, Loh CH, Yu YC, Uang SN, Shih TS, Chen HI [2005]. "Occupational bladder cancer in a 4,4 -methylenebis(2-chloroaniline) (MBOCA)-exposed worker." Environ Health Perspect *113*(6): 771-774.

Lochner C, Kinnear CJ, Hemmings SM, Seller C, Niehaus DJ, Knowles JA, Daniels W, Moolman-Smook JC, Seedat S, Stein DJ [2005]. "Hoarding in obsessive-compulsive disorder: clinical and genetic correlates." J Clin Psychiatry *66*(9): 1155-1160.

Loizou G, Spendiff M, Barton HA, Bessems J, Bois FY, d'Yvoire MB, Buist H, Clewell HJ, Meek B, Gundert-Remy U [2008]. "Development of good modelling practice for physiologically based pharmacokinetic models for use in risk assessment: the first steps." Regul Toxicol Pharmacol *50*(3): 400-411.

Lu FC, Kacew S [2002]. Lu's Basic Toxicology: Fundamentals, target organs and risk assessment. Vol.: CRC Press, pp.

Lubin JH, Boice JD, Jr., Edling C, Hornung RW, Howe G, Kunz E, Kusiak RA, Morrison HI, Radford EP, Samet JM, Tirmarche M, Woodward A, Xiang Yao S [1995]. "Radon-exposed underground miners and inverse dose-rate (Protraction enhancement) effects." Health Phys *69*(4): 494-500.

Maier A, Kohrman-Vincent M, Hertzberg R, Allen B, Haber LT, Dourson M [2012]. "Critical review of dose-response options for F344 rat mammary tumors for acrylamide - Additional insights based on mode of action." Food Chem Toxicol *50*(5): 1763-1775.

Marquart J, Brouwer DH, Gijsbers JH, Links IH, Warren N, van Hemmen JJ [2003]. "Determinants of dermal exposure relevant for exposure modelling in regulatory risk assessment." Ann Occup Hyg *47*(8): 599-607.

Martignoni M, Groothuis GM, de Kanter R [2006]. "Species differences between mouse, rat, dog, monkey and human CYP-mediated drug metabolism, inhibition and induction." Expert Opin Drug Metab Toxicol *2*(6): 875-894.

Mauderly JL [1997]. "Relevance of particle-induced rat lung tumors for assessing lung carcinogenic hazard and human lung cancer risk." Environ Health Perspect *105 Suppl 5*: 1337-1346.

Maudsley G, Williams EMI [1996]. "'Inaccuracy' in death certification - where are we now?" J Public Health Med *18*(1): 59-66.

McClellan RO [1997]. "Use of mechanistic data in assessing human risks from exposure to particles." Environ Health Perspect *105*(Suppl 5): 1363.

McLanahan ED, El-Masri HA, Sweeney LM, Kopylev LY, Clewell HJ, Wambaugh JF, Schlosser P [2012]. "Physiologically based pharmacokinetic model use in risk assessment—why being published is not enough." Toxicol Sci *126*(1): 5-15.

Meek ME, Renwick A, Ohanian E, Dourson M, Lake B, Naumann BD, Vu V, International Programme on Chemical S [2002]. "Guidelines for application of chemical-specific adjustment factors in dose/concentration-response assessment." Toxicology *181-182*: 115-120.

Mercer RR, Russell ML, Roggli VL, Crapo JD [1994]. "Cell number and distribution in human and rat airways." Am J Respir Cell Mol Biol *10*(6): 613-624.

Mercer RR, Scabilloni JF, Hubbs AF, Wang L, Battelli LA, McKinney W, Castranova V, Porter DW [2013]. "Extrapulmonary transport of MWCNT following inhalation exposure." Part Fibre Toxicol *10*: 38.

Mete F, Kilic E, Somay A, Yilmaz B [2012]. "Effects of heat stress on endocrine functions & behaviour in the pre-pubertal rat." Indian J Med Res *135*: 233-239.

Miller FJ [2000]. "Dosimetry of particles in laboratory animals and humans in relationship to issues surrounding lung overload and human health risk assessment: a critical review." Inhal Toxicol *12*(1-2): 19-57.

Moeschberger ML, Tordoff KP, Kochar N [2007]. "10 A Review of Statistical Analyses for Competing Risks." Handbook of Statistics *27*: 321-341.

Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF [1999]. "Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement." Lancet *354*(9193): 1896-1900.

Moher D, Schulz KF, Altman D [2001]. "The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials." JAMA *285*(15): 1987-1991.

Moon H, Kim HJ, Chen JJ, Kodell RL [2005]. "Model averaging using the Kullback information criterion in estimating effective doses for microbial infection and illness." Risk Anal *25*(5): 1147-1159.

Morales KH, Ibrahim JG, Chen C-J, Ryan LM [2006]. "Bayesian model averaging with applications to benchmark dose estimation for arsenic in drinking water." J Am Stat Assoc *101*(473): 9-17.

Morfeld P, Bruch J, Ngiewich Y, Chaudhuri I, Muranko H, Myerson R [2015]. Translation toxicology in setting occupational exposure limits for dusts and hazard classification- a critical evaluation of a recent approach to translate dust overload findings from rats to humans. Particle and Fibre Toxicology. Vol., pp.

Morrow PE [1988]. "Possible mechanisms to explain dust overloading of the lungs." Fundam Appl Toxicol *10*(3): 369-384.

Morrow PE, Muhle H, Mermelstein R [1991]. "Chronic Inhalation Study Findings as a Basis for Proposing a New Occupational Dust Exposure Limit." Int J Toxicol *10*(2): 279-290.

Moss O, Cheng Y [1989]. Generation and characterization of test atmospheres: particles. Concepts in Inhalation Toxicology. R. F. Henderson, R. McClellam and R. Henderson. New York, NY, Hemisphere Publishing Corp.**:** 85-120.

Muhle H, Bellmann B, Creutzenberg O, Heinrich U, Ketkar M, Mermelstein R [1990]. "Dust overloading of lungs after exposure of rats to particles of low solubility: comparative studies." J Aerosol Sci *21*(3): 374-377.

NAS [2017]. Controlled Human Inhalation-Exposure Studies at EPA. Vol. Washington DC: National Academies Press, pp. 159.

Naumann BD, Weideman PA [1995]. "Scientific basis for uncertainty factors used to establish occupational exposure limits for pharmaceutical active ingredients." Hum Ecol Risk Assess *1*(5): 590-613.

Needham LL, Calafat AM, Barr DB [2007]. "Uses and issues of biomonitoring." Int J Hyg Environ Health *210*(3–4): 229-238.

Nelson G [1992]. Gas mixtures: preparation and control. Vol.: CRC Press, pp.

Nieuwenhuijsen MJ [2010]. Exposure assessment in occupational and environmental epidemiology. Vol. Oxford ; New York: Oxford University Press, pp.

Nikula K, Vallyathan V, Green F, Hahn F [2000]. "Influence of dose on the distribution of retained particulate material in rat and human lungs." Proceedings of "Particulate Matter: 23-28.

Nilsson C, Raun K, Yan FF, Larsen MO, Tang-Christensen M [2012]. "Laboratory animals as surrogate models of human obesity." Acta Pharmacol Sin *33*(2): 173-181.

NIOSH [1982]. Comments to OSHA on Identification, Classification and Regulation of Potential Occupational Carcinogens. Cincinnati, OH, National Institute for Occupational Safety and Health.

NIOSH [1991]. Criteria for a Recomended Standard: Occupational Exposure to Ethylene Glycol Monomethyl Ether, Ethylene Glycol Monoethyl ether, and thier Acetates. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 91–119, pp.

NIOSH [2013a]. Criteria for a Recomended Standard: Occupational Exposure to Hexavalent Chromium. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 2013–128, pp.

NIOSH [1995a]. Criteria for a Recomended Standard: Occupational Exposure to Respirable Coal Mine Dust. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 95–106, pp.

NIOSH [1998]. Criteria for a Recomended Standard: Occupational Noise Exposure. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 98–126, pp.

NIOSH [2016a]. Criteria for a Reommended Standard: Occupational Exposure to Diacetyl and 2,3-Pentanedione. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 2016-111, pp.

NIOSH [2011]. Current Intelligence Bulletin 63: Occupational Exposure to Titanium Dioxide. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 2011-160, pp.

NIOSH [2013b]. Current Intelligence Bulletin 65: Occupational Exposure to Carbon Nanotubes and Nanofibers. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 2013-14, pp.

NIOSH [2017]. Current Intelligence Bulletin 68: NIOSH Chemical Carcinogen Policy. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication Number 2017-100, pp.

NIOSH [2016b]. Current Intelligence Bulletin: Health Effects of Occupational Exposure to Silver Nanomaterials Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, (External Review Draft - January 7), pp.

NIOSH [1995b]. Guidelines for air sampling and analytical method development and evaluation. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 95–117, pp.

NIOSH [1988]. Occupational Safety and Health Guideline for Benzene: Potential Human Carcinogen. Cincinnati, OH: National Institute for Occupational Safety and Health, pp. 1-6.

NIOSH [1987]. A recommended standard for occupational exposure to radon progeny in underground mines. Washington, DC: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 88-101, pp. 1-215.

NIOSH [1986a]. Testimony on OSHA's Proposed Rule: Occupation Exposure to Benzene. Oresented at the OSHA Informal Public Meeting, March 20, 1986. Washing DC, Occupational Safety and Health.

NIOSH [1986b]. Testimony to OSHA on the Proposed Rule: Occupational Exposure to Formaldehyde, May 5, 1986. Washington DC, Occupational Safety and Health Administration.

Noble RB, Bailer AJ, Park R [2009]. "Model-averaged benchmark concentration estimates for continuous response data arising from epidemiological studies." Risk Anal *29*(4): 558-564.

NRC [1956]. The biological effects of atomic radiation; a report to the public. Vol. Washington: National Academy of Sciences, National Research Council, pp.

NRC [2006]. Health risks from exposure to low levels of ionizing radiation : BEIR VII Phase 2. Vol. Washington, D.C.: National Academies Press, pp.

NRC [1988]. Health Risks of Radon and Other Internally Deposited Alpha-Emitters: BEIR IV

Vol. Washington (DC): National Academies Press, pp.

NRC [1981]. Mammalian models for research on aging. Vol. Washington, D.C.: National Academy Press, pp.

NRC [1983]. Risk Assessment in the Federal Government: Managing the Process. Vol. Washington (DC): National Academy Press, pp.

NRC [2009]. Science and Decisions: Advancing Risk Assessment. Vol. Washington (DC): National Academy Press, pp.

NTP [2015a]. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration: Office of Health Assessment and Translation (OHAT), Division of the National Toxicology Program, National Institute of Environmental Health Sciences, U.S. Department of Health and Human Services, pp. 94.

NTP [2015b]. Handbook for Preparing Report on Carcinogens Monographs. Washington DC: Office of the Report on Carcinogens, Division of the National Toxicology Program, National Institute of Environmental Health Sciences, U.S. Department of Health and Human Services, pp. 83.

NTP [2015c]. OHAT Risk of Bias Rating Tool for Human and Animal Studies Office of Health Assessment and Translation (OHAT), Division of the National Toxicology Program, National Institute of Environmental Health Sciences, U.S. Department of Health and Human Services, pp. 37.

Oberdörster G [1989]. "Dosimetric principles for extrapolating results of rat inhalation studies to humans, using an inhaled Ni compound as an example." Health Phys *57 Suppl 1*: 213-220.

Oberdörster G [1995]. "Lung particle overload: implications for occupational exposures to particles." Regul Toxicol Pharmacol *21*(1): 123-135.

Oberdörster G [1997]. "Pulmonary carcinogenicity of inhaled particles and the maximum tolerated dose." Environ Health Perspect *105 Suppl 5*: 1347-1355.

Oberdörster G, Ferin J, Soderholm S, Gelein R, Cox C, Baggs R, Morrow P [1994]. "Increased pulmonary toxicity of inhaled ultrafine particles: Due to lung overload alone?" Ann Occup Hyg *38*(inhaled particles VII): 295-302.

Oberdörster G, Oberdörster E, Oberdörster J [2005]. "Nanotoxicology: an emerging discipline evolving from studies of ultrafine particles." Environ Health Perspect *113*(7): 823-839.

29 USC Chapter 15 [December 29, 1970, As amended through January 1, 2004]. Occupational Safety and Health Act of 1970

OECD [1987]. Acute Dermal Toxicity, Test Guideline 402 Paris, France: Organisation for Economic Co-operation and Development (OECD), TG 402, pp. 7.

OECD [2009a]. Acute Inhalation Toxicity, Test Guideline 403 Paris, France: Organisation for Economic Co-operation and Development, TG 403, pp. 19.

OECD [2009b]. Chronic Toxicity Tests, Test Guideline No. 452. Paris: Organisation for Economic Co-operation and Development, TG 452, pp. 1-16.

OECD [2005]. Good laboratory practice : OECD principles and guidance for compliance monitoring. Vol. Paris: Organisation for Economic Co-operation and Development, pp.

OECD [2009c]. Subacute Inhalation Toxicity: 28-Day Study, Test Guideline No. 412. Paris: Organisation for Economic Co-operation and Development, pp. 1-14.

OECD [2009d]. Subchronic Inhalation Toxicity: 90-Day Study, Test Guideline No. 413. Paris: Organisation for Economic Co-operation and Development, pp. 1-15.

OECD [2015]. Test No. 404: Acute Dermal Irritation/Corrosion Paris, France: Organisation for Economic Co-operation and Development (OECD), TG 404, pp. 8.

OECD [1981a]. Test No. 410: Repeated Dose Dermal Toxicity: 21/28-day Study. Paris, France: Organisation for Economic Co-operation and Development (OECD), TG 410, pp. 8.

OECD [1981b]. Test No. 411: Subchronic Dermal Toxicity: 90-day Study. Paris, France: Organisation for Economic Co-operation and Development (OECD), TG 411, pp. 9.

OECD [2010]. Test No. 429: Skin Sensitisation Local Lymph Node Assay. Paris, France: Organisation for Economic Co-operation and Development (OECD), TG 421, pp. 20.

Oller AR, Oberdörster G [2016]. "Incorporation of dosimetry in the derivation of reference concentrations for ambient or workplace air: A conceptual approach." Journal of Aerosol Science *99*(Supplement C): 40-45.

Orsini N, Li R, Wolk A, Khudyakov P, Spiegelman D [2012]. "Meta-analysis for linear and nonlinear dose-response relations: examples, an evaluation of approximations, and software." Am J Epidemiol *175*(1): 66-73.

OSHA [1997]. Occupational Exposure to Methylene Chloride. D. o. L. Occupational Safety and Health Administration (OSHA). Washington, DC**:** 1494-1619.

Overmyer KA, Thonusin C, Qi NR, Burant CF, Evans CR [2015]. "Impact of anesthesia and euthanasia on metabolomics of mammalian tissues: studies in a C57BL/6J mouse model." PLoS One *10*(2): e0117232.

Ozasa K, Shimizu Y, Suyama A, Kasagi F, Soda M, Grant EJ, Sakata R, Sugiyama H, Kodama K [2012]. "Studies of the mortality of atomic bomb survivors, report 14, 1950-2003: An overview of cancer and noncancer diseases." Radiat Res *177*(3): 229-243.

Panchal SK, Brown L [2011]. "Rodent models for metabolic syndrome research." J Biomed Biotechnol *2011*: 351982.

Paquet F, Etherington G, Bailey MR, Leggett RW, Lipsztein J, Bolch W, Eckerman KF, Harrison JD, Icrp [2015]. "ICRP Publication 130: Occupational Intakes of Radionuclides: Part 1." Ann ICRP *44*(2): 5-188.

Park R, Rice F, Stayner L, Smith R, Gilbert S, Checkoway H [2002]. "Exposure to crystalline silica, silicosis, and lung disease other than cancer in diatomaceous earth industry workers: A quantitative risk assessment." Occup Environ Med *59*(1): 36-43.

Park RM, Bena JF, Stayner LT, Smith RJ, Gibb HJ, Lees PSJ [2004]. "Hexavalent chromium and lung cancer in the chromate industry: A quantitative risk assessment." Risk Anal *24*(5): 1099-1108.

Park RM, Bowler RM, Roels HA [2009]. "Exposure-response relationship and risk assessment for cognitive deficits in early welding-induced manganism." J Occup Environ Med *51*(10): 1125-1136.

Park RM, Maizlish NA, Punnett L, Moure-Eraso R, Silverstein MA [1991]. "A comparison of PMRs and SMRs as estimators of occupational mortality." Epidemiology *2*(1): 49-59.

Pauluhn J [2014]. "Derivation of occupational exposure levels (OELs) of low-toxicity isometric biopersistent particles: How can the kinetic lung overload paradigm be used for improved inhalation toxicity study design and OEL-derivation?" Part Fibre Toxicol *11*: 72.

Pauluhn J [2010]. "Multi-walled carbon nanotubes (Baytubes): approach for derivation of occupational exposure limit." Regul Toxicol Pharmacol *57*(1): 78-89.

Pauluhn J [2011]. "Poorly soluble particulates: searching for a unifying denominator of nanoparticles and fine particles for DNEL estimation." Toxicology *279*(1): 176-188.

Persad AS, Cooper GS [2008]. "Use of epidemiologic data in Integrated Risk Information System (IRIS) assessments." Toxicol Appl Pharmacol *233*(1): 137-145.

Phalen R, Stuart B, Lioy P [1988]. Rationale for and implications of particle size-selective sampling. Advances in Air Sampling, American Conference of Governmental Industrial Hygienists, Lewis Publishers, Chelsea, MI.

Phalen RF [1997]. Methods in inhalation toxicology. Vol. 3. Boca Raton, FL: CRC Press, pp. 176.

Phillips CV, Goodman KJ [2004]. "The missed lessons of Sir Austin Bradford Hill." Epidemiol Perspect Innov *1*(1): 3.

Piegorsch WW, Xiong H, Bhattacharya RN, Lin L [2013]. "Benchmark dose analysis via nonparametric regression modeling." Risk Anal.

Piegorsch WW, Xiong H, Bhattacharya RN, Lin L [2012]. "Nonparametric estimation of benchmark doses in environmental risk assessment." Environmetrics *23*(8): 717–728.

Poet TS, McDougal JN [2002]. "Skin absorption and human risk assessment." Chem Biol Interact *140*(1): 19-34.

Price O, Asgharian B, Miller F, Cassee F, de Winter-Sorkina R [2002]. Multiple path particle dosimetry model (MPPD v 1.0): A model for human and rat airway particle dosimetry Bilthoven, The Netherlands, National Institute for Public Health and the Environment (RIVM)

Prince MM, Gilbert SJ, Smith RJ, Stayner LT [2003]. "Evaluation of the risk of noise-induced hearing loss among unscreened male industrial workers." J Acoust Soc Am *113*(2): 871-880.

Raabe OG, Al-Bayati MA, Teague SV, Rasolt A [1988]. "Regional deposition of inhaled monodisperse coarse and fine aerosol particles in small laboratory animals." Ann Occup Hyg *32*(inhaled particles VI): 53-63.

Raftery AE [1995]. "Bayesian model selection in social research." Sociological methodology: 111-163.

Ramachandran G, Banerjee S, Vincent JH [2003]. "Expert judgment and occupational hygiene: Application to aerosol speciation in the nickel primary production industry." Ann Occup Hyg *47*(6): 461-475.

Rappaport SM [1991]. "Assessment of long-term exposures to toxic substances in air." Annals of Occupational Hygiene *35*(1): 61-122.

Redelings MD, Wise M, Sorvillo F [2007]. "Using multiple cause-of-death data to investigate associations and causality between conditions listed on the death certificate." American Journal of Epidemiology *166*(1): 104-108.

Relier C, Dubreuil M, Lozano Garcìa O, Cordelli E, Mejia J, Eleuteri P, Robidel F, Loret T, Pacchierotti F, Lucas S, Lacroix G, Trouiller B [2017]. "Study of TiO2 P25 Nanoparticles Genotoxicity on Lung, Blood, and Liver Cells in Lung Overload and Non-Overload Conditions After Repeated Respiratory Exposure in Rats." Toxicological Sciences *156*(2): 527-537.

Rhomberg LR, Goodman JE, Bailey LA, Prueitt RL, Beck NB, Bevan C, Honeycutt M, Kaminski NE, Paoli G, Pottenger LH, Scherer RW, Wise KC, Becker RA [2013]. "A survey of frameworks for best practices in weight-of-evidence analyses." Crit Rev Toxicol *43*(9): 753-784.

Rhomberg LR, Goodman JE, Haber LT, Dourson M, Andersen ME, Klaunig JE, Meek B, Price PS, McClellan RO, Cohen SM [2011]. "Linear low-dose extrapolation for noncancer heath effects is the exception, not the rule." Crit Rev Toxicol *41*(1): 1-21.

Rice C, Harris Jr RL, Lumsden JC, Symons MJ [1984]. "Reconstruction of silica exposure in the North Carolina Dusty Trades." Am Ind Hyg Assoc J *45*(10): 689-696.

Rice FL, Park R, Stayner L, Smith R, Gilbert S, Checkoway H [2001]. "Crystalline silica exposure and lung cancer mortality in diatomaceous earth industry workers: A quantitative risk assessment." Occup Environ Med *58*(1): 38-45.

Richardson DB, Cole SR, Chu H, Langholz B [2011]. "Lagging exposure information in cumulative exposure-response analyses." Am J Epidemiol *174*(12): 1416-1422.

Rinsky RA, Smith AB, Hornung R, Filloon TG, Young RJ, Okun AH, Landrigan PJ [1987]. "Benzene and leukemia. An epidemiologic risk assessment." N Engl J Med *316*(17): 1044-1050.

Rodricks JV, Gaylor DW, Turnbull D [2007]. Quantitative extrapolations in toxicology. Principles and Methods of Toxicology, Fifth Edition A. W. Hayes. Boca Raton, FL, CRC Press, Taylor & Francis Group**:** 453-474.

Rodrigues MA, Arezes P, Leão CP [2014]. "Risk Criteria in Occupational Environments: Critical Overview and Discussion." Procedia Soc Behav Sci *109*: 257-262.

Rothman KJ, Greenland S, Lash TL [2008]. Modern epidemiology. Vol. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins, pp.

Sackett DL [1979]. "Bias in analytic research." J Chronic Dis *32*(1-2): 51-63.

Sahu SC, Casciano DA [2009]. Nanotoxicity: From In Vivo and In Vitro Models to Health Risks. Vol.: John Wiley & Sons, pp.

Salem H, Katz SA [2014]. Inhalation toxicology. Vol. Boca Raton, FL: CRC Press, pp.

Sand S, Portier CJ, Krewski D [2011]. "A signal-to-noise crossover dose as the point of departure for health risk assessment." Environ Health Perspect *119*(12): 1766–1774.

Sanderson WT, Petersen MR, Ward EM [2001]. "Estimating historical exposures of workers in a beryllium manufacturing plant." American Journal of Industrial Medicine *39*(2): 145-157.

Sarangapani R, Teeguarden JG, Cruzan G, Clewell HJ, Andersen ME [2002]. "Physiologically based pharmacokinetic modeling of styrene and styrene oxide respiratory-tract dosimetry in rodents and humans." Inhal Toxicol *14*(8): 789-834.

Schlesinger RB [1985]. "Comparative deposition of inhaled aerosols in experimental animals and humans: a review." J Toxicol Environ Health *15*(2): 197-214.

Schubauer-Berigan MK, Anderson JL, Hein MJ, Little MP, Sigurdson AJ, Pinkerton LE [2015]. "Breast cancer incidence in a cohort of U.S. flight attendants." Am J Ind Med *58*(3): 252-266.

Schubauer-Berigan MK, Couch JR, Deddens JA [2017]. "Is beryllium-induced lung cancer caused only by soluble forms and high exposure levels?" Occupational and Environmental Medicine *74*(8): 601-603.

Schubauer-Berigan MK, Couch JR, Petersen MR, Carreón T, Jin Y, Deddens JA [2011]. "Cohort mortality study of workers at seven beryllium processing plants: Update and associations with cumulative and maximum exposure." Occupational and Environmental Medicine *68*(5): 345-353.

Schuhmacher-Wolz U, Kalberlah F, Oppl R, van Hemmen JJ [2003]. "A toolkit for dermal risk assessment: toxicological approach for hazard characterization." Ann Occup Hyg *47*(8): 641-652.

Schulz H, Brand P, Heyer NJ [2000]. Particle Deposition in the Respiratory Tract. Particle-lung Interactions. P. Gehr and J. Heyder. New York, Marcel Dekker, Inc.**:** 229-290.

Schwarz G [1978]. "Estimating the Dimension of a Model." Ann Stat *6*(2): 461-464.

SCOEL [2010]. Recommendation from the Scientific Committee on Occupational Exposure Limits for cadmium and its inorganic compounds: European Comission, pp. 26.

SCOEL [2008]. Recommendation from the Scientific Committee on Occupational Exposure Limits for formaldhyde: European Comission, pp. 46.

SCOEL [2011]. Recommendation from the Scientific Committee on Occupational Exposure Limits for nickel and inorganic nickel compounds: European Comission, pp. 46.

Seed J, Carney EW, Corley RA, Crofton KM, DeSesso JM, Foster PM, Kavlock R, Kimmel G, Klaunig J, Meek ME, Preston RJ, Slikker W, Jr., Tabacova S, Williams GM, Wiltse J, Zoeller RT, Fenner-Crisp P, Patton DE [2005]. "Overview: Using mode of action and life stage information to evaluate the human relevance of animal toxicity data." Crit Rev Toxicol *35*(8-9): 664-672.

Seeley MR, Tonner-Navarro LE, Beck BD, Deskin R, Feron VJ, Johanson G, Bolt HM [2001]. "Procedures for health risk assessment in Europe." Regul Toxicol Pharmacol *34*(2): 153-169.

Sehdev AES, Hutchins GM [2001]. "Problems with proper completion and accuracy of the cause-of-death statement." Arch Intern Med *161*(2): 277-284.

Seixas NS, Checkoway H [1995]. "Exposure assessment in industry specific retrospective occupational epidemiology studies." Occup Environ Med *52*(10): 625-633.

Seixas NS, Heyer NJ, Welp EA, Checkoway H [1997]. "Quantification of historical dust exposures in the diatomaceous earth industry." Ann Occup Hyg *41*(5): 591-604.

Seixas NS, Robins TG, Moulton LH [1988]. "The use of geometric and arithmetic mean exposures in occupational epidemiology." Am J Ind Med *14*(4): 465-477.

Shipp A, Lawrence G, Gentry R, McDonald T, Bartow H, Bounds J, Macdonald N, Clewell H, Allen B, Van Landingham C [2006]. "Acrylamide: Review of toxicity data and dose-response analyses for cancer and noncancer effects." Crit Rev Toxicol *36*(6): 481-608.

Sievert R, Failla G [1959]. "Recommendations of the International Commission on Radiological Protection." Health Phys *2*: 1-20.

Smith AH [1988]. "Epidemiologic input to environmental risk assessment." Arch Environ Health *43*(2): 124-129.

Smith CM, Christiani DC, Kelsey KT [1994]. Chemical risk assessment and occupational health: current applications, limitations, and future prospects. Vol.: Greenwood Publishing Group, pp.

Smith MT, Guyton KZ, Gibbons CF, Fritz JM, Portier CJ, Rusyn I, DeMarini DM, Caldwell JC, Kavlock RJ, Lambert PF, Hecht SS, Bucher JR, Stewart BW, Baan RA, Cogliano VJ, Straif K [2016]. "Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis." Environmental Health Perspectives *124*(6): 713-721.

Snipes MB [1989]. "Long-term retention and clearance of particles inhaled by mammalian species." Crit Rev Toxicol *20*(3): 175-211.

Sobels FH [1977]. "Some problems associated with the testing for environmental mutagens and a perspective for studies in "comparative mutagenesis"." Mutat Res *46*(4): 245-260.

Stahlhofen W, Rudolf G, James A [1989]. "Intercomparison of experimental regional aerosol deposition data." J Aerosol Med *2*(3): 285-308.

Stayner L, Dankovic D, Smith R, Steenland K [1998]. "Predicted lung cancer risk among miners exposed to diesel exhaust particles." Am J Ind Med *34*(3): 207-219.

Stayner L, Smith R, Bailer J, Gilbert S, Steenland K, Dement J, Brown D, Lemen R [1997]. "Exposure-response analysis of risk of respiratory disease associated with occupational exposure to chrysotile asbestos." Occup Environ Med *54*(9): 646-652.

Stayner L, Smith R, Thun M, Schnorr T, Lemen R [1992a]. "A dose-response analysis and quantitative assessment of lung cancer risk and occupational cadmium exposure." Ann Epidemiol *2*(3): 177-194.

Stayner L, Smith R, Thun M, Schnorr T, Lemen R [1992b]. "A quantitative assessment of lung cancer risk and occupational cadmium exposure." IARC Sci Publ(118): 447-455.

Stayner L, Smith RJ, Gilbert S, Bailer AJ [1999]. "Epidemiologic approaches to risk assessment." Inhal Toxicol *11*(6-7): 593-601.

Stayner L, Steenland K, Dosemeci M, Hertz-Picciotto I [2003]. "Attenuation of exposure-response curves in occupational cohort studies at high exposure levels." Scand J Work Environ Health *29*(4): 317-324.

Stayner LT, Dankovic DA, Smith RJ, Gilbert SJ, Bailer AJ [2000a]. "Human cancer risk and exposure to 1,3-butadiene--a tale of mice and men." Scand J Work Environ Health *26*(4): 322-330.

Stayner LT, Dankovic DA, Smith RJ, Gilbert SJ, Bailer J [2000b]. "Human cancer risk and exposure to 1,3-butadiene - A tale of mice and men." Scand J Work Environ Health *26*(4): 322-330.

Steenland K, Deddens J, Stayner L [1998]. "Diesel exhaust and lung cancer in the trucking industry: Exposure-response analyses and risk assessment." Am J Ind Med *34*(3): 220-228.

Steenland K, Deddens JA [2004]. "A practical guide to dose-response analyses and risk assessment in occupational epidemiology." Epidemiology *15*(1): 63-70.

Steenland K, Karnes C, Darrow L, Barry V [2015]. "Attenuation of exposure-response rate ratios at higher exposures a simulation study focusing on frailty and measurement error." Epidemiology *26*(3): 395-401.

Steenland K, Savitz DA [1997]. Topics in environmental epidemiology. Vol.: Oxford University Press, USA, pp.

Steenland K, Seals R, Klein M, Jinot J, Kahn HD [2011]. "Risk estimation with epidemiologic data when response attenuates at high-exposure levels." Environ Health Perspect *119*(6): 831-837.

Steenland K, Whelan E, Deddens J, Stayner L, Ward E [2003]. "Ethylene oxide and breast cancer incidence in a cohort study of 7576 women (United States)." Cancer Causes and Control *14*(6): 531-539.

Stewart PA, Blair A, Cubit DA, Bales RE, Kaplan SA, Ward J, Gaffey W, O'Berg MT, Walrath J [1986]. "Estimating Historical Exposures to Formaldehyde in a Retrospective Mortality Study." Appl Ind Hyg *1*(1): 34-41.

Stewart PA, Lees PS, Francis M [1996]. "Quantification of historical exposures in occupational cohort studies." Scand J Work Environ Health *22*(6): 405-414.

Stewart WF, Tonascia JA, Matanoski GM [1987]. "The validity of questionnaire-reported work history in live respondents." J Occup Med *29*(10): 795-800.

Stöber W, Einbrodt H, Klosterkötter W [1965]. Quantitative studies of dust retention in animal and human lungs after chronic inhalation. Inhaled Particles and Vapours II (Proceedings of an International Symposium organized by the British Occupational Hygiene Society, Cambridge, 28 Sept-1 Oct 1965). Cambridge: Pergamon Press, Symposium Publications Division.

Stöber W, Morrow PE, Morawietz G [1990]. "Alveolar retention and clearance of insoluble particles in rats simulated by a new physiology-oriented compartmental kinetics model." Fundam Appl Toxicol *15*(2): 329-349.

Stone KC, Mercer RR, Gehr P, Stockstill B, Crapo JD [1992]. "Allometric relationships of cell numbers and size in the mammalian lung." Am J Respir Cell Mol Biol *6*(2): 235-243.

Stratmann G, Sall JW, May LD, Loepke AW, Lee MT [2010]. "Beyond anesthetic properties: the effects of isoflurane on brain cell death, neurogenesis, and long-term neurocognitive function." Anesth Analg *110*(2): 431-437.

Streffer C, Bolt H, Follesdal D, Hall P, Hengstler JG, Jacob P, Oughton D, Prieß K, Rehbinder E, Swaton E [2004]. Low dose exposures in the environment: dose-effect relations and risk evaluation. Vol. 23: Springer Science & Business Media, pp.

Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB [2000]. "Meta-analysis of observational studies in epidemiology: A proposal for reporting." JAMA *283*(15): 2008-2012.

Sturm R, Hofmann W [2009]. "A theoretical approach to the deposition and clearance of fibers with variable size in the human respiratory tract." J Hazard Mater *170*(1): 210-218.

Su W-C, Cheng YS [2015]. "Estimation of carbon nanotubes deposition in a human respiratory tract replica." J Aerosol Sci *79*: 72-85.

Su WC, Cheng YS [2014]. "Carbon nanotubes size classification, characterization and nasal airway deposition." Inhal Toxicol *26*(14): 843-852.

Suvrathan A, Tomar A, Chattarji S [2010]. "Effects of chronic and acute stress on rat behaviour in the forced-swim test." Stress *13*(6): 533-540.

T Mannetje A, Fevotte J, Fletcher T, Brennan P, Legoza J, Szeremi M, Paldy A, Brzeznicki S, Gromiec J, Ruxanda-Artenie C, Stanescu-Dumitru R, Ivanov N, Shterengorz R, Hettychova L, Krizanova D, Cassidy A, Van Tongeren M, Boffetta P [2003]. "Assessing exposure misclassification by expert assessment in multicenter occupational studies." Epidemiology *14*(5): 585-592.

Tchiehe DN, Gauthier F [2017]. "Classification of risk acceptability and risk tolerability factors in occupational health and safety." Safety Science *92*: 138-147.

Teschke K, Kennedy SM, Olshan AF [1994]. "Effect of different questionnaire formats on reporting of occupational exposures." Am J Ind Med *26*(3): 327-337.

Teschke K, Olshan AF, Daniels JL, De Roos AJ, Parks CG, Schulz M, Vaughan TL [2002]. "Occupational exposure assessment in case-control studies: opportunities for improvement." Occup Environ Med *59*(9): 575-593; discussion 594.

Thygesen LC, Andersen GS, Andersen H [2005]. "A philosophical analysis of the Hill criteria." J Epidemiol Community Health *59*(6): 512-516.

Tielemans E, Heederik D, Burdorf A, Vermeulen R, Veulemans H, Kromhout H, Hartog K [1999]. "Assessment of occupational exposures in a general population: Comparison of different methods." Occup Environ Med *56*(3): 145-151.

Tooth L, Ware R, Bain C, Purdie DM, Dobson A [2005]. "Quality of reporting of observational longitudinal research." Am J Epidemiol *161*(3): 280-288.

Tran C, Buchanan D [2000]. Development of a biomathematical lung model to describe the exposure-dose relationship for inhaled dust among UK coal miners. Edinburgh, Scotland, UK: Institute of Occupational Medicine (IOM), IOM Report TM/00/02, pp. 54.

Tran CL, Buchanan D, Cullen RT, Searl A, Jones AD, Donaldson K [2000]. "Inhalation of poorly soluble particles. II. Influence of particle surface area on inflammation and clearance." Inhal Toxicol *12*.

647 F.2d 1189 [1980]. United Steelworkers of America v. F Marshall U.S. Court of Appeals, District of Columbia Circuit. United Steel Workers of America, AFL-CIO v. F. Ray Marshall.

UNSCEAR [2015]. Sources, effects and risks of ionizing radiation : UNSCEAR 2012 report to the General Assembly, with scientific annexes. Vol. New York: United Nations, pp.

Ury HK [1975]. "Efficiency of case control studies with multiple controls per case: continuous or dichotomous data." Biometrics *31*(3): 643-649.

Valentin J [2002]. "Basic anatomical and physiological data for use in radiological protection: reference values: ICRP Publication 89: Approved by the Commission in September 2001." Ann ICRP *32*(3-4): 1-277.

van Ravenzwaay B, Leibold E [2004]. "The significance of in vitro rat skin absorption studies to human risk assessment." Toxicol In Vitro *18*(2): 219-225.

Van Wendel De Joode B, Van Hemmen JJ, Meijster T, Major V, London L, Kromhout H [2005a]. "Reliability of a semi-quantitative method for dermal exposure assessment (DREAM)." J Expo Anal Environ Epidemiol *15*(1): 111-120.

Van Wendel De Joode B, Vermeulen R, Van Hemmen JJ, Fransman W, Kromhout H [2005b]. "Accuracy of a semiquantitative method for Dermal Exposure Assessment (DREAM)." Occup Environ Med *62*(9): 623-632.

Vanem E [2012]. "Ethics and fundamental principles of risk acceptance criteria." Safety Science *50*(4): 958-967.

von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP [2007]. "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies." Lancet *370*(9596): 1453-1457.

Wallace RB, Kohatsu N [2008]. Wallace/Maxcy-Rosenau-Last public health & preventive medicine. Vol. New York: McGraw-Hill Medical, pp.

Warheit DB, Kreiling R, Levy LS [2016]. "Relevance of the rat lung tumor response to particle overload for human risk assessment—Update and interpretation of new data since ILSI 2000." Toxicology *374*(Supplement C): 42-59.

Waters M, McKernan L, Maier A, Jayjock M, Schaeffer V, Brosseau L [2015]. "Exposure Estimation and Interpretation of Occupational Risk: Enhanced Information for the Occupational Risk Manager." J Occup Environ Hyg *12 Suppl 1*: S99-111.

Weed DL [2005]. "Weight of evidence: a review of concept and methods." Risk Anal *25*(6): 1545-1557.

Weingand K, Brown G, Hall R, Davies D, Gossett K, Neptun D, Waner T, Matsuzawa T, Salemink P, Froelke W, Provost JP, Dal Negro G, Batchelor J, Nomura M, Groetsch H, Boink A, Kimball J, Woodman D, York M, Fabianson-Johnson E, Lupart M, Melloni E [1996]. "Harmonization of animal clinical pathology testing in toxicity and safety studies. The Joint Scientific Committee for International Harmonization of Clinical Pathology Testing." Fundam Appl Toxicol *29*(2): 198-201.

Weitzel JN, Blazer KR, MacDonald DJ, Culver JO, Offit K [2011]. "Genetics, genomics, and cancer risk assessment: State of the Art and Future Directions in the Era of Personalized Medicine." CA Cancer Journal for Clinicians *61*(5): 327-359.

Wheeler M, Bailer AJ [2012]. "Monotonic Bayesian semiparametric benchmark dose analysis." Risk Anal *32*(7): 1207–1218.

Wheeler MW, Bailer AJ [2013]. "An empirical comparison of low-dose extrapolation from points of departure (PoD) compared to extrapolations based upon methods that account for model uncertainty." Regul Toxicol Pharmacol *67*(1): 75–82.

Wheeler MW, Bailer AJ [2007]. "Properties of model-averaged BMDLs: a study of model averaging in dichotomous response risk estimation." Risk Anal *27*(3): 659–670.

Wheeler MW, Shao K, Bailer AJ [2015]. "Quantile benchmark dose estimation for continuous endpoints." Environmetrics *26*(5): 363-372.

White E, Armstrong BK, Saracci R, Armstrong BK [2008]. Principles of exposure measurement in epidemiology : collecting, evaluating, and improving measures of disease risk factors. Vol. Oxford ; New York: Oxford University Press, pp.

White RH, Cote I, Zeise L, Fox M, Dominici F, Burke TA, White PD, Hattis DB, Samet JM [2009]. "State-of-the-science workshop report: issues and approaches in low-dose-response extrapolation for environmental health risk assessment." Environ Health Perspect *117*(2): 283.

WHO [1994]. Assessing human health risk of chemicals: derivation of guidance values for health-based exposure limits. Vol. Geneva, Switzerland: World Health Organization, pp. 1-73.

WHO [2005]. Chemical-specific adjustment factors for interspecies differences and human variability: guidance document for use of data in dose/concentration-response assessment. Geneva, Switzerland: World Health Organization (WHO), 2, pp. 96.

WHO [2000]. Evaluation and use of epidemiological evidence for environmental health risk assessment: WHO guideline document. Copenhagen: World Health Organization. Regional Office for Europe, EUR/00/5020369, pp. 38.

WHO [1977]. Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death: Based on the Recommendations of the Ninth Revision Conference, 1975 and adopted by the Twenty-Ninth World Health Assembly. Geneva, Switzerland: World Health Organization, pp. 387.

Wignall JA, Shapiro AJ, Wright FA, Woodruff TJ, Chiu WA, Guyton KZ, Rusyn I [2014]. "Standardizing benchmark dose calculations to improve science-based decisions in human health assessments." Environ Health Perspect *122*(5): 499–505.

Wong BA [1999]. Inhalation exposure systems design, methods, and operation. Toxicology of the Lung. D. E. Gardner, J. D. Crapo and R. O. McClellan. Philadelphia, PA, Taylor and Francis**:** 1-53.

Wong BA [2007]. "Inhalation exposure systems: design, methods and operation." Toxicol Pathol *35*(1): 3-14.

Woodruff TJ, Sutton P [2014]. "The navigation guide systematic review methodology: A rigorous and transparent method for translating environmental health science into better health outcomes." Environmental Health Perspectives *122*(10): 1007-1014.

Woodward M [2013]. Epidemiology: study design and data analysis. Vol.: CRC press, pp.

Woskie SR, Smith TJ, Hammond K, Schenker MB, Garshick E, Speizer FE [1988]. "Estimation of the diesel exhaust exposures of railroad workers: I. Current exposures." Am J Ind Med *13*(3): 381-394.

Yu C [1996]. "Extrapolation modeling of particle deposition and retention from rats to humans." Particulate science and technology *14*(1): 1-13.

Zaccai JH [2004]. "How to assess epidemiological studies." Postgrad Med J *80*(941): 140.

Zdeb MS [1977]. "The probability of developing cancer." Am J Epidemiol *106*(1): 6-16.

APPENDIX A: GLOSSARY

National Institute for Occupational Safety and Health

Term	Definition
absolute risk	The probability that a disease free individual will develop a given disease over a specified time given age, other risk factors and in the presence of competing risk.
adverse effect	Change in the morphology, physiology, growth, development, reproduction, or life span of an organism, system, or population that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress, or an increase in susceptibility to other influences.
agent	A chemical, biological, or physical entity that contacts a target.
aneuploidy	A change in chromosome number from the species' normal diploid or haploid number, other than an exact multiple of the haploid number (polyploidy).
apical effect	An observable outcome in a whole organism, such as a clinical sign or pathologic state, that is indicative of a disease state that can result from exposure to a toxicant.
attributable risk	The proportion of disease cases that is attributable to the exposure.
benchmark dose	A dose or concentration that produces a predetermined change in the response rate of an adverse effect relative to the background response rate of this effect.
benchmark response	A predetermined change in the response rate of an adverse effect relative to the background response rate of this effect.
biomarker	Indicator of changes or events in biological systems. Biological markers of exposure refer to cellular, biochemical, analytical, or molecular measures that are obtained from biological media such as tissues, cells, or fluids and are indicative of exposure to an agent.

Term	Definition
chronic exposure	A continuous or intermittent long-term contact between an agent and a target. (Other terms, such as "long-term exposure," and "protracted exposure" are also used.)
clastogenicity	The disruption or breakage of chromosomes, leading to sections of the chromosome being deleted, added, or rearranged.
confounding	The mixing of the effects from the exposure of interest with the effects from other factor(s) on the risk of the adverse effect.
cumulative risk	The aggregate human health risk to the target entity caused by the accumulation of risk from multiple stressors.
cytotoxicity	The harmful effects to cell structure or function that ultimately leads to cell death.
dose	Total amount of an agent administered to, taken up by, or absorbed by an organism, system, or (sub) population.
dose-response	The relationship between the amount of an agent administered to, taken up by, or absorbed by an organism, system, or population and the change developed in that organism, system, or population in reaction to the agent. Related term: *exposure-response*.
dose-response assessment	The analysis of the relationship between the total amount of an agent administered to, taken up by, or absorbed by the organism, system, individual, or population and the changes developed in that the organism, system, individual, or population in reaction to that agent. The products of the dose-response assessment are unbiased estimates of the risk per unit dose that are used in risk characterization.
excess relative risk	A measure of association equivalent to the relative risk -1.

Term	Definition
exposure	Contact between an agent and a target. Contact takes place at an exposure surface over an exposure period.
exposure assessment	The process of estimating or measuring the magnitude, frequency, and duration of exposure to an agent, along with the number and characteristics of the population exposed. Ideally, it describes the sources, pathways, routes, and the uncertainties in the assessment
exposure duration	The length of time over which continuous or intermittent contacts occur between an agent and a target.
exposure event	The occurrence of continuous contact between an agent and a target.
exposure frequency	The number of exposure events in an exposure duration.
exposure index	A measured or estimated quantity of exposure or dose.
exposure model	a conceptual or mathematical representation of the exposure process
exposure pathway	The course an agent takes from the source to the target
exposure route	The way in which an agent enters a target after contact (e.g., by ingestion, inhalation, or dermal absorption)
exposure scenarios	A combination of facts, assumptions, and inferences that define a discrete situation where potential exposures may occur. These may include source, exposed population, period of exposure, microenvironment(s), and activities. Scenarios are often created to aid exposure assessors in estimating exposure.

Term	Definition
exposure-response	The relationship between the intensity, frequency, or duration of exposure to a stressor or agent and the intensity, frequency, or duration of the subsequent biological response of the organism. Given varied usage of the terms 'dose' and 'exposure' in many settings, exposure-response and dose-response are often used interchangeably. Related terms: *concentration-response*, *dose-response*.
genotoxicity	A general description of all types of DNA or chromosome damage, such as breaks, adducts, mutations, chromosome aberrations, and aneuploidy.
hazard	The inherent property of an agent (or situation) having the potential to cause an adverse effect when an organism, system, or population is exposed to that agent.
hazard function	In survival analyses, the hazard function is the rate of failure at an instant in time, *t*, given that the individual survives up to *t*.
hazard identification	Identification of the type and nature of adverse effects that an agent has an inherent capacity to cause in an organism, system, or population. Hazard identification is the initiation stage of the risk assessment. The products of hazard identification are definitions of the agent and outcome used in dose-response analysis.
hazard ratio	In survival analysis it is the hazard function (rate) of one individual (e.g., the exposed) divided by another individual (e.g., the unexposed), typically holding all other predictors constant (i.e., a rate ratio).

Term	Definition
immediately dangerous to life or health (IDLH)	An exposure condition or environment that is likely to cause death or immediate or delayed permanent adverse health effects or prevent escape from such an environment. The IDLH values developed by NIOSH characterize these high-risk exposure concentrations and conditions.
information bias	A bias in the effect estimate that occurs from systematic inaccuracies in the measurement of either the exposure or the adverse effect.
intake	The process by which an agent crosses an outer exposure surface of a target without passing an absorption barrier, i.e., through ingestion or inhalation.
interpretation bias	A bias that arises from improper inference or speculation based on a naïve or deliberate lack of impartiality by the interpreter.
job-exposure matrix (JEM)	A cross-classification of jobs/tasks and exposure level spanning a specified period. The JEM is used to estimate exposure indices that vary by job and time.
key event	An empirically observable precursor step that is itself a necessary element of the mode of action or is a biologically based marker for such an element.
limit of detection (LOD)	For an analytical procedure, the lowest amount or concentration of the analyte that is reliably distinguishable from the absence of analyte (i.e., low false negative rate). For example, for air sampling methods, NIOSH defines the LOD as the mass of the analyte that gives a mean signal that is three standard deviations above the mean blank signal.

Term	Definition
limit of quantification (LOQ)	For an analytical procedure, the amount or concentration of the analyte at which quantitative results can be reported with a high degree of confidence. The 'high degree of confidence' is based on a set of acceptance criteria that are assay-specific. For example, for air sampling methods, NIOSH defines the LOQ as the larger of: a) the mass corresponding to the mean blank signal + 10 standard deviations of the blank signal, or b) the mass above which recovery is ≥75%.
lowest observed adverse effect level (LOAEL)	The lowest dose or concentration at which there is biologically significant increases in frequency or severity of adverse effects between the exposed population and its appropriate control group.
mechanism of action	The underlying biochemical interactions, usually at the molecular level, that lead to the mode of action at the cellular level and ultimately the expression of the adverse effect. The mechanism of action is a more detailed understanding and description of events than is meant by mode of action.
mode of action (MOA)	A sequence of key events and processes, starting with interaction of an agent with a cell, proceeding through operational and anatomical changes, and resulting in the adverse effect.
no observed adverse effect level (NOAEL)	The highest dose level at which there are no biologically significant increases in the frequency or severity of adverse effects between the exposed population and its appropriate control; some effects may be produced at this dose level, but they are not considered adverse or precursors of adverse effects observed.
occupational exposure level (OEL)	The allowable concentration or intensity of a hazardous agent in the worker's work environment over a period. Generally expressed as an 8-hour time weighted average or as a short-term exposure limit of 15 or 30 minutes.

Term	Definition
odds ratio	A measure of association in comparative studies, particularly case-control studies. It is the ratio of the odds that an outcome will occur given a particular exposure to the odds of the outcome occurring in the absence of that exposure.
pharmacokinetics	The study of the absorption, distribution, metabolism, and elimination of exogenous chemicals in biological systems.
point of departure (PoD)	The estimate of dose-response at an exposure in the low range of (or just below) the observable data where extrapolation bias is limited, and a variety of approaches is available for its estimation.
point of departure (PoD)	The estimate of dose response at an exposure in the low range of (or just below) the observable data, and a variety of approaches is available for its estimation. The simplest approach defines the PoD as the no observed adverse effect level (NOAEL) or the lowest observed adverse effect level (LOAEL) from an animal or human epidemiology study.
publication bias	A bias from an editorial preference for publishing particular findings, which distorts inferences made from available evidence.
random error	The variation of results and inferences from the truth, occurring only because of chance.
rate ratio	A measure of association that quantifies the relation between an exposure and a health outcome from an epidemiologic study, calculated as the ratio of incidence rates or mortality rates of two groups.

Term	Definition
relative risk	The ratio of the risk (disease probability) that is observed in the exposed (intervention) group to that observed in the unexposed (control) group. However, relative risk is used as a general term for measures of association on a relative scale, including risk ratio, rate ratio, hazard ratio, odds ratio, standardized incidence ratio, and standardized mortality ratio.
reliability	The extent to which multiple assessments are consistent.
risk assessment	The determination of the relationship between the predicted exposure and adverse effects in four major steps: hazard identification, dose–response assessment, exposure assessment and risk characterization. NIOSH tasks supporting the risk assessment process primarily involve hazard identification, dose–response assessment, and risk characterization; however, exposures are assessed during hazard identification.
risk-based decision	A risk management decision using risk assessment as the basis for decision-making.
risk characterization	The qualitative and, wherever possible, quantitative determination, including attendant uncertainties, of the probability of occurrence of known and potential adverse effects of an agent in workers under defined exposure conditions.
risk-informed decision	A risk management decision using risk assessment as an input to decision-making.
risk management	The managerial, decision-making, and control process intended to avert intolerable risk.

Term	Definition
risk ratio	A measure of association that quantifies the association between an exposure and a health outcome from an epidemiologic study, calculated as the ratio of incidence proportions of two groups.
short-term exposure limit (STEL)	The acceptable 15-minute TWA average exposure that should not be exceeded at any time during a workday.
similar exposure groups	Workers having the same exposure profile because of similarity and frequency of tasks performed.
standardized incidence ratio (SIR)	The ratio of the observed number of disease cases in the study population to the number of cases that would be expected, based on disease rates in the referent population that are applicable to the characteristics (e.g., age, race, gender, calendar period) in the study population.
standardized mortality ratio (SMR)	The ratio of the observed number of deaths in a study population to the number of deaths that would be expected, based on death rates in the referent population that are applicable to the characteristics (e.g., age, race, gender, calendar period) in the study population.
stressor	Any physical, chemical, biological, or psychosocial entity that can induce an adverse effect.
target	Any biological entity that receives an exposure or a dose (e.g., an organ, an individual, or a population).
underlying cause of death	The disease or injury, which initiated the train of events leading directly to death, or the circumstances of the accident or violence, which produced the fatal injury.
uptake	The process by which an agent crosses an absorption barrier.

Term	Definition
validity	The quality of being logically or factually sound; the extent to which the measure describes that which is being measured.

1

APPENDIX B: SOURCES OF ERROR

National Institute for Occupational Safety and Health

1

2

SELECTION BIAS

The term 'selection bias" is used to describe many biases that are themselves a distortion in the estimate of effect that results from the manner in which the study subjects are selected from the source population [Gail and Benichou 2000]. These include biases resulting in differential followup, recall, self-selection, volunteering or non-response, and sampling-frames. Selection bias is possible in all observational studies and particularly so in case-control studies because the outcome is known at study inception. For example, MacMahon et al. [1981b] conducted a hospital based case-control study that reported a strong association between coffee drinking and pancreatic cancer. Controls were selected from "… all other patients who were under the care of the same physician in the same hospital at the time of an interview with a patient with pancreatic cancer." [MacMahon et al. 1981b]. This selection process resulted in a large proportion of controls who presented mainly with gastrointestinal disorders; therefore, these patients may have been advised by physicians not to consume coffee [Feinstein et al. 1981; MacMahon et al. 1981a; Silverman et al. 1983]. The abnormally low odds of coffee consumption among controls would cause a spurious positive association between coffee intake and pancreatic cancer. This bias may have been avoided by selecting controls from patients hospitalized for conditions not requiring a change in diet [Silverman et al. 1983]. Primary control of selection bias is managed by study design. The avoidance of selection bias in case-control studies is accomplished by drawing cases and controls from the same study base; therefore, it is imperative that the study base be well defined before sampling. Other methods include maximizing participation rates, using randomized sampling protocols, and applying sound inclusion/exclusion criteria.

There is often overlap between confounding and selection bias; therefore, secondary control of selection bias can sometimes be achieved by treating identifying factors as confounders in analyses and controlling for confounding accordingly. For example, if union workers are more likely to participate in a study and be exposed compared to office workers, then partial control of the bias may be realized by including job information as a confounder. Finally, sensitivity analyses using an array of inclusion criteria can help characterize the potential for significant selection bias and define a data set for use in risk assessment.

The risk assessor should be able to recognize potential sources of selection bias based on his or her evaluation of the study design. The risk assessor should give more weight to studies that have best addressed this source of bias through design, control, and sensitivity analyses. The risk assessor should pay special attention to (non-nested) case-control studies, which are most vulnerable to selection biases.

When reviewing the design of studies for the potential for selection bias, risk assessors should consider the following questions:

- What study design was used and where does this design fall in the hierarchy for WoE? Preferred studies will provide a detailed description of the study design, which includes limitations that are inherent to the design.
- Has the study population been sufficiently described to determine potential differences between study and control groups (i.e., do inclusion criteria differ between groups)? Preferred studies will include a detailed description of the characteristics of the study and control groups.
- What methods were used to select study participants? How could those excluded from study have effected study results had they been included? Preferred studies will include a description of the exclusion and inclusion criteria used for study participation and the methods used to reduce the potential for bias.
- What steps were taken to maximize participation rates? Low participation is indicative of a potential for selection bias.
- Is participation non-differential with respect to exposure? Case-control studies are particularly vulnerable to differential selection to study and control groups with respect to exposure given that case status is known at enumeration.
- Is their significant loss to followup? Loss to followup is typically less than 10% in well-designed studies.

INFORMATION BIAS

Information bias, sometimes referred to as 'data collection bias', 'measurement error', or 'misclassification bias', is a distortion in the effect estimate that occurs when the measurement of either the exposure or the adverse effect is systematically inaccurate [Gail and Benichou 2000]. In this context, information bias is a study execution bias that is restricted to data on study participants (i.e., the sample population). Information biases may stem from errors in the measurement instrument (instrument bias), data source, (data source bias), the observer or investigator (observer bias), and/or the subject (subject bias). Given limitations in available data, observational studies are particularly prone to several sources of information bias. For example, exposure data can be biased when collected with prior knowledge of case status (as in a case-control study). If exposure is self-reported, then a recall bias (a form of subject bias) may result from differential self-reporting of exposure status among cases and control group when cases are aware of a potential association between exposure and their disease. If exposure data are collected by interview, then the interviewer must be blinded to case status to reduce the potential for an observer bias. Likewise, if measurement data are collected, then care must be taken to ensure that identical procedures were used for both cases and controls. In general, when assessing the presence of information bias in a study under review, the risk assessor should ask:

- Was the information on the adverse effect obtained in the same way for all comparison groups?
- Was the information on exposure?

Errors in the data are usually separated by data type, such that the term 'measurement error' is reserved for errors in continuous data and 'misclassification error' refers to errors in discrete data. Measurement errors of explanatory variables used in analyses are unavoidable, even in the best-designed studies. Risk assessors should have a firm understanding of the potential effects from these errors in studies selected for dose-response assessment; therefore, a detailed discussion is provided in the following section. This discussion is primarily in context of errors in the measurement of the exposure of interest; however, the concepts presented are shared by all data sources vulnerable to an information bias.

Measurement Error and Misclassification

In the context of exposure, measurement (observation) error refers to any discrepancy between the true exposure, *X*, and the imperfect measured value, *W*; thus, it is analogous to exposure misclassification[2]. By this strict definition, measurement error comprises both systematic and random components. Random errors are stochastic fluctuations in observed values around the true (but unknown) value without directional preference. Systematic error or bias refers to inaccuracies in measured values that are inherent to the measurement system. For example, a common source of systematic error in exposure estimates are methods used to report 'nondetects', i.e., measurements below a detection threshold [Helsel 2005]. In these cases, the true value lies somewhere between the null and the detection threshold. In practice, nondetects are typically recorded as either zero (likely underestimation of exposure), the limit of detection (LOD) (likely overestimation of exposure), or simply omitted (a bias in either direction depending of the use of the data). Here, the LOD is the lowest amount or concentration of the analyte that is reliably distinguishable from the absence of analyte. For example, in developing air-sampling methods, NIOSH defined the LOD as the mass of the analyte that gives a mean signal that is three standard deviations above the mean blank signal [NIOSH 1995]. Methods to account for nondetects can range from simple substitution (e.g., substituting with LOD/2 or LOD/$2^{0.5}$) to complex parametric and nonparametric statistical modeling [Helsel 2005; NCRP 2010].

In general, measurement error reduces statistical power for trend tests because of added variance and may bias effect measures in dose-response analyses. The influence on dose-response modeling results depends on the combination of error characteristics and model specification, and can range from negligible effects to a strong bias in either direction [Armstrong 1998; Nieuwenhuijsen 2010]. Furthermore, measurement error is often thought of only in terms of the primary predictor; however, risk assessors should be mindful that a dose-response relation could also be strongly influenced by measurement error in covariates that confound or mediate effects of interest.

[2]Although used to describe all measurement error, the term 'misclassification' is sometimes limited to errors in qualitative indices and replaced with 'misspecification' when referencing errors in numerical indices.

Risk assessors should be reasonably assured that the data selected for dose-response analyses are free of a potential for significant bias. This assurance is partly gained through rigorous adherence to exposure assessment methods designed to avoid bias, such as observance of the data hierarchy, blinding assessors to case status, using and comparing multiple indices, and validating estimates. Well-conducted epidemiologic studies typically pay careful attention to obvious sources of systematic error in exposure estimates, but analyses have been generally conducted without considering residual measurement error effects [Jurek et al. 2006]. This is because assessments of measurement errors often require elaborate tests of reliability and validity, which are infrequently performed, if even feasible. Furthermore, many investigators assume that random measurement error always induces bias toward a null association; therefore, they incorrectly conclude it cannot cause spurious positive findings. Consequently, information needed to account for measurement error in risk analyses may be lacking. When data are available, researchers have suggested methods for adjusting estimates to account for random error or assessing its potential effects in dose-response analyses [Carroll et al. 2006; French et al. 2004; Hoffman et al. 2007; Maldonado 2008; Mallick et al. 2002; Meliker et al. 2010; Schafer et al. 2001; Spiegelman and Valanis 1998; Stayner et al. 2007; Thomas et al. 1993a]. These methods can involve integrating validation data into regression calibration or likelihood based methods to produce adjusted estimates [Hoffman et al. 2007; Mallick et al. 2002; Schafer et al. 2001; Spiegelman and Valanis 1998] or using Monte Carlo simulation to predict a range of plausible estimates [French et al. 2004; Meliker et al. 2010; Stayner et al. 2007]. In the former case, risk assessors should be cautious of adjustments made based on inadequate information that could induce a potentially stronger bias relative to unadjusted values.

In summary, there may be few options available to risk assessors in regards to limiting the potential effects of measurement error in dose-response analyses. Nevertheless, it is important for risk assessors to have a fundamental understanding of measurement error and its associated effects so that they can better describe and account for the limitations in analyses that support quantitative risk assessment. There are a number of seminal works on measurement error and dose-response modeling that should be reviewed [Armstrong 1998; Carroll et al. 2006; Fuller 1987; Ron and Hoffman 1999; Thomas et al. 1993b]. Some general concepts are discussed below.

Differential versus Nondifferentially Error

Exposure measurement error is either differential or nondifferential contingent on its relation to the dependent variable (e.g., disease status). Error that is independent of case status (and other predictors) is said to be nondifferential. It is commonly thought that nondifferential error results in bias toward a null association, which is a proven condition of binary variables or continuous variables in which the magnitude and direction of the measurement error are independent of the true value (i.e., the 'classical' error model, see Table B1). However, there are examples of nondifferential error in polytomous and continuous exposure measures that induce bias away from the null [Dosemeci et al. 1990; Greenland and Gustafson 2006; Wacholder 1995].

Table B1. Direction of bias caused by nondifferential measurement error of the primary predictor variable.

Predictor Scale	Bias expected
Binary	Biases the effect measure toward a null association
Polytomous	Estimates of trend across adjacent categories are biased downward. Estimates from comparison of categories can be biased in either direction.
Numerical	Classical error biases regression coefficients toward zero. Berkson error (i.e., random error that is statistically independent from the observed variable), leads to little or no bias in coefficients in most regression models.

Differential error can result in serious bias in either direction. For example, workers diagnosed with leukemia may be more apt to report or may have more thorough histories of benzene exposure than workers who are cancer free. In this instance, leukemia cases will appear to have higher exposures, thus biasing the association between benzene exposure and leukemia away from the null. Differential exposure error is unlikely if exposure data are collected prior to the disease outcome or without prior knowledge of the hypothesized association. Therefore, the primary means to avoid differential error is to ensure that exposure assessment methods were blinded to case status and that case ascertainment is, to the extent practical, independent of exposure status. When using data from previous studies, risk assessors' should examine the study design for any potential weaknesses that may lead to differential measurement error. Common sources of exposure information that are vulnerable to

differential measurement error are self or proxy reports, medical records, and compensation records). Cautious interpretation of studies involving case ascertainment by differential diagnoses of 'occupational diseases' (e.g., silicosis, asbestosis, malignant mesothelioma, and chronic beryllium disease) is also warranted. There are also certain situations in which nondifferential exposure information can be restructured to induce differential misclassification [Dosemeci et al. 1990; Flegal et al. 1991; Wacholder et al. 1991]. For example, combining categories of a polytomous exposure variable or constructing exposure categories from continuous exposure data can result in differential measurement error.

Categorical Indices

Measurement error in qualitative data is typically described as probabilities of exposure misclassification. For example, the error in a dichotomous exposure index can be expressed by its probability of correctly classifying an exposed worker (i.e., sensitivity) and the probability of correctly classifying an unexposed worker (i.e., specificity). A matrix of misclassification probabilities can be used to describe errors in indices with more than two levels. Misclassification probabilities are generally determined in validity studies comparing exposure estimates for a sample of workers in the study to estimates derived from another source that is believe to be as precise or better. Random (nondifferential) measurement error in a dichotomous exposure variable will always attenuate its effect; i.e., suggest that the agent under study is less toxic than it truly is. Trends across ordered categories of polytomous exposure variables will also be attenuated by nondifferential measurement error; however; comparisons between categories can be biased in either direction [Armstrong 1998].

Error Models for Numerical Indices: Classical versus Berkson Error

Two approaches to modeling random measurement error for numerical data are classical and Berkson. Settings where observations are subject to random variation from factors such as instrument imprecision and recording errors may be amenable to a classical model of measurement error, e.g., $W=X+U$, where measurement error, U, is a random variable with mean zero, variance, σ_U^2, and is independent of X. The observed exposure is equal to the true (but unobserved) exposure plus some measurement error; therefore, average values obtained

from replicate measures are unbiased estimates of the true exposure but will always have greater variability than the true exposure. In contrast, the Berkson error model, expressed as $E(X|W=w) = w$, arises when a single estimate, w, is applied to several individuals who actually have differing values of the quantity being estimated that average to w. In this model, the true exposure is more variable than the observed exposure. For example, assigning the average measured concentration from an ambient air monitor to the group of workers can be modeled by Berkson error. Modeling of the measurement errors may be approached using additive or multiplicative structures under each of these approaches.

The error form is significant with regard to dose-response analyses. For example, consider the simple case of a univariate linear dose response model: $E(\boldsymbol{Y}) = \alpha + \beta_x \boldsymbol{X}$, where the regression of response variable, $\boldsymbol{Y}$, on the independent variable, $\boldsymbol{X}$ (with variance, σ_X^2), has parameters α and β. If $\boldsymbol{X}$ is unavailable and exposure measure, $\boldsymbol{W}$, with classical additive error (i.e., $\boldsymbol{W}=\boldsymbol{X}+\boldsymbol{U}$) is substituted the resulting regression model $E(\boldsymbol{Y}) = \alpha^* + \beta_W \boldsymbol{W}$ has the slope parameter $\beta_W = \beta_X \sigma_X^2/(\sigma_X^2 + \sigma_U^2)$, where the quantity $(\sigma_X^2 + \sigma_U^2)$ is the variance of the measured variable. The ratio of true to measured value variances [referred to by Fuller (1987) as the reliability ratio, λ] must be less than unity; therefore, classical error results in attenuation of the observed linear dose-response [Fuller 1987]. The degree of attenuation is relative to the quantity σ_U^2/σ_X^2, such that smaller measurement error or larger spread of true values reduce bias. For example, the effect of classical error in a cumulative dose estimate is likely less compared to a single measurement of the same magnitude, given that the error of the single measurement is larger relative to that of multiple measurements comprising the cumulative dose. In contrast, attenuation of linear regression coefficients does not result from Berkson error. Recall that for Berkson error, $E(\boldsymbol{X}|\boldsymbol{W}) = \boldsymbol{W}$, thus $E(\boldsymbol{Y}|\boldsymbol{W}) = \alpha + \beta_x \boldsymbol{W}$ and therefore the estimator (β_X) is not attenuated [Carroll et al. 2006].

These observations, as described for a linear response, are qualitatively equivalent in logistic and loglinear dose-response models. However, recall that the effect measure in a logistic model (e.g., relative risk, RR) is found by RR=exp(β). Thus, the quantitative relation for attenuation of the slope coefficient can be expressed by using the reliability ratio: $\beta_W = \beta_X^{\lambda}$.

Additive, Multiplicative, and Mixed Error Structures

Error structures (for numerical values) are generalized by two limiting cases: additive error, in which the variance is constant for different magnitudes of the measurand; and multiplicative error, in which the variance increases with increasing values of the measurand. The classical multiplicative error model can be expressed by $W=Xe^{U}$, such that there is additivity on the logarithmic scale [i.e., $\ln(W)=\ln(X)+U$]. Measurement error can be modeled using additive, multiplicative or a combination of each resulting in a mixed error structure. Replicate measurements of several occupational agents have shown a multiplicative error structure. As in the additive measurement error model, the increased variance from multiplicative error attenuates the observed dose-response; however, the effect is larger at higher exposures, resulting in the appearance of downward curvature with increasing values of the error-prone measurements of exposure [Carroll et al. 2006]. An attenuated response at higher exposure levels has been observed in numerous occupational studies and in simulations [Carroll et al. 2006; Stayner et al. 2003; Steenland et al. 2015]. Nevertheless, there is considerably less literature on accounting for multiplicative or mixed error structures in predictor variables of dose-response regression models. The subsequent effects on these models vary by structure; therefore, some notion of the error structure is important for understanding subsequent model limitations.

Errors in Confounders and Effect Modifiers

Generally, random measurement error in a confounder, in which the error is not correlated with other measures or the exposure of interest, tends to increase confounding from that covariate [Armstrong 1998]. This means that the effect measure of interest is likely to lie between the unadjusted value (crude measure) and a value obtained under complete control of confounding (i.e., residual confounding from incomplete control). The directionality of induced bias depends on the direction of the confounding effect. The amount of bias depends on the strength of the confounder and the reliability ratio. As in confounding, random measurement error in an effect modifier tends to attenuate its effect modification; therefore; the ability to observe risk difference among groups is diminished [Armstrong 1998].

Errors in Adverse Effect Definitions

Information bias is also plausible from misclassification or measurement error in the adverse effect. For example, consider a cancer incidence study of U.S. workers in which cases in the exposed population are ascertained from a single state registry. If incidence rates in the exposed group are compared to standardized national rates, then the resulting effect measure (e.g., SIR) is likely biased from underascertainment of cases due to some migration out of the state by the workforce. Thus, cancer incidence studies within the U.S. are improved with ascertainment involving multiple states. In this case, the misclassification is differential and the bias is likely toward a null association. As a similar example, consider a case control study in which the study population and adverse effect data are drawn from electronic health records (EHRs). As in the previous example, cases may be missed if diagnosed outside of the EHR catchment area (e.g., a single clinic or group of clinics) and the potential for error increases with decreasing catchment area size. In this scenario, the affluent workers in the study have a more flexible health insurance plan; therefore, they are more likely to be diagnosed outside of the catchment area (and be missed). These same workers may have less exposure because of their job assignment. Under these conditions, the misclassification is differential with respect to exposure. As previously discussed, the resulting bias can be in either direction [Wang et al. 2016].

Misclassification can also occur from differences in diagnostic criteria used for defining the adverse effect. These criteria can vary by data source and by time. For example, the ICD published by the World Health Organization has been the standard diagnostic tool used for epidemiology since the late 1940s. This manual has been revised 10 times since its inception, with each revision updating diagnostic codes for diseases based on changes in diagnostic criteria. The definitions of certain diseases (e.g., hematopoietic cancers) have dramatically changed over the course of the ICD; therefore, studies published at different times may not have comparable disease definitions. As another example, consider an adverse effect defined using data abstracted from medical records. The reliability and validity of data in each individual medical record is vulnerable to different interpretations of different scenarios and often by different observers [Worster and Haines 2004].

Researchers acquire adverse effect data using several different approaches, which are typically tailored to the response definition, data availability, and study feasibility. Thus, the adverse effect data may stem from direct measurements (e.g., lung function tests), existing health outcomes databases (e.g., National Death Index, disease registries, compensation databases), medical records (paper or EHRs), and patient (or proxy) self-report. These sources are not without error and the potential for bias is dependent on the magnitude of these errors. For example, EHRs appear to be a promising source of medical information suitable for risk assessment. However, data residing in these systems are inputted by imperfect systems and persons. Sources of misinformation associated with medical records are include physician misdiagnoses, flawed laboratory results, erroneous patient self-reporting, and others [Ash et al. 2004; Burnum 1989; Luck et al. 2000; Worster and Haines 2004]. Thus, data collected prospectively using study criteria that were defined *a priori* are likely to be superior to data abstracted from EHRs.

In summary, the potential for bias from mismeasurement of the adverse effect is reduced when case definitions and ascertainment are the same among comparison groups. Nondifferential misclassification of the adverse effect with respect to risk factor exposure will likely result in an underestimation of the effect (i.e., bias toward a null association), whereas differential misclassification may result in a bias in either direction. When selecting studies for data synthesis, the risk assessor should confirm consistency in adverse effect definition among comparison groups in data used for risk assessment. Studies with well-defined adverse effects that are consistent throughout observation should be given more weight. For example, data from a compulsory reporting system (e.g., cancer registry) is preferred to information gathered by self-report. Studies with poorly defined adverse effects should be avoided. Risk assessors must also consider limitations that are inherent to the sources of adverse effect data. The risk assessor must consider the potential bias in estimates that may result from errors in the source data and weight the evidence accordingly.

CONFOUNDING

With respect to causal inference, confounding has been described as a ‘mixing’ of the effects from extraneous factors (confounders) with the effect of interest [Checkoway et al. 2004]. This mixing occurs when the

comparison groups (e.g., exposed and unexposed workers) have differing background risks of disease. There are many definitions of confounders; however, perhaps the most complete is that suggested by McNamee (2003), who posited that a factor should be considered a confounder if three conditions are met:

1. The factor is a cause of the disease, or a surrogate measure of a cause, in unexposed people. Factors satisfying this condition are called risk factors.
2. The factor is correlated, positively or negatively, with exposure in the study population. If the study population is classified into exposed and unexposed groups, this means that the factor has a different prevalence in the two groups.
3. The factor is not affected by the exposure (i.e., does not reside on the causal pathway) [McNamee 2003].

Disease risk factors can comprise a wide array including demographic factors (age, sex, race), lifestyle factors (smoking habits, diet, and alcohol use), or exposures to other agents in the workplace or elsewhere. In planning the study, all known or suspected risk factors should be identified, especially those factors most apt to confound dose-response associations. This information is needed to achieve appropriate confounding control and characterize the potential influence on effect measures from residual confounding.

Methods to control for confounding are generally related to study design or analysis. Design methods are meant to ensure that the exposed group is comparable to or exchangeable with the referent group with respect to the potential confounders [Greenland et al. 1999]. Exchangeability is the concept that response distributions in exchangeable comparison groups are identical under the same exposure conditions. These methods include restriction, randomization (i.e., clinical trial), and matching on potential confounders. In practice, there is limited success in finding exchangeable comparison groups in observational studies; therefore, these studies tend to rely on analytical methods for controlling confounding, such as stratified analyses and multiple regression. For example, dose-response analysis in a longitudinal study may use Poisson regression to control for confounding effects of age (an important confounder for most chronic illnesses) on the exposure interest by either background stratifying on age or including age as a covariate in the model. Similarly, a nested case control study of the same

cohort may use conditional logistic regression with age (attained age of the case) as the time scale. Both approaches are used extensively in occupational epidemiology.

In general, reasonable control for important demographic risk factors (e.g., age sex, and race) and calendar period is achieved in most published epidemiologic studies. However, measures of association are still vulnerable to confounding effects from incomplete control of measured risk factors or from the lack of controlling unmeasured risk factors. Smoking is known to cause several types of cancer and nonmalignant disease. If smoking prevalence is also related to exposure status then the potential for confounding by smoking exists. The resultant bias could be in either direction (i.e., positive or negative confounding) depending on the smoking characteristics of the comparison populations. For example, consider that blue-collar workers tend to use tobacco products more than white-collar workers. If blue-collar worker are also more likely to be exposed than white-collar workers (a reasonable assumption in some workplaces) then smoking can be a correlate of exposure [Lee et al. 2004; Stellman et al. 1988]. Under these conditions, smoking could confound the effect of occupational exposure on a smoking related disease. The expected effect in this particular case is positive confounding of the exposure effect by smoking, which means the measure of association will be biased away from the null without control for smoking. Unfortunately, information on the smoking habits of workers in most longitudinal studies is rarely available; therefore, direct adjustment for confounding effects of smoking are seldom seen. Instead, researchers might use indirect methods for adjustment [Axelson and Steenland 1988; Richardson 2010]. In the example above, job descriptions could be used as a proxy for smoking. Socioeconomic status is a well-known proxy for many lifestyle factors, including smoking, which may confound a dose-risk relationship [Lantz et al. 1998; McFadden et al. 2008]. One could also examine alternative adverse effects that are strongly associated with the unknown confounder but not with the exposure of interest [Richardson 2010]. An observed (but unexpected) dose-risk relationship between the agent of interest and alternative adverse effects is indicative of residual confounding. At the very least, researchers should provide some information on the potential for significant bias because of incomplete control of known or suspected confounders.

Confounding can also occur in occupational studies that do not account for concomitant workplace exposures. For example, Sathiakumar et al. (2015) examined the relationship between styrene exposure and leukemia in a large pooled study of North American synthetic rubber workers [Sathiakumar et al. 2015]. Styrene-exposed workers were also exposed to 1, 3-butadiene, which is a known human leukemogen [Humans 2012]. Quantitative estimates of cumulative exposure to 1,3-butadiene and styrene were calculated. Statistical analyses used Cox proportional hazards regression models with age as the time scale and adjusting for race, year of birth, and plant. Modest positive dose-response associations between leukemia and cumulative exposures to both agents were observed in separate models; however, the independent effects of styrene exposure could not be determined because of its strong correlation with 1,3-butadiene. Thus, the carcinogenic effects of these agents in combination appear hopelessly entangled in these workers, and the dose-response observed for styrene could be due wholly, or in part, to unmeasured confounding by 1,3-butadiene.

Whether a study is a valid contributor to hazard identification depends on how well the published results address the potential for confounding. In turn, resultant datasets must also inform and support the analytical approaches used in in the subsequent dose-response assessment. Thus, the risk assessor should evaluate the adequacy for control of measured and unmeasured confounders in studies under review. When unmeasured risk factors are identified, the risk assessor should evaluate the steps taken by researchers to reduce the potential for significant bias from residual confounding by these risk factors. The risk assessor should also consider the potential for unknown risk factors and assess their potential impact on internal validity. In all instances, the risk assessor should give more weight to studies with measures of association that are least likely to be affected by residual confounding.

HEALTHY WORKER EFFECTS

Another important source of potential bias in occupational studies is healthy worker effects (HWE). These effects primarily stem from two points of selection: 1) into the study at the time of hire and 2) out of the workforce at time of termination, and as such are commonly referred to as the ‘hire effect’ and ‘survivor effect’, respectively [Arrighi and Hertz-Picciotto 1993; Arrighi and Hertz-Picciotto 1994; Fox and Collier 1976]. A third

aspect to HWE is the natural decline in health status with time since hire [Checkoway et al. 2004]. Given distinct differences in the sources of potential bias and methods available for control, risk assessors should consider these aspects separately as they are discussed in the following sections.

Hire effect

The healthy worker hire effect results from increased employment eligibility among healthier persons, which can be exacerbated by hiring practices that screen against poor health (e.g., pre-employment exams). These conditions can result in a group of interest (i.e., workers) who is in better overall health than the comparison group irrespective of exposure status. Hire effects are typically observed in external comparisons (e.g., SMR or SIR studies using the general population as referent); however, some employers have used medical screening information for job placement within the industry, which could bias results from internal comparisons. For external comparisons, the hire effect results in a deficit in risk compared to true effects, particularly in chronic diseases most associated with lifestyle factors (e.g., diabetes mellitus and cardiovascular diseases). An alternative comparison group, such as similar working population that is unexposed to the agent of interest, may reduce this bias. When job assignment is influenced by medical screening, the direction of the potential bias depends on the relationship between job assignment and exposure. Risk assessors need to be wary of the potential for strong hire effects in data from external comparisons. When available, data from internal comparisons should be preferred for dose-response analyses of working populations. When internal comparison data are available, the risk assessor should evaluate the potential for bias from continued medical surveillance.

Survivor Effect

The survivor effect occurs when healthy workers continue to work and unhealthy workers leave employment prematurely, or are reassigned to less hazardous work due to their poor health. A potential exacerbating factor is a possible health benefit from employment compared to unemployment, such as the beneficial effect of physical exertion in reducing cardiovascular risk. In any case, exposure is always a condition of employment, which may be conditional on exposure, health status, or both. The likely effect from these relationships is attenuation of the estimated dose-response.

Methods for mitigating survivor effects vary. Control of these effects in longitudinal studies tend to involve one or more factors: age at hire, employment duration, employment status, time-since hire, and age at risk [Checkoway et al. 2004]. Methods have typically involved confounding control by restriction, matching (stratifying) or by covariate adjustment. For example, controls have included restricting the analysis to participants alive after a minimum length of time since hire; adjusting for employment status as a time-dependent variable; and using time lags (exposure windows). However, it is now recognized that the nature of the healthy worker survivor effect may preclude complete control of the effect by traditional approaches [Buckley et al. 2015]. This is because the deleterious health effect from a prior exposure may affect employment status (i.e., violates the third aspect of a confounder, as shown in Figure B1). Of course, a key consideration is whether it is reasonable for exposure to influence employment status. For example, strong survivor effects are much less likely to occur in late onset adverse effects (e.g., malignant mesothelioma) compared to debilitating effects (or precursor effects) that present during employment years (e.g., occupational asthma). Thus, risk assessors must evaluate the severity and likelihood of survivor effects based on the spatial and temporal relationships between employment, exposure, and outcome.

When there is a potential for a strong survivor bias, studies have recently employed methods, such as g-estimation [Bjor et al. 2015; Chevrier et al. 2012; Naimi et al. 2014; Picciotto et al. 2016] and g-formula [Cole et al. 2013; Neophytou et al. 2016] in structural nested models or accelerated failure time models. Although promising, studies using these methods are currently sparse; therefore, the evidence available for hazard identification will likely be restricted to studies in which residual survivor bias is likely when prior exposure affects employment status. Nevertheless, these new methods may be well suited for dose-response modeling in the dose-response assessment.

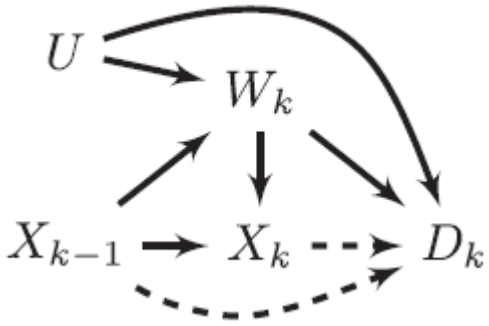

Figure B1. (Adapted from Buckley et al., 2015). Directed acyclic graphs of survivor bias in estimates of the association between cumulative occupational exposure (X_k) and disease (D_k) in a longitudinal study with time indexed by *k*. U_k is an unmeasured variable (eg, health status). Employment status (W_k) is a time-varying confounder affected by prior exposure (X_{k-1}).

A survivor bias may also result from underestimation of prevalent cases in cross-sectional studies when the adverse effect causes persons to leave employment or move to less hazardous jobs. For example, Eisen et al. [1997] identified significant selection bias in estimates of asthma prevalence in a cross-sectional study of workers exposed to metal working fluids. In that study, workers transferred to a job with less exposure because of the onset of asthma symptoms. This resulted in underestimating disease prevalence in those exposed and subsequently overestimating prevalence among unexposed persons at the time of the health survey (i.e., a negative dose response). However, a reanalysis of the data using exposure and disease status at the time of asthma onset instead of time of survey revealed significant excess risk. When evaluating occupational cross-sectional studies for risk assessment, risk assessors should determine whether there is a potential for the adverse effect to influence work status. Studies in which influence is likely should be avoided, unless the healthy worker effect has been adequately addressed in the study design and execution.

Length of followup

Although often given less attention, the length that a working population is followed in longitudinal studies is an important consideration when evaluating the potential for HWE. The strength of a healthy worker hire effect diminishes with increasing time since first employment; thus, the problem is partially addressed by

increased length of followup or stratifying by length of followup. Risk assessors should be wary of studies of chronic (or latent) adverse effects that have relatively short followup periods. For example, a recent cohort study of mortality patterns among paid Australian firefighters reported a statistically significant deficit in risk of cancer death (SMR=0.81 95% CI 0.72, 0.90); however, the average length of followup was less than 16 years [Glass et al. 2016]. In contrast, a comparable study of U.S. career firefighters reported excess cancer mortality (SMR=1.14; 95% CI: 1.10, 1.18) in a cohort with average followup of 29 years [Daniels et al. 2014]. The relatively short followup period in the Australian study is unlikely to counter the selection effects due to pre-employment health criteria for firefighters. In addition to person-years at risk, the percent deceased is a useful indicator of the length and quality of followup in cohort studies, especially in examinations of adverse effects that generally occur late in life (e.g., cancer). Using the previous example, less than 5% of the cohort of full-time Australian firefighters were deceased compared to over 40% of the U.S. firefighter cohort.

In the previous examples, we discussed the potential for selection bias toward a null association with decreased followup. Studies of chronic diseases characterized by a short latent period and short-lived risk after exposure may provide for excess risks that decrease with increasing length of followup. For example, leukemia risks that were attenuated with increased followup have been observed in follow-on studies of working populations exposed to benzene and ionizing radiation [Boice et al. 2011; Daniels et al. 2013; Rinsky et al. 2002; Silver et al. 2002]. Thus, the risk assessor must also consider effect modification by temporal factors that are associated with the length of followup.

OTHER POTENTIALLY IMPORTANT SOURCES OF BIAS

Publication and Interpretation Biases

A publication bias refers to an editorial preference for publishing particular findings, which distorts inferences made from available evidence. For example, a positive results bias may occur when authors and editors are more likely to publish positive findings compared to null findings. Publication bias can also occur when there is reluctance to publish disparate or controversial results, or when publication is driven by an emerging issue such that preliminary data are more likely to be published. Publication bias is plausible in all studies; however,

observational and experimental animal studies are more susceptible compared to randomized clinical trials [Easterbrook et al. 1991]. In all cases, published data can misinform on the consistency of evidence used for hazard identification. Risk assessors should be cognizant of the potential for publication bias and give appropriate attention to all findings, including those from negative studies. Attempts should be made to uncover relevant unpublished works.

Interpretation bias arises from improper inference or speculation based on a naïve or deliberate lack of impartiality by the interpreter. In this case, the 'interpreter' refers to study researchers, who interpret their findings at publication, or risk assessors, who translate findings for risk assessment. Research objectivity is always challenged given an ever-present interaction between data and judgement; therefore, interpretation is somewhat dependent on opinion, notion, or conviction [Kaptchuk 2003]. A smaller potential for significant interpretation bias is likely found in the peer-reviewed literature compared to trade journals and commercially funded technical reports. Information on potential conflicts of interest or disclosures can be useful in assessing the potential for interpretation bias. A willingness to examine alternative interpretations by investigators and risk assessors alike will lessen the potential for bias. Rigorous peer and public reviews also aid in avoiding interpretation bias. Nonetheless, it is often difficult to assess interpretation bias based on the information at hand, thus a bias cannot be ruled out.

Effect Modification and Interaction

The terms 'effect modification' and 'interaction' have been used interchangeably in the literature; the former seemingly preferred by epidemiologists and the latter by statisticians. It has been proposed that, in the strictest sense, these terms describe different phenomena [Vanderweele 2009]. Effect modification is described as condition in which the exposure-related effect varies by levels of an extraneous factor [Checkoway et al. 2004]. Typically, the extraneous factor is a descriptor of subpopulations (e.g., gender, race); therefore, effect modification may elucidate susceptibly differences in the population. For example, suppose a study reported an association between exposure to agent X and lung cancer in women but not in men. In this case, gender is the effect modifier of agent X for causing lung cancer. There is one intervention (exposure) and the susceptible

population is women. In contrast, an interaction specifically refers to the effect of two exposures together to be different (either more or less) than the combination of the two effects considered separately. Thus, an interaction describes the casual effects of the two exposures combined. For example, the joint effects of radon exposure and smoking status on lung cancer differs such that the excess relative risk per unit of radon exposure among nonsmokers is higher than that of smokers [Lubin et al. 1995]. In this case, there are two possible interventions (smoking and radon exposure). Interaction effects can range from profoundly antagonistic to strongly synergistic. Unfortunately, most studies available for risk assessment have not examined effect modification by factors other than by race or gender and information is usually insufficient to draw conclusions on potential interactions [Knol and VanderWeele 2012].

Random Error

Random error is the variation of results and inferences from the truth, occurring only because of chance. Effect measures are influenced by random variation in many components of an epidemiologic study. For example, a major contributor to random error in human studies is the process used to select study participants. This process is referred to as sampling and the random error contribution is known as sampling variation or sampling error [Rothman et al. 2008]. Random variation around true values related to estimates used in statistical models is another source of random error.

The common measure of random error in an estimation process is its variance, and the inverse of variance is a measure of statistical precision of the estimate. Precision can be improved by increasing the sample size, thus reducing the variance. This variance can also be reduced for a given sample size through design improvements; this is referred to as increasing study efficiency [Rothman et al. 2008]. Typically, the random error that is associated with the point estimate reported in a study is reflected by its associated confidence interval or p-value. In hazard identification, more weight is generally given to effect estimates with better precision (e.g., narrower confidence intervals). Nevertheless, estimate precision does not reflect a lack of bias from systematic errors. Moreover, random measurement error can also lead to biased estimates of the dose-response [Carroll et al. 2006].

For example, as previously discussed, when explanatory variables (e.g., dose) are measured with error, the regression coefficient in dose-response models is typically bias toward the null.

Unfortunately, deleterious effects of random error are rarely accounted for in epidemiologic studies, although some studies of health effects associated with ionizing radiation have made recent headway. In particular, regression calibration and Monte Carlo simulation have been used sparingly to account for uncertainty in dose-response analyses in studies relying on complex dosimetry systems subject to shared and unshared measurement error [Fearn et al. 2008; Pierce et al. 2008; Simon et al. 2015; Stram et al. 2015; Zhang et al. 2017].

APPENDIX B REFERENCES

Armstrong BG [1998]. "Effect of measurement error on epidemiological studies of environmental and occupational exposures." Occup Environ Med *55*(10): 651-656.

Arrighi HM, Hertz-Picciotto I [1993]. "Definitions, sources, magnitude, effect modifiers, and strategies of reduction of the healthy worker effect." J Occup Med *35*(9): 890-892.

Arrighi HM, Hertz-Picciotto I [1994]. "The evolving concept of the healthy worker survivor effect." Epidemiology *5*(2): 189-196.

Ash JS, Berg M, Coiera E [2004]. "Some Unintended Consequences of Information Technology in Health Care: The Nature of Patient Care Information System-related Errors." Journal of the American Medical Informatics Association *11*(2): 104-112.

Axelson O, Steenland K [1988]. "Indirect methods of assessing the effects of tobacco use in occupational studies." American Journal of Industrial Medicine *13*(1): 105-118.

Bjor O, Damber L, Jonsson H, Nilsson T [2015]. "A comparison between standard methods and structural nested modelling when bias from a healthy worker survivor effect is suspected: an iron-ore mining cohort study." Occup Environ Med *72*(7): 536-542.

Boice JD, Jr., Cohen SS, Mumma MT, Ellis ED, Eckerman KF, Leggett RW, Boecker BB, Brill AB, Henderson BE [2011]. "Updated mortality analysis of radiation workers at Rocketdyne (Atomics International), 1948-2008." Radiat Res *176*(2): 244-258.

Buckley JP, Keil AP, McGrath LJ, Edwards JK [2015]. "Evolving methods for inference in the presence of healthy worker survivor bias." Epidemiology *26*(2): 204-212.

Burnum JF [1989]. "The misinformation era: the fall of the medical record." Ann Intern Med *110*(6): 482-484.

Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM [2006]. Measurement error in nonlinear models : a modern perspective. Vol. Boca Raton, FL: Chapman & Hall/CRC, pp.

Checkoway H, Pearce N, Kriebel D [2004]. Research methods in occupational epidemiology. Vol. New York: Oxford University Press, pp.

Chevrier J, Picciotto S, Eisen EA [2012]. "A comparison of standard methods with g-estimation of accelerated failure-time models to address the healthy-worker survivor effect: application in a cohort of autoworkers exposed to metalworking fluids." Epidemiology *23*(2): 212-219.

Cole SR, Richardson DB, Chu H, Naimi AI [2013]. "Analysis of occupational asbestos exposure and lung cancer mortality using the G formula." American Journal of Epidemiology *177*(9): 989-996.

Daniels RD, Bertke S, Waters KM, Schubauer-Berigan MK [2013]. "Risk of leukaemia mortality from exposure to ionising radiation in US nuclear workers: a pooled case-control study." Occup Environ Med *70*(1): 41-48.

Daniels RD, Kubale TL, Yiin JH, Dahm MM, Hales TR, Baris D, Zahm SH, Beaumont JJ, Waters KM, Pinkerton LE [2014]. "Mortality and cancer incidence in a pooled cohort of US firefighters from San Francisco, Chicago and Philadelphia (1950-2009)." Occup Environ Med *71*(6): 388-397.

Dosemeci M, Wacholder S, Lubin JH [1990]. "Does nondifferential misclassification of exposure always bias a true effect toward the null value?" American Journal of Epidemiology *132*(4): 746-748.

Easterbrook PJ, Gopalan R, Berlin JA, Matthews DR [1991]. "Publication bias in clinical research." The Lancet *337*(8746): 867-872.

Eisen EA, Holcroft CA, Greaves IA, Wegman DH, Woskie SR, Monson RR [1997]. "A strategy to reduce healthy worker effect in a cross-sectional study of asthma and metalworking fluids." American Journal of Industrial Medicine *31*(6): 671-677.

Fearn T, Hill DC, Darby SC [2008]. "Measurement error in the explanatory variable of a binary regression: regression calibration and integrated conditional likelihood in studies of residential radon and lung cancer." Stat Med *27*(12): 2159-2176.

Feinstein AR, Horwitz RI, Spitzer WO, Battista RN [1981]. "Coffee and Pancreatic Cancer: The Problems of Etiologic Science and Epidemiologic Case-Control Research." JAMA: The Journal of the American Medical Association *246*(9): 957-961.

Flegal KM, Keyl PM, Nieto FJ [1991]. "Differential misclassification arising from nondifferential errors in exposure measurement." American Journal of Epidemiology *134*(10): 1233-1244.

Fox AJ, Collier PF [1976]. "Low mortality rates in industrial cohort studies due to selection for work and survival in the industry." Br J Prev Soc Med *30*(4): 225-230.

French NHF, Goovaerts P, Kasischke ES [2004]. "Uncertainty in estimating carbon emissions from boreal forest fires." Journal of Geophysical Research D: Atmospheres *109*(14).

Fuller WA [1987]. Measurement error models. Vol. New York: Wiley, pp.

Gail MH, Benichou J [2000]. Encyclopedia of epidemiologic methods. Vol. Chichester ; New York: Wiley, pp.

Glass DC, Pircher S, Del Monaco A, Hoorn SV, Sim MR [2016]. "Mortality and cancer incidence in a cohort of male paid Australian firefighters." Occup Environ Med.

Greenland S, Gustafson P [2006]. "Accounting for independent nondifferential misclassification does not increase certainty that an observed association is in the correct direction." American Journal of Epidemiology *164*(1): 63-68.

Greenland S, Robins JM, Pearl J [1999]. "Confounding and Collapsibility in Causal Inference." Stat Sci *14*(1): 29-46.

Helsel DR [2005]. Nondetects and data analysis : statistics for censored environmental data. Vol. Hoboken, N.J.: Wiley-Interscience, pp.

Hoffman FO, Ruttenber AJ, Apostoaei AI, Carroll RJ, Greenland S [2007]. "The Hanford thyroid disease study: An alternative view of the findings." Health Physics *92*(2): 99-111.

Humans IWGotEoCRt [2012]. "Chemical agents and related occupations." IARC Monogr Eval Carcinog Risks Hum *100*(Pt F): 9-562.

Jurek AM, Maldonado G, Greenland S, Church TR [2006]. "Exposure-measurement error is frequently ignored when interpreting epidemiologic study results." European Journal of Epidemiology *21*(12): 871-876.

Kaptchuk TJ [2003]. "Effect of interpretive bias on research evidence." British Medical Journal *326*(7404): 1453-1455.

Knol MJ, VanderWeele TJ [2012]. "Recommendations for presenting analyses of effect modification and interaction." International Journal of Epidemiology *41*(2): 514-520.

Lantz PM, House JS, Lepkowski JM, Williams DR, Mero RP, Chen J [1998]. "Socioeconomic factors, health behaviors, and mortality: results from a nationally representative prospective study of US adults." JAMA *279*(21): 1703-1708.

Lee DJ, LeBlanc W, Fleming LE, Gómez-Marín O, Pitman T [2004]. "Trends in U.S. smoking-rates in occupational groups: The National Health Interview Survey 1987-1994." Journal of Occupational and Environmental Medicine *46*(6): 538-548.

Lubin JH, Boice JD, Edling C, Hornung RW, Howe GR, Kunz E, Kusiak RA, Morrison HI, Radford EP, Samet JM, Tirmarche M, Woodward A, Yao SX, Pierce DA [1995]. "Lung cancer in radon-exposed miners and estimation of risk from indoor exposure." Journal of the National Cancer Institute *87*(11): 817-827.

Luck J, Peabody JW, Dresselhaus TR, Lee M, Glassman P [2000]. "How well does chart abstraction measure quality? A prospective comparison of standardized patients with the medical record." American Journal of Medicine *108*(8): 642-649.

MacMahon B, Yen S, Trichopoulos D, Warren K, Nardi G [1981a]. "Coffee and Cancer of the Pancreas." New England Journal of Medicine *304*(26): 1604-1606.

MacMahon B, Yen S, Trichopoulos D, Warren K, Nardi G [1981b]. "Coffee and cancer of the pancreas." N Engl J Med *304*(11): 630-633.

Maldonado G [2008]. "Adjusting a relative-risk estimate for study imperfections." Journal of Epidemiology and Community Health *62*(7): 655-663.

Mallick B, Hoffman FO, Carroll RJ [2002]. "Semiparametric regression modeling with mixtures of Berkson and Classical error, with application to fallout from the Nevada test site." Biometrics *58*(1): 13-20.

McFadden E, Luben R, Wareham N, Bingham S, Khaw KT [2008]. "Occupational social class, educational level, smoking and body mass index, and cause-specific mortality in men and women: a prospective study in the European Prospective Investigation of Cancer and Nutrition in Norfolk (EPIC-Norfolk) cohort." Eur J Epidemiol *23*(8): 511-522.

McNamee R [2003]. "Confounding and confounders." Occup Environ Med *60*(3): 227-234; quiz 164, 234.

Meliker JR, Goovaerts P, Jacquez GM, Nriagu JO [2010]. "Incorporating individual-level distributions of exposure error in epidemiologic analyses: An example using arsenic in drinking water and bladder cancer." Annals of Epidemiology *20*(10): 750-758.

Naimi AI, Cole SR, Hudgens MG, Richardson DB [2014]. "Estimating the effect of cumulative occupational asbestos exposure on time to lung cancer mortality: using structural nested failure-time models to account for healthy-worker survivor bias." Epidemiology *25*(2): 246-254.

NCRP [2010]. Radiation dose reconstruction : principles and practices. Vol. Bethesda, Md.: National Council on Radiation Protection and Measurements, pp.

Neophytou AM, Picciotto S, Costello S, Eisen EA [2016]. "Occupational Diesel Exposure, Duration of Employment, and Lung Cancer: An Application of the Parametric G-Formula." Epidemiology *27*(1): 21-28.

Nieuwenhuijsen MJ [2010]. Exposure assessment in occupational and environmental epidemiology. Vol. Oxford ; New York: Oxford University Press, pp.

NIOSH [1995]. Guidelines for air sampling and analytical method development and evaluation. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 95–117, pp.

Picciotto S, Ljungman PL, Eisen EA [2016]. "Straight metalworking fluids and all-cause and cardiovascular mortality analyzed by using G-estimation of an accelerated failure time model with quantitative exposure: Methods and interpretations." American Journal of Epidemiology *183*(7): 680-688.

Pierce DA, Væth M, Cologne JB [2008]. "Allowance for random dose estimation errors in atomic bomb survivor studies: A revision." Radiation Research *170*(1): 118-126.

Richardson DB [2010]. "Occupational exposures and lung cancer: Adjustment for unmeasured confounding by smoking." Epidemiology *21*(2): 181-186.

Rinsky RA, Hornung RW, Silver SR, Tseng CY [2002]. "Benzene exposure and hematopoietic mortality: A long-term epidemiologic risk assessment." Am J Ind Med *42*(6): 474-480.

Ron E, Hoffman FO [1999]. Uncertainties in radiation dosimetry and their impact on dose-response analyses. Vol. Bethesda, Md.: U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health, pp.

Rothman KJ, Greenland S, Lash TL [2008]. Modern epidemiology. Vol. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins, pp.

Sathiakumar N, Brill I, Leader M, Delzell E [2015]. "1,3-Butadiene, styrene and lymphohematopoietic cancer among male synthetic rubber industry workers--Preliminary exposure-response analyses." Chem Biol Interact *241*: 40-49.

Schafer DW, Lubin JH, Ron E, Stovall M, Carroll RJ [2001]. "Thyroid cancer following scalp irradiation: A reanalysis accounting for uncertainty in dosimetry." Biometrics *57*(3): 689-697.

Silver SR, Rinsky RA, Cooper SP, Hornung RW, Lai D [2002]. "Effect of follow-up time on risk estimates: a longitudinal examination of the relative risks of leukemia and multiple myeloma in a rubber hydrochloride cohort." Am J Ind Med *42*(6): 481-489.

Silverman DT, Hoover RN, Swanson GM, Hartge P [1983]. "The Prevalence of Coffee Drinking Among Hospitalized and Population-Based Control Groups." JAMA: The Journal of the American Medical Association *249*(14): 1877-1880.

Simon SL, Hoffman FO, Hofer E [2015]. "The two-dimensional Monte Carlo: a new methodologic paradigm for dose reconstruction for epidemiological studies." Radiat Res *183*(1): 27-41.

Spiegelman D, Valanis B [1998]. "Correcting for bias in relative risk estimates due to exposure measurement error: A case study of occupational exposure to antineoplastics in pharmacists." American Journal of Public Health *88*(3): 406-412.

Stayner L, Steenland K, Dosemeci M, Hertz-Picciotto I [2003]. "Attenuation of exposure-response curves in occupational cohort studies at high exposure levels." Scand J Work Environ Health *29*(4): 317-324.

Stayner L, Vrijheid M, Cardis E, Stram DO, Deltour I, Gilbert SJ, Howe G [2007]. "A Monte Carlo maximum likelihood method for estimating uncertainty arising from shared errors in exposures in epidemiological studies of nuclear workers." Radiation Research *168*(6): 757-763.

Steenland K, Karnes C, Darrow L, Barry V [2015]. "Attenuation of exposure-response rate ratios at higher exposures a simulation study focusing on frailty and measurement error." Epidemiology *26*(3): 395-401.

Stellman SD, Boffetta P, Garfinkel L [1988]. "Smoking habits of 800,000 American men and women in relation to their occupations." Am J Ind Med *13*(1): 43-58.

Stram DO, Preston DL, Sokolnikov M, Napier B, Kopecky KJ, Boice J, Beck H, Till J, Bouville A [2015]. "Shared dosimetry error in epidemiological dose-response analyses." PLoS One *10*(3): e0119418.

Thomas D, Stram D, Dwyer J [1993a]. "Exposure measurement error: Influence on exposure-disease relationships and methods of correction." Annual Review of Public Health *14*: 69-93.

Thomas D, Stram D, Dwyer J [1993b]. "Exposure measurement error: influence on exposure-disease. Relationships and methods of correction." Annu Rev Public Health *14*: 69-93.

Vanderweele TJ [2009]. "On the distinction between interaction and effect modification." Epidemiology *20*(6): 863-871.

Wacholder S [1995]. "When measurement errors correlate with truth: Surprising effects of nondifferential misclassification." Epidemiology *6*(2): 157-161.

Wacholder S, Dosemeci M, Lubin JH [1991]. "Blind assignment of exposure does not always prevent differential misclassification." American Journal of Epidemiology *134*(4): 433-437.

Wang L, Shaw PA, Mathelier HM, Kimmel SE, French B [2016]. "Evaluating risk-prediction models using data from electronic health records." Annals of Applied Statistics *10*(1): 286-304.

Worster A, Haines T [2004]. "Advanced Statistics: Understanding Medical Record Review (MRR) Studies." Academic Emergency Medicine *11*(2): 187-192.

Zhang Z, Preston DL, Sokolnikov M, Napier BA, Degteva M, Moroz B, Vostrotin V, Shiskina E, Birchall A, Stram DO [2017]. "Correction of confidence intervals in excess relative risk models using Monte Carlo dosimetry systems with shared errors." PLoS One *12*(4): e0174641.

APPENDIX C: EMERGING PRACTICES

National Institute for Occupational Safety and Health

OCCUPATIONAL EXPOSURE BANDING

Occupational Exposure Limits (OELs) play a critical role in protecting workers and emergency response personnel from exposure to dangerous concentrations of hazardous materials [Schulte et al. 2010]. In the absence of an OEL, determining the appropriate controls needed to protect workers from chemical exposures can be challenging. According to the EPA, the Toxic Substances Control Act (TSCA) Chemical Substance Inventory currently contains over 85,000 chemicals that are commercially available [EPA 2015] yet only about 1,000 of these chemicals have been assigned an authoritative (government, consensus, or peer reviewed) OEL. Furthermore, the rate at which new chemicals are being introduced into commerce significantly outpaces OEL development, creating a need for guidance on thousands of chemicals that lack reliable exposure limits [McKernan and Seaton 2014; Michaels 2014; Zalk and Nelson 2008]. Occupational exposure banding, also known as hazard banding or health hazard banding, is a systematic process that uses both qualitative and quantitative hazard information on selected health effect endpoints to identify potential inhalation-based exposure ranges or categories. The draft NIOSH occupational exposure banding process seeks to create a consistent and documented process to characterize chemical hazards so timely and well-informed risk management decisions can be made for chemicals lacking OELs [McKernan and Seaton 2014].

The concept of using hazard-based categories to communicate potential health concerns, alert workers to the need for risk management, and inform exposure control requirements is not new. Numerous hazard classification and category-based systems have seen extensive use in the occupational setting. Such systems are deeply embedded in occupational hygiene practice, particularly in the pharmaceutical industry, and are elements of well-developed, modern hazard communication programs (e.g., United Nations 2013 Globally Harmonized System of Classification and Labelling of Chemicals [2013]). The draft NIOSH occupational exposure banding process is distinguished from other hazard classification and category-based systems in several ways. These unique attributes of the NIOSH process include (1) a three-tiered system that allows users of varying expertise to utilize the process, (2) determination of potential health impacts based on eight toxicological endpoints separately, (3) hazard-based categories linked to quantitative exposure ranges, and (4) assessment of the process via extensive evaluation exercises to determine accuracy and repeatability.

To effectively apply the draft NIOSH occupational exposure banding process, it is important to understand the three tiers of the process. Each tier of the process has different requirements for data sufficiency, which allows a variety of stakeholders to use the process in many different situations. The most appropriate tier for banding depends on the availability and quality of the data, how it will be used, and the training and expertise of the user [McKernan and Seaton 2014]. While Tier 1 requires relatively little information and modest specialized training, each successive tier necessitates more chemical-specific data and more user expertise to assign an Occupational Exposure Band (OEB) successfully. A primary goal of Tier 1 is to give the user a quick summary of the most important health effects associated with exposure to the chemical of interest and quickly identify extremely toxic chemicals that should be considered for substitution or elimination. Tier 2 requires the user to examine a number of publicly available databases and extract relevant toxicological and weight-of-evidence data to be used in the NIOSH banding algorithm. Tier 3 employs expert judgment to critically evaluate experimental data and discern toxicological outcomes.

Another important component of the draft NIOSH occupational exposure banding process is the inclusion of five exposure bands. Occupational exposure banding uses limited chemical toxicity data to group chemicals into one of five bands ranging from A through E. These bands or OEBs, define the range of exposures expected to be protective of worker health [McKernan et al. 2016]. Band E is the most protective band for the most dangerous chemicals, while band A is the least protective for the least dangerous [McKernan et al. 2016]. One major benefit of occupational exposure banding is that the amount of time and data required to categorize a chemical into an OEB is significantly less than that required to develop an OEL.

The burden of worker and responder exposure to potentially hazardous chemicals that lack authoritative OELs is considerable, and the need for risk management for these chemicals is clear. Occupational exposure banding is one way to provide this type of guidance. An OEB provides more than a range of exposures that is expected to be protective of worker health. Rather, an OEB can be utilized to identify potential health effects and target organs, inform implementation of control interventions and preparedness plans, inform medical surveillance decisions, and provide critical information quickly.

EMERGING ALTERNATIVES TO ASSESSING APICAL ENDPOINTS

Apical and Non-apical Endpoints

An apical endpoint is an observable outcome in a whole organism, such as a clinical sign or pathologic state that is indicative of a disease state that can result from exposure to a toxicant [NRC 2007]. For risk assessment, it is typically the final stage of disease progression. Adverse effects are generally related to traditional apical endpoints such as death, reproductive failure, or developmental dysfunction [Villeneuve and Garcia-Reyero 2011]. In some instances, data on the apical effect are not available, therefore, the risk assessment may rely on a non-apical surrogate that lies on the adverse effect pathway (Figure C-1) between the molecular initiating event and the adverse effect (e.g., functional genomics and biomarkers). These sub-organism effects are sometimes referred to as precursor effects.

It has been suggested that the future of risk assessment is likely to shift away from toxicity testing of apical endpoints and move toward research evaluating biologically significant perturbations in toxicity pathways at earlier stages of the disease state. This research is anticipated to use a combination of computational biology and high-throughput *in vitro* tests of human cells and tissues [NRC 2007]. The advantages of *in vitro* non-apical toxicity testing is that it allows for: 1) broad coverage of chemicals, chemical mixtures, outcomes, and life stages, 2) reduced cost and time necessary of testing, 3) fewer animals used and less harm to animals, and 4) the development of a more robust scientific basis for assessing health effects from hazardous agents [NRC 2007].

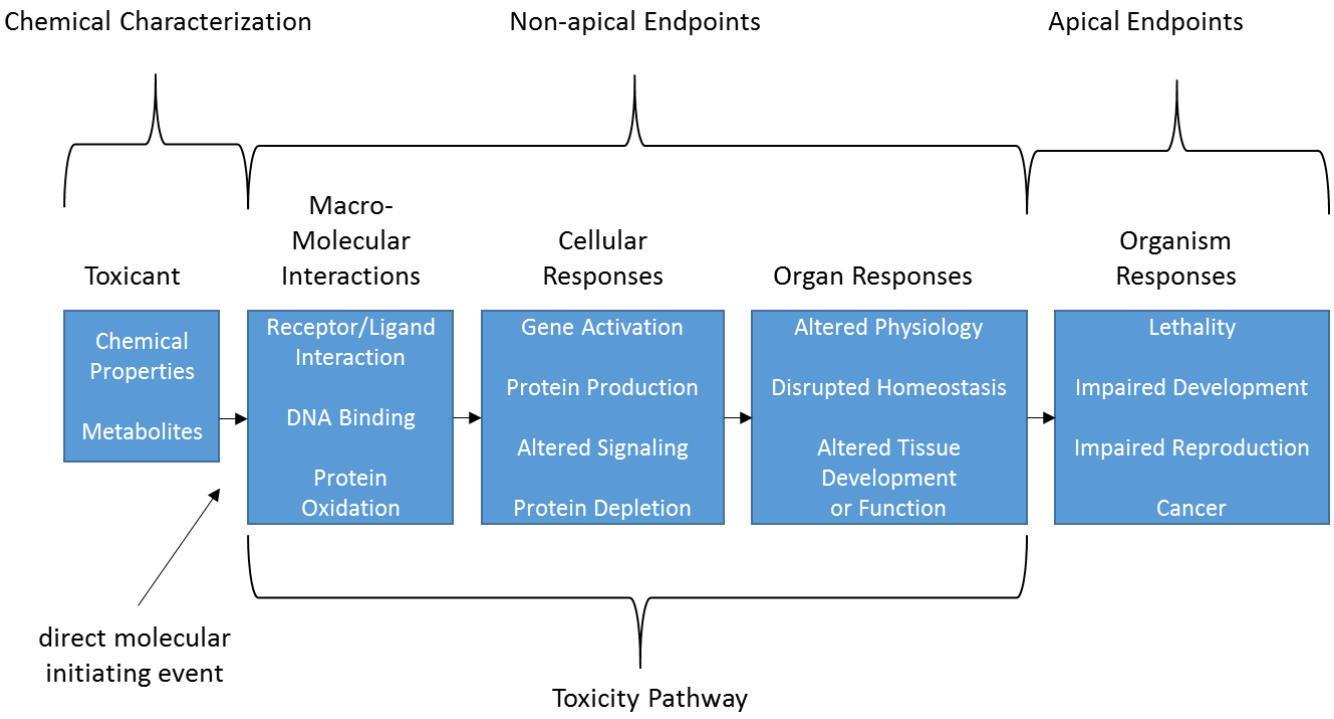

Figure C2. Diagram of the adverse outcome pathway. A chemical (or its metabolites) interacts with a molecular target to initiate leading a sequence of higher order effects to produce an adverse effect relevant to risk assessment. (Adapted from Ankley et al. [2010])

Biomarkers

A biomarker is defined as a characteristic that is objectively measured and evaluated as an indicator of normal physiologic processes, pathologic processes, pharmacologic responses to a therapeutic intervention, or susceptibility [Atkinson et al. 2001; Schulte 1993]. In the context of an adverse effect, the biomarker refers to a biological analyte that predicts the individual's disease state. By definition, biomarkers include conventional measures, such as blood pressure, blood cholesterol, and enzyme levels; however, recent advancements have focused on cellular, genetic, and molecular markers that are sought as screening tools for early diagnosis of a severe disease (e.g., lung cancer and cardiovascular disease). The utility of defining sets of responses based on multiple genomic, transcriptomic, proteomic, and metabolomic markers and processes and/or second messenger and other biochemical pathways is an evolving area of work [Cote et al. 2016].

Ideally, risk assessors prefer to measure early indicators of serious health effects, rather than wait for frank expression of disease. For example, lung cancer is a rare (<60 cases per 100,000 person-years) and serious

(<20% survival after 5-years from diagnosis) adverse effect observed predominantly in ages 65 years or older and at later stages of disease progression [Howlader et al. 2016]. Epidemiologic studies of occupational lung cancer require large populations who are observed over a long period to assure adequate statistical power for typical effect sizes observed. A biomarker intended for early indication of lung cancer could act to relax some of these design requirements. Research suggests that exhaled breath contains organic compounds from metabolic processes that can vary between healthy subjects and subjects with lung cancer; making it a potentially viable biomarker for early onset of disease [Dent et al. 2013]. If the relationships between dose, the exhaled breath condensate analytes of interest, and lung cancer can be adequately characterized, then exhaled breath condensate may also be a useful response quantity in future dose-response analyses of lung carcinogens.

Quantitative risk assessment of biomarkers is an area of active research and development and has only been successfully used for risk assessment in very limited situations [Cote et al. 2016; Poirier 2016]. As such information evolves, the risk assessor must be prepared to consider whether such exposure-biomarker associations are useful relationships to model in occupational risk assessment.

USE OF GENETICS AND EPIGENETICS IN RISK ASSESSMENT

A growing body of literature demonstrates that genetic and epigenetic factors condition biological responses to occupational and environmental hazards or serve as targets of them. Generally, genetic and epigenetic data might be used as endpoints in hazard identification, indicators of exposure, effect modifiers in exposure assessment and dose-response modeling, descriptors of mode of action, characterization of toxicity pathways. Vast amounts of genetic and epigenetic data may be generated by high-throughput technologies (described in Section 4.0 of this Appendix). Ideally, these data can be useful for assessing variability and reducing uncertainty in extrapolations, and help identify previously unidentified biological perturbations that may be of interest in risk assessment [Schulte et al. 2015].

One of the most critical areas to understand in the incorporation of genetic and epigenetic information in risk assessment is in the area of gene-environment interactions. The term “gene-environment” interaction can involve a range of interpretations of joint effects, including the risk of a single genotype across a range of environmental

exposures, or the risk of exposure across a range of genotypes. There are several approaches available to evaluate the impact of gene-environment interactions. Many potential methods are reviewed in Schulte et al. [2015].

Future risk assessments may involve acquired changes in the somatic genome or changes in the epigenetics, which comprises the factors influencing expression of the genome. Techniques for addressing these require deep knowledge of mechanisms of action of toxic agents, and well-defined experimental designs to address specific risk assessment questions. The elucidation of perturbations in genetic and epigenetic information on human health is likely to be a rich area for future risk assessment. A framework for organizing the research around these types of risk assessment questions can be found in Schulte et al. [2015]. One area where some progress has been made in developing genetics for quantitative risk assessment is in the use of high throughput analyses, as described in the next section.

MOLECULAR TOXICOLOGY AND HIGH THROUGHPUT ANALYSIS

Because of the lack of full toxicology data on many chemicals, NIOSH is investigating the utility of high throughput screening and in vitro short-term tests for occupational risk assessment. In the past decade, there has been an exponential increase in the publication of new toxicity data focusing on genomic analysis using high-through put screening and in-vitro short-term exposure. The paradigm for assessing chemical risks to human health is rapidly changing because of the availability of this toxico-genomic information and because of an increased understanding in the gene-environment interactions.

In 2007, the U.S. Environmental Protection Agency (EPA) requested the National Research Council (NRC) to conduct a complete review of toxicity-testing methods and strategies. NRC presented its long-range vision and strategy to advance toxicity testing [NRC 2007]. By recognizing the importance of NRC's vision, several federal agencies (EPA, National Institutes of Environmental Health Sciences/National Toxicology Program, National Institutes of Health and the Food and Drug Administration) formed a collaborative program known as Toxicity Testing in the 21st Century, (Tox 21). This program uses high-throughput screening methods and computational toxicology approaches to screen, rank, and prioritize chemicals for further testing and assessment. The Tox 21 program has screened more than 10,000 chemicals using approximately 70 *in vitro* cell-

based assays with 15-point dose response at the NIH Chemical Genomics Center using innovative robotic technology [Kavlock et al. 2009]. In addition to Tox 21, the EPA's Toxicity Forecaster, simply known as ToxCast which is part of Tox 21 has generated data for over 1,800 subset of chemicals from Tox21 inventory by expanding into more biological endpoints. ToxCast screens chemicals for dose-related changes in at least six-doses in over 700 high-throughput assays (both cell-based and cell-free) and 300 signaling pathways that cover a wide-range of cell responses [Richard et al. 2016]. The EPA's "Next Generation Risk Assessment: Recent Advances in Molecular, Computational, and Systems Biology", describes how new molecular, computational and systems biology data and approaches could better inform risk assessment [EPA 2014]. Overall, the screening data generated by these programs are used to predict the toxicity of chemicals and to prioritize the chemicals that need further comprehensive toxicity evaluation. In addition, the results from high-throughput analysis could be used in adverse outcome pathway (AOP) analysis, although the specifics for these have not yet been well worked out [Tollefsen et al. 2014].

Thomas et al. [2011; 2013] demonstrated a high degree of correlation between the BMD values for transcriptional changes and the corresponding apical endpoint changes in male Sprague Dawley and F344 rats and in female B6C3F1 mice exposed to various chemicals. The author went on to suggest that the transcriptional points of departure values could be used as a potential surrogates for both cancer and non-cancer points of departure. Kuppusamy et al. [2015] and Alyea et al. [2012] demonstrated that concordance exists between the changes in epigenetics and apical endpoints. Schulte et al. [2015] discussed the utilization of genetic and epigenetic data in occupational health risk assessment.

Applications of the molecular toxicology approach could include screening out problematic chemicals; identifying critical in vivo testing; prioritization of data-poor chemicals; replacing traditional testing with more efficient alternatives. Although the current effort in the utilization of molecular toxicology data looks promising, additional data and methods of analysis are needed. In order for these efforts to continue, a strong collaboration between agencies is needed.

QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIPS

The literature on Quantitative Structure Activity Relationship (QSAR, for a review of this topic see Roy et al. [2015]) models is vast, and has been widely applied in pharmaceutical research and risk assessment. In order to predict a response in untested chemicals, QSAR models link chemical, physical, and structural properties to a biological outcome using a mathematical model. Currently, no human health risk assessment has been based solely on a QSAR analysis; however, as the future of toxicity testing is moving away from animal testing [NRC 2007], such approaches may become more common. Ideally, QSAR approaches can link fundamental chemical properties to adverse outcome pathways, and eventually, whole organism response (e.g., cancer, death) [Ankley et al. 2010]. As the use of QSAR modeling in risk assessment is an emerging discipline, general guidelines outlined below should be followed; however, as the science for this field is still in its infancy, it is stressed that these guidelines are general and the individual approach should be tailored to the situation.

The first issue in the use of QSAR modeling in a risk assessment framework is the adverse outcome predicted and its relevance to human health. To date the majority of QSAR models focus on prediction of single outcomes such as the median lethal dose (i.e., LD50) from chemical structural properties. Such outcomes are often a gross measure of toxicity and say little about low levels of exposure. Others, which compute the lowest observed adverse effect level, or equivalent endpoint [Mumtaz et al. 1995], may be directly applicable to the risk assessment, but require strong assumptions that should be carefully reviewed. QSAR modeling is under development to predict the entire dose-response curve, which would provide additional information on toxicity. In sum, care should be taken when choosing the endpoint for a risk assessment. If the endpoint is a gross measure of toxicity, it may be useful to bin a chemical based upon its relative toxicity but unreasonable to provide an exposure level in the nature of an occupational exposure limit (OEL). Predicting the entire dose-response curve may have additional applications for quantitative risk assessment. Any QSAR based risk assessment should start out with exploring the limitations of the model and the predicted endpoint *a priori* and subsequent assessment should carefully consider these limitations.

Once the endpoint/model is chosen, it is important to assess the validity of the QSAR model. This is usually done in a statistical analysis that analyzes the prediction in terms of a leave one out (or leave many out)

hold out analysis. Here, the model is fit to a reduced data set and the held out data are predicted. Such analyses provide a useful tool to measure the accuracy of the model within the context of the entire data set tested. Note that chemicals beyond the scope of the dataset will be less likely to behave as predicted. The model should have a high degree of accuracy in prediction for the chemicals of interest, where accuracy is defined relative to the analysis at hand. Further, the model should be validated and a sensitivity analysis including various plausible assumptions and defaults for the model structure should be performed. Finally, as the estimates are based upon a limited or no data, the preliminary nature of the assessment should be stressed. If new data are made available that suggest the chemical is more or less toxic, the risk assessment should be updated with the new data within a reasonable timeframe.

NANOMATERIALS RISK ASSESSMENT

Overview

Given the large and growing number of engineered nanomaterials (ENMs) with limited data, as for other emerging and existing substances produced or used in the workplace, alternative test strategies (i.e., a toxicological approach other than primary animal testing), such as high throughput screening and *in vitro* exposures, may help to fill the gaps by providing data that could be used in validated hazard and risk assessment models. [Kuempel et al. 2012].

Dose Normalization *in Vitro* and *In Vivo*

As risk assessments begin to rely largely on *in vitro* data and *in silico* modeling, accurate description of dose both *in vitro* and *in vivo* will be key to evaluating these dose-response relationships and validating alternative test strategies for use in risk assessment [Gangwal et al. 2011; Oberdörster 2012]. Many *in vitro* studies have used doses that are much higher than occupationally equivalent lung doses [Gangwal et al. 2011]. Such studies could be useful for hazard identification and screening evaluations but may over predict the *in vivo* response. *In vitro* studies are also limited in the cell types represented and interactions among cells.

A challenge in quantifying the dose-response relationships *in vitro* is estimating the effective dose, i.e., the dose that reaches the target cells. The particle surface area doses to cells can differ significantly at a given

mass concentration (µg/ml), due to the differences in the specific surface area (m^2/g) of particles of different sizes and to differences in the sedimentation and diffusion properties of particles in liquid-based systems [Hinderliter et al. 2010]. The *In Vitro* Sedimentation, Diffusion and Dosimetry (ISDD) model was developed to account for differences in settling velocity in the liquid media based on particle size, density, and specific surface [Hinderliter et al. 2010]. Adjusting the *in vitro* dose to estimate the total surface area of nanoparticles that reaches cells in the petri dish was shown to better correlate with acute *in vivo* endpoints [Hinderliter et al. 2010; Teeguarden et al. 2007].

Correlation of *In Vitro* and *In Vivo* Responses

Several studies of ENMs have shown good correlation between the *in vitro* and acute *in vivo* inflammation-related responses to poorly soluble particles [Donaldson et al. 2008; Rushton et al. 2010; Zhang et al. 2012]. The dose metric in these studies differed, including comparison of either the total particle surface area to the total cell surface area *in vitro* or *in vivo* (cm^2/cm^2) [Donaldson et al. 2008], the response per unit particle surface area [Rushton et al. 2010], or the area under dose-response curve [Zhang et al. 2012]. The steepest portion of the dose-response slope showed the best correlation of *in vitro* with *in vivo* responses [Han et al. 2012; Rushton et al. 2010]. Responses included cell-free generation of reactive oxygen species (ROS), rat lung epithelial cell release of lactate dehydrogenase (LDH) or induction of protein oxidation endpoints, and rat pulmonary inflammation measured as polymorphonuclear leukocyte (PMN) response in bronchoalveolar lavage fluid (BALF) after intratracheal instillation (IT) exposure to different ENMs.

Pulmonary fibrosis *in vivo* (in rodents) and fibrosis-related markers *in vitro* (in rodent or human lung cells) have been shown to be correlated with exposure to some ENMs. Specifically, the activation of the NLRP3 inflammasome and pro-fibrogenic endpoints *in vitro* or fibrosis *in vivo* have been associated with exposure to carbon nanotubes [Hamilton et al. 2009; Li et al. 2013; Sager et al. 2014; Wang et al. 2012; Wang et al. 2011]. With further validation, an *in vitro* inflammasome activation assay may be useful for assessing the potential for chronic adverse effects of carbon nanotubes and other ENMs.

Alternative Methods for Nanomaterials

Comparative Potency Estimation

One promising use of alternative test strategies data is comparative potency analyses between nanomaterials and benchmark materials for use in the development of OEBs [Kuempel et al. 2012]. Benchmark materials can serve as points of references for comparison to ENMs. Benchmark materials are well-characterized substances within biological mode-of-action categories for which the health hazards are well known and quantitative risk estimates have been (or could be) developed [Kuempel et al. 2012; Nel et al. 2013]. Possible benchmark materials to evaluate inhalation hazards may include fine crystalline silica, asbestos, and ultrafine titanium dioxide and/or carbon black [Oberdörster et al. 2005]. These comparative toxicity analyses would be conducted *in vitro* for a set of ENMs, along with benchmark particles (including positive and negative controls or references), to which the new materials could be compared. The *in vitro* to *in vivo* dose-response relationships would be validated for the benchmark materials in specific assays. A parallelogram approach [Schoeny and Margosches 1989; Sobels 1977; Sutter 1995] has been used for comparative toxicity and risk estimation, and has been proposed for use in setting provisional OELs of pharmaceutical intermediates [Maier 2011]. Such comparative approaches could be used in deriving initial OELs or OEBs for individual ENMs or groups of ENMs [Kuempel et al. 2012].

The use of *in vitro* dose-response data to estimate a POD directly has been proposed, using methods similar to those used for *in vivo* data, including adjustment of the POD by uncertainty factors (initially until more evidence is available) [Crump et al. 2010].

Hazard Classification/Clustering

NIOSH and others are exploring methods to utilize physicochemical properties, such as particle size, shape, solubility, crystal structure, and chemical composition as predictors of a material's hazard potency, such as tested in high-throughput cellular studies and validated in limited rodent studies. Potency is the inverse of dose (i.e., higher potency substances are those with a lower dose associated with an adverse effect). In these ongoing analyses, NIOSH is investigating the dose-response relationships and substance-specific physicochemical data,

and using statistical learning methods such as Random Forests to identify groups of similarly behaving materials with respect to hazard potency [Drew et al. 2017]. The adverse outcomes of interest include pulmonary inflammation, fibrosis, cancer, and systemic effects associated with inhaled nanoscale particles. In current analyses of acute pulmonary inflammation, a set of 16 microscale and nanoscale particles in a training data set have been grouped into four potency clusters, including three groups for nanoscale particles, which are 4 – 175 times more potent than a fourth group containing a microscale reference particle. These analyses illustrate proof of concept for grouping particles by pulmonary hazard potency [Drew et al. 2017].

Next steps are to evaluate an *in vitro* dataset of some of the same materials as in the *in vivo* dataset to investigate the possible utility of *in vitro* studies of cellular responses to particle exposure that are involved in the *in vivo* mechanism of activation of pulmonary inflammation associated with particle exposure, including cytokine, gene transcription, and cell toxicity endpoints. Ultimately, it is envisioned that extended and validated analyses will be used as a framework to develop initial OEL categories or OEBs as hazard inputs into nanomaterial control banding tools [Drew et al. 2017; Kuempel et al. 2012].

Validation

A key challenge to utilizing alternative test strategies data is the development and application of validation criteria. Validation would include evaluation of variability across laboratories and selected assays of reference particles. Such evaluations for ENMs have shown considerable inter-laboratory variability in dose-response relationships for the same ENMs across laboratory [Bonner et al. 2013], especially in the *in vitro* assays [Xia et al. 2013].

To facilitate the validation and implementation of alternative test strategies data, standard sets of particle descriptors, dose metrics, and response parameters are needed to compare biological the mode-of-action and dose–response relationships across different studies [Kuempel et al. 2012]. *In vitro* data could be used in a tiered toxicology testing such that selected materials (e.g., highest and lowest toxicity within a category) in the *in vitro* assays would go on for *in vivo* testing.

NON-CHEMICAL AND CUMULATIVE RISK ASSESSMENT

With the exclusion of ionizing radiation, quantitative risk assessment of non-chemical stressors has received little attention in the risk assessment community. Moreover, risk assessment methods have largely focused on a single stressor, although risks often involve complex exposures to multiple stressors from multiple routes and pathways. Recently; however, there has been interest in assessing risks from non-chemical stressors separately and in combination with chemical exposures. For example, the National Research Council has recommended that risk assessors consider exposures to both chemical and non-chemical stressors as sources of cumulative risk [NRC 2009]. Such stressors can include physical, operational, and psychosocial domains. Examples include work stress, heat stress, noise exposures, and vibrational exposures.

Research into risk assessment methods for nonchemical stressors and cumulative risks is ongoing [Lentz et al. 2015; Lewis et al. 2011]. For example, NIOSH has developed methods similar to chemical risk assessment to assess some non-chemical hazards, such as ionizing radiation and noise [NIOSH 1998; NIOSH 1987]. Expanding these methods to include: other non-chemical hazards, the joint effects of multiple stressors, and the contribution of non-occupational stressors to occupational risk are areas of interest in NIOSH risk assessment.

APPENDIX C REFERENCES

Alyea RA, Moore NP, LeBaron MJ, Gollapudi BB, Rasoulpour RJ [2012]. "Is the current product safety assessment paradigm protective for epigenetic mechanisms?" J Pharmacol Toxicol Methods *66*(3): 207-214.

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, Serrrano JA, Tietge JE, Villeneuve DL [2010]. "Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment." Environ Toxicol Chem *29*(3): 730-741.

Atkinson AJ, Jr., Colburn WA, DeGruttola VG, DeMets DL, Downing GJ, Hoth DF, Oates JA, Peck CC, Schooley RT, Spilker BA, Woodcock J, Zeger SL [2001]. "Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework." Clinical Pharmacology and Therapeutics *69*(3): 89-95.

Bonner JC, Silva RM, Taylor AJ, Brown JM, Hilderbrand SC, Castranova V, Porter D, Elder A, Oberdorster G, Harkema JR, Bramble LA, Kavanagh TJ, Botta D, Nel A, Pinkerton KE [2013]. "Interlaboratory evaluation of rodent pulmonary responses to engineered nanomaterials: the NIEHS Nano GO Consortium." Environ Health Perspect *121*(6): 676-682.

Cote I, Andersen ME, Ankley GT, Barone S, Birnbaum LS, Boekelheide K, Bois FY, Burgoon LD, Chiu WA, Crawford-Brown D, Crofton KM, DeVito M, Devlin RB, Edwards SW, Guyton KZ, Hattis D, Judson RS, Knight D, Krewski D, Lambert J, Maull EA, Mendrick D, Paoli GM, Patel CJ, Perkins EJ, Poje G, Portier CJ, Rusyn I, Schulte PA, Simeonov A, Smith MT, Thayer KA, Thomas RS, Thomas R, Tice RR, Vandenberg JJ, Villeneuve DL, Wesselkamper S, Whelan M, Whittaker C, White R, Xia M, Yauk C, Zeise L, Zhao J, DeWoskin RS [2016]. "The Next Generation of Risk Assessment Multiyear Study-Highlights of Findings, Applications to Risk Assessment and Future Directions." Environ Health Perspect.

Crump KS, Chen C, Louis TA [2010]. "The future use of in vitro data in risk assessment to set human exposure standards: challenging problems and familiar solutions." Environ Health Perspect *118*(10): 1350-1354.

Dent AG, Sutedja TG, Zimmerman PV [2013]. "Exhaled breath analysis for lung cancer." Journal of Thoracic Disease *5*(SUPPL.5): S540-S550.

Donaldson K, Borm PJ, Oberdorster G, Pinkerton KE, Stone V, Tran CL [2008]. "Concordance between in vitro and in vivo dosimetry in the proinflammatory effects of low-toxicity, low-solubility particles: the key role of the proximal alveolar region." Inhal Toxicol *20*(1): 53-62.

Drew NM, Kuempel ED, Pei Y, Yang F [2017]. "A quantitative framework to group nanoscale and microscale particles by hazard potency to derive occupational exposure limits: Proof of concept evaluation." Regul Toxicol Pharmacol *89*: 253-267.

EPA [2014]. Next Generation Risk Assessment: Incorporation of Recent Advances in Molecular, Computational, and Systems Biology Washington, DC: National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency (EPA), EPA/600/R-14/004, pp. 1-196.

EPA. (2015). "Toxic Substances Control Act (TSCA) Chemical Substance Inventory." Retrieved June 6, 2016, from https://www.epa.gov/tsca-inventory.

Gangwal S, Brown JS, Wang A, Houck KA, Dix DJ, Kavlock RJ, Hubal EA [2011]. "Informing selection of nanomaterial concentrations for ToxCast in vitro testing based on occupational exposure potential." Environ Health Perspect *119*(11): 1539-1546.

Hamilton RF, Wu N, Porter D, Buford M, Wolfarth M, Holian A [2009]. "Particle length-dependent titanium dioxide nanomaterials toxicity and bioactivity." Part Fibre Toxicol *6*: 35.

Han X, Corson N, Wade-Mercer P, Gelein R, Jiang J, Sahu M, Biswas P, Finkelstein JN, Elder A, Oberdorster G [2012]. "Assessing the relevance of in vitro studies in nanotoxicology by examining correlations between in vitro and in vivo data." Toxicology *297*(1-3): 1-9.

Hinderliter PM, Minard KR, Orr G, Chrisler WB, Thrall BD, Pounds JG, Teeguarden JG [2010]. "ISDD: A computational model of particle sedimentation, diffusion and target cell dosimetry for in vitro toxicity studies." Part Fibre Toxicol *7*(1): 36.

Howlader N, Noone AM, Krapcho M, Miller D, Bishop K, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA [2016]. SEER cancer statistics review, 1975–2013. Bethesda, MD: National Cancer Institute, pp.

Kavlock RJ, Austin CP, Tice RR [2009]. "Toxicity testing in the 21st century: implications for human health risk assessment." Risk Anal *29*(4): 485-487; discussion 492-487.

Kuempel ED, Castranova V, Geraci CL, Schulte PA [2012]. "Development of risk-based nanomaterial groups for occupational exposure control." J Nanopart Res *14*: 1029.

Kuppusamy SP, Kaiser JP, Wesselkamper SC [2015]. "Epigenetic Regulation in Environmental Chemical Carcinogenesis and its Applicability in Human Health Risk Assessment." Int J Toxicol *34*(5): 384-392.

Lentz TJ, Dotson GS, Williams PRD, Maier A, Gadagbui B, Pandalai SP, Lamba A, Hearl F, Mumtaz M [2015]. "Aggregate Exposure and Cumulative Risk Assessment-Integrating Occupational and Non-occupational Risk Factors." Journal of Occupational and Environmental Hygiene *12*: S112-S126.

Lewis AS, Sax SN, Wason SC, Campleman SL [2011]. "Non-chemical stressors and cumulative risk assessment: An overview of current initiatives and potential air pollutant interactions." International Journal of Environmental Research and Public Health *8*(6): 2020-2073.

Li R, Wang X, Ji Z, Sun B, Zhang H, Chang CH, Lin S, Meng H, Liao YP, Wang M, Li Z, Hwang AA, Song TB, Xu R, Yang Y, Zink JI, Nel AE, Xia T [2013]. "Surface charge and cellular processing of covalently functionalized multiwall carbon nanotubes determine pulmonary toxicity." ACS Nano *7*(3): 2352-2368.

Maier MS [2011]. "Setting occupational exposure limits for unstudied pharmaceutical intermediates using an in vitro parallelogram approach." Toxicol Mech Methods *21*(2): 76-85.

McKernan L, Seaton M [2014]. "The banding marches on: NIOSH proposes a new process for occupational exposure banding." The Synergist(May 2014): 44-46.

McKernan L, Seaton M, Gilbert S [2016]. "The NIOSH Decision Logic for OEBs: Applying Occupational Exposure Bands." The Synergist(March 2016).

Michaels D [2014]. Press Release: Request for information on updating OSHA's chemical permissible exposure limits. Washington, DC, Occupational Safety and Health Administration.

Mumtaz MM, Knauf LA, Reisman DJ, Peirano WB, DeRosa CT, Gombar VK, Enslein K, Carter JR, Blake BW, Huque KI, et al. [1995]. "Assessment of effect levels of chemicals from quantitative structure-activity relationship (QSAR) models. I. Chronic lowest-observed-adverse-effect level (LOAEL)." Toxicol Lett *79*(1-3): 131-143.

Nations U [2013]. Globally Harmonized System of Classification and Labelling of Chemicals (GHS). New York and Geneva: United Nations, ST/SG/AC.10/30/Rev.5, pp. 536.

Nel AE, Nasser E, Godwin H, Avery D, Bahadori T, Bergeson L, Beryt E, Bonner JC, Boverhof D, Carter J, Castranova V, Deshazo JR, Hussain SM, Kane AB, Klaessig F, Kuempel E, Lafranconi M, Landsiedel R, Malloy T, Miller MB, Morris J, Moss K, Oberdorster G, Pinkerton K, Pleus RC, Shatkin JA, Thomas R, Tolaymat T, Wang A, Wong J [2013]. "A multi-stakeholder perspective on the use of alternative test strategies for nanomaterial safety assessment." ACS Nano *7*(8): 6422-6433.

NIOSH [1998]. Criteria for a Recomended Standard: Occupational Noise Exposure. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 98–126, pp.

NIOSH [1987]. A recommended standard for occupational exposure to radon progeny in underground mines. Washington, DC: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 88-101, pp. 1-215.

NRC [2009]. Science and Decisions: Advancing Risk Assessment. Vol. Washington (DC): National Academy Press, pp.

NRC [2007]. Toxicity testing in the 21st century: A vision and a strategy. Vol. Washington, DC: National Academies Press, pp. 1-216.

Oberdörster G [2012]. "Nanotoxicology: in vitro-in vivo dosimetry." Environ Health Perspect *120*(1): A13; author reply A13.

Oberdörster G, Maynard A, Donaldson K, Castranova V, Fitzpatrick J, Ausman K, Carter J, Karn B, Kreyling W, Lai D, Olin S, Monteiro-Riviere N, Warheit D, Yang H, Group IRFRSINTSW [2005]. "Principles for characterizing the potential human health effects from exposure to nanomaterials: elements of a screening strategy." Part Fibre Toxicol *2*: 8.

Poirier MC [2016]. "Linking DNA adduct formation and human cancer risk in chemical carcinogenesis." Environ Mol Mutagen *57*(7): 499-507.

Richard AM, Judson RS, Houck KA, Grulke CM, Volarath P, Thillainadarajah I, Yang C, Rathman J, Martin MT, Wambaugh JF, Knudsen TB, Kancherla J, Mansouri K, Patlewicz G, Williams AJ, Little SB, Crofton KM, Thomas RS [2016]. "ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology." Chem Res Toxicol *29*(8): 1225-1251.

Roy K, Kar S, Das RN [2015]. Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment. Vol.: Academic press, pp.

Rushton EK, Jiang J, Leonard SS, Eberly S, Castranova V, Biswas P, Elder A, Han X, Gelein R, Finkelstein J, Oberdorster G [2010]. "Concept of assessing nanoparticle hazards considering nanoparticle dosemetric and chemical/biological response metrics." J Toxicol Environ Health A *73*(5): 445-461.

Sager TM, Wolfarth MW, Andrew M, Hubbs A, Friend S, Chen TH, Porter DW, Wu N, Yang F, Hamilton RF, Holian A [2014]. "Effect of multi-walled carbon nanotube surface modification on bioactivity in the C57BL/6 mouse model." Nanotoxicology *8*(3): 317-327.

Schoeny RS, Margosches E [1989]. "Evaluating comparative potencies: developing approaches to risk assessment of chemical mixtures." Toxicol Ind Health *5*(5): 825-837.

Schulte PA [1993]. "Use of biological markers in occupational health research and practice." J Toxicol Environ Health *40*(2-3): 359-366.

Schulte PA, Murashov V, Zumwalde R, Kuempel ED, Geraci CL [2010]. "Occupational exposure limits for nanomaterials: state of the art." Journal of Nanoparticle Research *12*(6): 1971-1987.

Schulte PA, Whittaker C, Curran CP [2015]. "Considerations for Using Genetic and Epigenetic Information in Occupational Health Risk Assessment and Standard Setting." J Occup Environ Hyg *12 Suppl 1*: S69-81.

Sobels FH [1977]. "Some problems associated with the testing for environmental mutagens and a perspective for studies in "comparative mutagenesis"." Mutat Res *46*(4): 245-260.

Sutter TR [1995]. "Molecular and cellular approaches to extrapolation for risk assessment." Environ Health Perspect *103*(4): 386-389.

Teeguarden JG, Hinderliter PM, Orr G, Thrall BD, Pounds JG [2007]. "Particokinetics in vitro: dosimetry considerations for in vitro nanoparticle toxicity assessments." Toxicol Sci *95*(2): 300-312.

Thomas RS, Clewell HJ, 3rd, Allen BC, Wesselkamper SC, Wang NC, Lambert JC, Hess-Wilson JK, Zhao QJ, Andersen ME [2011]. "Application of transcriptional benchmark dose values in quantitative cancer and noncancer risk assessment." Toxicol Sci *120*(1): 194-205.

Thomas RS, Wesselkamper SC, Wang NC, Zhao QJ, Petersen DD, Lambert JC, Cote I, Yang L, Healy E, Black MB, Clewell HJ, 3rd, Allen BC, Andersen ME [2013]. "Temporal concordance between apical and transcriptional points of departure for chemical risk assessment." Toxicol Sci *134*(1): 180-194.

Tollefsen KE, Scholz S, Cronin MT, Edwards SW, de Knecht J, Crofton K, Garcia-Reyero N, Hartung T, Worth A, Patlewicz G [2014]. "Applying Adverse Outcome Pathways (AOPs) to support Integrated Approaches to Testing and Assessment (IATA)." Regul Toxicol Pharmacol *70*(3): 629-640.

Villeneuve DL, Garcia-Reyero N [2011]. "Vision & strategy: Predictive ecotoxicology in the 21st century." Environmental toxicology and chemistry / SETAC *30*(1): 1-8.

Wang X, Xia T, Duch MC, Ji Z, Zhang H, Li R, Sun B, Lin S, Meng H, Liao YP, Wang M, Song TB, Yang Y, Hersam MC, Nel AE [2012]. "Pluronic F108 coating decreases the lung fibrosis potential of multiwall carbon nanotubes by reducing lysosomal injury." Nano Lett *12*(6): 3050-3061.

Wang X, Xia T, Ntim SA, Ji Z, Lin S, Meng H, Chung CH, George S, Zhang H, Wang M, Li N, Yang Y, Castranova V, Mitra S, Bonner JC, Nel AE [2011]. "Dispersal state of multiwalled carbon nanotubes elicits profibrogenic cellular responses that correlate with fibrogenesis biomarkers and fibrosis in the murine lung." ACS Nano *5*(12): 9772-9787.

Xia T, Hamilton RF, Bonner JC, Crandall ED, Elder A, Fazlollahi F, Girtsman TA, Kim K, Mitra S, Ntim SA, Orr G, Tagmount M, Taylor AJ, Telesca D, Tolic A, Vulpe CD, Walker AJ, Wang X, Witzmann FA, Wu N, Xie Y, Zink JI, Nel A, Holian A [2013]. "Interlaboratory evaluation of in vitro cytotoxicity and inflammatory responses to engineered nanomaterials: the NIEHS Nano GO Consortium." Environ Health Perspect *121*(6): 683-690.

Zalk DM, Nelson DI [2008]. "History and evolution of control banding: A review." Journal of Occupational and Environmental Hygiene *5*(5): 330-346.

Zhang H, Ji Z, Xia T, Meng H, Low-Kam C, Liu R, Pokhrel S, Lin S, Wang X, Liao YP, Wang M, Li L, Rallo R, Damoiseaux R, Telesca D, Madler L, Cohen Y, Zink JI, Nel AE [2012]. "Use of metal oxide nanoparticle band gap to develop a predictive paradigm for oxidative stress and acute pulmonary inflammation." ACS Nano *6*(5): 4349-4368.