JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Report of the State-of-the-Science Workshop: Evaluation of Epidemiological Data Consistency for Application in Regulatory Risk Assessment

Submitted to the United States Environmental Protection Agency

June 30, 2011

Introduction and Background

Epidemiological Data and Regulatory Risk Assessment

Epidemiological data have often had a key role in the assessment of risks associated with exposure to chemicals and pollutants and for development of regulatory limits in the environment and in occupational settings. Examples include the development of occupational health standards and health risk values in the United States for benzene (OSHA 1987; U.S. EPA 2003), as well as for the U.S. National Ambient Air Quality Standards for Particulate Matter (U.S. EPA 2006). The strengths and weaknesses, as well as the overall value of the use of epidemiological data in regulatory risk assessment have been widely discussed in the scientific and public health policy literature.

Epidemiological data have been criticized as too often flawed by quality issues and incompletely controlled sources of bias (see, for example, Graham et al. 1995). Epidemiological studies of environmental agents involving typical ambient levels of exposure have been particularly characterized as uninformative or especially susceptible to bias and uncontrolled confounding because the target for estimation is often relatively small risk ratios that are dismissed as "weak associations" (Gamble and Lewis 1996). Proponents of use of epidemiological data, while acknowledging the limitations of observational studies, advance its strengths; the investigation of the effects of real exposures as received by the general population, the characterization of an effect across the full range of susceptibility in the population and, most significantly, the direct relevance of epidemiologic evidence to public health (Whittemore 1986; Gordis 1988; Hertz-Picciotto 1995; Burke 1995; Samet, Schnatter, and Gibb 1998). In addition, the ability to ascertain relatively low relative risks has improved with advances in exposure assessment and study design methodologies.

Methodological challenges in the use of epidemiological data in quantitative risk assessment, and the need to apply modern biostatistical techniques as well as appropriately present results from risk assessments that utilize epidemiological studies, have also been noted (Nurminen et al. 1999; Stayner et al. 1999; Schwartz 2002; Ryan 2003). As epidemiological and risk assessment techniques have become more sophisticated, a growing body of literature has developed to address the use of epidemiological studies in risk assessment. Guidelines for the conduct of epidemiological research and criteria/frameworks for evaluation and use of epidemiological studies in risk assessments have been offered to strengthen the evidence base used in public health policy decision-making (IARC 1991; Hertz-Picciotto 1995; Auchter 1995; Federal Focus Inc. 1996; Federal Focus Inc. 1999; WHO-Europe 2000; U.S. EPA 2005; Goldbohm et al. 2006; Swaen 2006; Vlaanderen et al. 2008). The criteria or frameworks provided by this literature are intended to improve the quality and validity of the studies themselves, as well as the risk analyses in which they are used.

In response to the need for improved epidemiological methods for the quantitative application of epidemiological data in risk assessment, there have been significant improvements in the collection of exposure and health data. However, fundamental challenges remain for identifying hazards and quantifying exposure-response relationships for a broad spectrum of pollutants and chemicals.

Evaluating Consistency in Epidemiological Data

In evaluating whether epidemiological data support the inference of a causal association for the purposes of regulatory-related risk assessment, a key uncertainty stems from variability in defining and operationalizing the concept of consistency across studies. Assessments of data consistency are often a controversial component of regulatory-related risk assessments, and contradictory determinations may result from varying stakeholder perspectives. Key stakeholders involved in, or affected by, the process of assessing consistency in epidemiological data for regulatory applications include federal, state and local health and environmental protection and occupational health agencies, affected industries, national and local public health and environmental organizations, researchers, as well as workers and the communities directly affected by regulatory risk management decisions (e.g., Superfund site clean-up).

The issue of evaluating consistency of results of epidemiological studies has been discussed for more than 50 years. In their seminal article "Smoking and lung cancer: recent evidence and a discussion of some questions", Cornfield et al. (1959) discuss the logical consistencies and inconsistencies in the body of epidemiological literature on smoking and lung cancer. The U.S. Surgeon General's 1964 Report on Smoking and Health (US Surgeon General 1964) discusses consistency in the direction and magnitude of associations, although a definition of what was meant by "consistency" was not offered. Hill (1965) considered observed associations as consistent if repeatedly found by different persons, in different places, circumstances and times. Hill added that he "would … put a good deal of weight upon similar results reached in quite different ways", for example, retrospective and prospective studies (i.e., case-control and cohort studies). A specific association observed in multiple studies of diverse design provides greater confidence that the association is real and not an artifact due to any particular methodological weakness, error, confounder or modifier that may be present in any one particular study.

Discussion of epidemiological study consistency in contemporary risk assessment usage has expanded and become more nuanced as the field of epidemiology has matured, and as experience in using epidemiological data for regulatory purposes has accrued. For example, understanding of the potential for gender-based differences in susceptibility to a potential endocrine-disrupting chemical can provide a framework for viewing differences in observed effects among studies that include varying proportions of males and females – that is, there may be biological reasons why one would not expect to see the same effect when studied in different populations. Similarly, differences in the sensitivity and specificity of exposure measures would be expected to lead to differences in observed estimates among studies. The 2006 Criteria Document for Ozone (U.S. EPA 2006a), for example, includes an explanation of why heterogeneity in effect estimates is to be expected among studies utilizing different ozone exposure metrics. The Ozone Criteria Document notes that "consideration of consistency and heterogeneity of effects are appropriately understood as an evaluation of the similarity or general concordance of results, rather than an expectation of finding quantitative results within a very narrow range."

The consistency of a body of scientific evidence is also evaluated on the existing understanding of the biological basis or plausibility for how an agent causes or contributes to a disease process (U.S. EPA 2005, IARC 2006). In other words, is the body of epidemiological evidence

consistent with other evidence such as that from laboratory animal studies? The IARC methodological preamble also discusses the concept of consistency in the context of when evidence for causal association is lacking, i.e., in cases when there is little or no evidence of association.

Although the idea of evaluating consistency of findings across a diverse collection of epidemiological studies is central to evaluating that body of evidence for supporting causal inferences, the identified variability in definition and formal evaluation methods strongly suggest the need for constructive approaches to consistently and transparently evaluate consistency. Researchers involved with epidemiological studies need to understand how best to design, analyze and report their study data to maximize their utility in the regulatory process. This workshop report provides recommendations toward filling this identified void.

Workshop on Evaluating Consistency in Epidemiological Data

In response to the need to improve approaches to assessing consistency in epidemiological data for application in regulatory risk assessments, the Johns Hopkins Risk Sciences and Public Policy Institute (RSPPI) organized a workshop to identify and discuss key methodological issues, and to develop recommendations for qualitative and quantitative approaches to addressing those issues. The workshop, held in Baltimore, Maryland on September 23-24, 2010, was co-sponsored by the U.S. Environmental Protection Agency, the National Institute of Environmental Health Sciences, and the National Institute for Occupational Safety and Health, with additional support provided by Health Canada. A multi-disciplinary approach was utilized for the workshop, involving invited experts from fields including epidemiology, risk assessment, exposure assessment, biological sciences, biostatistics, and science policy. In order to assure a broad spectrum of views on this issue, the invited participants were drawn from representatives affiliated with academia, industry, government, and the public interest sectors (Table 1 provides a list of workshop participants). A background information document was developed by the RSPPI in consultation with, and contributions from, the Workshop Organizing Committee and was provided in advance to the workshop participants with case study examples that illustrated the issues to be considered at the workshop.

Table 2 provides a copy of the workshop agenda. Following an opening plenary session in which several individual perspectives on evaluation of epidemiological data consistency for regulatory application were presented and an open discussion of the issues and approaches for evaluating epidemiological data consistency, the participants were divided into three assigned sub-groups for discussion of the key issues identified below. Information contained in the background document was utilized to varying degrees during the break-out group discussions. Results of the sub-group discussions were then reported and discussed in a concluding plenary session. This document summarizes the workshop discussions of the key issues identified for the scope of the workshop, and presents the key findings and recommendations from those discussions. While there was general consensus regarding the findings and recommendations discussed below except where explicitly noted, no formal process (e.g., voting) regarding unanimity of views was undertaken.

Workshop Discussions

General Issues

One of the underlying issues addressed in all of the workshop discussions concerns the characterization of study results. There was general agreement that that the results of a study should not be defined on the basis of the presence or absence of statistical significance, and that the question of the consistency of epidemiological data is not answered by counting the number of studies that have "positive" or "negative" results. Lack of statistical significance is not synonymous with "no effect", and it is important to distinguish between studies that demonstrate no effect and studies that would be better described as being "inconclusive" or "uninformative." This understanding comes from the discussion of statistical testing, power, precision, and interpretation of study results that has been a core part of epidemiology for more than 30 years (Rothman 1978; Rothman 2010). Thus considering only these statistical issues, a study with a relative risk estimate close to 1.0 with tight confidence intervals could reasonably be described as showing no effect, but a small study with a similar point estimate but much wider confidence intervals would be better described as being inconclusive rather than "null" or "negative."

Another underlying issue addressed in the workshop discussions concerns the distinction between heterogeneity and inconsistency within the context of evaluation of study results. Heterogeneity represents variation in observed effects that can be expected based on differences in populations (e.g., different effects of an endocrine-disrupting chemical between men and women), and variation in observed effects that could be explained by differences in study designs (e.g., a very weak or null effect seen with an "ever/never" exposure classification compared with a stronger effect seen with a more sensitive and specific exposure measure). However, there may be inconsistency reflecting variation in observed effects that cannot be explained by modifying factors (e.g., sex, genetics). Distinguishing between these concepts is central to an evaluation of the consistency of study results.

One other core concept that was central to the workshop discussions concerns the difference between hazard identification and exposure-response modeling, two of the core components of risk assessment (NRC 1983). Approaches to assessing consistency across epidemiological study data must be considered through the lens of their application in each of these components. Epidemiological data from studies that may not provide quantitative exposure-response functions may still be qualitatively informative in assessing consistency for the purpose of hazard identification. The evaluation of the consistency of results, or the evaluation of the contribution of potential sources of heterogeneity to variability in results, is a goal of the kind of systematic review that is part of hazard identification. For example, identification of the limitations in the exposure assessment component of epidemiological studies may provide a basis for understanding apparent inconsistencies in the results of epidemiological findings for a given chemical exposure.

Other specific issues considered at the workshop with respect to evaluating consistency among epidemiological studies for regulatory applications included: variation in outcome definition, exposure assessment methodology, critical periods of exposure and follow-up period, and how

these variations should be considered when comparing results among studies, as well the definition and identification of an exposure-response trend. Finally, the workshop also considered approaches for evaluating large bodies of epidemiological evidence with respect to determining consistency of findings. Each of these topics is discussed in more detail below.

Topic A: How do we consider variation in outcome definition in interpreting the consistency of results across studies?

- How can variation in study findings on potentially related health outcomes be evaluated?
- How can the quality of the disease definition (i.e., reliability and validity, or refinement by subtype) be considered when evaluating consistency (or variation) in effect measures among studies?

*Summary of Issues: S*ome types of diseases and early states of disease in particular may be difficult to define or measure. Different studies may measure different functional tests or disease markers, which may or may not be considered adverse outcomes. With improved understanding of the etiologic pathways and overall biological basis for disease, more recent epidemiological studies often use upstream markers of the disease process as defined outcomes rather than apical endpoints. For example, in assessing the relationship of air pollution exposure to exacerbation of cardiovascular disease and cardiovascular mortality, epidemiological studies have examined endpoints such as heart rate variability, dysrhythmic susceptibility, and cardiac repolarization as well as several hematological outcome measures. In some situations, there may be evidence of an abnormality across studies, but there is variation in what specific abnormality is seen (even if some of the same tests are used across studies). A question arises as to whether these observations are consistent because there is evidence of damage across the studies, or inconsistent because the results for specific tests differ among the studies. It can be difficult to interpret the results of multiple epidemiological studies that examine a range of effects acting on the same physiological system when these effects may not be coherent with one another. For example, cardiovascular system responses to air pollution range from cardiovascular-related mortality to incidence of myocardial infarction and non-fatal cardiovascular-related hospitalizations to numerous cardiovascular function parameters. A question arises as to the degree of coherence to expect across these outcomes when determining the consistency of an effect.

A different type of challenge arises when the definitions or classification criteria for a disease differ across studies or change (e.g., become more refined) over time. For example, WHO (2008) revised its classification system for acute myeloid leukemia but the older French-American-British (FAB) system is still in use (American Society of Clinical Oncology 2010). In this situation, how should we evaluate consistency between older and more recent studies?

Workshop Discussion: There was recognition by workshop participants that the selection of outcome measurement is often driven by study feasibility for assessment of population-level health outcomes rather than selection of the clinical "gold standard" that may be specified for individual-level health assessments. The purpose of the study also determines the choice of outcome measure. For example, a study designed to measure mechanisms of action may use a

measure of sperm damage or sperm concentration, but a study focusing on broader population impact issues may use a measure of fertility or time to pregnancy.

In general workshop participants noted that more information is often needed regarding the sensitivity and specificity of outcome measurements, including biomarkers. Validation information, if available, can offer a basis for comparing outcome assessment methods or outcome scales to each other. However for historical publications, which may be all that are available, such validation information is often not available and professional judgment is required.

There was a recognition that larger or more robust relationships may be found with more sensitive "upstream" markers of disease when compared to endpoints reflecting clinical expression of a defined disease state but that these differences may not indicate inconsistent results. The issue of whether the upstream endpoints represent an adverse health effect or might be a reversible, transient effect was noted, though this issue was not discussed in detail as it was deemed beyond the scope of this workshop.

When considering a series of study results for a given health outcome with several related outcome measurement methods (e.g., lung function or kidney function) or histological sub-types (certain cancers), an important consideration involves the issue of "lumping" together of study results as opposed to "splitting" or stratifying the data. If possible, the decision on how to analyze study results for a specific outcome or subcategory of outcome should depend on the known biological mechanism of the hazard. For instance, some measures of kidney function reflect glomerular disease, and some reflect proximal tubular disease. It may be reasonable to examine these as two sets of measures, but this "splitting" approach is based on the assumption that there is a clear delineation between the categories. This assumption may not hold, particularly under different stages of disease development and progression.

Another example concerns histological typing used with cancer analyses since there may be etiological factors for different histological subtypes. However, there was recognition that the decision regarding approaches to assessing related outcome measures or disease sub-categories in historical data may be limited by a lack of necessary information in the original study reports. Further, it was noted that outcome groupings based on the historical literature are a function of the manner in which data were presented, and more current understanding of the disease biology may indicate that such groupings are no longer appropriate and should no longer be considered homogenous and therefore assessed separately.

Topic B: How do we consider variation in exposure measurement in interpreting the consistency of results across studies?

- How can we account, in a formal and transparent manner, for an expected attenuation of an effect estimate due to non-differential exposure misclassification when evaluating the consistency of results from studies using different types of exposure assessment methodologies?
- What are the differences in approaches to the issue of variation in exposure measurement that can be used in situations in which numerous studies focusing on a specific type of exposure are available, and situations in which there are relatively

few studies of a specific agent, perhaps within a larger collection of studies with more general exposure assessments?

- What criteria could be applied in selecting specific data points (e.g., exposure groups) for the evaluation of the consistency of data among studies?

Summary of Issues: Differences in exposure assessment techniques across studies may create difficulties in understanding an apparent inconsistency of results. For instance, for a given chemical exposure, some occupational cohort studies categorize workers into broad groups based upon job title, while other studies incorporate individual or area measurements of specific exposures. In other studies, individual measurements would more closely account for differences in worker tasks, time periods, and location. Moreover, studies may also differ in the extent to which information is available on worker history and other potential confounders, including individual behaviors such as smoking and prior work history. Data may be analyzed with methods that take into account the measurement error in the exposure estimates. In addition, differences in effects or in the statistical significance of results might be observed based on the exposure categories selected for the study. An "ever" exposed category disregards ranges of exposure, which may be biologically useful in explaining results. Consequently, in reviewing evidence, attention needs to be given to identifying exposure assessment approaches and harmonizing exposure categorization across studies to the extent possible, whether for qualitative or quantitative assessment.

Workshop Discussion: The accuracy of the exposure assessment methodology used in an epidemiological study is an important determinant of overall study quality. For most exposures, the "gold standard" of having data across the biologically relevant time window is not achievable. Exposure measurements are often based on proxies for this gold standard measurement, e.g. a biomarker that may not represent exposure over the relevant time period, or ambient pollutant measurements representing most, but not all, sources of exposure and which do not capture variation among individuals in exposure or response. This reliance on proxies introduces uncertainty into the analysis in terms of the extent to which a proxy is a valid substitute for actual exposure. The participants noted that a variety of other factors, e.g. changes in regulatory measures, changes in economic conditions that impact emissions, changes in manufacturing processes/controls, can impact exposure levels and therefore can affect the exposure-response relationship over time. Given this context, variation in exposure measurement should not focus only on measurement error, but should also consider the contribution of these other dimensions of exposure assessment to variation in observed results.

A major part of the discussions of this topic focused on the feasibility of assessing the extent to which exposure assessment methodologies contributes to heterogeneity of results among studies. A variety of approaches, of varying degrees of complexity, could be used. One suggested approach was to stratify studies by key characteristics of the exposure assessment approach to consider the impact of potential exposure misclassification; adjustment for the observed attenuation may be possible. Modeling could assist in determining the extent to which different misclassifications or measurement errors may influence risk estimates.

Another important concept with respect to exposure and variability in results concerns the need to clearly understand the exposure range that is being considered within, and between, studies. It

is reasonable to expect to see different exposure-response relationships within different exposure ranges. For example, a stronger effect may be estimated in a study that incorporates a wider range of exposures (and thus a greater contrast between the "exposed" and referent categories) than in a study with a more limited exposure range. Thus, comparison of "high" exposure categories across studies (e.g., a meta-analysis of "high" versus "low" comparisons across studies) can be problematic since the different studies may incorporate different range values for their exposure categories. A more valid comparison may be to use this type of comparison to evaluate whether the "high" versus "low" comparison gives a stronger effect estimate than an "ever" versus "never" comparison among the same set of studies.

Topic C: What is the definition of a trend, and how should a trend be identified in practice?

- How might the shape of exposure-response function/gradients (e.g., monotonic, reaching a plateau or a more stair-case type pattern) inform assessment of a trend?
- Should a statistical test be the basis for deciding if a trend is present? If so, what considerations should be used in choosing the test and the level of statistical significance to be used? If a statistical test was not presented in a published paper (or if the optimal test was not conducted), what options for statistical testing are available to someone evaluating the data?
- How can differences among studies in the quality of the exposure assessment be transparently and reasonably incorporated into the evaluation of the presence/strength/shape of the observed exposure-response trend?

Summary of Issues: The presence of an exposure-response gradient is an important consideration within the Hill framework for evaluating causality (Hill, 1965; U.S. EPA 2005). If risk increases at higher levels of exposure, alternative explanations other than causality become less tenable. However, results from several occupational cohort mortality studies suggest that under certain circumstances, an exposure-response function/gradient may be demonstrated by a nonlinear relationship (Stayner et al., 2003). In addition, the observed form of the exposure-response relationship may be affected in complicated ways by exposure measurement error, population selection, and modeling approaches. How do we interpret this collective evidence, particularly when the range of exposures covered differs between the studies (i.e., some studies cover a wider range or higher exposures)?

Workshop Discussion: Workshop participants agreed that an expectation of monotonic increasing risk with increasing exposure is a reasonable consideration with respect to assessment of a causal association. They also stressed, however, that the underlying (true) exposure-response curve can have a variety of shapes, even within the general category of a monotonic increasing curve, including a linear pattern as well as "hockey stick" and a plateauing curve. Each of these curves may make biological sense (e.g., a hockey stick pattern reflecting a threshold type of response, and a plateau reflecting saturation of a key metabolic activation step). A flattening or downturn in the exposure-response function curve at high exposure may also reflect the underlying biology e.g., as would be seen with a significant competing risk. An additional difficulty in interpreting trends in epidemiological studies is that because of sources of bias and error, the observed exposure-response may differ from the underlying exposure-response, and thus the absence of a linear exposure-response within a study is not in itself strong evidence for the absence of a causal association.

Given these issues, participants recommended a variety of approaches to the assessment of trends within a study. These approaches ranged from "describe, don't test", to use of formal statistical tests assuming linearity on a particular scale across all exposures, to decomposing curves into linear and nonlinear components. The advantage of formal statistical tests is that they provide quantitative support to qualitative and subjective descriptions. The advantage of a descriptive approach is that it increases the information provided to the reader, and can support different interpretative perspectives: Is the relationship consistent with biological understanding? Is there evidence for bias from exposure measurement error? This framework would encourage study authors to provide information that would be useful in addressing the likelihood of these different explanations. The sparseness of the data should also be considered; it may be more appropriate to say "these data do not provide a basis for describing the exposure-response relationship" than to say "these data indicate there is [or is not] a trend."

Another issue concerns the comparison of trends across studies. Observed exposure-response patterns can differ among studies, particularly among studies with different exposure ranges. For studies in which the exposure range is relatively narrow, or when the shape of the exposure-response function within a study is relatively flat, a trend in the exposure-response function may only be observed when studies spanning a wider exposure range are combined. Depending on the details of the exposure measures used in the various studies, it may be possible to use meta-analysis and meta-regression approaches to obtain an overall estimate of trend and to understand differences among studies.

Topic D: What factors should be considered when evaluating the consistency of findings across varying lengths of follow-up or exposure windows?

- When two or more studies of the same cohort are available, with different lengths of follow-up, what are the considerations (i.e., type of disease, mechanism of disease, age-interactions) for determining the follow-up window that is most relevant to the risk assessment question?
- How can differences among studies in the length of follow-up or exposure windows be transparently and reasonably incorporated into the evaluation of consistency of observed effects?

Summary of Issues: Depending on the nature of the outcome and exposure, the length of the follow-up period could have a considerable impact on the results. Often, data from occupational (or other cohorts) are analyzed at multiple points during follow-up. There is a potential for risk estimates to vary over follow-up, reflecting changing patterns of exposure and underlying exposure-response time dynamics. Several studies have found that differences in time-related exposure metrics vary the exposure-response effect or trend that is observed, with effects that are seen earlier not being observed later, or effects only emerging after the passage of a greater period of time (Beane Freeman et al. 2009; Mundt et al. 2000). In using the findings of epidemiological research to characterize trends in risk or exposure-response over time, trends might be explored in one or more time dimensions: time since follow-up began, time since exposure, chronological age, and calendar time. Risks might plausibly vary across each of these scales and such variation might be relevant in the development of models for exposure-response relationships. An example of the complexity that can occur with time-related measures can be

seen in the analysis of radon-induced lung cancer. The NAS Biological Effects of Ionizing Radiation (BEIR) VI Committee had access to a large and rich data set, created by merging the data from 11 cohorts of underground miners (NRC 1999). The final risk model showed a decline in relative risk with increasing time since exposure and with increasing attained age. This time-varying model differs from more typical analyses that provide evidence on exposure-response that is cross-sectional in time and reflective of a particular point of follow-up.

Workshop Discussion: Workshop participants agreed that trends may change with time without detracting from the validity of the exposure-response gradient observed at one time. When analyzing an exposure-response relationship over a long period of time, an attenuation of risk due to a dilution effect from increased person-years observed or from depletion of the study population due to death or recovery from disease can be expected. In addition, the susceptible population may have been removed from the study, or the biological mechanism may vary with increasing lengths of follow-up.

There was considerable discussion concerning the way in which different types of biological mechanisms would result in different observed effects in situations with different lengths of follow-up. The types of mechanisms include initiation, promotion, initiation and promotion, combinations of effects, and short-term effects (e.g., such as can be seen immediately after pregnancy). Consequently, it is highly important to understand the underlying biological mechanisms involved in the specific exposure-disease under study where this information is available, as an understanding of these underlying mechanisms could assist in explaining trends over time, as well as potentially identify susceptible populations. However, participants also noted that the epidemiological observations often provide important insights into the nature of the underlying biological mechanism.

Substantial datasets may be needed for analyzing time-varying exposure-response (see BEIR VI example above). Where adequate data are available, studies could be systematically assessed to analyze exposure-response relationships for different follow-up periods. In doing so, it would be important to distinguish the study follow-up period from the outcome latency period. Sometimes necessary time-related details are lacking; publications of cohort studies may not adequately document key information on length of follow-up, exposure windows and related changes over time (particularly when reporting on later follow-up periods).

Topic E: What approaches can be considered in evaluating large bodies of epidemiological evidence with respect to determining consistency of findings?

- What could be the basis for selection of a "weight of evidence" (inclusion of all available study results) versus a "strength of evidence" (selection of the "best quality" or "most informative" studies) approach to evaluating consistency across epidemiological study data?
- What criteria could be applied in selecting studies for inclusion and for selecting specific data points (e.g., subgroups or exposure groups) in assessments of epidemiological data consistency for each of these approaches?
- How should factors such as variation in study design, study population, differing exposures to pollutant mixtures (ambient and occupational exposures), and mode of action information be considered?

Summary of Issues: Two general approaches have been used in efforts to summarize and assess large amounts of scientific information or to determine the basis for a causal relationship between chemicals/pollutants and health effects. One approach for summarizing large amounts of information for a causal assessment is a weight of the evidence approach (WOE). Formal meta-analysis with weighting of studies by size (i.e., inverse of study variance) could be considered a WOE approach. IARC and the U.S. EPA also use a WOE approach for assignment of cancer classifications. A second approach for summarizing information is the strength of the evidence (SOE) approach. The SOE approach selects studies for inclusion in the science review and/or causal assessment based on the quality of the study typically utilizing a set of criteria as the basis for the inclusion decisions. Review documents of epidemiology studies on substances where a large body of studies exists (e.g., U.S. EPA Integrated Science Assessments for criteria air pollutants such as particulate matter and ozone) sometimes use this approach. When there are many studies, these SOE "qualitative reviews" often summarize information from a select group of large high quality studies sometimes referred to as "informative studies". Decisions regarding study selection are embedded in various approaches to the analysis of data from multiple epidemiological studies.

Workshop Discussion: Participants indicated a preference for an inclusive approach to study selection for use in assessing relatively large bodies of studies, rather than excluding certain studies (while acknowledging consideration of study quality). Concern was raised that excluding studies could appear as if the data are being manipulated to match the original hypothesis. All types of studies, such as prospective or retrospective studies, ought to be included since each study type has its own set of strengths and weaknesses. Including different study types may assist in balancing potential problems associated with exposure or outcome biases. Stratification may also be used to note differences in outcomes by study design or variable of interest. The most important factor is communicating what criteria were used for weighting or choosing studies, and providing justification for the criteria. Emphasis should be placed on core principles that are applied with best judgment to study analyses, rather than to set rules.

As a first step towards core principles for WOE in handling large collections of studies, the participants discussed a framework for the initial presentation of epidemiological literature, which typically occurs in the hazard identification in the form of a qualitative, narrative review. The presentation should begin with a discussion of the disease under study, preclinical and clinical phases and issues regarding disease ascertainment. Similarly, issues pertaining to the specific exposure under study (e.g., exposure conditions and exposure measures) should also be outlined. Each study can then be reviewed and critiqued within this context. The critique can assess how relevant factors (i.e., outcome and exposure measures, sources of bias) affect the understanding of that study's findings within the larger dataset.

Issues of efficiency may require some studies to be emphasized when analyzing very large databases. If feasible as defined by the available studies and risk management needs, exposure-response analyses could include a combined analysis (e.g., meta-analysis, pooled analysis) from multiple studies. Study selection may also be based on the availability of data relevant to specific risk management needs, e.g., protection of a specific susceptible population. Consistency for studies should be addressed after the studies have been chosen or weighted according to the

criteria selected by the risk assessor/assessment team and informed by the risk management objectives.

The application of quality criteria has historically been utilized in clinical study systematic reviews with rarer application in environmental epidemiological studies. Workshop participants discussed the potential application of study quality criteria in the context of qualitative systemic reviews, as well as the basis for quality criteria scoring in the context of quantitative approaches to data consistency assessment. Defining, *a priori*, criteria for a "good" study, as well as possibly weighting those criteria, can be challenging. The use of quality criteria and a scoring framework in Turner et al. (2010) and Wigle et al. (2009) for assessment of epidemiological study data were discussed as case study examples, though it was noted that there was a movement away from the approach of quantitative scoring based on qualitative criteria for systematic reviews of clinical trials (Whiting et al. 2005). In practice, it is seldom that a study fulfills all of the chosen criteria, and it can be difficult to distinguish the failure of a study to fulfill a specific criterion from the failure of a report to provide enough details to allow the correct scoring of a specific criterion. However, the studies that are chosen for inclusion in a multi-study assessment could be qualitatively weighted to emphasize study designs (e.g., subject selection criteria, exposure assessment methods, statistical analysis approaches) that are methodologically more accepted or validated.

Workshop Discussions: Approaches and Tools for Consistency Evaluation

Qualitative and quantitative analytic approaches

Workshop participants discussed the potential utility of a variety of qualitative and quantitative tools and approaches for potential application in assessing epidemiological data consistency. Qualitative approaches discussed included analysis of the quality of individual studies as well as systematic synthesis across multiple study results. Quantitative tools considered included meta-analysis, pooled data analysis, meta-regression, trend tests, and quality criteria scoring.

The qualitative approaches build on the framework described in the previous section for the presentation of each study within the context of relevant issues concerning disease ascertainment, exposure assessment, and other issues that could affect the observed effects. Specifically, the following should be considered:

1. Study Attributes:
    a) What is the health outcome under investigation? Are there differences in case definition, case mix or type of outcome data (incidence vs. mortality; preclinical vs. clinical disease state) that might affect risk estimates compared with other studies?

    b) What are the exposures of the groups under comparison (in term of routes, levels and timing), what exposure monitoring data are available, how is exposure assigned to subjects, and how would sources of bias associated with exposure assignment affect observed risk estimates in comparison with other studies?

c) What potential confounders are strongly associated with both exposure and outcome and considered important, how well have these been taken into account in the study's design or statistical analysis, and to what extent might residual confounding cause the risk estimate to be over- or underestimated? In thinking about possible confounding, a good starting point is to consider known causes of the disease (especially those carrying high relative risks) and factors that are strongly associated with the exposure under study (might these factors be causes of the disease?).

d) What other potential major sources of bias might have caused the risk estimate to be over- or underestimated, and to what extent?

2. Considering these attributes, how different are the risk estimates among the studies? To what extent could these differences be attributable to differences in identified effect modifiers (i.e., heterogeneity of effects due to population differences)? How likely are they to be explained by differences in exposure measures? Is there a level of heterogeneity in risk estimates that is unlikely to be attributed to differences in study design or population characteristics/subject selection (i.e., true inconsistency)?

Meta-analysis (in conjunction with a systematic review) and pooled analysis can be useful for deriving summary risk estimates from multiple studies that are considered or can be made substantially similar (homogeneous or consistent) (Gordon et al.1998). Stratified meta-analysis is used when relevant studies are heterogeneous and obtaining an overall summary estimate is not advised. Stratified meta-analysis further subdivides the study set into homogeneous strata before estimating risk by strata and represents an approach to describing various sources of heterogeneity that may be useful to inform a regulatory risk assessment. Meta-regression modeling is another statistical tool that can be used to explore contributors to heterogeneity in terms of study-level covariates (Morton et al. 2004). In addition to supporting evaluation of consistency, these data combination tools can address a potential limitation of dose-response assessment practice, namely selection of one study for derivation of regulatory reference levels or cancer slope factors. A key issue that was not addressed in detail at this workshop was the feasibility of these types of analyses within the context of risk assessment, and the types of situations in which this effort would be useful.

Workshop Findings and Recommendations

Findings

Given the range of hypotheses, study designs, study populations, exposure assessment methodologies, outcome definitions and analytic techniques, the workshop participants clearly expressed that heterogeneity among reported results was to be expected – even when there might be a consistent underlying effect of exposure on the risk of a particular outcome. It was noted that a variety of factors may affect the observed heterogeneity of a group of studies, including: the type and ranges of exposure experiences and circumstances, for example, occupational (workers) as opposed to ambient environment (general population); the measurement methods and details of reporting the exposure metric, including the accuracy and the relevance (does it represent exposure during the critical period, etc.) of the exposure metric; the temporality of

exposure assessment with respect to time of outcome assessment; the length of follow-up for disease incidence or mortality; the extent of misclassification of outcomes due to changes in disease detection or definitions over time, and the statistical power of a study to detect true effects particularly in situations in which the magnitude of effects is low. Other sources of bias may also introduce heterogeneity and thus alter the perception of consistency including selection bias (differences between who is eligible to participate and who actually participates), misclassification of exposures, outcomes, and confounders; confounding by known causes of the outcome that are also associated with the exposure of interest; and differences in populations or susceptibility as related to outcome (i.e., effect modification).

However, the participants were clear that there was a difference between heterogeneity and inconsistency. While heterogeneity and inconsistency may be related, they need to be separated and addressed individually. The workshop participants considered that much of the heterogeneity might be explained by differences in study designs and applied methods and, therefore, only when the effects of these considerations had been evaluated and separated from the observed results could any remaining differences be thought of as potentially representing either imprecision or inconsistency. The statistical attributes for heterogeneity can then be defined and tested, whereas inconsistency suggests that there are unexplained differences between studies.

A distinction was made between qualitative and quantitative evaluation of consistency. For qualitative evaluations, expert judgment needs to be applied to ascertain the potential impacts of study design elements on the observed effects. It was recommended that this be done in a weight of evidence approach in which all studies were available for evaluation and input but supplemented with stratified evaluation of the observed evidence by factors thought to potentially explain the observed heterogeneity. Meta-analyses can be useful in some situations for quantitative evaluations of heterogeneity with attention to the explanation of heterogeneity and not simply the amalgamated effect size from every reported study. Stratified meta-analyses by important determinants of any heterogeneity may offer the clearest distillation of the underlying direction and magnitude of potential health effects.

Overall it is important to determine how "inconsistency" is defined. The direction of effects estimates is important, and often data tend to be more consistent than inconsistent with each other. Instead of focusing on whether the data are consistent or inconsistent (binary outcome), the analysis could focus on the extent to which the data are heterogeneous (utilizing quantitative methods such as meta-regression when appropriate), in what direction, and with how much precision. The workshop participants suggested that the observation and evaluation of 'consistency' and 'inconsistency' was thought to fall on a continuum. Current reporting styles tend to emphasize a checklist mentality as consistent or inconsistent, so a better method of communication is needed to express the distribution of results falling on a continuum on consistency.

The need for substantial improvement in both the quantity and quality of exposure measurement data available in environmental epidemiological studies was highlighted. The potential for exposure-related measurement error to contribute to apparent inconsistencies among study results (i.e., heterogeneity, rather than actual inconsistency) was specifically noted. Stratification

of studies by exposure characteristics such as quality and validity to determine if study results are attenuated was suggested. Likewise, such stratification by other potential determinants of heterogeneity would be helpful.

The need for overall improved or increased access to epidemiological data in general, and particularly to datasets with sufficient detail to evaluate consistency, was noted. One frustration voiced by some workshop participants was that relevant information, such as quantitative exposure measures, stratified analyses or adequate detail on follow-up procedures for cohort studies, often is missing from peer-reviewed articles. One recommendation suggested to address lack of detail in publications was a concerted outreach effort to editors of journals that regularly publish environmental and/or occupational health exposure and epidemiology studies to inform them of key types of information needed to evaluate and apply epidemiological findings to regulatory policies, such as details regarding exposure range represented by the study.

A major area of discussion focused on the developing need for a biologically-based approach to interpreting differences across study results as well as in assessing the quality of study designs and analyses. The participants emphasized the value of understanding the biology of the disease development and progression process in the selection of exposure metrics, identification of related health outcomes, and interpretation of exposure-response relationships. It was also noted that the development of information regarding the biological basis for diseases was typically an evolving process, and that epidemiologic studies also contribute to this process. Such a biologically-based approach requires information from a variety of scientific/medical disciplines, and therefore the concept of utilizing multi-disciplinary teams when assessing epidemiological data for consistency was recommended when the relevant data could be available for such comprehensive review.

Workshop participants also noted a need for further exploration of key concepts related to length of follow-up through case examples based on existing literature or through development of simulation modeling approaches. Example issues include the effect of depletion of susceptible populations on effect estimates over time, and the effect of a specific form of time-varying exposure-response on the observed results under different lengths of follow-up. Participants noted that consideration should be given to the uncertainty associated with such estimates and to the utilization of time-dependent models.

Recommendations

- Do not define consistency based on presence or absence of statistical significance.
    - Lack of statistical significance does not mean “no effect.”
    - Distinguish between negative and inconclusive results.

- Distinguish between heterogeneity and inconsistency across study results.
    - Heterogeneity represents real variation and can be expected based on differences in study design, population, exposures, etc.
    - Inconsistency may be rooted in study differences in design and analysis, differences in exposure assessment approaches and to bias or error rather than reflecting true differences in study results.

    - Quantitative methods are available to assess heterogeneity, but use of these methods depends on many considerations including data availability, data reporting, data quality and other considerations such as risk management goals. As such, assessment of data inconsistency will likely involve qualitative as well as quantitative approaches.

- Systematic reviews and meta-analyses can strengthen risk assessments.
    - Qualitative and quantitative approach can be useful; qualitative approaches can benefit from a framework for presenting epidemiological studies that is based upon understanding of the disease and exposure measurement issues under study.
    - A weight of evidence approach including all studies is generally preferred; subsetting/stratification of the full study set will assist in investigating heterogeneity and inconsistency.
    - Specific attention to exposure assessment is warranted – consider systematic review of exposure measures and characteristics.

- An understanding of the underlying biology and disease progression will inform many aspects of the consistency evaluation including expectations of dose-response shape, assessment of exposure, outcome latency, and appropriate length of follow-up.

- Multidisciplinary team (including exposure scientists, epidemiologists, clinicians, and others as appropriate) can provide insights in factors that may contribute to "heterogeneity" and "inconsistency" in epidemiological evidence.

- Conduct empirical research to evaluate the utility of the recommended tools.

References

American Society of Clinical Oncology. Leukemia –Acute Myeloid. Available: http://www.cancer.net/patient/Cancer+Types/Leukemia+-+Acute+Myeloid+-+AML?sectionTitle=Subtypes [accessed February 26, 2011].

Auchter TG. The reliable use of epidemiology studies in regulatory risk assessments. In *The Role of Epidemiology in Regulatory Risk Assessment,* Graham, JD (Ed.). New York: Elsevier Publishers, Inc., 1995.

Burke, TA. The proper role of epidemiology in regulatory risk assessment: a regulators perspective. In *The Role of Epidemiology in Regulatory Risk Assessment,* Graham, JD (Ed.). New York: Elsevier Publishers, Inc., 1995.

Beane Freeman LE et al. Mortality from lymphohematopoietic malignancies among workers in formaldehyde industries: the National Cancer Institute Cohort. J Natl Cancer Inst 2009;101:751-761.

Cochrane Collaboration. Cochrane Reviews. Accessed July 2010. Available at: http://www.cochrane.org/reviews/clibintro.htm

Cochrane Collaboration. Glossary of terms in the Cochrane Collaboration: Systematic Review http://www.cochrane.org/sites/default/files/uploads/glossary.pdf [accessed December 2, 2010].

Cornfield J et al. Smoking and lung cancer: recent evidence and a discussion of some questions. JNCI 1959; 22:173-203

Federal Focus, Inc. Principles for Evaluating Epidemiologic Data in Regulatory Risk Assessment. *Developed by an Expert Panel at a Conference in London, England,* October 1995. Washington, DC: Federal Focus, Inc; 1996.

Federal Focus, Inc. Epidemiology in Hazard and Risk Assessment. A partial review of the "London Principles", with recommendations for revisions and additions, by an expert panel. Federal Focus, Inc. 1999.

Gamble JF, Lewis RJ. Health and respirable particulate (PM10) air pollution: a causal or statistical association? Environ Health Perspect. 1996 Aug; 104(8): 838-850.

Goldbohm RA, Tielemans E, Heederik D, Rubingh CM, Dekkers S, Willems MI, and Kroese ED. Risk estimation for carcinogens based on epidemiological data: A structured approach, illustrated by an example on chromium. Regulatory Toxicology and Pharmacology 2006; 44: 294-310.

Gordis, L., ed. *Epidemiology and Health Risk Assessment.* New York: Oxford University Press, 1988.

Gordon I, Boffetta P, Demers P. A Case Study Comparing a Meta-Analysis and a Pooled Analysis of Studies of Sinonasal Cancer among Wood Workers. 1998. Epidemiology 9:518-524.

Graham, JD, Paustenbach, DJ and Butler, WJ . Epidemiology and risk assessment: Divorce or marriage? In *The Role of Epidemiology in Regulatory Risk Assessment,* Graham, JD (Ed.). New York: Elsevier Publishers, Inc., 1995.

Hertz-Picciotto I. Epidemiology and Quantitative Risk Assessment: A Bridge from Science to Policy. American Journal of Public Health 1995; 85: 484-491.

Hill AB. 1965. The environment and disease: association or causation. Proc R Soc Med 58:285-300.

International Agency for Research on Cancer. IARC monographs on the evaluation of carcinogenic risk of chemicals to man. 1991. International Agency for Research on Cancer; Geneva, Switzerland.

International Agency for Research on Cancer. 2006. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Preamble. Available: http://monographs.iarc.fr/ENG/Preamble/CurrentPreamble.pdf [accessed November 2, 2010].

Morton SC, Adams JL, Suttorp MJ, Shekelle PG. Meta-regression Approaches: What, Why, When, and How? Technical Review 8, AHRQ Publication No. 04-0033. Rockville, MD: Agency for Healthcare Research and Quality. March 2004.

Mundt KA et al. Historical cohort study of 10,109 men in the North American vinyl chloride industry, 1942-72: update of cancer mortality to 31 December 1995. J Occup Environ Med 2000; 57: 774-781.

National Research Council Committee on the Institutional Means for Assessment of Risks to Public Health. Risk Assessment in the Federal Government: Managing the Process. Washington, D.C.: National Academies Press; 1983.

National Research Council. Health Effects of Exposure to Radon: BEIR VI. Washington, DC: National Academies Press; 1999.

National Research Council. Toxicity Testing in the Twenty-first Century: A Vision and a Strategy. Washington, DC: National Academies Press; 2007.

National Research Council. Science and Decisions: Advancing Risk Assessment. Washington, D.C.: National Academies Press; 2009.

Nurminen M, Nurminen T, and Corvalán CF. Methodologic Issues in Epidemiologic Risk Assessment. Epidemiology 1999; 10: 585-593.

Office of the Surgeon General of the United States. Report of the Surgeon General's Advisory Committee on Smoking and Health. 1964. Smoking and Health, Public Health Service Publication No. 1103, Washington, DC. Available: http://www.surgeongeneral.gov/library/reports/index.html [accessed November 2, 2010].

Office of the Surgeon General of the United States. The Health Consequences of Smoking: A Report of the Surgeon General. Dept. of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; Washington, D.C. 2004. Available at: http://www.cdc.gov/tobacco/data_statistics/sgr/2004/index.htm [accessed November 2, 2010].

Rothman KJ. A show of confidence. N Engl J Med. 1978;299(24):1362-1363.

Rothman KJ. Curbing type I and type II errors. Eur J Epidemiol. 2010;25(4):223-224.

Ryan L. Epidemiologically based environmental risk assessment. Statistical Science 2003 18(4): 466-480.

Samet JM, Schnatter R, and Gibb H. Invited Commentary: Epidemiology and Risk Assessment. American Journal of Epidemiology 1998; 148: 929- 936.

Samet JM. 2009. Data: To Share or Not to Share? Epidemiology 20(2):172-174.

Samet JM. 2010. To Register or Not to Register? Epidemiology 21(5):610-611.

Schwartz J. The Use of Epidemiology in Environmental Risk Assessment. Human and Ecological Risk Assessment 2002; 8(6): 1253-1265.

Stayner L, Smith RJ, Gilbert S, and Bailer AJ. Epidemiological Approaches to Risk Assessment. *Inhalation Toxicology* 1999; 11: 593-601.

Stayner L. Attenuation of exposure-response curves in occupational cohort studies at high exposure levels. Scand J Work Environment Health. 2003; 29: 317-324.

Swaen GMH. A framework for using epidemiological data for risk assessment. Human and Experimental Toxicology 2006; 25: 147-155.

Szklo M and Nieto FJ. Chapter 8 Quality Assurance and Control in Epidemiology: Beyond the Basics. Jones and Bartlett Publishers. Boston. 2007. Pages 297-347.

Turner MC, Wigle DT, Krewski D. Residential Pesticides and Childhood Leukemia: A systematic review and metaanalysis. Environ Health Perspect. 2010; 118: 33-41.

U.S. EPA. Integrated Risk Information System: Benzene (CASRN 71-43-2). 2003. Available: http://www.epa.gov/iris/subst/0276.htm [accessed March 30, 2011].

U.S. EPA. Guidelines for carcinogen risk assessment. U.S. Environmental Protection Agency, Washington DC, EPA/630/P-03/001B, 2005.

U.S. EPA. Air Quality Criteria for Ozone and Related Photochemical Oxidants: Volume I of III. *U.S. Environmental Protection Agency*, Washington, DC, EPA 600/R-05/004aF, 2006a.

U.S. EPA. National Ambient Air Quality Standards for Particulate Matter. 40CFR50. 2006b.

U.S. EPA. Integrated Science Assessment for Particulate Matter (Final Report). *U.S. Environmental Protection Agency*, Washington, DC, EPA/600/R-08/139F, 2009.

U.S. OSHA. Occupational Safety and Health Standards for Benzene. 29CFR1910.1028. 1987.

Vlaanderen J, Vermeulen R, Heederik D, and Kromhout H. Guidelines to Evaluate Human Observational Studies for Quantitative Risk Assessment. European Union Network of Excellence: Integrated Risk Assessment Group. Environmental Health Perspectives 2008; 116: 1700-1705.

Vlaanderen J, Moore LE, Smith MT, Lan Q, Zhang L, Skibola CF, Rothman N, Vermeulen R. 2010. Application of OMICS technologies in occupational and environmental health research; current status and projections. Occup Environ Med. 2010 67(2):136-143.

Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. BMC Medical Research Methodology 2005; 5:19.

Whittemore AS. Epidemiology in Risk Assessment for Regulatory Policy. J Chron Dis 1986; 39(12): 1157-1168.

Wigle DT, Turner MC, Krewski D. A systematic review and meta-analysis of childhood leukemia and parental occupational pesticide exposure. Environ Health Perspect. 2009; 117: 1505-1513.

World Health Organization-Europe. Evaluation and Use of Epidemiological Evidence for Environmental Health Risk Assessment. Guideline Document. World Health Organization. 2000. EUR/00/5020369.

World Health Organization. Classification of Tumours of Haematopoietic and Lymphoid Tissues. IARC Press: Lyon, 2008.

Table 1

State of the Science Workshop: Evaluation of Epidemiological Data Consistency For Application In Regulatory Risk Assessment

September 23-24, 2010 Baltimore, MD

Participants

Thomas Burke, Ph.D., Johns Hopkins University (co-chair)*
Jonathan Samet, MD, MPH, University of Southern California (co-chair)*
Thomas Bateson, Sc.D., MPH, US Environmental Protection Agency*
Aaron Blair, Ph.D., MPH, National Cancer Institute
David Coggon MD, Southampton General Hospital
Glinda Cooper, Ph.D., US Environmental Protection Agency*
Elizabeth Delzell, D.Sc., University of Alabama at Birmingham
Kay Dickersin, Ph.D., Johns Hopkins University
Elizabeth Fontham MPH, Dr.PH, Louisiana State University
Bruce Fowler, Ph.D., Agency for Toxic Substances and Disease Registry, Centers for Disease Control and Prevention
Mary Fox, Ph.D., Johns Hopkins University*
Freya Kamel, Ph.D., National Institute of Environmental Health Sciences
Ellen Kirrane, Ph.D., US Environmental Protection Agency
Daniel Krewski, MHA, MSc, Ph.D., U of Ottawa
Germaine Buck Louis, Ph.D., National Institute for Child Health and Human Development
Charles Poole, Ph.D., University of North Carolina
Ruthann Rudel, MS, Silent Spring Institute
Jennifer Sass, Ph.D., Natural Resources Defense Council
Cheryl Siegel Scott, MSPH, US Environmental Protection Agency
Jack Siemiatycki, Ph.D., U Montreal
Thomas Smith, Ph.D., MPH, Harvard University
Leslie Stayner, Ph.D., U of Illinois
Patricia Stewart, Ph.D., National Cancer Institute
J. Morel Symons, Ph.D., DuPont Haskell Laboratory for Health and Environmental Sciences
Elizabeth A. Whelan, Ph.D., National Institute of Occupational Safety and Health
Ronald White, MST, Johns Hopkins University*
Michael Wright, Sc.D., MPH US Environmental Protection Agency

Observers

Krista Christensen, Ph.D., MPH, US Environmental Protection Agency
Barbara Glenn, Ph.D., MPH, US Environmental Protection Agency
Karen Hogan, Ph.D.US Environmental Protection Agency
Jennifer Jinot, Ph.D. US Environmental Protection Agency
Patricia Murphy, Ph.D., MPH, US Environmental Protection Agency
Molini Patel, Ph.D., MPH, US Environmental Protection Agency

*Workshop Organizing Committee

Table 2

State of the Science Workshop: Evaluation of Epidemiological Data Consistency For Application In Regulatory Risk Assessment

Mt. Washington Conference Center
Baltimore, MD September 23-24, 2010

Agenda

Day 1

1:00 – 1:45PM Participant introductions; workshop objectives and format; issue overview – *Participants/Co-chairs*

1:45 – 2:45PM Stakeholder panel: Perspectives on evaluation of epidemiologic data consistency for regulatory application – *(Cooper, Symons, Fowler)*

2:45 – 3:00PM Break

3:00 – 4:00PM Synthesis of panel presentations/discussion- *Co-chairs*; Discussion of issues and approaches to developing criteria for evaluating epidemiologic data consistency – *Co-chairs/Participants*

4:00 – 6:00PM Break-out session #1 *All*

6:30 – 7:30PM Reception

7:30 – 9:00PM Dinner

Day 2

8:00 – 10:00AM Break-out session #2 *All*

10:00 – 10:15AM Break

10:15AM – 12:15PM Break-out session #3 *All*

12: 15 – 1:15PM Lunch

1:30 – 2:30PM Reports from break-out sessions – *Break out session leaders/Reporters*

2:30 – 4:00PM Discussion of break out session results/Workshop findings and recommendations – *All*

4:00 – 4:30PM Next steps – C*o-chairs/Participants*

4:30PM Adjourn