June 11, 2013

Dr. Andrew Rooney
NIEHS/NTP
Deputy Director, OHAT
P.O. Box 12233, Mail Drop K2-04
Research Triangle Park, NC 27709
Submitted via email to: Andrew.rooney@nih.gov and through online system

Regarding: Comments on the Draft OHAT Approach and BPA and PFOA/PFOS Protocols for Systematic Review and Evidence Integration

Dear Dr. Rooney:

The American Chemistry Council (ACC)[1] and ACC's Center for Advancing Risk Assessment Science and Policy (ARASP)[2] appreciate the opportunity to provide comments to the National Institute of Environmental Health Sciences (NIEHS) National Toxicology Program (NTP) Office of Health Assessment and Translation (OHAT) on 1) the Draft February 2013 OHAT Approach (the draft Approach), 2) the Draft Protocol for Systematic Review to Evaluate the Evidence for an Association Between Bisphenol A (BPA) Exposure and Obesity (the draft BPA Protocol) and 3) the Draft Protocol for Systematic Review to Evaluate the Evidence for an Association Between Perfluorooctanoic Acid (PFOA) or Perfluorooctane Sulfonate (PFOS) Exposure and Immunotoxicity (the draft PFOA/PFOS Protocol).[3]

[1] The American Chemistry Council (ACC) represents the leading companies engaged in the business of chemistry. ACC members apply the science of chemistry to make innovative products and services that make people's lives better, healthier and safer. ACC is committed to improved environmental, health and safety performance through Responsible Care®, common sense advocacy designed to address major public policy issues, and health and environmental research and product testing. The business of chemistry is a $760 billion enterprise and a key element of the nation's economy. It is the largest exporting sector in the U.S., accounting for 12 percent of U.S. exports. Chemistry companies are among the largest investors in research and development. Safety and security have always been primary concerns of ACC members, and they have intensified their efforts, working closely with government agencies to improve security and to defend against any threat to the nation's critical infrastructure.

[2] ARASP is a coalition of 19 organizations focused on promoting the development and application of up-to-date, scientifically sound methods for conducting chemical assessments. ARASP supports science based chemical assessments that utilize objective, transparent data acquisition and evaluation criteria and advocates for the use of mode of action data in risk assessment.

[3] (1) Draft OHAT Approach for Systematic Review and Evidence Integration for Literature-based Health Assessments – February 2013 (http://ntp.niehs.nih.gov/NTP/OHAT/EvaluationProcess/DraftOHATApproach_February2013.pdf); (2) Draft Protocol for Systematic Review to Evaluate the Evidence for an Association Between Bisphenol A (BPA) Exposure and Obesity. April 9, 2013 (http://ntp.niehs.nih.gov/NTP/OHAT/EvaluationProcess/BPAProtocolDraft.pdf); (3) Systematic

ACC and ARASP have been actively engaged in reviewing and providing scientific information to inform chemical assessments. Consistent with recommendations from the National Research Council in 2011,[4] we strongly support standardized approaches for conducting weight of evidence evaluations. We have a significant interest in an NTP/NIEHS hazard evaluation process that considers all relevant scientific data in a transparent, unbiased, and rigorous manner.

The NTP/NIEHS is developing an approach that will inform the scientific basis for important future evaluations of substances conducted by OHAT, NTP/NIEHS, and other government assessment programs. As such, the draft Approach, and the other associated documents, constitute economically significant guidance. As a consequence, the documents are subject to the requirements of the Office of Management and Budget (OMB) Final Bulletin for Agency Good Guidance Practices.[5] In addition, consistent with OMB's Memorandum on Guidance for Regulatory Review,[6] such significant guidance documents should be subject to review by the Office of Information and Regulatory Affairs (OIRA) under Executive Orders 12866 and 13563.

We support the open process in which NTP/NIEHS is soliciting comments on the draft documents and look forward to further opportunities to actively engage with NTP/NIEHS as the staff work to incorporate public comments into the draft documents. In particular, we look forward to reviewing a response to comments document once all public comments have been evaluated and appropriately incorporated into revised draft documents. After all public comments are addressed by NTP/NIEHS, we suggest that revised draft documents be submitted to OMB for coordinated interagency review before revised drafts are released to the public for further input.

The accompanying attachment (ACC and ARASP Comments on the draft Approach and Protocols for Systematic Review and Evidence Integration) contains our detailed comments and recommendations. Notably, we have significant concerns with the limited transparency of the draft documents and substantive concerns with the overarching systematic review and evidence integration approach including: a lack of consideration of exposure levels, an inappropriate emphasis on risk of bias, the marginalization of mechanistic and mode of action information throughout the process, a lack of justification for factors that inform confidence in the body of evidence, an assumption that the identified associations imply causality, and a misleading choice of terminology to inform risk communication. We discuss each of these areas in more detail in the accompanying comments.

Review to Evaluate the Evidence for an Association Between Perfluorooctanoic Acid (PFOA) or Perflurooctane Sulfonate (PFOS) Exposure and Immunotoxicity, April 9, 2013 (http://ntp.niehs.nih.gov/NTP/OHAT/EvaluationProcess/PFOS_PFOA_ImmuneProtocolDraft.pdf).

[4] National Research Council, Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde (2011). Available at: https://download.nap.edu/catalog.php?record_id=13142

[5] See OMB, Final Bulletin for Agency Good Guidance Practices (Jan. 18, 2007). Available at: http://www.whitehouse.gov/sites/default/files/omb/assets/regulatory_matters_pdf/m07-07.pdf.

[6] See OMB, Guidance for Regulatory Review (Mar. 4, 2009). Available at: http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_fy2009/m09-13.pdf.

We appreciate the opportunity to comment on the draft documents. Please feel free to contact either one of us by email (Nancy_Beck@americanchemistry.com or Kimberly_Wise@americanchemistry.com) or phone (202-249-7000) with any questions.

Sincerely,

[Redacted]

[Redacted]

Nancy Beck, Ph.D., DABT
Senior Director
Regulatory & Technical Affairs

Kimberly, Wise, Ph.D.
Senior Director
Chemical Products & Technology
ARASP

cc: Linda Birnbaum, Ph.D.

COMMENTS ON THE NATIONAL INSTITUTE OF ENVIRONMENTAL HEALTH SCIENCES NATIONAL TOXICOLOGY PROGRAM OFFICE OF HEALTH ASSESSMENT AND TRANSLATION (OHAT)

DRAFT OHAT APPROACH FOR SYSTEMATIC REVIEW AND EVIDENCE INTEGRATION FOR LITERATURE-BASED HEALTH ASSESSMENTS – FEBRUARY 2013

DRAFT PROTOCOL FOR SYSTEMATIC REVIEW TO EVALUATE THE EVIDENCE FOR AN ASSOCIATION BETWEEN BISPHENOL A (BPA) EXPOSURE AND OBESITY - APRIL 9, 2013

SYSTEMATIC REVIEW TO EVALUATE THE EVIDENCE FOR AN ASSOCIATION BETWEEN PERFLUOROOCTANOIC ACID (PFOA) OR PERFLUROOCTANE SULFONATE (PFOS) EXPOSURE AND IMMUNOTOXICITY - APRIL 9, 2013

Submitted by
The American Chemistry Council
and
The Center for Advancing Risk Assessment Science and Policy

June 11, 2013

I. Introduction

The American Chemistry Council (ACC)[1] and its Center for Advancing Risk Assessment Science and Policy (ARASP)[2] appreciate the opportunity to provide comments to the National Institute of Environmental Health Sciences (NIEHS) National Toxicology Program (NTP) Office of Health Assessment and Translation (OHAT) on 1) the Draft February 2013 OHAT Approach (the draft Approach), 2) the Draft Protocol for Systematic Review to Evaluate the Evidence for an Association Between Bisphenol A (BPA) Exposure and Obesity (the draft BPA Protocol) and 3) the Draft Protocol for Systematic Review to Evaluate the Evidence for an Association Between Perfluorooctanoic Acid (PFOA) or Perfluorooctane Sulfonate (PFOS) Exposure and Immunotoxicity (the draft PFOA/PFOS Protocol). ACC and ARASP support the development of a systematic review approach that meets the highest standards of scientific integrity, transparency, and peer review. Below we provide specific comments on the draft Approach and the draft BPA[3] and PFOA/PFOS[4] Protocols.

II. Comments on the Draft February 2013 OHAT Approach (the draft Approach)

Consistent with recommendations from the National Research Council in 2011,[5] we strongly support standardized approaches to conducting weight of evidence evaluations. We have a significant interest in an NTP/NIEHS hazard evaluation process that considers all relevant scientific data in an unbiased and rigorous manner. OHAT's draft systematic review approach represents a positive step towards developing a transparent process for evaluation. However, we have significant concerns with the current draft Approach that should be addressed. Many of the shortcomings of the draft Approach carry forward into the draft BPA and draft PFOA/PFOS Protocols and thus will also need to be

[1] The American Chemistry Council (ACC) represents the leading companies engaged in the business of chemistry. ACC members apply the science of chemistry to make innovative products and services that make people's lives better, healthier and safer. ACC is committed to improved environmental, health and safety performance through Responsible Care®, common sense advocacy designed to address major public policy issues, and health and environmental research and product testing. The business of chemistry is a $760 billion enterprise and a key element of the nation's economy. It is the largest exporting sector in the U.S., accounting for 12 percent of U.S. exports. Chemistry companies are among the largest investors in research and development. Safety and security have always been primary concerns of ACC members, and they have intensified their efforts, working closely with government agencies to improve security and to defend against any threat to the nation's critical infrastructure.

[2] ARASP is a coalition of 19 organizations focused on promoting the development and application of up-to-date, scientifically sound methods for conducting chemical assessments. ARASP supports science based chemical assessments that utilize objective, transparent data acquisition and evaluation criteria and advocates for the use of mode of action data in risk assessment.

[3] Draft Protocol for Systematic Review to Evaluate the Evidence for an Association Between Bisphenol A (BPA) Exposure and Obesity - April 9, 2013

[4] Systematic Review to Evaluate the Evidence for an Association Between Perfluorooctanoic Acid (PFOA) or Perflurooctane Sulfonate (PFOS) Exposure and Immunotoxicity - April 9, 2013

[5] National Research Council, Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde (2011). Available at: https://download.nap.edu/catalog.php?record_id=13142

addressed to become consistent with an improved draft Approach. We set forth below suggestions to improve the draft Approach.

A. Overall, we were disappointed to see that many of the details that were released to the NTP Board of Scientific Counselors (BSC) Working Group in August 2012 have been removed from the draft Approach. The rationale for having less standardization in the draft Approach, and to instead include the details in the protocols is not clear. While developing a protocol and making it available for public comment before a review is conducted is valuable, if NTP/NIEHS programs will evaluate multiple substances (including chemicals, mixtures or physical substances) and endpoints, for comparison and prioritization purposes, it is important that the same approach be used for each evaluation. Therefore we support having consistency and standardization in the draft Approach and having minimal variation in the protocols for each review. The standardized criteria for literature searches, determinations of study relevance, the evaluation of the quality, including evaluation of risk of bias of individual studies, and other details regarding how studies will be evaluated and utilized should be consistent for all substances so that each review meets similar scientific standards.

B. The draft Approach is described by NTP/NIEHS as being developed to assess evidence of adverse effects and to provide opinions on whether substances may be of concern given what is known about current exposure levels. However, in the draft Approach, as well as in the draft Protocols, the evaluation is conducted completely without the context of the exposure levels. It will be most helpful to all stakeholders if NTP/NIEHS develops hazard identification conclusions that are correlated to specific dose levels. For instance, simply stating that substance X causes an effect provides interested stakeholders with only a partial conclusion. It is more important to understand the exposures/doses at which substance X causes an effect. Without this exposure/dose information, it will be difficult for the public to understand the implications of the hazard conclusions. Indeed while the draft Approach notes that each topic is formulated based on PECO principles (Population of interest, Exposure or Intervention, Control or comparator group, and Outcomes of interest), in the draft Approach the project-specific details that are recommended for each protocol do not include explicit consideration of exposure/dose. The draft Approach is also inconsistent with the recommendations provided by the NTP BSC in the December 11, 2012 Draft Working Group Report, which states (at page 2): "Consideration of relevant human exposure levels including specific and susceptible populations should inform the scope of the topic."[6] It is imperative that the scope of the draft Approach include an exposure/dose component.

C. When selecting studies for inclusion and exclusion in the systematic review (Step 2), the draft Approach states (at page 2): "a basis for excluding those studies may be described *a priori* in the protocol." Consistent with the Draft Working Group Report recommendations regarding having clear criteria in advance of conducting a review, *a priori* criteria for excluding and including studies should be clearly described. This should not be an optional part of the draft Approach or

[6] See Draft Report of the NTP Board of Scientific Counselors Working Group on Evaluating NTP's Approach for Reaching Conclusions for Literature-Based Evidence Assessment, December 11, 2012, [hereinafter referred to as the Draft Working Group Report]. Available at: http://ntp.niehs.nih.gov/NTP/About_NTP/BSC/2012/December/DraftReport_LiteratureBasedEvi.pdf.

Protocol. The draft Approach notes, also at page 2, that "any reference possibly meeting the inclusion criteria is retrieved for full text review." The draft Approach should clarify how information that "possibly" meets criteria should be used. A lack of clarity on the criteria used to include information could potentially result in a reliance on material that does not meet standards, thereby undermining the validity of the systematic review approach.

D. NTP/NIEHS may want to consider incorporating a "request for relevant data and information" step that is particularly focused on receiving negative data that was not subsequently published in the peer reviewed literature. Such studies, if well conducted, could be very informative to a systematic review, yet because there are no significant positive findings, they are often not published. Consistent with plans in the draft Protocols, NTP/NIEHS could have this information peer reviewed before they rely upon it.

E. NTP/NIEHS has not yet shared a copy of the template that will be used in Step 3 for extracting data from studies. To improve transparency these templates should be released for public comment and peer review.

F. In Step 4 and throughout the draft Approach, there is too strong an emphasis on risk of bias and a lack of a sufficient emphasis on a robust evaluation of study quality and relevance. We recognize that risk of bias is an important component of study quality evaluation; however, it is a term that is only common in the epidemiological community and much of the wording and approach used in asking risk of bias questions is not relevant to animal studies. As noted by the Draft Working Group (WG) Report, some members "were concerned that this could create a bias against animal studies because they felt many studies would unjustifiably be downgraded based on RoB [risk of bias]; other WG members were concerned that this would result in RoB in animal studies going undetected or characterized inappropriately by reviewers."[7] Instead of focusing on risk of bias, we suggest that the draft Approach provide criteria for a full review of all aspects of study quality and relevance for epidemiological information as well as toxicological information (including *in vitro*, *in vivo* and mechanistic information). Commonly accepted protocols and approaches currently exist, and are used for evaluating animal and mechanistic information. This information was presented by Dr. Richard Becker at the March 20, 2013 webinar hosted by NTP.[8] Similar methods specifically designed for evaluating quality of human observational epidemiology studies have also recently become available (Money et al. 2013).[9] NTP/NIEHS should incorporate these scientifically accepted protocols for evaluating toxicological evidence into the draft Approach and lessen the emphasis on risk of bias.

---

[7] See Draft Working Group Report at page 12.

[8] See presentation entitled "Data Quality in Toxicology Studies: A Key Element in Systematic Review for Evaluating Chemical Risks. Available at: http://ntp.niehs.nih.gov/NTP/OHAT/EvaluationProcess/Presentations/March2013/Becker20130320_508.pdf. See also: Klimisch HJ, Andreae E and Tillmann U. (1997). A systematic approach for evaluating the quality of experimental and ecotoxicological data. Regul.Toxicol. Pharmacol. 25:1-5; and ECETOC, JACC report #55 on Linear Polydimethylsiloxanes which incorporates the modified and expanded the justification phrases for each Klimisch reliability category. Available at: http://www.ecetoc.org/jacc-reports.

[9] See Money CD, Tomenson JA, Penman MG, Boogaard PJ, Lewis RJ. (2013). A systematic approach for evaluating and scoring human data. Regul. Toxicol. Pharmacol. 66:241-247.

G. In Step 5, the draft Approach appears to focus on the confidence in the body of evidence, without consideration for determining the confidence of each individual study. The draft Approach should clearly allow for explicit determinations in the confidence of each individual study, based on its quality and relevance to the scope of the draft Approach, before moving towards a determination in the confidence in the full body of evidence.

H. The draft Approach does not put sufficient emphasis on animal data, *in vitro* information, and mechanistic information. The BSC noted the importance of this information when extracting data from studies,[10] yet the draft Approach does not appear to incorporate this information. Page 3 of the draft Approach suggests that human and non-human data be considered separately. This suggestion runs counter to an integrative approach that appropriately considers all information. When evaluating and integrating evidence, NTP/NIEHS should incorporate a discussion of biological plausibility, as well as mechanistic and mode of action (MOA) information to assess the relevance of animal findings to humans and in evaluating associations observed in epidemiology studies. Indeed, MOA should serve as the central organizing principle in the draft Approach. Hypothesized MOAs should be articulated as part of the problem formulation stage. Consistent with established best practices of systematic evidence-based reviews, NTP/NIEHS should employ a consistent weight of evidence framework, based on specific hypothesized MOAs to allow data from laboratory experiments, epidemiological investigations, and cutting-edge mechanistic research to be integrated in a manner that provides a robust understanding of the MOA and the potential hazards and risks that exposures to a substance could pose to humans.

I. Page 3 of the draft Approach states "Conclusions developed in the subsequent steps of the approach are based on the evidence with the highest confidence." It is unclear from this statement how the draft Approach will treat conflicting information. In a truly integrated approach, all information should be appropriately weighed based on its quality and relevance scores. Further clarification from NTP/NIEHS is needed.

J. The rationale for the key features used to determine initial confidence is not justified and would allow study confidence to be inappropriately elevated. The draft Approach has chosen four key study design features to use in determining initial confidence in the body of evidence. No justification or rationale is provided for the selection of these four features. However, it appears that these four study design features are based on the Grading of Recommendations Assessment, Development and Evaluation (GRADE) and Agency for Healthcare Research and Quality (AHRQ) approaches; GRADE and AHRQ are not appropriate in this context because they are designed for medical/health care-related data, not toxicology and observational epidemiology data used in human health hazard and risk evaluations. The choice of applying GRADE and AHRQ has the consequence of taking cross-sectional studies that would typically, and appropriately, be characterized as low confidence, due to their study design, and allow them to receive a moderate or high confidence rating if they contain three or four of the features NTP/NIEHS has selected. By applying the four key factors NTP/NIEHS has arbitrarily identified, many databases consisting of predominantly cross-sectional studies will receive a higher final confidence rating. During the

[10] See Draft Working Group Report at page 8, which states "Initially the NTP should incorporate more weight on non-apical studies."

April 23, 2013, web-based informational meeting, Dr. Thayer acknowledged that this was a deviation from GRADE, but noted that it was necessary to allow NTP/NIEHS "more ability to discriminate between studies." Dr. Thayer's rationale, however, does not support assigning low quality studies higher quality ratings than they deserve. Moreover, confidence ratings will be inconsistent with those typically applied by the epidemiologic community. The draft Approach, by design, inappropriately allows a body of evidence that should have a low rating to become elevated to a higher confidence rating. This bias in the draft Approach, through the arbitrary selection of four key factors, should be corrected so that cross-sectional studies only reach a higher confidence rating on the limited rare occasion. Furthermore, it is not clear why the draft Approach is applying individual study design feature criteria to a body of evidence, not individual studies.

K. When discussing indirectness, the draft Approach states (at page 4) "We recognize that exposure level is an important factor when considering the relevance of study findings to human health effects at known human exposure levels. In OHAT's evaluation process, this consideration occurs after hazard identification as part of reaching a 'level of concern' conclusion, where the health effects are interpreted in the context of what is known regarding the extent and nature of human exposure." As noted above, we disagree with this approach and believe that the hazard identification findings should be made in the context of an exposure level. Without such a context the utility of the hazard finding is extremely limited, particularly for informing public health concerns.

L. When looking at factors that increase the confidence rating in the body of evidence, as noted by the Draft Working Group Report (see page 18), there should be clear consideration of factors such as mechanistic information, pharmacokinetic data, MOA and other factors that can increase the confidence rating in a body of evidence. As this information should be considered in every protocol, it should be a factor in the draft Approach. For example, NTP/NIEHS should consider how EPA's Integrated Risk Information System (IRIS) program incorporates estimates of internal dosimetry, such as a human equivalent dose (HED), into their hazard evaluations. In addition, it is not clear why factors such as dose-response and confounding are not similarly used to downgrade a confidence rating. More clarity is needed as to why NTP/NIEHS chose each particular aspect for upgrading and downgrading confidence.

M. In Step 6 the draft Approach states "Although the conclusions describe associations, a causal relationship is implied and the ratings describe the level of evidence for health effects in terms of confidence in the association or the estimate of effect determined from the body of evidence." While the Bradford-Hill considerations are related to the process of evaluating the confidence in the body of evidence, NTP/NIEHS should not conclude a causal relationship is implied by the confidence in the association. Following the draft Approach does not meet the criteria for establishing causality and we note that the Draft Working Group Report (see page 22) was not in agreement about the use of the word causality in the definitions.[11] While the draft Approach does not have the term causality in the definitions, it seems equally inappropriate to state that association implies causation. The assumption of causality should be removed from the draft Approach.

[11] See Draft Working Group Report at Page 22 which states: "The WG was not in agreement about the use of the word *causality* in the definitions."

N. In Step 7, the draft Approach integrates evidence to develop hazard identification conclusions. We delineate below three substantive scientific and risk communication concerns in this section.

1) As mentioned previously, these hazard identification conclusions must be made in the context of an exposure level to have public health utility.

2) The draft Approach incorrectly uses four hazard identification conclusion categories. The greatest concern is with the category labeled "suspected to be a hazard to humans." The term "suspected" is typically used to imply that something is surmised to be true or probable or likely to occur. However, in the draft Approach, NTP/NIEHS uses the term for categories of only low to moderate evidence (moderate human evidence and low non-human evidence or a mix of low human evidence and moderate non-human evidence). Thus the term "suspected" will be misleading to the public regarding the true level of evidence that exists. A term that represents weak evidence (e.g., possibly, not likely, or limited) would be much more appropriate and should be used in place of the term "suspected." Such wording inappropriately introduces bias into the scientific evaluation process. In addition, it is not clear why NTP/NIEHS is also creating a new category labeled as "presumed." NTP/NIEHS should consider using terminology that the public health community is already familiar with (e.g., "likely" as is used in the 2005 EPA Guidelines for Carcinogen Risk Assessment , or "reasonably anticipated" as is used by the NTP Report on Carcinogens).

3) Important information, such as mechanistic and *in vitro* data are treated as "other relevant data." As discussed previously, the framework should use this information, and other important and critical MOA information, as a key organizing principle for evaluating evidence, not solely as a modifier to a primary classification.

III. Comments on the Draft Protocol for Systematic Review to Evaluate the Evidence for an Association Between Bisphenol A (BPA) Exposure and Obesity (the draft BPA Protocol)

The draft BPA Protocol was released with a separate document entitled "Appendix 2: Guidance for Assessing Risk of Bias in the BPA-Obesity Systematic Review." As this Risk of Bias Appendix is similar to the Appendix associated with the draft PFOA/PFOS Protocol,[12] our comments on these appendices are delineated in a separate section below.

Many of the concerns with the draft BPA Protocol mirror our concerns with the draft Approach. Therefore most of the concerns noted above apply here and will not be reiterated below. In particular, some of the major concerns (elaborated in more detail in comments on the draft Approach) that should be corrected include:

[12] Appendix 2: Appendix 2: Risk of Bias Guidance for BPA Exposure and Obesity Protocol (http://ntp.niehs.nih.gov/NTP/OHAT/EvaluationProcess/Appendix2BPA_Draft.pdf); Appendix 2: Risk of Bias Guidance for PFOA or PFOS Exposure and Immunotoxicity Protocol (http://ntp.niehs.nih.gov/ntp/OHAT/EvaluationProcess/Appendix_2_PFOA_PFOS_RiskOfBias.pdf)

- The lack of a focus on a specific exposure range in Step 1 and throughout the draft Protocol;
- The strong focus on risk of bias and the arbitrary determination of key criteria versus other criteria for determining study quality in Step 4;
- The arbitrary use of four key features to elevate the confidence in the body of evidence in Step 5;
- The implication that even though conclusions describe associations, the embedded Bradford Hill considerations allow for the determination that association is equal to causation in Step 6; and
- The misleading choice of terminology for the hazard identification category labels in Step 7.

Fundamentally, for the BPA Protocol, NTP/NIEHS should first establish the hypothesis by which BPA's activation of nuclear receptors (presumably ERα and/or ERβ) interact at a biochemical, cellular and organ level to generate key events that are biologically plausible for leading to the apical outcome of obesity. The proposed MOA can then be tested to see if the literature in fact supports the MOA. NTP/NIEHS should build the case for an adverse outcome pathway in which observed changes are truly adverse and can culminate, provided sufficient dose, in the apical outcome of obesity. To simply catalogue papers that report biochemical, cellular and tissue changes that could impact obesity, is not enough.

Secondly, and as mentioned above in the draft Approach comments, it is essential that exposure be considered as a hazard determinant. This is of high importance for BPA considering its well-described extensive first pass conjugation to its estrogenically inactive glucuronide and/or sulfate metabolites and the relatively low (< 1 µg/kg/day) dietary ingestion. In order to establish BPA as a cause of obesity:

1. Internal dosimetry assessments for BPA require linking urinary BPA-glucuronide (BPAG) levels to blood/serum BPA aglycone concentrations, using PBPK modeling.
2. Estimated blood BPA aglycone levels must be evaluated against the estrogen receptor MOA to see if in fact background dietary BPA exposures could have resulted in BPA aglycone blood levels sufficient to activate the ER.
3. Provided estrogenically active BPA aglycone levels are established, the weight-of-evidence must proceed to link an ER activation response (an initial molecular event) to key events that culminate in the adverse apical outcome of obesity (e.g., adipose tissue inflammation that leads to obesity in laboratory animals treated with BPA via a relevant, oral/dietary route of exposure). In general, *in vitro* and *in vivo* biological responses upstream from the apical adverse effect may be considered as precursor events. But a precursor event is not the same as the adverse effect, and data pertaining to precursor events must be used in accordance with understanding the relationship of the precursor to the apical adverse effect.. This requires knowledge of dose response and dose-dependent transitions as biochemical, cellular and organ responses progress from homeostatic responses, through adaptive responses and then to adverse responses. All of this goes towards supporting the biological plausibility of the proposed mode of action and key events leading to the hazard or health outcome under evaluation.. This exercise requires careful attention to analytical challenges when measuring BPA in biological media, such as blood or urine (e.g., urinary

concentrations near the limit of detection and the problem with sample contamination). Absent this level of proof, hazard identification is not possible for a weak estrogenic molecule like BPA.

Additional concerns with the draft BPA Protocol on aspects not addressed in the draft Approach are discussed below:

A. In Step 1, the draft BPA Protocol states (at page 3) that "Animal and *in vitro* studies without a concurrent control will be excluded." This approach may be scientifically sound in certain instances, but not in others. Therefore, development and use of such criteria needs to ensure that important information, including mechanistic and pharmacokinetic information that informs MOA are not excluded. All relevant and reliable studies, including those that inform MOA, should be included.

B. Step 1 discusses types of exposure information, including "indirect measures such as job title." Job title is often a poor surrogate for true exposures, especially chemical specific exposures. Including such weak and limited information, without discussing the associated uncertainties and limitations, can lead to improper characterizations of exposure. We suggest that this information not be included unless there are data to show that this is information is a reasonable surrogate for exposure.

C. The discussion of types of outcomes that will be evaluated in Step 1 needs significant improvements. No clear definition of obesity is provided and it is impossible to determine what the dose level of concern will be for the evaluation. In addition, the draft BPA Protocol arbitrarily excludes body weight as a basis for eligibility for animal studies, based apparently on a previous literature review. The exclusion appears to be because previous evaluations of BPA on body weight were negative (as discussed on page 4 of the draft BPA Protocol). Our understanding is that the goal of the systematic review is to objectively evaluate all the data. Subjectively excluding endpoints that may show a negative association defeats the purpose of conducting an objective systematic review. This is a serious flaw in the BPA Protocol in that it inappropriately assumes that it has been established with scientific certainty that body weight in animal studies is not a scientifically validated metric for evaluating obesity. The validity of such an assumption would need to be established by a scientifically robust evaluation (a systematic review) before it can be used in the proposed BPA Protocol. NTP/NIEHS cannot simply assert that certain metrics are scientifically valid and others are invalid.

D. Step 2 of the draft BPA Protocol notes that only publicly available information will be considered by NTP (see page 8 of the draft BPA Protocol). Yet, Appendix 2 relies heavily on the Krauth et al. study (it is cited 12 times), which is **not** publicly available. NTP/NIEHS should rely only on the publicly available information.

E. The discussion on page 4 notes that all publications must be peer reviewed, however on page 8, the draft BPA Protocol notes that if something is not peer-reviewed NTP will have it peer-reviewed. If Step 1 is excluding non-peer reviewed literature, what is the systematic process through which NTP is acquiring non-peer reviewed literature? Additionally, what will be

NTP's process for having a study peer-reviewed (e.g., will internal staff conduct the review, if external reviewers will be selected, how will those reviewers be selected)?

F. The transparency in Step 3 should be improved. The draft BPA Protocol notes that the results of the data extraction will be made publicly available "when the evaluation has been completed." NTP/NIEHS should make these results available to inform the peer review of the draft evaluation, **before** it is completed. In addition, after each critical step in the process (Steps 2-7), as NTP/NIEHS develops these first case studies, there should be a pause to seek public input before continuing further in the systematic review process.

G. Table 2 of the draft BPA Protocol seems to be missing elements that are important for a risk of bias determination (e.g., random assignment, blindness). Further clarity is needed regarding why certain elements are in the data extraction tables and others are missing. The process used by NTP/NIEHS to develop this table is not clear.

H. In Tables 4 and 5, NTP/NIEHS refers to key criteria as more important than other criteria. No rationale is provided for this differentiation. In addition, the BSC Draft Working Group Report (at page 10) explicitly recommended against using a subset of questions as major risk of bias questions.[13] The draft Protocol provides no clarity regarding how these key criteria will be used, compared to the other criteria. Consistent with the Draft Working Group Report, the use of these key criteria should be eliminated.

I. In Step 4, NTP/NIEHS inappropriately rejects the use of the ToxRTool claiming that it is a tool that mainly assesses reporting quality. While some aspects of the ToxRTool do indeed assess reporting quality, other aspects focus on important issues relating to study design and quality. Regardless, none of these aspects, including reporting quality, should be rejected simply because they are not consistent with the epidemiological approach for risk of bias determinations. NTP/NIEHS appears to be inappropriately wedded to focusing on risk of bias. As noted previously, commonly accepted protocols and approaches currently exist and are in use for evaluating the quality of animal and mechanistic information. NTP/NIEHS should refrain from fitting such evidence into an epidemiological paradigm and use the currently existing tools to transparently and objectively evaluate the quality of non-human data and information. The onus is on NTP/NIEHS to demonstrate why such internationally accepted approaches are not suitable for use.

J. Figures 2, 3 and 4 in the draft BPA Protocol are misleading and serve as poor communication tools. The use of box and whisker plots, in a manner similar to a formal meta-analysis, can lead to an inappropriate visualization of the overall weight of the evidence. Without including information on the quality of studies, reviewers and users could inappropriately draw

[13] See the Draft Working Group Report which states, at page 10, "The WG suggested dropping the designation of a subset of questions as *major* risk of bias questions. The WG recognized NTP was attempting to use the *major* questions as a means of excluding lower quality studies as the basis for conclusions. The WG was split on the question of excluding studies. However, the WG did not support a pre-defined subset of RoB questions as being more definitive compared to other questions or to use these pre-defined subset of *major* RoB to exclude studies for every systematic review that might be undertaken by the NTP." (emphasis added by NTP/NIEHS)

conclusions under the assumption that each study presented is of equal strength and quality. For example, with the proposed system, a plot with five poor quality studies using methods that have not been scientifically validated showing "positive" results would appear to "out-weigh" a high-quality scientifically solid study using well validated test methods that shows a "negative" result or "no difference from control." Thus, such figures would constitute a biased presentation of the scientific data. Therefore, such figures should not be used unless the quality of the study and the validity of the test methods are also captured integrally in such figures.

K. Step 5 in the draft BPA Protocol notes that changes were made to refine the GRADE approach to accommodate the need to integrate data from multiple evidence streams. Unfortunately, no further clarity regarding the specific refinements is presented. NTP/NIEHS should transparently document and provide reviewers with a table that explains what changes were made to the GRADE approach.

L. When determining the confidence in the body of evidence, the current approach appears to incorporate all the studies, regardless of their risk of bias ratings. For comparison purposes, it would be helpful for the draft BPA Protocol to present the results of a determination of confidence that relies only upon studies of the highest quality to see if this changes the outcomes. This transparent presentation of the two approaches will allow all readers and reviewers to understand the impacts that poorer quality data may have on the confidence in the body of evidence.

M. It is not clear in Step 5 (at page 45) why statistical power of studies will only be considered if there are inconsistencies in the findings across studies. More clarity on this approach is needed.

N. We agree with page 48 of the draft BPA Protocol that treats gavage studies as having similar relevance to feeding studies because most human exposure is oral. However, it is well established that gavage dosing leads to a bolus effect that can impact the kinetics of exposure, which can lead to markedly different responses compared to similar oral exposures when administered in food or water.[14] As previously discussed regarding the critical importance of BPA's first-pass metabolism to estrogenically inactive conjugates, NTP/NIEHS should ensure that the role of metabolism is appropriately considered when considering the quality of oral gavage studies. In this context, the need to ensure that an oral dosage give rises to estrogenically active BPA aglycone blood levels is a fundamental proof point for accepting the positive findings of both animal, as well as human evidence. In general, NTP/NIEHS should ensure that 1) the role of first-pass metabolism is appropriately considered when considering the quality of oral studies compared to other routes of administration and 2) kinetic differences arising from bolus doses from gavage studies are considered. This is particularly important for BPA.

[14] See, for example, La DK, Schoonhoven R, Ito N, Swenberg JA. (1996). The effects of exposure route on DNA adduct formation and cellular proliferation by 1,2,3-trichloropropane. Toxicol Appl Pharmacol. 140(1):108-14.

O. When discussing imprecision in the draft BPA Protocol, more clarity is needed regarding how statistical power will be considered. At page 50, the draft BPA Protocol notes that significantly underpowered studies may be omitted from the confidence rating phase, however they may be used in a sensitivity analysis. The role of statistical significance and power is also not clear in Table 17.

P. More clarity is needed regarding how findings relating to publication bias will be incorporated into the systematic review. More importantly, Section 5 lists many factors that may upgrade or downgrade the confidence rating; however, there is no clear systematic approach for incorporating all of these factors. Without such a standard protocol, the systematic review will be less objective.

Q. On page 52, regarding publication bias the reference to "industries" should be dropped. It is inappropriate for NTP/NIEHS to single out industries as being separate from authors. Certainly there are many different types of authors (including those from industries, academia, non-profits and government) and conflict of interest determinations are necessarily fact specific. Furthermore, there is a need to appropriately distinguish conflict of interest from bias, recognizing that all authors, irrespective of affiliation, will have certain biases. This distinction is discussed in the Keystone Report[15] and in the NAS guidelines.[16]

R. When discussing dose-response, the draft BPA Protocol suggests that a non-monotonic dose-response should be upgraded when the pattern is consistent with expectations. This approach, which upgrades or downgrades something based on prior beliefs is inappropriate for a systematic review which is meant to be objectively driven by the science, not individual beliefs. This is another serious flaw in the draft BPA Protocol, and again inappropriately assumes that such non-monotonic dose responses have been established with scientific certainty. The validity of such an assumption would need to be established by a scientifically robust evaluation (a systematic review) before it can be used in the proposed BPA Protocol. NTP/NIEHS cannot simply assert that certain metrics are scientifically valid and others are invalid. In fact, serious scientific questions have been raised in numerous peer reviewed publications, including those authored by CERHR panels and EPA scientists[17,18,19] about many of the purported low-dose / non-monotonic effects. NTP/NIEHS should establish a MOA for how BPA allegedly induces an adverse outcome pathway, as a function of exposure, that leads to obesity. It is within the MOA that key events based on presumed non-monotonic behavior should be evaluated.

---

[15] Dealing with Conflict of Interest and Bias in Scientific Advisory Panels, and Improving Systematic Scientific Reviews (2012). Available at: https://www.keystone.org/images/keystone-center/spp-documents/Health/Research%20Integrity%20Rountable%20Report.pdf

[16] The National Academies Policy on Committee Composition and Balance and Conflicts of Interest for Committees Used in the Development of Reports. May 12, 2003.

[17] Rhomberg LR and Goodman JE. (2012). Low-dose effects and nonmonotonic dose–responses of endocrine disrupting chemicals: Has the case been made? Regul. Toxicol. Pharmacol. 64: 130–133.

[18] Chapin et al. (2008). NTP-CERHR Expert Panel Report on the Reproductive and Developmental Toxicity of Bisphenol A. Birth Defects Research (Part B) 83:157–395.

[19] Ryan et al. (2010). In Utero and Lactational Exposure to Bisphenol A, in contrast to Ethinyl Estradiol, Does not Alter Sexually Dimorphic Behavior, Puberty, Fertility and Anatomy of Female LE Rats. Toxicol. Sci. 114(1): 133-148.

S. To be useful, the examples in Table 19 need to be described. The description should include whether or not the example is representative of a study that should be upgraded.

T. In Step 7, NTP/NIEHS should delete the sentence, "NTP does not require mechanistic or mode of action data in order to reach hazard identification conclusions, although when available this type of evidence may be used to raise (or lower) the level of the hazard identification conclusion when the evidence is 'strong' (Figure 8)." NTP's intention of eliminating MOA and the failure to consider an adverse outcome pathway approach diminishes the scientific validity and reliability of NTP/NIEHS's intention to evaluate BPA as a cause of obesity. The NTP/NIEHS approach should be revised, as discussed previously in these comments, so that mechanistic and mode of action data are key elements considered of primary importance in a systematic review.

IV. Comments on the Draft Protocol for Systematic Review to Evaluate the Evidence for an Association Between Perfluorooctanoic Acid (PFOA) or Perfluorooctane Sulfonate (PFOS) Exposure and Immunotoxicity (the draft PFOA/PFOS Protocol)

The draft PFOA/PFOS Protocol is nearly identical to the draft BPA Protocol in its overarching approach. Therefore, all the comments noted above for the draft BPA Protocol and the draft Approach document are relevant to the draft PFOA/PFOS Protocol and should be similarly addressed as NTP/NIEHS works to revise the protocols.

V. Comments on Appendix 2: Guidance for Assessing Risk of Bias in the BPA-Obesity Systematic Review (the BPA ROB Appendix) and Appendix 2: Guidance for Assessing Risk of Bias in the in the Protocol for Evaluating the Evidence for an Association Between PFOA and PFOS Exposure and Immunotoxicity (the PFOA/PFOS ROB Appendix)

These two appendices (the Risk of Bias (ROB) Appendices) [20] are nearly identical. The only substantive difference we noticed had to do with evaluating whether the researchers adjusted or controlled for other exposures that are anticipated to bias results. In the BPA ROB Appendix, it is addressed under performance bias, whereas in the PFOA/PFOS ROB Appendix it is considered part of confounding bias. NTP/NIEHS should explain the differential treatment in the two appendices.

Substantive concerns in the ROB Appendices should be addressed before NTP/NIEHS finalizes the associated draft protocols. A summary of our concerns and recommendations for improvement are provided below.

[20] Appendix 2: Appendix 2: Risk of Bias Guidance for BPA Exposure and Obesity Protocol (http://ntp.niehs.nih.gov/NTP/OHAT/EvaluationProcess/Appendix2BPA_Draft.pdf); Appendix 2: Risk of Bias Guidance for PFOA or PFOS Exposure and Immunotoxicity Protocol (http://ntp.niehs.nih.gov/ntp/OHAT/EvaluationProcess/Appendix_2_PFOA_PFOS_RiskOfBias.pdf).

A. The ROB Appendices, as also noted in our comments on the draft Approach, place an overemphasis on risk of bias and an under emphasis on a robust evaluation of study quality and relevance for all data streams. We recognize that risk of bias is an important component of study quality evaluation; however, it is a term that is only common in the epidemiological community and much of the wording and approach used in asking risk of bias questions is not relevant to animal studies. Instead of focusing on risk of bias, we suggest that NTP/NIEHS provide criteria for a full review of all aspects of study quality and relevance for epidemiological information as well as toxicological information (including *in vitro*, *in vivo* and mechanistic information). Commonly accepted protocols and approaches currently exist, and are in use for evaluating animal and mechanistic information.

B. The ROB Appendices fail a basic transparency test in that one of the key studies relied upon by NTP/NIEHS (Krauth et al.) is not publicly available for review. Cited 12 times, this non-peer reviewed paper appears to be one of the key justifications for how NTP/NIEHS has adopted risk of bias to address non-human evidence. Without the ability to review this non-peer reviewed paper, it is impossible to evaluate the scientific underpinnings of the approach taken by NTP/NIEHS. NTP/NIEHS should revise these appendices without relying on this paper or alternatively re-open a public comment period once it is available for public review.

C. Further clarity is needed regarding how risk of bias differs from quality.

D. The ROB Appendices should clearly describe how a high risk of bias will impact the use of a study in a systematic review.

E. In discussing selective reporting bias, the ROB Appendices mention that there is a bias if pre-specified outcomes are not reported. We support this approach; however, the ROB Appendices should also consider the lack of pre-specified outcomes as a factor that increases risk of bias. This should be explicitly discussed in the ROB Appendices.

F. In the BPA ROB Appendix, in regards to performance bias (question 7 on page 13 of the BPA ROB Appendix), and also selective reporting bias (question 14 on page 27 of the BPA ROB Appendix), a significant strength of guideline studies is a detailed and documented study plan in place before the start of the study. A crucial parameter for questions 7 ("Did deviations from the study protocol impact the results?") and 14 ("Were all measured outcomes reported?") is detailed knowledge on the study protocol. As mentioned in the draft Approach, the materials and methods section in publications is not a surrogate for an *a priori* study protocol and, consequently, it is incorrect to routinely indicate "probably low risk of bias" when no inconsistency is reported in the Material and Methods compared to Results of a publication.[21]

[21] For example, the evaluation of Ferguson et al. (2011) (evaluated in the BPA ROB Appendix) concerning the questions "*Were all measured outcomes reported?*" is answered "*yes, primary outcomes discussed in methods were presented in results section with adequate level of detail for data extraction*." This evaluation did not take into account that data on the siblings in this study are reported in later publications by Ferguson et al. (2012) and Cao et al. (2013). In case no *a priori* study plan is mentioned in a publication, questions 7 and 14 should be answered "n/a".

VI. Necessary Next Steps

As noted in our cover letter to these comments, we support the open process in which NTP/NIEHS is soliciting comments on the draft documents and look forward to further opportunities to actively engage with NTP/NIEHS as the staff work to incorporate public comments into the draft documents. Based on the substantive concerns found during our review, further public engagement and review is necessary before the NTP/NIEHS begins to implement the draft Approach and corresponding Protocols. In particular, we look forward to reviewing a response to comments document once all public comments have been evaluated and appropriately incorporated into revised draft documents. After all public comments are addressed by NTP/NIEHS, we suggest that revised draft documents be submitted to OMB for coordinated interagency review, consistent with EO 12866 and EO13563, before these updated drafts are released to the public for further input.

NTP/NIEHS is on the way to developing an systematic review and evidence integration approach that will inform the scientific basis for important future evaluations of substances conducted not only within NIEHS, but also by other government assessment programs. Consequently, it is critically important that it meet the highest scientific standards. The comments we have provided should help bring NTP/NIEHS closer to meeting these necessarily high standards for transparency and scientific rigor.