Regulatory Toxicology and Pharmacology 67 (2013) S86–S93

Contents lists available at SciVerse ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Peer consultation on relationship between PAC profile and toxicity of petroleum substances

Jacqueline Patterson *, Andrew Maier, Melissa Kohrman-Vincent, Michael L. Dourson

Toxicology Excellence for Risk Assessment (TERA), USA

ARTICLE INFO

Article history:
Available online 19 November 2012

Keywords:
Peer consultation
Peer review
PAC (polycyclic aromatic compound)
Petroleum HPV chemical challenge
ARC
QCAR
HBPS
Quantitative composition activity relationship

ABSTRACT

An expert peer consultation panel reviewed a report by the PAC Analysis Task Group, which hypothesized that systemic, developmental, and reproductive toxicity observed in repeated-dose dermal toxicity studies was related to polycyclic aromatic compound (PAC) content. Peer consultations seek to solicit scientific and technical input from experts on the scientific basis and merits of the subject report. This peer consultation panel included nine scientists with expertise in petroleum chemistry, biostatistics, toxicology, risk assessment, structure activity, and reproductive and developmental toxicology. The panel evaluated the technical quality of the PAC report and provided recommendations for improving the statistical and biological approaches. The PAC report authors revised their methods and documentation, which are published elsewhere in this supplement. A review of the post peer consultation manuscripts confirmed that many of the key suggestions from expert panel members were considered and incorporated. In cases where the PAC report authors did not fully incorporate panel suggestions from the peer consultation, they have provided an explanation and support for their decision. This peer consultation demonstrates the value of formal engagement of peers in development of new scientific methods and approaches.

1. Introduction

As part of the EPA Chemical Challenge Program, the Petroleum High Production Volume (HPV) Testing Group[1] has submitted test plans to the United States Environment Protection Agency (US EPA) for 13 petroleum substances categories. For seven of these (aromatic extracts, asphalt, crude oil, gas oils, heavy fuel oils, lubricating oil base stocks, and waxes and related materials) a relationship between toxicity and the polycyclic aromatic compound (PAC) content was asserted or implied. The PAC Analysis Task Group of the Petroleum HPV Testing Group (Task Group) prepared a report ("An Investigation into the Relationship between the Polycyclic Aromatic Compound Content and Acute, Repeat-Dose, Developmental, and Reproductive Toxicity of Petroleum Substances"), which underwent an expert peer consultation in October 2007. The report hypothesized that systemic, developmental, and reproductive toxicity observed in repeated-dose dermal toxicity studies are associated with PAC content and that the PAC profile (aka Aromatic Ring Class [ARC] profile) can then be used to predict a dose–response estimate for the repeated-dose and developmental endpoints of an untested petroleum stream.

The Task Group collected toxicity and analytical study reports on high-boiling petroleum substances (HBPS) (final boiling points are ⩾ approximately 650 °F [343 °C]) from the Petroleum HPV Testing Group's member companies and analyzed these data in order to address two key questions:

(1) Are there quantitative relationships between PAC content[2] of petroleum substances and their critical effects as identified in repeat-dose, developmental, and reproductive toxicity studies?
(2) Can the critical effects/levels of untested petroleum substances, as would be identified in an OECD 422 study, be predicted using their PAC content?

The American Petroleum Institute (API) asked Toxicology Excellence for Risk Assessment (TERA) to develop a peer consultation process to ensure expert input on these key questions. This paper describes the process and results of the 2007 peer consultation:

* Corresponding author. Address: Toxicology Excellence for Risk Assessment (TERA), 2300 Montana Ave., Suite 409, Cincinnati, OH 45211, USA. Fax: +1 513 542 7487.
E-mail address: patterson@tera.org (J. Patterson).

[1] The Petroleum HPV Testing Group (PHPTVG) is an unincorporated group of manufacturers affiliated by contractual obligation to fund a voluntary data disclosure and toxicity testing program on certain petroleum-related chemical substances in response to EPA's HPV Challenge Program. The American Petroleum Institute (API) manages the Groups' activities.

[2] At the time of the peer consultation meeting, the term "PAC content" was used. The authors subsequently referred to this as the "ARC profile" and use that term throughout the other publications in this supplement.

0273-2300/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.yrtph.2012.11.005

including topics discussed, panel recommendations, and key scientific issues and resolutions. Other papers in this supplement (i.e., Murray et al., 2013; Nicolich et al., 2013; Roth et al., 2013; Simpson et al., this issue) describe the PAC evaluation/determination method, underlying data, statistical validation, predictive modeling results, and possible applications of the results. These four papers reflect the method and information reviewed by the expert panel, as well as improvements and extensions to the 2007 documentation. Three additional papers in this supplement were not part of the 2007 peer consultation (i.e., Gray et al., 2013; McKee et al., 2013; Murray et al., 2013).

2. Methods

The purpose of a peer consultation is to solicit scientific and technical input from expert scientists who have the requisite experience to evaluate the scientific basis and merits of the subject report. In a peer consultation the experts' opinions are sought while still developing the work product in order to include the experts' input to improve the product. Peer consultation differs from a peer review, where the final work product is presented and experts are asked to judge its adequacy and reach a group consensus on the appropriateness of the conclusions. This peer consultation, therefore, gathered the opinions of technical experts, but the panel was not asked to reach a consensus group opinion.

Table 1 describes four key principles that define a successful peer review or consultation: scientific robustness, selection of appropriate expertise, independence, and transparency (Meek et al., 2007; TERA, 2012). TERA designed this peer consultation with the four principles, as well as using guidance for peer review provided by the US Office of Management and Budget (OMB, 2004), the National Academy of Sciences (NAS, 2003), and the US EPA (US EPA, 2006).

2.1. Development of charge questions for the expert panel

Table 2 is a list of the focused questions on relevant, critical scientific issues that formed the "charge" to the expert panel and provided direction and scope for the panel's review and discussions and also ensured the robustness of the process. To build the charge, TERA first asked the PAC Analysis Task Group to identify key questions and issues about the proposed method. TERA supplemented these suggestions with additional questions on key decision points and potential controversial scientific issues with the proposed method, as well as questions regarding development and use of the models. TERA developed the charge questions to ensure a comprehensive panel discussion and included open-ended questions to allow panelists to raise additional issues or concerns.

2.2. Selection of expert panel

Review of the draft report and development of the charge questions was essential to determining the key types of expertise needed to provide a robust scientific review of the report. Because the most important goal in organizing a peer consultation is to engage scientists with the needed expertise on the subject and issues, knowledgeable individuals from a variety of organizations and those with backgrounds in petroleum chemistry, statistics, toxicology, alternatives to animal toxicity testing, risk assessment, structure activity, and the HPV Challenge Program were included, providing essential expertise and a broad range of perspectives (see Table 3). To ensure a diverse list of candidates to choose from, TERA asked the sponsors and several interested parties (i.e., US EPA, Environmental Defense Fund, and Physicians Committee for Responsible Medicine) for suggestions of experts. These groups were contacted based on their history of active involvement in the HPV Challenge Program. TERA independently identified additional candidates through literature and Internet searches, an in-house database, and contacts in the field. TERA was solely responsible for the final selection of the panel members. The experts served as individuals, representing their own personal scientific opinions, and were not on the panel as representatives of their companies, agencies, funding organizations, or other entities with whom they are associated.

Independence and transparency of the peer consultation were supported by careful identification and disclosure of conflicts of interest and biases. Each candidate completed a questionnaire to identify activities, financial holdings, or affiliations that could pose a real or perceived conflict of interest or bias. TERA closely reviewed information regarding candidates' financial and other ties to the issue under discussion or the funding group. The completed questionnaires were reviewed by TERA staff to identify potential conflicts and/or biases and evaluate whether these situations would hinder the candidate from objective and full participation in the peer consultation discussions. TERA, as the independent entity selecting the panel, determined that none of the selected panel members had a financial or other type of conflict of interest based on this documented COI evaluation procedure.

Every scientist and expert has biases resulting from his or her education, training, and experience (Lord et al., 1979; Wilson et al., 1993). TERA identified sources of potential biases for each candidate (e.g., public opinions or past employers/affiliations) and evaluated these to determine if the biases would interfere with the expert's ability to constructively and objectively participate in the discussions and the peer consultation. Because this was a peer consultation, rather than a peer review, and multiple perspectives and opinions were sought, TERA did not attempt to exclude scientists who had taken public positions on the subject of toxicity testing of petroleum substances or the HPV Challenge Program. Those knowledgeable in these areas were considered an asset to the panel discussions. Expertise in petroleum chemistry and toxicology was essential for this peer consultation and TERA found that individuals with this type of expertise are almost always employed by petroleum companies, are with firms that provide technical support for petroleum companies, or are individuals who receive research support from such companies.

Several of the panel experts had previously worked for petroleum companies or had received contract or research support from petroleum companies. However, none of these experts worked on the PAC method project. TERA determined that the past work of these individuals did not constitute a conflict of interest because the panel members did not have any financial interests in the outcome of this peer consultation, nor would the past work bias the experts in evaluating the PAC method and results.

To ensure transparency, biographical sketches and relevant information regarding conflict of interest and bias were included in the meeting materials and final meeting report (TERA, 2008).

2.3. Peer consultation meeting

Table 4 lists the materials provided to the panel several weeks prior to the meeting so that they could thoroughly review and reflect on the subject documents before the meeting. To maintain a transparent process, the meeting was open to the public to attend in person and available for viewing in real-time via a webcast over the Internet. Observers were invited to submit comments or questions, which were then considered and addressed by the panel. Throughout the process, a website provided a central source for information about the peer consultation, including the Task Group's report, background information, charge questions, and TERA's final meeting report. Through this site, interested parties

Table 1
Four key principles for a peer review or peer consultation.[a]

Independence: Freedom from institutional or ideological bias and conflicts of interest, as well as distance from the development of the work product under review. At early stages of project scoping, data and issue identification, independence may not be as critical as one seeks critical knowledge and insights. At later stages of peer review, independence is critical.
Inclusion of appropriate expertise: Participation of highly qualified "peers," those who are qualified through training and experience to offer scientific opinions on the questions and issues at hand. A broad range of backgrounds and affiliations (e.g., government, academia, industry, environmental or public interest groups, consulting) can provide diverse scientific perspectives.
Transparency: Peer involvement activities should be conducted in a transparent manner, so that those both within and external to the process can evaluate how the activity was organized and conducted, and judge for themselves the adequacy and credibility of the process and the results. Information on the basis for and nature of the important decisions made during the process of conducting a review, particularly the basis for selection of the reviewers and sufficiently detailed record of the experts' deliberations and basis for conclusions and recommendations, should be publicly available.
Robust scientific process: Robustness is dependent on involvement of appropriate and sufficient number of experts, broad and specific charge questions that address the critical areas of controversy; sufficiently detailed subject report and review materials to facilitate a complete and efficient review; and thorough documentation of the review process and results. A robust scientific process that addresses the first three principles will contribute to robust scientific results and work products.

[a] Adapted from Meek et al. (2007) and TERA (2012), www.tera.org/peer.

could access key information about the peer consultation process and its results.

The meeting facilitator established ground rules for the discussions, which were limited to the panel members. Task Group authors were available to answer panel questions and were allowed to ask the panel members clarifying questions as needed. To ensure independence of the panel and transparency of the deliberations, all parties were asked to refrain from discussing issues related to the subject matter of the peer consultation with panel members prior to the meeting or during the meeting breaks. Several comments were provided by observers and were incorporated into the meeting report (TERA, 2008).

The Task Group authors made brief presentations of their work in three areas – selection and evaluation of data sets and endpoints, identification and characterization of the relationships between PAC content and toxicity in tested and untested petroleum substances, and validation of the statistical methods and use of results. Panel members discussed available data, the proposed approach, models, results, and conclusions during the two-day meeting. The panel discussed each of the charge questions and offered their individual opinions and suggestions for improvements.

2.4. Meeting report

A TERA scientist prepared a draft meeting report summarizing the panel's discussions and the panel members' individual opinions. The draft report was reviewed by the panel for completeness and accuracy. The Task Group authors also reviewed the draft report, but they were limited to commenting only on the clarity and completeness of their own presentations and on the statements they made during the discussions.

3. Results

The panel discussion centered on the charge questions (Table 2) and is summarized in detail in the meeting report (TERA, 2008). The intent of this paper is to highlight the key scientific issues raised by the panel on the PAC approach, and the responsiveness of the authors to the experts' recommendations. Throughout the meeting, panelists offered suggestions for report improvements that would help the reader better understand the authors' work and to make that work more transparent. More specifically, panelists suggested that more information was needed to provide context for the model, including more information about the HPV program and how the modeling results would be used in that program. The panel also suggested that the authors needed better documentation of the model form and make the domain of the model very clear, as well as more clearly identify limitations and inappropriate uses of the model. In addition, the panelists made specific suggestions related to providing a roadmap in the document that links the related data sets to the results of the analyses and the conclusions derived from those analyses.

3.1. Selection and evaluation of data sets and endpoints

Panel members discussed the selection and evaluation of the data sets and endpoints used to build the statistical regression model. The panelists were generally satisfied with the analytical technique used to determine the PAC profile of categories for the petroleum substances covered by the approach. This analytical technique is described in Nicolich et al. (2013). In the original document and during the peer consultation, the method was referred to as PAC-2 Method, but has since been labeled Method II in the other papers of this supplement. The authors found there were more study data utilizing Method II than the other analytical methods. Panelists found Method II promising and the relationships with toxicity compelling. However, the panelists recommended that the authors better explain the validation procedures and the performance of the overall method in order to better support their selection of this analytical method.

In developing the statistical models, the authors assumed that PACs are the source of toxicity for the identified group of petroleum substances. The panel discussed both the accuracy of this assumption and the usefulness and adequacy of the available supporting data. Panelists discussed the selection of data from the available toxicological studies that were used to build the models. In particular, they discussed the number of data points, the appropriateness of criteria used for excluding data, and the use of biological considerations in selection of the endpoints used for modeling. In general, the panelists found the data set selection procedure and range of endpoints selected to be appropriate. Although, individual panel members recommended further evaluation of some endpoints, they questioned why existing oral mouse studies were not used. The authors explained that the dermal exposure is the relevant route of exposure for petroleum substances and that the vast majority of studies were conducted by the dermal route of exposure; therefore they developed the models for dermal exposure only. Panelists noted the need for more analysis and information regarding maternal toxicity to determine whether the observed developmental effects were associated with maternal toxicity or if the substances are selective developmental toxicants. Panel members suggested that the authors address maternal toxicity as a potential endpoint and conduct analyses to determine if developmental toxicity was associated with or occurred at the same exposure levels that produced signs of maternal toxicity. The panelists discussed dermal irritation and its potential influence

Table 2
Charge to panel.

These charge questions formed the basis for the panel's discussions over a two-day period.

A. Selection and evaluation of data sets and endpoints

1. Is the Method 2 analytical procedure described in this project a reliable and accurate method of determining the PAC profile of the categories of petroleum substances referenced in this report?
2. The authors made an assumption that PACs are the source of toxicity for petroleum substances that contain PACs. Did the authors have adequate sampling to test and support this? Was the decision to conduct the statistical analyses based on PAC analysis Method 2 supported by the data? Could data sets from studies using other analytical methods contribute significantly to the analyses? Does the description of the chemical composition of petroleum substances support the conclusion that toxicity evaluation methods based on individual PAH content cannot be used?
3. Discuss the criteria and procedures used for identification, inclusion, and exclusion of toxicological data sets for the modeling. Were the criteria and procedures fully described and are they defensible? Is it reasonable to assume that all the relevant data have been collected and accurately compiled and analyzed? Are there other data sets that should have been considered? Did the exclusions make statistical and biological sense and were the impacts of such exclusions adequately explored? Would the procedures used in data set selection generate any bias in the results?
4. Were the procedures for selecting biological endpoints for modeling adequately communicated? Were the methods adequate to identify all key relevant endpoints? Was the final selection of endpoints fully justified by the data including the analyses conducted? Should other endpoints be considered based on the current state of knowledge of PAH and petroleum stream toxicity?
5. Are there other important issues to discuss regarding the selection and evaluation of data sets?

B. Identifying and characterizing the relationships between PAC content and mammalian toxicity (SIDS endpoints) and determining if they could be used to predict the toxicity of untested petroleum substances

6. Discuss whether the statistical methods were appropriate and adequate and if the procedures were implemented correctly. Would other valid statistical approaches yield different results? Would alternative model development approaches have improved the models and results? Were the authors' conclusions regarding the models clearly articulated and justified by the results?
7. The authors identified the toxicologically meaningful degree of change for each general toxicological endpoint (body weight and liver weight, hematology changes, thymus weight change). For each of the endpoints discuss whether these values are toxicologically and physiologically meaningful and if the best value was chosen. Are the observations consistent with what is known about petroleum toxicity? Are there other important issues regarding the relationship between PAC content and general systemic toxicity?
8. The authors identified the toxicologically meaningful degree of change for each developmental toxicity endpoints. For each of the endpoints discuss whether these values are toxicologically and physiologically meaningful and if the best value was chosen. Are the observations consistent with what is known about petroleum toxicity?
9. The authors conclude that the pre-defined change (PD_x) for developmental toxicity will be a reasonably good predictor of the PD_x for reproductive toxicity. Is this conclusion valid?

C. Validation of methods and use of the results

10. The authors present predicted dose–response curves and compared these to actual results of the study from which the information had been derived. How accurately do the predicted dose–response curves fit the observed data, and how do the predicted PD_x effect levels compare with the endpoint Lowest Observed Adverse Effect Levels/Lowest Observed Effect Levels (LOAELs/LOELs) observed in the actual studies?
11. Discuss the model validation methods. Could additional validation approaches be used to enhance confidence in the model?
12. Are the conclusions reached by the authors regarding utility of the models for interpolation versus extrapolation justified? Are they presented definitions and procedures adequate to identify data sets that can be accurately predicted by the proposed models?
13. Are the conclusions in Volume 2 Section 5 biologically plausible, supported by the data, and do they reflect sound statistical analysis?
14. Discuss the models' strengths and limitations. Were they clearly identified and the implications well described? Are there ways to ameliorate the weaknesses? Is the documentation transparent and complete? Are uncertainties in the approach fully articulated? Are there suggestions for improving the presentation of the analyses or information?
15. Can the models that were developed be used to predict repeated-dose and developmental toxicity of PAC-containing petroleum substances for purposes of the HPV program?
16. Are there other important issues regarding model development, validation, and use?

on maternal health status, reproduction, and fetal effects. They suggested that the authors further explore this to determine if direct effects on the skin are an important endpoint to include in evaluation of PAC-associated reproductive failure or developmental toxicity.

The panel explored the possibilities and limitations of using alternative approaches to predict toxicity, such as relative potency, an additivity approach of the individual component's toxicity, or using parallel analysis to individual PAHs. Several panelists noted that the analytical methods could not support these types of component approaches. These petroleum streams contain thousands of individual isomers of PACs, in comparison to the relatively simple mixtures of PAHs, for which relative potency factors have been estimated. Panelists also discussed that PACs interact in ways that are not fully understood and that an integrated approach as proposed by the authors is limited to demonstrating association and not causation – because the complex toxicity mechanisms are unclear.

3.2. Identifying and characterizing the relationships between PAC content and mammalian toxicity to predict the toxicity of untested petroleum substances

The Task Group authors started with a simple linear regression model and attempted numerous transformations of independent and dependent variables to increase statistical fit of the model (Nicolich et al., 2013). A factor analysis was conducted in an attempt to optimize the slate of variables in each model, but this approach was discarded as there was no gain in predictive ability. Ultimately, the selected variables were based on comparative

Table 3
List of peer consultation panel members (affiliations are for identification purposes only and reflect affiliation as of October 2007).

Peer consultation panel members	
Dr. Caroline Baier-Anderson Environmental Defense	Dr. Sati Mazumdar University of Pittsburgh
Dr. John DeSesso Noblis, Inc.	Dr. Andrew Nicholson Geomega
Mr. Stephen D. Emsbo-Mattingly New Fields Environmental Forensics Practice, LLC	Dr. Robert Scala Consultant
Dr. David Gaylor Gaylor and Associates, LLC	Dr. Calvin Willhite State of California
Dr. Andrew Maier Toxicology Excellence for Risk Assessment	**Facilitator** Dr. Michael Dourson Toxicology Excellence for Risk Assessment

Table 4
Documents reviewed by the peer consultation panel.

Peer review documents are available at: http://www.tera.org/peer/API/APIWelcome.htm
Volume 1 – Description of United States High Production Volume Program
• Appendix A – Supportive Information
Volume 2 -Investigation into the Relationship between PAC Content & Acute, Repeat-Dose, Development, & Reproductive Toxicity of Petroleum
• Appendix 1 – PAC: Nomenclature and Analysis
• Appendix 2 – Company Report/Studies
• Appendix 3 – ID of Biological Endpoints for Mathematical Characterization of DR Curve
• Appendix 4 – Biological Endpoints for Which Data Were Captured from Study Reports
• Appendix 5 – Summary of Analytical Data
• Appendix 6 – Statistical Evaluation of Data and Model Development
• Appendix 7 – Utility of the model(s) for Predictive Purposes
• Appendix 8 – Reproductive Toxicity
• Appendix 9 – Observed and Predicted DR curves
• Appendix 10 – Raw Data
• Appendix 11 – Commentary on Concordance/Lack of Concordance Between Endpoints Selected for Modeling and Data from Other Reviews

goodness of fit and general considerations related to PAC content and relevant biological variables (e.g., body weight was used as a predictive or independent variable for several of the repeated-dose models related to organ weights) that were expected to be important predictors of the response.

Panel members discussed the statistical methods used to build the models and the alternative methods that were considered. Panelists raised questions about the general model structures and suggested that the authors better describe the various approaches and exercises they used to develop the models and provide a more explicit description of how alternative modeling approaches improve or worsen model fits. Panelists raised issues regarding handling of control response, and use of mean response data from each input study dose group versus individual data. They noted that, with so many parameters, multicolinearity is important and suggested that additional presentation of mixed model results and sensitivity analyses would be helpful. They also suggested that a non-mathematical explanation of the models be included to help some readers understand these difficult concepts.

The Task Group authors used the models to predict predefined levels of response (PDR_x), which were selected based on informed professional judgment (e.g. 10% change from controls). The panelists discussed the definition of "toxicologically meaningful degree of change" and the basis for the response percentages selected for the evaluation, questioning the support for selection of one percentage value over another. Panelists had concerns about the biological and statistical basis for the authors' selections and concluded that the choice of what degree of change that is biologically significant is beyond the scope of the authors' charge and gets into risk assessment applications beyond the intent of the HPV program. Panelists suggested that the authors need not identify specific toxicologically-meaningful values in their report; rather, they should demonstrate the approach with multiple or hypothetical values and let the user decide what value is most appropriate for the situation of interest. The development of an approach based on control response variability also could be considered so as to define an abnormal, not necessarily adverse, range for effects that are measured on a continuous scale. This approach would be most consistent with current use of dose response estimates in risk assessment, when a clear level of adversity has not been defined. Then, PDR_x could be defined as a specified proportion of individuals in the abnormal range. Panelists concluded that the selection of PDR_x values, use of the PDR_x rather than benchmark response, and interpretation of results will likely be controversial, and these issues should be presented separately from model development, so as to not detract from the models themselves.

Panelists discussed reproductive toxicity and noted that the data set was inadequate to support the conclusion that developmental toxicity is more sensitive than, and likely a good predictor of, reproductive toxicity. The panel discussed the minimum data set that would be needed to predict potential reproductive toxicity and how studies might be best designed. The panel did acknowledge that current practice for the purposes of the HPV Challenge Program is to use reproductive organ pathology from repeat-dose studies as part of the basis for such screening determinations.

3.3. Validation of statistical models

Model validation involved three phases: data splitting techniques; randomized pairing of independent and dependent variable sets; and application of similar toxicological endpoints between the models where such data were available. The authors reported that the model did very well on all three validation[3] steps (Nicolich et al., 2013).

The panel discussed the validation methods used and explored alternative methods for further validation or confirmation of the models. Some panelists thought validating or confirming a model with additional, previously unused, data (i.e., data that had not been used in the development of the model or its parameters, and data from sources that did not use analytical Method II) were needed. Some panelists thought that new data or studies are needed, while others suggested looking for confirmatory or existing published or proprietary data. Other panelists were comfortable with the existing validation work, but suggested the authors might build the developmental toxicity models with a portion of the available data and test it with the remaining data. Further sensitivity analyses of the repeat-dose and developmental models were recommended by a number of panelists, while others did not think additional sensitivity analyses would improve the models or change the results.

The concepts of interpolation and extrapolation were discussed at length. The authors reported that the models worked well for interpolated substances, but not as well for extrapolated substances (Nicolich et al., 2013). Three elements were evaluated when considering interpolation and extrapolation: aromatic ring content, dose range, and boiling point range. The authors clarified that, to be considered an interpolation, the 7-ring profile for a new substance has to be inside one of the profiles of the substances used to build the model. To be considered an extrapolation, any one component of the new substance's 7-ring profile has to fall outside of all of the corresponding components of the existing profiles upon which the model was built. The authors noted that the

[3] Note that Nicolich and colleagues use the terms *corroborate* or *evaluate*, rather than *validate,* in their paper in this supplement; however, the Task Group authors used *validation* during the 2007 peer consultation and so use of that term is retained here.

interpolation concept also includes the consideration of dose. In addition, the authors noted that the models only applied to petroleum substances whose upper range of final boiling points are $\geqslant$650 °F. Panelists noted that use of the interpolation and extrapolation concepts contributes to the strength and validity of the models, and that the authors identified important limitations of the models based on testing of interpolated versus extrapolated data.

3.4. Use of results for High Production Volume (HPV) screening assessments

The panel discussed whether the models can be used to predict repeat-dose and developmental toxicity of PAC-containing petroleum substances for purposes of the HPV program. Some panelists thought that if this work is for screening and prioritization in HPV, the current effort may be sufficient without further modification for the developmental and repeated-dose endpoints. Others panel members noted that additional validation of the general toxicity models and further work related to assessing relationships between maternal toxicity, dermal irritation, and developmental effects would be needed before the models could be used for HPV screening purposes. Some panel members felt that additional data related to potential reproductive toxicity would also be needed before making conclusions regarding that endpoint.

The panelists cautioned that when presenting the modeled predictions in the context of HPV, it should be made explicit which values are calculated from the models, with the concept and meaning of interpolation and extrapolation carefully explained. Panel members strongly cautioned that the current model and results are not appropriate to use in definitive quantitative risk assessments (e.g., to estimate effect levels for US EPA Reference Dose estimation). The panel suggested that the model documentation clearly state this.

4. Discussion

TERA convened an expert peer consultation panel to provide comments and recommendations on a proposed statistical regression model approach for predicting selected biological responses to certain untested petroleum streams. In general, the panelists were positive and comfortable with the models, the general approach, and the statistical analyses that were done. Many thought the approach showed great promise and some stated that they thought it was close to being ready for regulators, while others thought the model and confirming assays should first be published as a scientific paper in the open literature, to gain wider exposure and input. This supplement responds to that recommendation. As a result of the peer consultation meeting, the PAC Analysis Task Group authors updated their original report (PAC, 2008). The papers in this supplement (Murray et al., 2013; Nicolich et al., 2013; Roth et al., 2013; Simpson et al., this issue) reflect those revisions, as well as additional data and analyses that extend the body of work that was considered by the peer consultation panel.

An author team's careful consideration of peer comments and recommendations is the purpose for conducting a peer consultation or review. Therefore, to consider a document "peer-reviewed" one ought to demonstrate that the document authors carefully considered the reviewers' recommendations. The authors need not adopt all recommendations, but they should be able to demonstrate that they evaluated key recommendations and provide rationales if recommendations were not adopted. A robust and transparent peer input effort should include a disposition of the most significant comments so readers can judge the authors' responsiveness to peer input. Peers can provide very valuable input and recommendations, but they may also make recommendations that upon later investigation may not be appropriate, or the authors may not be willing to adopt for policy or practical reasons. A post-review evaluation would provide interested parties with assurances that the peer consultation or review recommendations were seriously considered by the authors and incorporated into the resulting product; thus impacts the ability of risk managers or policy makers to be informed users of the final work product.

A review of the manuscripts on the PAC method that are in this supplement (Murray et al., 2013; Nicolich et al., 2013; Roth et al., 2013; Simpson et al., this issue) finds that many of the key suggestions from panel members have been considered and incorporated. In addition, the authors obtained data on two high boiling petroleum substances (from rat developmental and repeat-dose toxicity studies) that had not been used to develop the models. The ARC profiles of these substances fell within the domain of the models and the authors used the ARC profile for each sample and developed predictions which they compared with the observed values for further validation of their models. Other key modeling issues raised by the peer consultation panel (e.g., a lack of sensitivity analyses and presentation of mixed model results) were addressed. In addition, Nicolich et al. (2013) improved the descriptions of modeling methods and approaches such as interpolation versus extrapolation and provide a more transparent discussion of the model development and validation. The other papers in this supplement clearly describe each step of model development, discuss the various validation approaches, and include a sensitivity analysis which was absent from the initial documents that underwent peer consultation. Panel members had suggested confirmatory assays, and several new studies have been obtained and included for this additional confirmation.

The discussions of the PDR_x concept (Murray et al., 2013; Nicolich et al., 2013; Roth et al., 2013; Simpson et al., this issue) have been significantly revised and refined and address many of the concerns and recommendations of the peer consultation panel. For example, the peer consultation panel suggested that the authors utilize multiple or hypothetical values for x when presenting the PDR_x concept, rather than attempt to identify a toxicologically meaningful value for each endpoint.

Another key suggestion was to exclude reproductive endpoints from the models due to inadequate data, and to prevent reliance on the untested hypothesis that developmental endpoints are more sensitive than reproductive effects. Reproductive endpoints are discussed in Murray et al. (2013), but these reproductive data are clearly not used as part of the model. To address the hypothesis that developmental endpoints are more sensitive than reproductive effects, additional analyses were conducted and are described in Murray et al. (2013). Based on that analysis, the authors concluded that reproductive studies appear unnecessary if there are adequate developmental and repeat-dose toxicity studies available.

In some cases the revised approach described in this supplement did not fully incorporate peer consultation panel member suggestions, but in each of these cases the authors have provided a rationale as to why that decision was made. Several issues that were not addressed could be viewed as extensions or refinements of the methods that might increase its utility for risk assessment purposes beyond the stated objective of meeting (i.e., specific data requirements of the HPV Challenge Program). For example, extensions of the work to better understand the relationships between observed effects and dermal irritation or maternal toxicity was suggested by the panelists. While the value in better understanding the role of these effects was noted by the authors, such refinements were not viewed as necessary to use the current approach for HPV Challenge Program purposes with the health protective assumption that any effect seen (even if secondary to other high

dose stresses) was related to the chemical dosing. For dermal irritation, which was a common finding among the studies, the authors noted that lack of consistent and rigorous reporting of this endpoint within the available toxicity studies precluded its incorporation of this consideration into the model. The panel spent considerable time discussing the potential importance of maternal toxicity in determining if observed developmental effects represented selective developmental toxicity or were secondary results from maternal stress. Although maternal toxicity is an important consideration for assessing causality and mechanisms, the authors noted that, for the purposes of this model, the goal was limited to assessing associations between PAC content and the onset of developmental effects, without regard to underlying mechanisms. This approach was viewed as appropriate and health protective for hazard screening purposes.

Panel members also suggested some aspects of the modeling construct that might be altered to better represent the uncertainty or confidence bounds associated with the estimates. However, the authors noted that, due to the nature of the analyses, best fit estimates were probably more reliable for use in a screening tool for prediction of new untested materials since they are less sensitive to model changes driven by individual data sets. However, the authors addressed this issue completing and documenting additional sensitivity analyses.

The primary charge to the peer consultation panel was to evaluate the scientific and technical validity and support of the model in the context of its application as a tool for the HPV Challenge Program. Panelists agreed with the authors that interpretation and use of the model must be carefully described to prevent misuse of the model or results. The panel cautioned that the PAC method is a correlational study, and mechanistic-based conclusions should be avoided. One panelist noted that the model does a good job of estimating various biological effects based on PAC profile, but this does not mean the PAC (or ARC) profile is the cause of toxicity, an unknown factor may be involved. The panel discussed attempting to assess the biological significance of any level of the relationships, and agreed that it would be inappropriate to attempt to identify the specific parameter that contributes the most to any relationship. Therefore, panelists recommended that the authors should be careful to not imply or claim causation and that this qualification be stated unambiguously. The Task Group authors have explicitly communicated this qualification that the PAC method is built on statistical regression models based on associations, that they cannot be used to imply or prove causality of effect, nor are the models appropriate to use for quantitative risk assessment (see for example, Nicolich et al., 2013).

The peer consultation of the PAC method demonstrates the value of using a high quality peer consultation process. The Task Group authors were provided with specific and general comments from appropriately qualified experts to better develop and support their proposed method. Multiple aspects ensure the robustness of the process: development of a comprehensive and objective list of charge questions, selection of qualified experts with appropriate backgrounds, maintenance of independence from the sponsors and authors, and attention to transparency so that others may judge the quality of the process and the results. Suggestions were gathered in a formal and organized way to assist the Task Group authors' efforts to improve their method and ensure its scientific soundness/defensibility. This project and peer consultation demonstrate the value of formal engagement of peers in development of new scientific methods and approaches. The development and implementation of such peer input approaches is likely to have a large potential impact on chemical hazard and risk assessments. The need to develop robust mechanisms to vet and validate alternative methods for developing assessments with limited data sets has been a key focus of regulatory toxicologists. By conducting a public and transparent peer consultation, the scientific and regulatory communities can better judge the quality and objectivity of the results.

Conflict of interest

The American Petroleum Institute (API) provided TERA with financial support for the peer consultation meeting and preparation of the manuscript. The authors have no other conflicts of interest or competing interests.

Role of the funding source

The American Petroleum Institute (API), on behalf of the Petroleum High Production Volume Testing Group, provided TERA with financial support for the peer consultation meeting and preparation of the manuscript. This manuscript and underlying analyses were independently prepared by the authors without influence from API or anyone else. API and authors of other papers in this series reviewed a draft manuscript and provided editorial comments. The conclusions of the manuscript are solely those of the authors.

Acknowledgments

The American Petroleum Institute (API), on behalf of the Petroleum High Production Volume Testing Group, provided TERA with financial support for the peer consultation meeting and preparation of the manuscript. This manuscript and underlying analyses were independently prepared by the authors without influence from API or anyone else. API and the authors of other papers in this supplement reviewed a draft manuscript and provided editorial comments. The conclusions of the manuscript are solely those of the authors.

References

Gray, T.M., Simpson B.J., Nicolich, M.J., Murray, F.J., Verstuyft, A.W., Roth, R.N., McKee, R., 2013. Assessing the mammalian toxicity of high-boiling petroleum substances under the rubric of the HPV program. Regul. Toxicol. Pharmacol. 67 (2S), S1–S3.

Lord, C.G., Ross, L., Lepper, M.R., 1979. Biased assimilation and attitude polarization. J. Pers. Soc. Psychol. 37 (11), 2098–2109.

McKee, R., Schreiner, C., Nicolich, M.J., Gray, T.M., 2013. Genetic toxicity of high-boiling petroleum substances. Regul. Toxicol. Pharmacol. 67 (2S), S75–S85.

Meek, M.E., Patterson, J., Strawson, J., Liteplo, R., 2007. Engaging expert peers in the development of risk assessments. Risk Anal. 27 (6), 1609–1621.

Murray, F.J., Roth, R.N., Nicolich, M.J., Gray, T.M., Simpson, B.J., 2013. The relationship between developmental toxicity and aromatic-ring class profile of high-boiling petroleum substances. Regul. Toxicol. Pharmacol. 67 (2S), S46–S59.

NAS (National Academy of Sciences), 2003. Policy and Procedures on Committee Composition and Balance and Conflicts of Interest for Committees used in the Development of Reports, May. Available at: <http://www.nationalacademies.org/coi/index.html> (accessed 05.03.12).

Nicolich, M.J., Simpson, B.J., Murray, F.J., Roth, R.N., Gray, T.M., 2013. The development of statistical models to determine the relationship between aromatic-ring class profile and repeat-dose and developmental toxicities of high-boiling petroleum substances. Regul. Toxicol. Pharmacol. 67 (2S), S10–S29.

OMB (Office of Management and Budget), 2004. Final Information Quality Bulletin for Peer Review, December. Available at: <http://www.whitehouse.gov/omb/memoranda/fy2005/m05-03.pdf> (accessed 05.03.12).

PAC Analysis Task Group, 2008. The Relationship Between the Aromatic Ring Class Content and Selected Endpoints of Repeat-Dose and Developmental Toxicity of High-Boiling Petroleum Substances, March 31. American Petroleum Institute, Washington, DC. Available at: <http://www.petroleumhpv.org/pages/publications.html> (accessed 05.03.12).

Roth, R., Simpson, B.J., Nicolich, M.J., Murray, F.J., Gray, T.M., 2013. The relationship between repeat-dose toxicity and aromatic-ring class profile of high-boiling petroleum substances. Regul. Toxicol. Pharmacol.

Simpson, B., Nicolich, M.J., Roth, R.N., Murray, F.J., Gray, T.M., this issue. Application of statistical models to characterize the repeat-dose and developmental toxicity of high-boiling petroleum substances. Regul. Toxicol. Pharmacol.

TERA (Toxicology Excellence for Risk Assessment), 2008. Report of the Peer Consultation on Relationship Between PAC Profile and Toxicity of Petroleum

Substances. Toxicology Excellence for Risk Assessment, Cincinnati, OH. Available at: <http://www.tera.org/peer/API/APIWelcome.htm> (accessed 05.03.12).

TERA (Toxicology Excellence for Risk Assessment), 2012. How TERA Organizes and Conducts Independent Peer Review and Consultation Meetings. Toxicology Excellence for Risk Assessment, Cincinnati, OH. Available at: <http://www.tera.org/peer/PeerProcess.html> (accessed 05.03.12).

US EPA (US Environmental Protection Agency), 2006. Peer Review Handbook, third ed. EPA/100/B-06/002. Science Policy Council, Washington, DC. Available at: <http://www.epa.gov/peerreview/pdfs/Peer%20Review%20HandbookMay06.pdf> (accessed 05.03.12).

Wilson, T.D., DePaulo, B.M., Mook, D.G., Klaaren, K.J., 1993. Scientists' evaluations of research: the biasing effects of the importance of the topic. Psychol. Sci. 4, 322–325.