Comments on US EPA Next Generation Risk Assessment Report (September 2013 External Review Draft) Docket ID: EPA-HQ-ORD-2013-0680

Prepared by
Julie E. Goodman, Ph.D., DABT
Catherine Petito Boyce, S.M.
Daniella M. Pizzurro, Ph.D.
Lorenz R. Rhomberg, Ph.D., FATS

Prepared for
Texas Commission on Environmental Quality
MC-168
P.O. Box 13087
Austin, TX 78711-3087

November 15, 2013

GRADIENT
www.gradientcorp.com
20 University Road
Cambridge, MA 02138
617-395-5000

Table of Contents

1 Introduction

In its recent external review draft report, *Next Generation Risk Assessment: Incorporation of Recent Advances in Molecular, Computational, and Systems Biology* (US EPA, 2013a; hereinafter referred to as the NexGen report), the US Environmental Protection Agency (US EPA) presents valuable contributions to understanding the roles that evolving toxicity testing methods and associated interpretative techniques can play in assessing the risks associated with chemical exposures. In particular, drawing upon the expertise of individuals addressing chemical exposures and risks in a variety of settings, the prototype[1] analyses documented in the NexGen report offer useful – and needed – opportunities for synthesizing and reflecting on currently available new data types within specific applications. As recognized in the NexGen report, these analyses provide contexts for exploring how results from new study types can contribute to chemical risk evaluations (*e.g.*, proof-of-concept and value-of-information assessments), limitations in currently available data and interpretative techniques (*e.g.*, decision considerations for data applications), and directions for future research that will most effectively fill identified data gaps and enhance the usefulness of new data types.

While the prototype and other analyses presented in the NexGen report amply illustrate the promise of new toxicity test systems, they also reflect the many challenges yet to be surmounted before such data can be widely and reliably incorporated into risk assessment decisions, even for data-rich chemicals (such as those studied in the Tier 3 prototypes). In particular, as observed in the report, "[l]ogistical and methodological challenges in interpreting and using newer data and methods in risk assessment…remain significant." Still, there are a number of ways that the usefulness and scientific foundation of the report should be enhanced. For example, as an initial step to strengthen the overall context for understanding the risk assessment implications of the prototype analyses, the report should discuss the key risk assessment paradigm changes reflected in and implied by the new testing methodologies. In addition, the evaluations presented should more thoroughly address essential key factors that underlie critical review of toxicity information [including weight-of-evidence (WoE) evaluations], such as data relevance, endpoint adversity, and data quality. Moreover, the scientific soundness of the NexGen analyses should be improved by better documentation of the processes used to compile the literature reviewed in the prototypes and conduct the analyses based on that literature. Finally, the NexGen report should provide a clearer, more specific roadmap for guiding future research.

The remainder of these comments discuss ways in which the NexGen analyses could be enhanced to refine their usefulness for guiding future research and risk assessment applications. These issues are first discussed more generally and then illustrated in detail using examples from the ozone case study. These comments conclude with a summary of recommendations for strengthening the NexGen report, as well as future research efforts.

[1] The NexGen report presents analyses for three Tiers of prototypes, defined as: "Tier 3—major scope decision-making (considerable data indicating high hazard or widespread exposures); Tier 2—limited decision-making (limited exposure potential or limited hazard potential or data); and Tier 1—prioritization and screening (very little or no traditional data for chemicals known to be in commerce)."

2 Risk Assessment Paradigm Context for NexGen Evaluations

Although the NexGen report provides some context for the risk assessment evaluations considered in the report (*e.g.*, in Section 2 *Preparation for Prototype Development*), it would benefit from stronger grounding in the underlying changes in risk assessment paradigms that are inherent in the acceptance and application of new data and methodologies. Such considerations would highlight the importance of the critical toxicity data review elements discussed in Section 3 of these comments and would provide a valuable foundation for considering and prioritizing data gaps and future research needs.

As recognized in the NexGen report and reviewed in Rhomberg (2010) and National Research Council (NRC) (2007), evolving risk assessment methodologies and underlying developments in toxicity testing approaches present numerous opportunities to enhance our understanding of chemical toxicity. Chief among the potential advantages offered by new toxicity testing approaches and new data types are the ability to conduct testing that is less expensive, less time consuming, and less resource intensive than traditional toxicity testing. In particular, new approaches present the potential to reduce or replace animal studies for risk assessment purposes (as discussed in Scholz *et al*., 2013). Moreover, because of the "high-throughput" and low cost of many of the new tests, it is practical to examine many more test conditions (*e.g.*, to test more dose levels to better evaluate dose-response relationships; to test lower, more environmentally relevant doses where test systems are often more sensitive than traditional methods; to use model systems that are most relevant to human health; to test different patterns of exposure over time; to test effects of combinations of agents; or to evaluate interindividual variability). Because of such features, new toxicity testing approaches offer the possibility of evaluating more chemicals more efficiently, and conducting chemical screening on a greater scale. New toxicity testing techniques also offer the potential to enhance our scientific understanding of chemical-specific modes of action (MoAs)[2] and to transfer such insights to a broader spectrum of chemicals.

The evolution of toxicity testing techniques – and the changes in perspective regarding MoAs and other indicators of toxicity that accompany new data – also requires consideration of changes to standard risk assessment paradigms that accompany such data. Most notably, as discussed in Rhomberg (2009), the current risk assessment paradigm primarily works from observations of apical responses (*e.g.*, adverse effects observed in traditional animal or epidemiological studies) to explore underlying mechanisms of toxicity. By contrast, the risk assessment paradigm inherent to the new data types reverses this process and begins with studies of underlying mechanistic elements, working from there to evaluate apical effects that could result. To fully comprehend the risk assessment implications of this shift in perspective requires a thorough understanding of the biological control processes reflected in the available data and analyses (*e.g.*, sufficient knowledge regarding how statistical analyses and predictive profiles relate to apical effects of concern in humans).

In particular, interpretation of new data types requires a central focus on what constitutes sufficient perturbation of normal processes to yield adverse apical effects. As reviewed in Rhomberg (2011), connections between process perturbations and apical events can be viewed as a cascade of causative

[2] Although "mode of action" is the term generally used to describe a mechanistic understanding of the effect of a chemical on human health, the NexGen report states that it instead uses the term "mechanism of action" in accordance with the NRC report, *Science and Decisions: Advancing Risk Assessment* (2009). The term mode of action is used in these comments.

processes, with the outputs of earlier processes constituting the causes of later ones. Such processes are inherently and markedly non-linear; *i.e.*, processes reflecting continuous variation in causal factors are translated into discontinuous change-of-state outcomes. These discrete changes of state and underlying processes can be seen as either a series of interconnected control processes (from a systems-theory viewpoint) or as failure modes of adaptive processes (from a catastrophe-theory viewpoint). Factors to be considered in such evaluations include identification of key events, how sequences of events relate to each other, the persistence or independence of events, and factors leading to dose-response observations (*e.g.*, interindividual variation or event accumulation). Such considerations provide a valuable and necessary perspective for understanding the roles that process perturbations can play in apical effects, defining adverse effects, and evaluating dose-response relationships and their overarching implications.

In considering the implications of the changing risk assessment paradigm for interpretation of new toxicity data, it is also important to maintain a perspective on long- *vs*. short-term uses and goals for such data (*e.g.*, as discussed in Chiu *et al*., 2013). For example, in the short-term, gene expression changes may be used as markers of toxicity pathways that have been identified previously based on traditional toxicity data. In the future, such data will evolve in their application as investigative tools used to identify potential adverse outcome pathways that have yet to be established. The way in which the data are used, accompanying uncertainties, and dependence of the research on existing knowledge differ between these two uses of the data. Clearly, routine and reliable application of new toxicity data types in settings requiring a high degree of scientific certainty and rigor – and routine acceptance of such applications by the risk assessment community – will require far more extensive analysis of such methods than has yet occurred. However, as the test methodologies and risk assessment applications evolve, the new data types can play other useful roles (*e.g.*, as screening tools, biomarkers, or approaches for diagnoses and characterizing modes of action). They can also provide support to dose-response analyses, interspecies extrapolations, and evaluations of inter-individual variability (*e.g.*, Rhomberg, 2010; Burgess-Herbert and Euling, 2013). As discussed below, the NexGen report could help achieve long-term goals for use of these data by more clearly defining such goals and specific research needed to reach them.

3 Recommended Enhancements to NexGen Report

The evaluations presented in the NexGen report would benefit from more thorough consideration and acknowledgement of certain essential key factors that underlie critical review of toxicity information (including WoE evaluations, as reviewed in Rhomberg *et al.*, 2013), particularly within the prototype evaluations. Factors that merit particular emphasis in the NexGen evaluations include the relevance of the data derived from new toxicity test systems for human health risk assessment, the extent to which the endpoints under consideration reflect adverse effects, and the quality of the data derived from various applications of the new test systems. Such considerations are necessary to ensure that the types of evaluations presented in the NexGen report – and the conclusions and recommendations derived from those evaluations – are grounded in scientifically sound, representative reviews of the available data. The report would also be improved by providing more thorough discussions of the bases for the prototype assessments, fundamental uncertainties underlying the analyses, and representativeness of the assessments for other chemicals and settings.

3.1 Data Relevance

An essential concept underlying risk assessment applications of new toxicity data is the relevance of the data being reviewed to the adverse health effect that is the ultimate effect of interest (and not simply as an indication of adaptive or homeostatic perturbations). The evaluation of the data relevance encompasses aspects of the effects and exposure conditions of the test system as well as characteristics of the receptors of interest (*e.g.*, whether subpopulations with enhanced susceptibility exist). This concept also requires consideration of the degree to which the available data demonstrate that a specific toxicant causes, rather than is simply associated with, a particular outcome or interest. For example, in some cases, observed changes associated with a toxicant (*e.g.*, associations in a toxicogenomics network) may not be indicative of adverse effects, but instead reflect a change that plays a role in *preventing* adverse effects (Audouze *et al*., 2013).

In light of the pervasive and fundamental impacts of this issue on risk assessment conclusions, data relevance warrants thorough and systematic consideration in each of the NexGen report analyses.

Relevance of Test Systems

As reflected in the NexGen report, the types of data generated by new toxicity testing methodologies promote the development of complex model systems that consider the many factors and pathways that may contribute to the apical effect. Risk assessment evaluations of the resultant test data must identify the most likely pathway(s) whereby the measured endpoint would contribute to the apical outcome, as well as other features of such pathway(s) that affect the progression from the measured endpoint to the apical outcome (*e.g.*, whether a given level of effect may be mitigated by homeostatic processes; *e.g.*, Rhomberg, 2010).

In addition, to provide perspective on the degree to which a specific toxicant may contribute to or cause a specific apical outcome, the evaluations should consider available data regarding other factors that affect the measured endpoint or pathway, including other toxicants as well as potential non-chemical stressors (*e.g.*, lifestyle factors). For example, Ornish *et al*. (2008, 2013) conducted a pilot study of the impacts of lifestyle changes (*i.e.*, changes to exercise, sleep, stress, and social support patterns) on telomerase

activity and telomere length in men diagnosed with low-risk prostate cancer. Telomere length has an evolving use as a marker for cancer and mortality risk and disease progression in some types of cancer. In the pilot study, these researchers observed greater telomere length in the lifestyle intervention group, suggesting impacts of lifestyle factors on this biological measure. These types of findings suggest the need for a broader perspective on potential contributing factors to test system endpoints, the likely relative roles of various contributing factors to complex biological systems, and the implications of test results obtained in studies of specific toxicants. Relevant information for assessing such factors is not available for all test systems, which affects interpretation of the reliability and significance of observed results.

In addition to developing a sound model of the underlying biological processes relating the measured endpoint to the apical outcome, the relevance of the test system must also be assessed from the perspective of the scientific foundation of the test system itself. Such evaluations should consider how well the implications of the test results are understood, how widely the test has been applied and validated, and the confidence in the significance and robustness of the results. Of particular importance for assessing these evolving methodologies is providing a sound perspective on the current understanding of the consistency of the test results when applied in different test settings or for different chemicals. Similarly, it is important to consider the degree to which conclusions regarding the toxicity and risk associated with a specific chemical may be affected by the choices of test systems that have been applied. In other words, might conclusions regarding the biological actions associated with a specific chemical differ with the use of different test systems? This perspective considers not only the specific data set(s) selected for review, but also important aspects of the context for those data (such as other similar data that were not included in the review) and other relevant test options.

In several places, the NexGen report briefly acknowledges the substantial information requirements for reliably identifying test system relevance and applicability. For example, in a discussion in the Tier 3 benzene prototype regarding the use of molecular data patterns to screen chemicals for their potential to cause similar mechanistic disruptions (and apical outcomes), the report notes that "[a]nchoring of the molecular patterns to apical outcomes, considerable systems biology knowledge, and high-quality data…appear necessary to define the disease signature against which data-limited chemicals could be compared." Similarly, the report recognizes that "some test systems might better predict the potency of a chemical to disrupt normal biology than predict the specific adverse outcome resulting from that disruption." The report also acknowledges that "molecular signatures involve dynamic relationships among adaptive and nonadaptive processes that will require additional research to understand fully." However, because the issue of test system relevance is at the heart of evaluations of the usefulness of new toxicity data for risk assessment purposes, it warrants more comprehensive and rigorous discussion and review in the specific prototype examples presented in the NexGen report.

Relevance of Exposure Levels/Routes

A second major feature of assessing test relevance is the relevance of the exposure levels, routes, and settings to the exposures that would be expected in potential receptor populations. One aspect of this evaluation is the degree to which the exposure levels applied in the test settings are representative of or relevant to likely levels in typical human exposure settings. Frequently, exposure levels in test systems are greater than those typically encountered; thus, potential differences in response associated with high doses *vs*. those for low doses must be evaluated. Potential differences in response associated with chronic *vs*. acute exposure conditions should also be considered.

Similarly, tests such as *in vitro* studies of cultured cells clearly represent exposure settings that differ substantially relative to the exposure conditions that would be encountered by the same cells in their normal *in vivo* setting. One of the most obvious differences is that means of exposure to the substance of interest differs. Moreover, the isolation of the cultured cells from the normal *in vivo* environment

inherently prevents observation of interactions with other cells or biological processes that could influence the types of effects observed in the test systems, *e.g.*, processes that could mitigate adverse impacts. For example, in a recent report, US EPA (2013b) concluded that although non-monotonic dose-response curves are not unexpected *in vitro*, they "are not commonly identified in estrogen, androgen, or thyroid systems *in vivo* and are rarely seen in apical endpoints after low-dose and/or long-term exposure." Also, as recognized in the benzene prototype analysis in the NexGen report, factors such as differences in metabolism and cell types can complicate comparisons of *in vivo* and *in vitro* data. These important contextual issues must also be reviewed rigorously when interpreting results from new types of toxicity tests.

Evidence of Dose-Response Relationships

Another important factor in assessing the relevance of data to risk assessment evaluations is the degree to which a dose-response relationship is evident in the available data. This factor has long been recognized as a fundamental feature of systematic causation evaluations (*e.g.*, Hill, 1965), and continues to play an important role in such evaluations (*e.g.*, Goodman *et al*., 2010a). In the absence of such information, a causal relationship between exposure to a specific toxicant and a specific outcome cannot be confidently inferred, and increased uncertainty exists regarding the relevance of the data to the risk evaluations of interest. As noted below for the ozone case study, dose-response information was either unavailable or not evaluated in a number of the analyses presented in the NexGen report, weakening the analyses and the conclusions drawn from them.

Approaches to Addressing Susceptible Populations

A related issue in assessing the relevance of test data is evaluating whether any population subgroups exist that have heightened susceptibility to adverse effects. As with any other aspect of evaluating the risk assessment implications of testing data, the process of identifying susceptible populations and quantifying, if possible, the degree to which their sensitivity is enhanced must be comprehensive and rigorous. This aspect of assessing test relevance requires a sound understanding of the characteristic(s) of the identified subgroup that heighten their sensitivity to adverse effects and specific ways by which those characteristics may influence measured test endpoints and the progression from the test endpoints to the apical outcome of interest. As noted in Rhomberg (2009), a greater understanding of variability in population subgroups and individuals has led to a greater need to understand the impacts of that variability on disease processes and, in particular, to quantify how such variation can affect the probability and magnitude of adverse impacts.

Although the NexGen report briefly mentions this issue, little documentation is provided to support the identification of specific subgroups as "susceptible/sensitive" or address the influences of specific aspects of susceptibility on development of apical outcomes (see, *e.g.*, the comments on the ozone Tier 3 prototype analysis discussed in Section 4.2 below). Such issues warrant more detailed discussion in the NexGen report.

3.2 Adversity of Endpoint

Another related overarching issue that merits consideration in the NexGen report is the adversity of the effects being assessed in specific analyses, an issue that is related to data relevance but more specifically focused on the selected endpoint of concern. Again, although this issue is briefly identified as important in the report (*e.g.*, in Section 2.4 *Recurring Issues in Risk Assessment*), more detailed discussion of this topic is needed in the prototype evaluations. Aspects of this issue that should be addressed include the degree to which the selected outcome measure is appropriate for assessing risks of adverse health effects

and the strength of the support for a linkage between the selected outcome and an "adverse phenotypic outcome in an individual and population relevant to risk assessment" (US EPA, 2013a). Such analyses must also consider the influence of homeostasis and other adaptive responses and repair processes on the development of the identified adverse effect(s) (*e.g.*, as discussed in Goodman *et al*., 2010a).

3.3 Study Quality

Another aspect of study review that should be discussed more extensively in the NexGen report is the quality of the studies addressed in the specific prototype evaluations, as well as how considerations of data and study quality played a role in identifying studies for inclusion in the prototypes or study interpretation. The NexGen report identifies data quality as a criterion for selecting studies that were included in the prototype evaluations and indicates that the prototypes were identified, in part, on the basis of availability of "Multiple, high-quality studies" (*e.g.*, as listed in Box 3 in Section 3 *The Prototypes*). Moreover, the report also acknowledges the essential role of data and model quality on analysis conclusions. For example, within the ozone case study, the report observes that "the accuracy of [modeling] predictions depends on the extent and quality of the data used as inputs and on the technical quality of the model itself." However, within the documentation of the specific prototypes, the discussion of the role of study and data quality in selecting the prototypes, identifying the specific studies for evaluations, and interpreting the study findings should be enhanced. This documentation should reflect the aspects of critical review that were undertaken by the individuals conducting the prototype analyses. As discussed further in the following section, numerous frameworks have been developed to guide scientifically sound, systematic review of toxicity data, including critical aspects of data quality to address.

3.4 Comprehensiveness of Data Review/Transparency

In addition to providing more information on study quality, the NexGen report – and the ability of readers and reviewers to better assess the scientific validity of the conclusions and recommendations presented therein – could be enhanced by better documentation of the overall process used to select studies for inclusion in the prototypes, analyses conducted using those studies, and overarching perspectives applied in interpreting the prototype results. Such efforts should draw upon established frameworks that have been developed for conducting rigorous toxicity data reviews, which address issues such as study selection, study review, evidence integration, and conclusion development (*e.g.*, as reviewed in Rhomberg *et al*., 2013, which surveys approximately 50 WoE frameworks, and discussed in Goodman *et al*., 2013a; Linkov *et al*., 2009).

Thorough Review of Available Literature

The prototype-specific discussions in the NexGen report provide little documentation of the processes used to search and compile the evaluated literature. Although the report states that it is "not intended to be a comprehensive review of all available data that might be used in a risk assessment," the literature selection process should be systematic, better documented, and more transparent. Furthermore, the report should provide a more complete perspective regarding the representativeness of the reviewed literature. For example, it does not discuss what resources were searched, the time frame or extent of any searches, or specific search terms that were used. The report also does not discuss inclusion or exclusion criteria (*i.e.*, what factors were considered when selecting studies to include in or exclude from the prototype analyses). In particular, the report provides little perspective regarding the degree to which the selected studies reflect the available literature (*e.g.*, what portion of the available literature is represented or how the approaches used or reported results of the selected studies compare with other available studies). The

report could also be improved by additional discussion of alternative explanations for the observations reviewed in the prototypes (and the relative strength of the scientific support for the alternative conclusions) and whether studies reporting negative or null results were identified or considered in the prototype analyses. Without such information, it is impossible to determine the scientific validity and representativeness of the conclusions reached and recommendations made based on the prototype analyses (*e.g.*, conclusions that a particular analysis represents a "proof-of-concept").

Thorough Documentation of Analyses Performed

The documentation of the analyses conducted in the prototype evaluations could be similarly enhanced, *e.g.*, regarding specific approaches and assumptions that were applied, how conclusions were reached, and how alternative conclusions were excluded. The NexGen report currently omits important underlying information; *e.g.*, as discussed in Section 4.4 below, the ozone case study could be improved by providing certain assumptions included in the physiologically based pharmacokinetic (PBPK) models, as well as exposure levels tested in certain key studies. Again, these omissions undermine the usefulness of the report and the readers' ability to assess the scientific validity of the conclusions reached.

Evaluation of Consistency of Study Results

In addition to providing more complete documentation of the "mechanics" of the prototype data collection and analysis, the NexGen report could also be strengthened by placing the analyses in a more complete and informative context. In particular, it would be useful for the report to discuss the degree to which the observations in the reviewed studies are consistent with the understanding of the underlying biological processes. For example, where different research groups are applying the same test system, do the studies yield consistent results? Moreover, where study results are available for multiple measures (*e.g.*, several hormone levels or specific genes), do the relative results make biological sense [*e.g.*, do the measures all change in same (or expected) direction relative to each other]? For example, such an evaluation was included in a WoE analysis of associations between dioxin exposures and several biomarkers of thyroid function during early development (Goodman *et al*., 2010b). Again, this type of evaluation would strengthen the scientific foundation for assessing the NexGen report findings.

3.5 Approaches to Acknowledging and Addressing Uncertainty

The NexGen report would benefit from a more thorough acknowledgment and discussion of the uncertainties inherent in the current state of knowledge regarding new toxicity testing systems and analysis methods, as well as the implications of these uncertainties for risk assessment applications and future research needs. Although the report acknowledges the widespread insufficiency of current new toxicity data for the envisioned new risk assessment applications (even for relatively data-rich chemicals), such observations are not always adequately discussed in the context of the chemical-specific analyses. For example, the introductory discussion in the report notes that "many" of the currently available studies are "insufficient for the [new risk assessment] applications" explored in the report and that a "robust understanding and full implementation of new methods in general practice" will take substantial further research. Furthermore, the report notes that many of new scientific research areas (*i.e.*, molecular, computational, and systems biology) are "in their infancy" with respect to risk assessment applications. Similarly, one of the key overall conclusions drawn from the Tier 3 evaluations is that "even among the most well studied chemicals, very few chemicals had the type and quality of data needed for exploring the use of new data types in risk assessment. There are needs for systematic review criteria for new data types, adherence to standards of experimental and statistical practices in data generation and analyses, and thoughtful consideration of variability and uncertainty to improve the utility of new data types for risk assessment."

The report acknowledges some specific fundamental uncertainties in the prototype analyses: *e.g.*, the insufficiency of available dose-response information for polycyclic aromatic hydrocarbon (PAH) compounds and uncertainties regarding the mechanisms or existence of certain potentially ozone-induced signaling pathways. The report also notes that dose-response analyses were not undertaken for transcriptional changes ascribed to potential ozone effects. Conclusions drawn based on currently available data and analyses should be tempered by recognition of these types of fundamental knowledge gaps.

The report should also better acknowledge the fact that differing levels of uncertainty may be acceptable for different Tier applications; *i.e.*, the degree of uncertainty that can be tolerated in applying new toxicity data in screening programs (*e.g.*, Tier 1) must be substantially reduced for data applications in quantitative risk assessment applications (*i.e.*, Tier 3). This type of concept could be more explicitly incorporated into report components such as Table 10 (*Problem Formulation Table*), which includes an arrow indicating the need for "Increasing Evidence" as analysis goals move from Tier 1 to Tier 2 to Tier 3. This arrow could indicate the need for "Reduced Uncertainty" as well. Similar modifications could be made in Figures 2 and 5, as well as the text associated with these figures and Table 10. Moreover, the NexGen report should also more explicitly acknowledge that the development of sufficient, suitable data to support rigorous risk assessment applications will require an interplay between considerations of short-term utility and applications of specific data types with long-run ultimate goals (and maintaining progress toward those goals), as discussed in Rhomberg (2010).

3.6 Application of Case Studies to Other Situations/Chemicals

A final opportunity for strengthening the usefulness of the prototype analyses is to include more detailed discussion of the application of the specific evaluations for other chemicals and situations. Such discussions could address the robustness of conclusions reached, the types of chemicals and settings for which they might be relevant, and limitations on their applicability to other settings.

3.7 Other Issues

Markers of Toxicity *vs*. Investigations into Adverse-Outcome Pathways

An important distinction exists between interpreting gene expression changes as markers of known toxicity pathways and using them to investigate potential adverse-outcome pathways that have not been established and/or validated. The first case, which involves markers of known toxicity pathways, uses new information to interpret older understandings; the second is what the NRC report "Toxicity Testing in the Twenty-First Century: A Vision and a Strategy" envisioned as an ultimate goal for gene expression data (NRC, 2007). The way in which data are used, the uncertainties, and the dependence on existing knowledge differ between these two uses of the data.

With conventional MoA information, it is sometimes hard to determine which changes are directly causal of subsequent steps in a pathway of pathophysiological progression (*e.g.*, Pagan *et al*., 2007). This determination is even more challenging with new types of data because a thorough understanding has not been established to distinguish between molecular-level changes that are indicative of toxicity processes and those that demonstrate reactions of the body to other physiological processes. That is, some gene expression changes might simply be markers of cellular attempts to repair or accommodate stresses or damage, or attempts to survive and compensate for direct damages of the agent; however, these are not

the toxicity process itself, and may never lead to the apical endpoint of interest (Goodman *et al*., 2010a). Therefore, caution must be used when attempting to apply this type of data to a particular risk assessment.

As a result of these challenges, there are two ways that gene expression changes can be employed: (1) as markers for (at least partly) understood pathways and processes (*e.g.*, markers of cell division as a reaction to cytotoxicity), and (2) as purely empirical multivariate data that are associated with eventual toxicity processes only by statistical data-mining that does not produce or even aim at particular mechanistic understanding. The second approach may allow prediction of apical outcomes in conventional toxicity experiments based on pattern recognition and patterns of correlation of gene expression changes at specific loci, even if the reasons for the predictivity are not known. Many of the initial applications of massively parallel determinations (such as gene array studies) are of this type; however, such evaluations are only as useful as the degree to which they can be correlated with conventional toxicity studies and the degree to which these conventional studies are accurate predictors of human risk.

Low-Dose Extrapolation

As discussed in Section 4.3 of the NexGen report, a major challenge in risk assessment methodology is characterizing the expected response to low exposure levels of toxicants – *i.e.*, levels the public is most likely to encounter. The NexGen report correctly states that determining a point-of-departure (POD) on the dose-response curve is essential to human exposure guidelines for environmental agents. The NexGen report points to suggestions from NRC (2009, as cited in US EPA, 2013a) pertaining to adjustment factors applied to the POD based on the "expected" behavior of the exposure-response curve at low exposure levels, as well as the influence of background exposures and disease rates on the shape of the exposure-response curve at low exposure levels. Caution should be used, however, when deciding which methods and assumptions to use in modeling low-level exposure-response.

Rhomberg *et al*. (2011) discuss considerations for the "additivity to background" argument, for example, explaining the pitfalls of using a model developed to evaluate genotoxic carcinogens for the purposes of noncancer toxicity dose-response assessment. In the latter situation, Rhomberg *et al*. (2011) identify the challenge of translating modest degrees of underlying variation into discrete differences between healthy and diseased states. Ultimately, knowledge of endogenous levels of the toxicant under study, background levels of other stressors, background incidence of disease, relevant biological/physiological pathways, and biological mechanisms for coping with toxicant stressors is needed to justify assumptions made for such low-level exposure-response modeling on a case-by-case basis.

The concept of low-dose exposure-response has also been discussed in recent comments on the European Commission's recommended plan for regulation of endocrine-disrupting chemicals (Dietrich *et al*., 2013). The authors similarly recommend a substance-specific investigation of exposure and adverse effects for use in risk assessment, as opposed to blanket assumptions of non-monotonic dose-response that undermine consideration of threshold and safe limits of potential toxicants.

Low-dose extrapolation is one of many major risk assessment challenges. The use of new data types may help inform this issue but, as described in other sections, uncertainties and knowledge gaps about these new data may also further complicate the risk assessment process. More research is needed to address this complex but important issue.

4 Illustrative Case Study Example – Ozone

The NexGen report states that ozone was selected as a Tier 3 prototype pollutant because of the abundant data available regarding its effects on human health, specifically regarding lung injury and inflammation. The report asserts that the "mechanisms [of ozone toxicity] are well understood," and thus the available database of ozone-related studies provides an ideal resource to accomplish several goals: to demonstrate proof of concept for use of molecular biology data to inform human risk assessment, develop decision considerations for using such data, and explore the value of different types of information.

The ozone prototype focuses on ozone-induced pathways that potentially lead to inflammation. The report describes the adverse outcome pathway (AOP)-studies conducted by US EPA, which include *in vivo* studies that exposed young, healthy volunteers to filtered air and ozone (0.30 ppm); *in vitro* studies using a subset of cultured airway epithelial cells from filtered air-exposed individuals; and assessments of signaling pathways altered by exposure of cultured airway epithelial cells to ozone.

The ozone prototype represents an encouraging start to meeting US EPA's first goal, *i.e.*, where ozone could serve as a model pollutant to demonstrate proof-of-concept for use of molecular biology data in risk assessment. In particular, expanded evaluations of the ozone data within an AOP-based paradigm for risk assessment purposes could provide an opportunity to apply an increased understanding of homeostatic and perturbed biological pathways within risk assessment frameworks.

However, several essential discussions are limited or missing throughout the ozone prototype, which undermines the report's merit and validity of certain scientific conclusions. In several instances, the NexGen report does not provide a robust exploration of the literature or demonstrate how well understood the mechanisms of ozone toxicity actually are. For example, when introducing the framework for the ozone prototype, the NexGen report states that "several human studies characterize inflammation at multiple ozone concentrations," but these studies are not cited or discussed further. In addition, the report does not adequately discuss the merits and limitations of the studies conducted in the prototype work or those referenced to support its methods and choice of measured endpoints. When explored in greater depth, the literature on ozone-induced lung inflammation contains several studies that have inconsistent results, use ozone levels that are not reflective of environmental exposures, and fail to measure endpoints that directly correlate to adverse health effects. Addressing these concerns in the evaluations presented in the NexGen report would improve the scientific foundation of the analyses, as well as the implications for drawing broader conclusions regarding the use of new types of data in risk assessments. Exploring the inconsistencies and contrasting data in literature, as well as addressing issues concerning endpoint correlation with adverse health effects, will also increase the usefulness of such analyses in assessing health risks associated with ozone. Such analyses, however, may also demonstrate that the effects of ozone are not as clearly established as US EPA states.

US EPA has not yet attained its goal in this prototype of proof-of-concept of applications of molecular biology data for this pollutant. Similarly, more work is needed for US EPA to meet its second and third goals. To better achieve these goals, the NexGen report evaluations should be augmented to more thoroughly discuss considerations for using and weighing the value of new types of information in risk assessment applications. This type of discussion begins in Sections 5 and 6 of the NexGen report, which mention lessons learned from the prototypes, but it falls short of adequately addressing the greater implications of the uncertainty and limitations inherent to the new tools and types of data being used. Below, we provide an overview of illustrative issues identified in the ozone case study. This overview is

followed by a more detailed discussion of specific issues and recommendations for improving the scientific validity of the analyses presented in the NexGen report and their usefulness for informing risk evaluations and identifying future research needs.

4.1 Overview of Major Issues

The ozone Tier 3 prototype evaluation presented in the NexGen report illustrates a number of key opportunities for improvement (as discussed in Sections 2 and 3 of these comments). Most notably, in the current version of the NexGen report, the usefulness of US EPA's ozone model and the conclusions drawn based on that model are hampered by the insufficient context for its analyses. In particular, by failing to thoroughly discuss the full set of available data and test methods for the health effects being evaluated, US EPA undermines the robustness of the data set selected for prototype evaluation. US EPA needs to address missing information on the methods used and provide a discussion of uncertainties in the model before this tool can be used to accurately predict toxicity of ozone or any other pollutant. Listed below are specific issues that merit additional discussion and review in the NexGen report's ozone prototype evaluation.

- **Relevance of test systems and exposure levels:** The report should provide an expanded discussion of the relevance of the test systems used in the evaluated studies, as well as the environmental relevance of the tested ozone levels.
- **Basis for susceptibility models:** The studies used to discuss susceptibility to ozone health effects make several inferences about gene-environment interactions (*e.g.*, GSTM1 null genotype) and other conditions (*e.g.*, asthma) and their potential connection to increased susceptibility to ozone-induced lung injury. The nature of and scientific support for these potential associations should be discussed and documented to a greater degree in the report.
- **Relevance of transcriptional gene changes to human health:** The analyses presented in the NexGen report potentially overstate ozone-induced effects on lung injury because the report does not adequately document the actual effects that alterations in gene networks have on the manifestation of lung inflammation and subsequent injury. The report should clarify how signaling pathways affected by ozone were selected with respect to their relevance to adverse effects on lung function, the correlation between pathways altered *in vitro* and those affected by *in vivo* exposure, and the connection of altered pathways to those affected in individuals with lung injury.
- **Sufficiency of clear dose-response evidence:** Clear dose-response evidence for markers of injury and gene expression/pathway changes over a range of doses is essential for adequately concluding that an ozone-specific effect has occurred, as well as for accurately extrapolating the potential effects associated with lower, more environmentally relevant doses based on the higher doses typically used in the test systems. Such dose-response assessments are missing from the *in vivo* and *in vitro* portions of the report. For example, while there are understandable limitations for multiple bronchial cell collections from exposed human patients, *in vitro* dose-response analyses should be feasible and are recommended to make better conclusions about the causality of any effects by ozone.
- **Approaches used to address uncertainty and model validation:** The report would be enhanced if it provided greater confirmation and validation of the methods underlying certain assumptions. In particular, the report should expand the discussion of the procedures used to normalize the dose of ozone delivered to cultured cells and humans using labeled oxygen experiments; the parameters, procedures, and results of the PBPK analyses and models; and how uncertainty was addressed in such models and assumptions.

These topics are discussed in more detail in the following sections.

4.2 Data Relevance

Relevance of Test Systems

The ozone case study benefits from use of data from an established and commonly used test system of primary human bronchial epithelial cells that are cultured after collection from exposed and unexposed individuals. Such primary epithelial cells have been used to study inflammatory and oxidative stress-related effects of other toxicants, including air pollutants like particulate matter, as well as various nanomaterials (Russo *et al*., 2005; Fujii *et al*., 2001; King *et al*., 1998). As this system uses primary cells from human subjects, it is also likely to more closely reflect *in vivo* response than bronchial cell lines that have been transformed and altered in various ways. These cells also offer the advantage of being able to serve as their own controls; *i.e.*, results from exposed cultured cells can be compared to those in unexposed cells from the same subject.

As with any system and model, a number of factors should be considered when interpreting study results. Similar to the discussions presented in Klein *et al*. (2011) for this and comparable *in vitro* systems, the NexGen report should discuss information necessary to assess the relevance of the test system for the apical outcome of interest, the system's benefits and limitations, and the sufficiency of the current understanding regarding the reliability, implications, and interpretation of the test results. For example, the NexGen report should discuss whether this test system of primary bronchial epithelial cells is the only and/or best way to study lung inflammation (as compared to cultured cell lines, for instance, or methods other than air-liquid interface cultures). Also, to what degree are data available indicating consistency among studies using the same system and different test chemicals (*e.g.*, other air pollutants *versus* unrelated toxicants)? Are data available from other relevant test systems and, if so, do the results from those tests lead to similar conclusions regarding ozone effects? Would interpretations of the degree of toxicity, or specific toxic effects, be expected to differ depending on the test system used? In addition, does any available information indicate whether any observed changes in the epithelial cells occur due to the collection process? To help evaluate this last question, it may be useful to see if there are any biomarker data that could be collected *in vivo* or with pathology endpoints from tissue samples that could be compared with observations in the cultured cells. A discussion addressing these questions would enhance understanding of the strength, relevance, and applicability of interpretations and conclusions based on the test data.

Relevance of Exposure Levels/Routes

When constructing a tool that analyzes changes to relevant biological pathways, working within a concentration range relevant to actual environmental exposures is critical. Many contaminants elicit different biological responses at different exposure levels, and the mechanistic response at a very high exposure level is often not reflective of the response to smaller exposures. For these reasons, relevant levels of ozone need to be used in this study to make appropriate conclusions about the pathways that are, or might be, affected in realistic human exposure scenarios. Alternatively, if data from higher exposure levels are to be used in risk assessments, issues associated with extrapolating such findings to predict potential effects at lower exposures must be thoroughly explored.

In the *in vivo* portion of the Tier 3 prototype study, the NexGen report describes an ozone concentration of 0.30 ppm as "a relevant concentration of ozone." It is important to note, however, that this concentration is four times the National Ambient Air Quality Standard (NAAQS) for ozone (0.075 ppm)

and not reflective of environmental levels in any US cities (US EPA, 2013c). In addition, concentrations tested in the subsequent airway epithelial *in vitro* studies were not clearly identified in the NexGen report. Specifically, in describing certain dosing levels, the NexGen report states, "These cells were exposed to concentrations of ozone that had been shown (from the results of $^{18}O_3$ experiments) to be comparable to the dose of ozone encountered by airway epithelial cells following a specified *in vivo* exposure." The NexGen report describes this approach as acceptable, but it should provide additional documentation to support this conclusion (*e.g.*, a sample calculation of exactly how the *in vitro* concentrations were derived for this particular study).

Exposure route relevance is also an important factor to consider. Humans are primarily exposed to ozone in the environment *via* inhalation. While this helps to justify the evaluation of lung function and injury in response to ozone, it also raises questions regarding differences that may exist between procedures used to expose epithelial cells *in vitro* and cell exposures *in vivo*. Assuming there is no effect of the collection process itself (highlighted in the previous section), this issue is partly addressed in the prototype *in vivo* studies, where cells from exposed *versus* unexposed individuals are analyzed. For the *in vitro* tests, however, cells are exposed to ozone outside of the body, and this setting may yield a different exposure than what is experienced *in vivo*. This issue is discussed further by Klein *et al*. (2011), who suggest alternative *in vitro* methods that use submerged cultivation of cells instead of the air-liquid interface, or multi-cell co-culture models that more closely resemble *in vivo* situations.

Models and test systems that can more accurately represent chronic exposures would also be useful. In this way, effects of long-term ozone exposure could be assessed by using lower, more relevant levels of ozone over a longer time period (which is more reflective of actual human exposures). If a cell culture model was used for this purpose, it would be important to address issues associated with long-term cultures and any potential effects inherent to the prolonged *in vitro* setting. As with any model, these systems would need to be validated against some higher standard, such as epidemiology evidence or *in vivo* exposures.

Evidence of Dose-Response Relationships

Evidence of dose-response relationships is another important component of evaluating data relevance. The NexGen report states that microarray and quantitative proteomics were used to "identify and define pathways affected by ozone in [the epithelial] cells." However, these changes were not assessed over a range of doses in any of these studies. To confidently conclude that the observed effects (*e.g.*, changes in gene expression or biomarker levels) are actually caused by ozone, establishing a dose-response relationship is essential. Testing multiple concentrations on human subjects *in vivo* is not ideal since there is a relatively invasive procedure involved for bronchial cell collection. However, it is feasible and recommended to test a range of ozone concentrations in *in vitro* studies to more reliably draw conclusions regarding whether any observed effects are caused by ozone and if they occur at environmentally relevant concentrations. Moreover, low-dose extrapolations require a range of concentrations as a foundation for appropriate calculations. If the *in vitro* models are appropriately validated (as previously discussed), useful dose-response assessments could be conducted and applied to future quantitative risk assessments for ozone.

Approaches to Addressing Susceptible Populations

In the NexGen report, US EPA states that "performing a larger study of variability and susceptibility would be possible by recruiting and including specific populations." To appropriately identify specific research needs, however, it is necessary to conduct a more comprehensive and rigorous review of the literature addressing susceptible populations. The NexGen report indicates that susceptible populations

may include individuals in specific age ranges or with asthma or GSTM1 null genotypes, but it references only one or two studies for each of these populations. In contrast, in recent comments on the US EPA's Integrated Science Assessment on health effects of ozone, Goodman and Sax (2012) concluded that while there is substantial evidence supporting increased susceptibility in elderly populations, the case for increased susceptibility in younger individuals was much less clear. Thus, a more balanced and comprehensive review of the literature on the topic of susceptible populations is needed to provide a scientifically sound foundation for identifying future research needs.

The models discussed in the Susceptibility section of the NexGen report also can prove useful for rapid, *in vitro* screening of chemicals and vulnerable populations (*e.g.*, airway epithelial cells obtained from asthmatics, or nasal epithelial cells as alternatives to bronchial cells that require an invasive collection procedure). However, these would similarly need to be validated as representative of their respective populations before being used for more quantitative or rigorous human health risk assessment.

4.3 Adversity of Endpoint

An essential component of the applications of new toxicity data explored in the NexGen report is to clearly link the endpoints that will be measured in the test systems to adverse human health effects. The NexGen report briefly discusses this issue in Section 5, where it acknowledges the need for new methods to be adequately demonstrated as reliable and relevant in the context of the current state of knowledge. In Section 6, the report mentions challenges related to gaps in knowledge of the intermediate steps of perturbed pathways and their relation to an apical disease outcome. These types of discussions are necessary and should be expanded to include detail on how such data related to altered pathways and/or intermediate factors (*e.g.*, biomarkers of exposure or effect) will be weighed with respect to what can be definitively said about their direct impact on adverse health effects. For example, increased biomarkers of inflammation and markers of cell injury can be useful tools for observing effects of *in vivo* exposure to ozone. However, it is critical to adequately address the specificity and accuracy of such biomarkers (as indicators of exposure and toxicity) in relation to the pollutant of interest, as well as any confounding factors that may impact biomarker levels (Goodman *et al*., 2010a).

The genomic and proteomic analyses discussed in the ozone case study are less clear in their relevance to adverse human health effects. While such limitations are mentioned briefly in the latter sections of the NexGen report, the link between such genomic and proteomic changes and the adverse effect of interest (*i.e.*, lung injury) is not discussed. The case study in the NexGen report focuses on lung inflammation and injury, with the assumption that increased inflammation is indicative of an adverse effect on lung function. The report needs to expand on how increases in these inflammatory pathways ultimately translate into definitive damage to lung function. To demonstrate macro-level changes and impairments, it is essential to correlate these gene-level alterations with cellular or tissue morphology endpoints.

The evaluation of these endpoints in the NexGen report should consider whether all genes altered by ozone exposure (*e.g.*, that are up-or down-regulated) are modified in ways that are consistent with the understanding of underlying biological processes and how any inconsistent findings should be interpreted. The report should also address whether the selected pathways are affected equally by *in vivo* and *in vitro* exposures. Alternatively, if not, the report should discuss the extent to which only those genes changed as a result of *in vivo* exposure are included for *in vitro* or pathway analyses. These types of questions could be clarified by explaining how changes in one system but not the other would be reconciled. Such evaluations should also consider how these changes compare with homeostatic conditions of any fluctuations in levels or alterations in response to other external factors, including stress or diet. Another relevant contextual issue that merits review in the NexGen analysis is whether any of the genes activated in these pathways have been studied separately in the literature with respect to alterations resulting from

ozone exposure. These types of evaluations will be especially useful for future use of the test system for other, less-studied toxicants.

A recent collection of papers furthers this discussion of omics-based biomarkers, highlighting the challenges inherent to these new technologies in their use for linking environmental exposures and disease (Kyrtopoulos, 2013; McHale *et al*., 2013; Vineis *et al*., 2013). In particular, Kyrtopoulos (2013) provides a thorough review of the use of omics data in several examples of substance-specific population studies (*e.g.*, tobacco smoke and PCBs), as well as a discussion of challenges and limitations specific to linking exposure with disease through omics profiles.

4.4 Comprehensiveness of Data Review and Study Quality

Systematic Review of Available Literature

Ozone-induced inflammation is documented in the literature, although not without conflicting results. A comprehensive discussion of such studies is missing from the ozone case study, which undermines confidence in the conclusions presented in the NexGen report and potentially leads to overstating the applicability and usefulness of the new test system data. In many instances in the ozone prototype analysis, US EPA references specific studies to justify methods used (*e.g.*, use of $^{18}O_2$ to normalize ozone dose between *in vivo* and *in vitro* exposures) or certain measured endpoints (*e.g.*, specific biomarkers of inflammation compared to those reported in published studies). In most of these instances, however, there is no reference to the degree to which the selected studies are representative of the available literature or the quality of the studies that are used. To address this key omission, the NexGen report should provide more information on the process by which the specific studies referenced were chosen. This discussion should briefly address other available methods and studies that were not used in their analyses, together with reasons as to why they were excluded.

A similar issue arises in the discussion of susceptible populations, where only one or two studies are listed as a reference for a particular conclusion. This section is particularly lacking in citations to relevant studies to support concerns for specific population subgroups – making it challenging to assess the validity of the claims or the degree to which they reflect the available literature. The same criteria used to evaluate literature indicating a chemical hazard should be applied when attempting to define sensitive populations. A thorough review of such literature is essential to ensure a sound perspective on the full spectrum of relevant information.

Thorough Documentation of Analyses Performed

In a number of instances, the ozone case study inadequately documents the analyses performed or literature that was used to support various aspects of US EPA's approach. As mentioned above, assumptions included in the PBPK models, as well as ozone levels tested in the *in vitro* studies, were unclear and not stated explicitly in the text. This deficiency should be addressed, and the NexGen report should also specifically provide greater detail regarding the standards used to measure and evaluate changes in relevant biomarkers, genes/proteins, and pathways altered by ozone exposure.

Evaluation of Consistency and Coherence of Study Results

The NexGen report does not fully discuss the consistency or coherence of results from US EPA studies with other studies in the literature that evaluate similar or related endpoints. Some endpoints, such as alterations in genes involved in inflammatory pathways, are not discussed in the context of existing

studies at all. For other endpoints, the NexGen report addresses consistency and coherence in a limited way. For example, in the *in vivo* study section, the report discusses comparing specific biomarkers of inflammation with those measured in one other study in the scientific literature, but it does not put the results of either study in the context of the larger body of literature.

The NexGen report should discuss whether any relevant genes that are altered in US EPA studies are altered in the same way in other studies using this or other test systems (*i.e.*, in the expected direction relative to some other factor). With respect to inflammation, the report should closely evaluate the up-regulation of genes controlling relevant cytokines or other markers of inflammation, and how their alterations compare to expected effects in established cases of disease/injury-causing inflammation. Similarly, the report should discuss the choice of the endpoint from a specific test system within the context of the lung inflammation literature, justifying its association with documented cases of lung inflammation and injury.

Although the NexGen report is not intended to be a comprehensive review of the literature, it should place the studies presented in the context of the existing literature.

4.5 Approaches to Acknowledging and Addressing Uncertainty

The NexGen report identifies strategies that are used to reduce uncertainty when comparing results from *in vitro* and *in vivo* studies. For example, the report discusses using model systems in which both *in vitro* and *in vivo* data are available to "validate how well pathway information from the former can predict human responses to toxicants." Such approaches will help support sound conclusions regarding the relevance and accuracy of any *in vitro* assays predicting *in vivo* responses to ozone (or any toxicant). For the *in vitro* and signaling pathway analyses, the NexGen report discusses the ability to compare *in vitro* and *in vivo* response from cells from the same person for similar exposures. This type of approach can reduce a key source of uncertainty because each person serves as his or her own control. Such comparisons help to validate the use of *in vitro* assays in place of *in vivo* studies in the future.

The NexGen report also discusses coupling the results of quantitative *in vitro* assays with appropriate PBPK modeling to estimate tissue doses (and ultimately, to perform hazard identification and dose-response assessment). However, such models require many assumptions (*e.g.*, regarding exposure routes, exposure kinetics, and compartmentalization within the body), especially in settings where formal PBPK modeling has not been completed. To allow an informed perspective on the strengths and limitations of such modeling, the NexGen report should provide more detail regarding the sources of such information, reliability of any studies from which this information is derived, and implications of identified uncertainties for interpreting the results of the risk analyses.

4.6 Application of Case Study to Other Situations/Chemicals

While mechanistic studies and systems biology have many possible applications in risk assessment, the NexGen report should provide a realistic assessment of study limitations and how they may affect the conclusions that can be drawn regarding specific prototype toxicants and models. Some of these limitations are addressed in the ozone case study analyses. For example, the NexGen report acknowledges the multitude of complications involved in attempting to link one ubiquitous environmental pollutant with certain effects when many factors induce similar cardiopulmonary inflammation. In other instances, limitations are noted but not fully considered when presenting the conclusions [*e.g.*, in the discussion of the use of downstream mechanistic data on ozone-induced reactive oxygen species (ROS) generation]. The NexGen report acknowledges that "[w]hether the ROS produced following ozone

exposure actually activates downstream pathways *via* the mechanism shown in Figure 12 is unknown"; *i.e.*, it is not actually known if the pathways associated with ozone-induced inflammation in these studies [*i.e.*, NF-kB and extracellular signal-related kinase (ERK) signaling pathways] are elicited by ozone-induced ROS production. Regardless, this assumption is still used in its systems biology approach, with no mention of how such limitations would be addressed in the conclusions drawn from the analyses. In the absence of empirical data supporting the ability for ozone-induced ROS to upregulate these same pathways, these assumptions should not be included in the model for ozone without a larger discussion of uncertainties.

Evidence supporting similar mechanisms of action between ozone and other pollutants is necessary to warrant the use of this model for other toxicants (*i.e.*, to achieve one of the key goals identified by US EPA). One step toward achieving this goal would include a discussion of similar and dissimilar endpoints relevant to the different toxicants, including the relevance of the inflammatory pathways and biomarkers that are assessed in the case of ozone. Cases where affected pathways differ will need to be adequately addressed before moving forward to apply this model to less-studied pollutants.

4.7 Conclusions

Overall, the NexGen report provides a promising start to using new types of data for human health risk assessment in the Tier 3 prototype with ozone. However, US EPA begins this endeavor with the assumption that the relationship between ozone-induced lung inflammation and lung injury is well established. The subsequently presented framework operates under this assumption and provides data supporting these pathways. A major drawback to this approach is that many uncertainties and knowledge gaps are overlooked. While the NexGen report is not meant to be a full systematic review of the literature of ozone-induced lung inflammation and injury, the assessments incorporated into its studies and analyses need to be representative of the science surrounding this topic. Addressing the inconsistencies and issues discussed here could demonstrate that this ozone case study, and perhaps the case studies for other Tier 3 prototype toxicants, is far more uncertain than is acceptable for a prototype built "by 'reverse engineering' from known public health risks" to verify the use of new types of data. Without a sound basis for the conclusion that the health risk is known, any further efforts to verify these models and methods for future risk assessment is unwarranted.

5 Conclusions

As amply illustrated by the NexGen report, new toxicity testing methods and analytical approaches offer many opportunities for enhancing our understanding of toxicological processes and expanding our capabilities to assess health risks posed by chemical exposures. As detailed in these comments, however, the usefulness of the NexGen analyses for improving risk assessment and guiding future research should be enhanced by key modifications:

- The report (particularly the prototype analyses) should reflect greater consideration of certain essential concepts that underlie sound and systematic WoE reviews of toxicity information, including data relevance, adversity of effect, evidence of causation, data quality, and fundamental sources of uncertainty. Based on consideration of these factors, the report should more explicitly identify the types of analyses that are feasible with currently available data and should evaluate when new data (despite limitations) may reflect an improvement for risk assessment decision-making (relative to having no or limited data or more traditional data sources).
- The report should better acknowledge that the degree of uncertainty that is "acceptable" in various data sets will vary by the specific Tier/application of the data or, alternatively, that the suitability of different approaches for use in different Tiers/applications will vary depending on the uncertainty inherent in the data.
- The report should also more extensively recognize that suitable applications for various testing and analysis approaches will vary depending on their inherent uncertainty and will evolve over time as the knowledge base and experience with their applications expands.

The report could also serve as a more useful guidepost to future research directions by strengthening several key areas of research recommendations, including:

- More strongly highlighting the recognized needs in the NexGen report for stronger frameworks and standards to guide study design, data generation, and data analysis to enhance the usefulness of generated data. Such frameworks should address:
  - Approaches to enhance the consistency of data collection, documentation, and analysis to aid future data synthesis. Such guidance could help avoid the problems noted in the PAH prototype, where inconsistencies in all three aspects of data generation and study performance greatly impeded efforts in the analyses to synthesize available data.
  - Approaches for systematic review criteria for new data types. As discussed in Goodman *et al*. (2013b), development of such criteria (*e.g.*, for application of meta-analysis techniques) is not only useful for the specific support provided for study design and interpretation, but also for fostering a focused perspective among researchers on the specific study elements that need to be comparable across studies to allow study synthesis.
- Encouraging efforts to identify and prioritize specific mechanisms of action on which to focus research. As illustrated by the NexGen report prototypes, substantial data gaps exist – even in the data available for the relatively data-rich Tier 3 prototypes. The advancement of applications of new data types in risk assessment might benefit from focused research efforts to fill the identified information gaps for several selected mechanisms of action and continue to expand our understanding of what is possible with more data and which types of data are more or less useful.

- Providing perspective on the long-term goals for the use of new toxicity data (and how they may differ from short-term data uses) and identifying specific research areas to help achieve those goals.

Although the new toxicity data types reviewed in the NexGen report will require substantial additional work before they can be used reliably in rigorous risk assessment applications, they clearly offer exciting opportunities for advancing toxicological science and risk assessment understanding. With the types of enhancements encouraged in these comments, the NexGen report could serve an influential role in achieving the promise of these new research approaches.

References

Audouze, K; Juncker, AS; Roque, FJ; Krysiak-Baltyn, K; Weinhold, N; Taboureau, O; Jensen, TS; Brunak, S. 2010. "Deciphering diseases and biological targets for environmental chemicals using toxicogenomics networks." *PLoS Comput. Biol.* 6(5):e1000788.

Burgess-Herbert, SL; Euling, SY. 2013. "Use of comparative genomics approaches to characterize interspecies differences in response to environmental chemicals: Challenges, opportunities, and research needs." *Toxicol. Appl. Pharmacol.* 271(3):372-385.

Chiu, WA; Euling, SY; Scott, CS; Subramaniam, RP. 2013. "Approaches to advancing quantitative human health risk assessment of environmental chemicals in the post-genomic era." *Toxicol. Appl. Pharmacol.* 271(3):309-323.

Dietrich, DR; von Aulock, S; Marquardt, H; Blaauboer, B; Dekant, W; Kehrer, J; Hengstler, J; Collier, A; Gori, GB; Pelkonen, O; Lang, F; Nijkamp, FP; Stemmer, K; Li, A; Savolainen, K; Hayes, AW; Gooderham, N; Harvey, A. 2013. "Scientifically unfounded precaution drives European Commission's recommendations on EDC regulation, while defying common sense, well-established science and risk assessment principles." *Regul. Toxicol. Pharmacol.* doi:10.1016/j.yrtph.2013.07.001.

Fujii, T; Hayashi, S; Hogg, JC; Vincent, R; Van Eeden, SF. 2001. "Particulate matter induces cytokine expression in human bronchial epithelial cells." *Am. J. Respir. Cell Mol. Biol.* 25(3):265-271.

Goodman, JE; Dodge, DG; Bailey, LA. 2010a. "A framework for assessing causality and adverse effects in humans with a case study of sulfur dioxide." *Regul. Toxicol. Pharmacol.* 58:308-322.

Goodman, JE; Kerper, LE; Petito Boyce, C; Prueitt, RL; Rhomberg, LR. 2010b. "Weight-of-evidence analysis of human exposures to dioxins and dioxin-like compounds and associations with thyroid hormone levels during early development." *Regul. Toxicol. Pharmacol*. 58(1):79-99.

Goodman, JE; Prueitt, RL; Sax, SN; Bailey, LA; Rhomberg, LR. 2013a. "Evaluation of the causal framework used for setting National Ambient Air Quality Standards." *Crit. Rev. Toxicol*. doi:10.3109/10408444.2013.837864.

Goodman, JE; Petito Boyce, C; Sax, SN; Beyer, LA; Prueitt, RL. 2013b. "Rethinking Meta-analysis: Applications for Air Pollution Data and Beyond." Presented at Methods for Research Synthesis: A Cross-Disciplinary Workshop, Harvard Center for Risk Analysis, Cambridge MA, October 3.

Goodman, JE; Sax, SN. 2012. "Comments on the Integrated Science Assessment for Ozone and Related Photochemical Oxidants (Third External Review Draft)." Report to American Petroleum Institute. 125p., August 15.

Hill, AB. 1965. "The environment and disease: Association or causation?" *Proc. R. Soc. Med.* 58:295-300.

King, C; Brennan, S; Thompson, PJ; Stewart, GA. 1998. "Dust mite proteolytic allergens induce cytokine release from cultured airway epithelium." *J. Immunol.* 161(7):3645-3651.

Klein, SG; Hennen, J; Serchi, T; Blomeke, B; Gutleb, AC. 2011. "Potential of coculture in vitro models to study inflammatory and sensitizing effects of particles on the lung." *Toxicol. In Vitro* 25(8):1516-1534.

Kyrtopoulos, SA. 2013. "Making sense of OMICS data in population-based environmental health studies." *Environ. Mol. Mutagen*. 54(7):468-479.

Linkov, I; Loney, D; Cormier, S; Satterstrom, FK; Bridges, T. 2009. "Weight-of-evidence evaluation in environmental assessment: Review of qualitative and quantitative approaches." *Sci. Total Environ.* 407(19):5199-5205.

McHale, CM; Zhang, L; Thomas, R; Smith, MT. 2013. "Analysis of the transcriptome in molecular epidemiology studies." *Environ. Mol. Mutagen*. 54(7):500-517.

National Research Council (NRC). 2007. "Toxicity Testing in the Twenty-first Century: A Vision and a Strategy."National Academies Press, Washington, DC. 115p.

Ornish, D; Lin, J; Daubenmier, J; Weidner, G; Epel, E; Kemp, C; Magbanua, MJ; Marlin, R; Yglecias, L; Carroll, PR; Blackburn, EH. 2008. "Increased telomerase activity and comprehensive lifestyle changes: A pilot study." *Lancet Oncol.* 9(11):1048-1057.

Ornish, D; Lin, J; Chan, JM; Epel, E; Kemp, C; Weidner, G; Marlin, R; Frenda, SJ; Magbanua, MJ; Daubenmier, J; Estay, I; Hills, NK; Chainani-Wu, N; Carroll, PR; Blackburn, EH. 2013. "Effect of comprehensive lifestyle changes on telomerase activity and telomere length in men with biopsy-proven low-risk prostate cancer: 5-year follow-up of a descriptive pilot study." *Lancet Oncol*. 14(11):1112-1120.

Pagan, I; Haber, LT; Rhomberg, LR; Goodman, JE; Dodge, DG; Foureman, GL. 2007. "Development of a Pathophysiological Progression Model for Selected Endpoints." Presented at Society for Risk Analysis Annual Meeting, San Antonio, TX, December 8-12.

Rhomberg, LR. 2009. "Risk assessment in the 21st Century: Changes wrought by changing science." *Risk Anal*. 29(4):488-489.

Rhomberg, LR. 2010. "Toxicity testing in the 21st century: How will it affect risk assessment?" *J. Toxicol. Environ. Health B* 13:361-375.

Rhomberg, LR. 2011. "Interpreting Dose-response Information on Intermediate Stages of Causal Ccascades in Toxicity Mode of Action." Presented at Society of Toxicology 50th Annual Meeting, Washington, DC, March 8.

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.

Russo, P; Catassi, A; Cesario, A; Imperatori, A; Rotolo, N; Fini, M; Granone, P; Dominioni, L. 2005."Molecular mechanisms of hexavalent chromium-induced apoptosis in human bronchoalveolar cells." *Am. J. Respir. Cell Mol. Biol.* 33(6):589-600.

Scholz, S; Sela, E; Blaha, L; Braunbeck, T; Galay-Burgos, M; García-Franco, M; *et al*. 2013. "A European perspective on alternatives to animal testing for environmental hazard identification and risk assessment." *Regul. Toxicol. Pharmacol.* doi:10.1016/j.yrtph.2013.10.003.

US EPA. 2013a. "Next Generation Risk Assessment: Incorporation of Recent Advances in Molecular, Computational, and Systems Biology (External review draft)." National Center for Environmental Assessment (NCEA), EPA/600/R-13/214A. Accessed on October 11, 2013 at http://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=259936#Download, 142p., September.

US EPA. 2013b. "State of the Science Evaluation: Nonmonotonic Dose Responses as They Apply to Estrogen, Androgen, and Thyroid Pathways and EPA Testing and Assessment Procedures (Draft)." Available at http://epa.gov/ncct/download_files/edr/NMDR.pdf, June.

US EPA. 2013c. "Air Trends: Ozone." Accessed on November 6, 2013 at http://www.epa.gov/airtrends/ozone.html. Page last updated September 3, 2013.

Vineis, P; van Veldhoven, K; Chadeau-Hyam, M; Athersuch, TJ. 2013. "Advancing the application of omics-based biomarkers in environmental epidemiology." *Environ. Mol. Mutagen*. 54(7):461-467.