National Toxicology Program

Peer Review of the Draft NTP Research Report on the CLARITY-BPA Core Study

April 26, 2018

National Institute of Environmental Health Sciences
Research Triangle Park, NC

Peer-Review Report

Table of Contents

I. Attendees

Peer Review Panel Chair
David Dorman, North Carolina State University

Peer-Review Panel
Tracie Bunton, Eicarte LLC
Michael Conner, Global Blood Therapeutics, Inc.
Thomas Rosol, Ohio University
Janet Tooze, Wake Forest School of Medicine
Kimberley Treinen, Sunovion Pharmaceuticals, Inc.

NTP Board of Scientific Counselors Representative
Norman Barlow, Johnson & Johnson (by webcast)

National Institute of Environmental Health Sciences (NIEHS) Staff
Brian Berridge
Chad Blystone
John Bucher
Sue Fenton
Virginia Guidry
Jerry Heindel (retired)
Michelle Hooth
Troy Hubbard
David Malarkey
Barry McIntyre
Georgia Roberts
Kristen Ryan
Thad Schug
Keith Shockley
Vicki Sutherland
Suramya Waidyanatha
Nigel Walker
Atlee Watson
Mary Wolfe

Federal Agencies
Luísa Camacho, NCTR
Barry Delclos, NCTR
Robert P. Felton, NCTR
Gonçalo Gamboa da Costa, FDA
Cynthia Hines, NIOSH
Greg Olson, NCTR
William Slikker, NCTR

Contract Staff to NIEHS
Amy Brix, EPL
Canden Byrd, ICF
Kate Helmick, ICF
Jeanne Luh, ICF
Retha Newbold, Kelly Government Solutions (retired NIEHS)
Kelly Shipkowski, ICF

Public Attendees
In-person public attendees listed below. The peer-review meeting was webcast. Viewers of the webcast are not identified except for those who presented oral remarks by phone.

Scott Belcher, North Carolina State University
Kenkichi Fujii, Kao Corporation
Steven Hentges, American Chemistry Council
Ernie Hood, Bridport Services
Cécile Michel, ANSES (presented oral comments by phone)
Heather Patisaul, North Carolina State University
Robert Skoglund, Covestro
Laura Vandenberg, University of Massachusetts Amherst (presented oral comments by phone)

II. Welcome and Introductions

The peer review of the Draft NTP Research Report on the CLARITY-BPA Core Study was convened April 26, 2018, in Rodbell Auditorium, Rall Building, National Institute of Environmental Health Sciences (NIEHS), Research Triangle Park, North Carolina. Dr. David Dorman served as chair. The other peer-review panel members were Drs. Tracie Bunton, Michael Conner, Thomas Rosol, Janet Tooze, and Kimberley Treinen. Dr. Norman Barlow attended by webcast as the NTP Board of Scientific Counselors liaison. Interested public attended the meeting in person or watched the proceedings via webcast.

Dr. Dorman welcomed everyone to the meeting and asked all in-person attendees to introduce themselves. Dr. Brian Berridge added his welcome to the attendees, thanked the panel for their service, and noted that this peer-review meeting focused on the outcomes of the CLARITY-BPA Core Study, with additional Grantee Studies data to be integrated with the Core Study data later. Designated Federal Official Dr. Mary Wolfe read the conflict of interest statement and asked panel members to sign updated conflict of interest forms. Dr. Dorman presented the peer-review meeting format. Slide presentations for the meeting are available on the NTP Website (https://ntp.niehs.nih.gov/go/846162).

III. Presentation on the Overall CLARITY-BPA Program

Dr. Nigel Walker briefed the peer-review panel on the overall CLARITY-BPA program. He presented background information on bisphenol A (BPA), the current U.S. regulatory landscape, the historical NTP context, and the historical regulatory context, including the key recommendations the Food and Drug Administration (FDA) Science Board issued in 2008, which had recommended further research to address the central question of BPA developmental toxicity.

In response to the FDA Science Board's recommendation, NIEHS and FDA developed the CLARITY-BPA program, which stands for Consortium Linking Academic and Regulatory Insights on BPA Toxicity.

Dr. Walker emphasized this peer-review meeting was convened to review the Core Study, a 2-year chronic study conducted under good laboratory practices (GLP) at FDA/NCTR (National Center for Toxicological Research). Additional NIEHS Grantee Studies, which used siblings born to Core Study females and exposed to the same doses used in the Core Study, would add to the body of knowledge by focusing on a range of molecular, structural, and functional endpoints not usually addressed in guideline-compliant GLP studies. Dr. Walker noted that the CLARITY-BPA program would culminate with an integration report that interprets findings from both the Core Study and the Grantee Studies.

Dr. Walker provided details about the Core Study, conducted using NCTR Sprague-Dawley rats that are sensitive to estrogens and BPA. Rats were exposed to five doses of BPA or two doses of ethinyl estradiol, a classic estrogen. The study incorporated 56 exposure groups, 7 times the number in a standard NTP 2-year rat study. The BPA dose-response range was 10,000×, a much higher range than "guideline" chronic studies that usually cover a 4× to 10× dose range.

III.A Questions for Clarification

Dr. Bunton asked for more information about how the Grantee Studies data would be integrated with the Core Study data. Dr. Walker replied that the idea is to integrate the Core Study and

Grantee Studies data and determine whether effects occurring at earlier time points, at lower doses, or in potentially more sensitive endpoints could be matched with effects in siblings, sibling groups, and exposure groups. The approach for integration could be interpretive using different statistical approaches, although the plan has not yet been fully formulated.

IV. Peer Review of the Draft NTP Research Report on the CLARITY-BPA Core Study

IV.A Charge

Dr. Wolfe presented the charge to the panel:

- Review and evaluate the scientific and technical elements of the CLARITY-BPA Core Study and its presentation.
- Comment on the NTP's interpretations of the data.

IV.B Overview of Core Study Design

Dr. Luísa Camacho from NCTR presented an overview of the Core Study design.

The aim of the study was to characterize the toxicological potential of BPA following a perinatal-only or a chronic exposure in male and female rats exposed to a wide range of BPA levels.

Dr. Camacho provided details about the animal housing and animal feed used in the study and the presence of xenoestrogens and BPA in the feed. She described the BPA and ethinyl estradiol (EE_2) test articles.

The vehicle control group was exposed to 0.3% aqueous carboxymethylcellulose. Five BPA groups were used: 2.5, 25, 250, 2,500, and 25,000 µg/kg bw/day, and two EE_2 groups were included: 0.05 and 0.5 µg/kg bw/day. Dr. Camacho detailed the dosing administration method and the dosing regimen and rationale. She discussed animal allocation and breeding and provided information about in-life endpoints, including vaginal cytology endpoints, 1-year interim sacrifice endpoints, and 2-year terminal sacrifice endpoints. She described the study's histopathological evaluations and statistical analysis, including that of the histopathology data. Finally, she addressed the issue of background exposure to BPA, including sensitivity analysis aimed to determine the impact of potential unintended exposure to BPA. The overall findings of the sensitivity analysis indicated that a potential unintended exposure of the animals to BPA did not influence the outcome of the study.

IV.B.1 Questions for Clarification

Dr. Tooze noted that Dr. Camacho had stated litter was the unit of analysis. She expressed confusion on that point because, within litter and sex, the animals were assigned at weaning to the different dosing groups. Dr. Camacho said that because the males and females were analyzed separately, by using just one male or female per litter, the litter could be considered the unit of analysis.

Dr. Treinen asked about dosing and the rinsing of the dosing apparatuses between doses, specifically what was used to flush the dosing apparatuses. Dr. Camacho said that before the start of the study, various washing procedures were assessed and the final choice was a 1-minute wash with water followed by a 1-minute wash with ethanol and then another 1-minute wash with water.

Dr. Dorman asked if there was any indication of historical BPA levels and their potential impact on historical controls. Dr. Camacho said that was not determined, but historically, polycarbonate cages likely were used. She added that the cages and drinking water bottles do not leach much BPA and that most of the background exposure is from feed. The levels in the feed used in the studies were consistently low, and the assumption is that would be the case with historical controls, as well. Dr. Walker noted the same type of feed had been used in the previous endocrine studies conducted at NCTR that were used as the historical controls.

Dr. Bunton asked about the rationale for having the interim sacrifice group. Dr. Camacho said that it was to enable assessment of additional endpoints, including clinical chemistry, hematology, and organ weights, which are not as informative at the 2-year time point. The interim group also enabled the progression of lesions to be assessed. Dr. Walker added that several of the grantees were investigating more chronic endpoints, and to have grantee animals at a 1-year time point, Core Study animals at that time point also were needed. Dr. Bunton asked if other groups were earmarked for the grantee institutions. Dr. Camacho described the distribution of animals to the researchers, with animals first allocated to the Core Study and then to Grantee Studies based on individual study design or rationale.

IV.C Results and Preliminary Conclusions

Dr. Barry Delclos from NCTR summarized the results and preliminary conclusions of the Core Study. He divided his presentation into three sections, and provided an opportunity for the panel to ask clarifying questions after each section.

IV.C.1 Results and Preliminary Conclusions, Part One

Dr. Delclos first described the study protocols, including F_0 gestation and litter parameters, and the two F_1 dosing arms—the continuous-dose arm and the stop-dose arm. He discussed survival, including that with BPA, no statistically significant effects were observed in any phase of the study. For EE_2, a statistically significant effect was observed in the low-dose preweaning female cohort. He discussed several other survival parameters, including reasons for early removals/deaths after 1 year of age.

Several gestation and litter parameters showed no significant treatment effects. For body weights, no statistically significant effects were observed in males exposed to BPA or EE_2; in females, statistically significant effects of BPA and EE_2 treatments were few and confined to restricted periods in intermediate dose groups. Dr. Delclos presented additional data on body weights in several study groups.

Vaginal opening was monitored in 26 females per dose group. Neither BPA nor EE_2 affected the timing of vaginal opening at the doses tested, although the high EE_2 treatment had clear effects on the estrous cycle.

Dr. Delclos also reported data on organ weight, clinical chemistry, hematology, and sperm parameters collected at the interim sacrifice. Neither BPA nor EE_2 treatments affected male organ weights, but both agents had statistically significant effects on several clinical chemistry and hematology endpoints. The differences were not judged to be adverse effects. Neither agent appeared to affect sperm parameters.

IV.C.2 Questions for Clarification

Dr. Conner noted the absence of stopping rules based on survival for the study, which appeared to follow the NTP practice, versus the FDA approach.

Dr. Tooze asked what criteria were used for considering something an adverse effect, particularly regarding the hematology results, where there were significant differences that were not considered adverse. Dr. Delclos said no strict criteria were used, but the findings were in the normal range for clinical pathology data and the judgment was based on the magnitude of the effect and the opinion of the pathologist.

Dr. Dorman asked Dr. Delclos if he had a sense the clinical pathologists were using the normal ranges derived for the strain of rats at NCTR to make their determinations. Dr. Delclos indicated their determinations were based on the minimal magnitude of the differences across dose groups rather than on reference ranges.

IV.C.3 Results and Preliminary Conclusions, Part Two

Dr. Delclos presented the results of the histopathological examinations. The narrative report, based on the diagnoses of the study pathologist, indicated that many of the identified lesions were common in aging Sprague-Dawley rats with variable incidences across dose groups. Dr. Delclos described the statistical analyses used and the concurrent and past NCTR control incidences. He provided details about the incidences of neoplasms and nonneoplastic lesions in the female animals, in the interim and terminal sacrifice groups.

In the females:

- The high dose of EE_2 induced multiple effects, particularly in female reproductive organs, including increases in mammary gland and pituitary neoplasms.
- Statistically significant effects of BPA were observed in multiple tissues in various dose groups, but a coherent pattern of effects was difficult to discern.

IV.C.4 Questions for Clarification

Dr. Rosol asked about the pharmacokinetics of EE_2 in the chronic study. Dr. Delclos described a previous pharmacokinetic study of BPA and EE_2 performed in conjunction with the NCTR 90-day study, in which animals were dosed on multiple postnatal days. The pattern of measured blood levels of EE_2 in the study were similar to those of BPA, with much higher levels of unconjugated active compound in younger animals than in older animals.

Dr. Tooze asked why the incidence in the report was identified as simple incidence, since the p-values were based on the Poly-3 test, which also provides incidence. Dr. Delclos said that was the common practice in NTP reports, and that researchers did compare Poly-3 incidence to historical controls.

IV.C.5 Results and Preliminary Conclusions, Part Three

Dr. Delclos continued his presentation of the histopathological results for males and provided an overall summary of all results.

In the males:

- Effects of EE_2 were minimal, in contrast to the females.

- As with females, statistically significant effects of BPA were observed in multiple tissues in various dose groups, but the magnitude of most effects was small and a coherent pattern of effects was generally difficult to discern.
    - The incidence of hyperplasia in the pars distalis of the pituitary was increased at the highest BPA dose in both the continuous- and stop-dose study arms.
    - Effects in the epididymis but not in the testes were observed at the high dose of BPA.

Dr. Delclos summarized the Core Study's results and preliminary conclusions:

- BPA had effects that were distinguishable statistically from background, but with questionable biological relevance.
- The effects were largely not dose-responsive and often occurred in only one BPA dose group, and responses in the stop-dose and continuous-dose study arms did not present a clear pattern.
- The observed variations in endpoints, particularly below the highest BPA dose, were within the range of normal biological variation.
- Some of the observed effects at the highest BPA dose (25,000 µg BPA/kg bw/day) might be treatment related (pituitary hyperplasia and epididymal effects in males, effects in uterus and vagina in females).
- EE_2 produced clear adverse effects in females.

IV.C.6 Questions for Clarification

Noting that the study was designed to assess dose-response, not hazard identification, Dr. Dorman asked if, any hypotheses regarding dose-response and the nature of the curve, monotonic or non-monotonic, had been made a priori. Dr. Delclos replied no statistical analysis was designed to look for non-monotonic dose responses. Dr. Walker added no a priori expectations were formulated because, in covering such a wide range of doses, researchers were focused more on the margin of exposure than on the form of the dose-response curve.

Dr. Camacho responded to Dr. Rosol's earlier question regarding the pharmacokinetics of EE_2 in the chronic study. She said that internal EE_2 was not measured in the 1- and 2-year animals because relevant data were collected in the 90-day study.

IV.D Oral Public Comments

Dr. Laura Vandenberg from the University of Massachusetts Amherst was the first oral public commenter, speaking by telephone. She said that what is missing from the Core Study report is that it is part of a wider discussion about how to identify new methods and new endpoints for regulatory hazard assessment. She said NTP's role in each step was less clear after the meeting's presentations thus far. Also, that the study was supposed to be a dose-response study was not clear in the report. She approved of the opportunity for academics to participate in a project using tissues from GLP-compliant laboratories, although found the failure to acknowledge the role of academics to be a shortcoming. She expressed concern about endpoints not analyzed in the report, such as breeding success. She found that the numerous effects observed at low doses interesting. She posited that stress could be an important consideration, potentially altering response to BPA, and provided data on body weight differences between the continuous- and stop-dose arms to support that assertion. She cited the lack of exposure data as a serious concern and found the proposed use of historical controls problematic.

Cécile Michel spoke on behalf of ANSES, the French Agency for Food, Environmental and Occupational Health and Safety. She said the report was missing information, including discussion of how the pups were exposed to BPA. She mentioned her written comments submission included two questions regarding exposure. The first inquiry was regarding potential exposure related to housing conditions. The second concerned differences in exposures in the groups due to use of two formulations, potentially affecting systemic bioavailability. For the statistical analysis, she believes the five loading groups could be treated as independent studies. She said that the negative effects observed in estradiol exposure in positive controls were surprising. She expressed further concerns about some of the methods used with the controls. She asked for more discussion of effects on mammary glands.

IV.E Presentation of Peer-Review Comments

Dr. Dorman introduced the peer-review comment section of the meeting. The reviews were guided by a series of questions:

- *Question 1*: Please comment on whether the information presented in the draft report, including presentation of data in any tables and figures, is clearly and objectively presented. Please suggest any improvements.
- *Question 2*: Please comment on the study design and conduct used in these studies (e.g., animal model, window and route of exposure, utility of the stop-dose arm).
- *Question 3*: Please comment on whether the different statistical approaches have been appropriately applied. As part of your evaluation, please comment on whether the Relative Treatment Effect (RTE) method, which is not typically used in NTP studies, was useful in evaluating the results.
- *Question 4*: Please comment on whether the report sufficiently and appropriately compares the results of BPA with the results of the reference estrogen used in the study.

IV.E.1 Peer-Review Comments for Reviewer Question 1

Please comment on whether the information presented in the draft report, including presentation of data in any tables and figures, is clearly and objectively presented. Please suggest any improvements.

Dr. Bunton was the peer reviewer assigned to lead the discussion on Question 1.

She found the report comprehensive for a guideline-compliant study with primarily toxicological endpoints. She felt, overall, the information was clearly and objectively presented. She especially liked Tables 1 and 2. She said she would have preferred all BPA data be presented before the estrogen data, which is not how it was approached in the report; the estrogen data are currently presented between the two BPA studies. She cited inconsistency in the subject headers and said she had previously provided additional editorial comments.

Dr. Treinen concurred with Dr. Bunton's approval of the tables. She also agreed with adjusting the order of the EE_2 data presentation.

Dr. Conner said that including comments cross-referencing findings between stop-dose and continuous-dose groups would be helpful to make the data easier to interpret.

Dr. Tooze agreed that the presentation of data and results in the tables and figures was clear, particularly Table 1. She expressed concern, however, about the data presentation in the discussion. First, she was concerned about attributing statistically significant results as

biological variability. The purpose of statistical testing is to determine the probability of observing a particular result when accounting for variability, and she did not feel comfortable dismissing some of the statistically significant findings as biological variability. Second, she questioned the interpretation of the neoplasm findings. She noted that the conclusions were based on comparisons to historical controls, and she was unclear on the use of the term "marginally higher" as the language for comparison. The term was not defined, and she believed the percent differences in some of the incidence values were high enough to surpass "marginal." She suggested that the report be clear when reporting incidence results, specifically whether they are based on simple incidence or the Poly-3 test. Third, she noted that the report was clear that the statistical comparisons were designed to be made within the arms. Qualitatively comparing between arms in the report was appropriate, but due to the difference in sample size and statistical power between the two arms, they should not be compared statistically.

Dr. Dorman noted that in the written comments panelists submitted, a recurring concern is related to statistical testing and how biological variability is handled with respect to changing the interpretation of the statistical results.

Responding to those concerns, Dr. Delclos agreed that a more stringent p-value could have and perhaps should have been set. Given that NCTR used a liberal p-value and a large number of endpoints were involved, however, they felt discussing the statistically significant results in light of the pathologist's comments regarding biological variability was appropriate. He said the authors would review how this was discussed in the report. He noted that within the 2-year arm, the stop-dose and continuous-dose groups had equal sample sizes.

Mr. Felton, study statistician from NCTR, said with a study like this, where the chance of many false positives was high, many factors had to be considered; for example, effects in historical controls were considered in teasing out real effects, which could indicate whether a certain outcome might be a false positive.

Dr. Gamboa da Costa from FDA noted that a standard NTP approach was followed to put the data into perspective. The study was seven-fold larger than a typical cancer bioassay, he observed, and when using a p-value of <0.05 with four different statistical approaches to analyze the data, the likelihood of finding false positive data was increased. Using much stricter a priori statistical conditions might have led to a loss of biologically significant data. The method used was a much better approach, he stated—to allow more false-positive data to appear in the statistics, while trying ascertain some sense of biological significance.

Dr. Treinen felt there was a lack of discussion of biological plausibility, providing the uterine polyp findings as an example. She said she would like to see more discussion on what would have been expected in the 2-year versus the 1-year data.

Dr. Conner noted the likely presence of some false positives and how they should be approached in terms of weight of evidence. He referred specifically to the adenocarcinomas in the 2.5 µg BPA/kg bw/day stop-dose group, and whether they were biologically relevant. That result could be compared with the continuous-dose group, looking at whether it correlated with pre-neoplastic findings in the same dose group, which would make the result more likely to be a biological effect. Consulting the historical controls is important, not to discount elements, but to place findings in perspective. Taking a weight-of-evidence approach is important when considering low-dose effects.

Dr. Rosol called for a clear tabulation of which changes were felt to be "compound mediated," along with discussion of the subsequent biological relevance (or lack thereof). If the conclusion is that a particular effect is not compound mediated, that also should be discussed. He felt that eliciting that type of succinct information from the report is difficult. He approved of areas in the summaries where uncertainty was expressed and felt they should have been explained further.

Dr. Conner said whether an effect is treatment related is faced frequently in pharmaceutical research. He noted that several findings in the report would likely end up in a category of "uncertain relationship to treatment" and said that more clarification of what is considered treatment-related would help with interpretation.

Dr. Rosol added he would like more discussion of the expectations, or predictions, of exposure for both BPA and estrogen. Such discussion would be helpful to the reader in understanding what is expected in terms of blood levels and other findings.

Dr. Dorman noted several comments are related to reporting of incidences. He asked the panel if they had more comments on the statistical comparisons between the arms of the study.

Dr. Walker asked whether the concern about reporting of incidences referred to the discussion section of the report. Dr. Tooze confirmed, noting that incidences were reported clearly in the tables, but less so in the narrative sections. Dr. Delclos asked for further clarification of Dr. Tooze's concerns. Dr. Dorman asked whether the concern was qualitative discussion between the stop-dose and continuous-dose groups, or quantitative. Dr. Camacho observed that the two arms had not been statistically compared, and comparison in the discussion was only qualitative. Dr. Tooze noted the potential differences in statistical power between the stop-dose and continuous-dose groups. Dr. Camacho clarified that the sample size was the same between the two arms at each sacrifice time point. Dr. Tooze asked if comparisons were made across sample sizes. Dr. Walker said the only differences in sample size would be between the 1-year and 2-year sacrifices and between the BPA and EE_2 groups, not between the different arms.

IV.E.2 Peer-Review Comments for Reviewer Question 2

Please comment on the study design and conduct used in these studies (e.g., animal model, window and route of exposure, utility of the stop-dose arm).

Dr. Conner was the first peer reviewer assigned to lead the discussion for Question 2.

He noted the complexity of the study, which had several purposes as outlined by Dr. Conner:

- To establish potential effects of exposure to BPA on both male and female rats starting in utero and extending through 1 or 2 years of age.
- To establish effects of BPA in rats exposed in utero and the perinatal period through weaning.
- To establish whether BPA has effects on age-related findings in rats.
- To establish that the rat model used in this study is sensitive to estrogenic effects by including animals directly exposed to ethinyl estradiol.

Based on those purposes, Dr. Conner thought the study was adequately designed. The use of multiple groups was a strength of the design, as they allowed qualitative comparisons to help understand the weight-of-evidence approach for some of the findings. Another strength was the significant effort to minimize the presence of background BPA; the approach taken to deal with

the challenge was appropriate. He also thought the number of animals was adequate for the purpose.

Dr. Bunton was the second peer reviewer assigned to lead the discussion for Question 2.

She stated that she had not understood what the overall question to be addressed by the study was, and why the study was designed the way it was. The objective (or hypothesis) for the study should have been clearly stated in the report. A more straightforward reproductive and developmental study design (DART approach) might have yielded cleaner answers than dosing throughout the lifetime. The current approach might be better for a compound without a hormonal influence, as hormonal changes occur throughout development. With a hormonally active substance, different doses could have considerably different effects on reproductive cycling or developmental events, or both, as developmental cycles are passed through (i.e., neonatal, juvenile, adult, senescence). Therefore, key effects could be missed as the developmental stages progress. The design makes comparisons with the academic groups difficult, as they apparently are investigating more limited developmental events. Therefore, how the data from the Grantee Studies tie in with the data from the Core Study will be interesting. A potential concern, she noted, is the lack of plasma exposure data, with a lack of information about steady state. She called for adding justification to the report for why the interim groups were included.

Dr. Delclos agreed that the purpose of the study should be more clearly stated, although redesigning the study is not possible. He said that NCTR has extensive pharmacokinetic data for the strain of rat used for these studies, and they can make an estimate about the internal dose. He noted that the information on exposure can be explained more clearly in the report. Dr. Camacho added that the dosing and vehicle were the same in the 90-day and 2-year studies, and therefore, the internal dosimetry data from the 90-day study should be applicable to the 2-year study.

Dr. Walker noted, although this study was being reviewed in isolation, it is part of the continuum of data from the many studies involving BPA. He said the challenge is in deciding how far to go with biological interpretations of statistically significant effects.

Dr. Bunton again asked for clarification on the overall objective of this study. Dr. Bucher said that the intent was to explore the bounds of the power of standard guideline studies to detect endocrine-active agents that might be working through dose-response scenarios that are not fully understood. The challenge was to design a study sensitive enough to detect what should be detected, and not to overdesign to the point where answering the overall question is not realistic.

Dr. Conner asked whether the Core Study report was likely to be revised once the Grantee Studies data were available. Dr. Walker replied, historically NTP does not revise reports, but they have integrated different data streams in some instances. Previous reports are not changed though, even if the conclusions change. He said that the integration report is a subsequent step to this study that will incorporate all available data at that point, including potential changes to conclusions.

Dr. Bunton addressed the value of the stop-dose group. She found that that group was a unique aspect of the study design and yielded material of interest. Dr. Bunton indicated that, upon observing an increased incidence of adenocarcinomas in the mammary gland of the female animals in the 2.5 µg/kg bw/day stop-dose group, she had attempted to see if concurrent effects had been observed in other organs in that arm of the study that could correlate with that

change. She looked for effects in the hypothalamic-pituitary-gonadal axis (e.g., changes in the ovary or hyperplasia) and found none. She said the hypothalamic-pituitary-gonadal axis also should have been affected if the mammary gland had been impacted in a meaningful, biologically significant way. Dr. Bunton said she made several different comparisons and found little consistency of effects, noting that these data might make more sense when the grantee information is received.

Dr. Treinen said she was not enamored with the inclusion of the stop-dose group and found doing so made the data difficult to interpret. She felt that not many differences were observed between the short-term and long-term exposure data and said that she would like to see more integration and discussion on how that influenced data interpretation.

Dr. Dorman noted several of the public commenters had raised the issue of having no stop-dose group for the positive control, the EE_2 group. He asked the panel if they found that to be a concern.

Dr. Treinen said it was a concern, as relating the positive control to the observed effect was difficult, particularly with the lack of a positive control for the non-monotonic effect on adenocarcinomas. Dr. Walker explained that logistics precluded including that in the design, in particular, the large number of animals already required for the study. Dr. Treinen noted that another dose group had been included, the 0.05 EE_2 group, which turned out to be uninformative. She suggested that those animals could have been devoted to a stop-dose approach. Dr. Walker said the team had discussed that point extensively. Dr. Delclos noted that some tough choices had to be made at the beginning of the study design. Dr. Dorman encouraged the report's authors to consider the panel's comments when presenting the strengths and limitations of the positive control group, and whether not having a stop-dose group in the EE_2 group limited the ability to interpret certain data.

Dr. Walker noted BPA pharmacokinetics is well known and, once the exposure is stopped, internal levels of BPA quickly decrease.

Dr. Rosol acknowledged no EE_2 stop-dose group was included due to resource issues, but looking at the overall data, he felt having one was not as necessary as it might have been. He said the only finding in the stop-dose group that he found interesting was the cystic follicles observed in the high-dose BPA stop-dose group, and that all other findings did not raise his concern. Determining whether estrogen had the same effect would have been interesting, he noted. Dr. Conner agreed with Dr. Rosol, stating he thought the EE_2 group was only included to demonstrate sensitivity of the model, so a stop-dose group was not completely necessary.

Dr. Bucher explained that when the study was designed, the appropriateness of the positive control was intensely discussed, because much evidence at the time indicated BPA worked through a non-estrogen pathway. He said, in hindsight all would agree that the EE_2 arm was built in as a measure of sensitivity of the model, not as a mimic where one would expect to see identical effects. Dr. Treinen suggested that that point be included in the report discussion.

IV.E.3 Peer-Review Comments for Reviewer Question 3

Please comment on whether the different statistical approaches have been appropriately applied. As part of your evaluation, please comment on whether the Relative Treatment Effect (RTE) method, which is not typically used in NTP studies, was useful in evaluating the results.

Dr. Tooze was the peer reviewer assigned to lead the discussion for Question 3.

She found that the statistical approaches in the study were appropriately applied and reasonable for the data structure. For most endpoints, treatment groups were compared with adjustment using Dunnett's test. She said the relative treatment effects (RTE) method was useful, particularly in examining the potential presence of a non-monotonic trend. Dr. Tooze found the RTE method an appropriate test for the types of data in the study.

Mr. Felton from NCTR asked if the lack of adjustments in the pathology analyses had been discussed. Dr. Tooze wondered why they were not adjusted and said directly addressing that in the report would be helpful. Dr. Rosol asked if NTP ever adjusted for multiple comparisons in their studies. Dr. Walker said that NTP typically does not, but that the weight-of-evidence approach controls the false positive rate to the $p < 0.05$ level. Dr. Gamboa da Costa noted that the study was seven-fold larger than normal studies where no adjustment is made for multiple comparisons; for that reason, the critical criteria to analyze the data from a biological plausibility standpoint should be even higher.

IV.E.4 Peer-Review Comments for Reviewer Question 4

Please comment on whether the report sufficiently and appropriately compares the results of BPA with the results of the reference estrogen used in the study.

Dr. Treinen was the first peer reviewer assigned to lead the discussion for Question 4.

She felt that the data were very clearly presented and discussed. She referenced her previous comments concerning the lack of a stop-dose EE_2 group and the 0.05 low-dose group as being uninformative. Otherwise, the model was effective in showing estrogen-related effects, she noted.

Dr. Rosol was the second peer reviewer assigned to lead the discussion for Question 4.

He said he was intrigued that estrogen alone was so carcinogenic, even though the exposure level was rather temporary. He felt that the 1-year time point was especially important and critical for comparing the estrogen to BPA for potential BPA estrogenic effects. He suggested that point be emphasized in the discussion.

Dr. Walker asked Dr. Rosol if he felt the comparison had not been adequately addressed in the discussion. Dr. Rosol said it did not seem "front and center" to him.

IV.F Panel Discussion and Recommendations

The panel considered the scientific interpretations included in the Draft NTP Research Report on the CLARITY-BPA Core Study.

Dr. Dorman explained the process to the panel. He said that he would first ask the panel whether any members would wish to move to accept all seven of the draft scientific interpretations as written, and to second the motion, followed by a vote. If no second were stated or if the panel voted against the motion, he would call on the panel to consider each interpretation individually. He noted that NTP had asked the panel not to rewrite the interpretations, but instead to provide information about how the interpretation should be changed if necessary.

Dr. Conner moved to accept the seven interpretations as written. The motion was not seconded, and so it failed, leading the panel to consider each of the seven interpretations individually. Dr. Dorman mentioned that he would conduct a non-binding straw vote to begin each consideration,

to get a sense of the panel's thinking and perhaps streamline the process. Following discussion, a formal vote would be taken.

IV.F.1 Draft Scientific Interpretation #1

BPA produced minimal effects at varying doses that were distinguishable only statistically from background in this study, particularly below 25,000 µg BPA/kg bw/day.

Dr. Dorman conducted a straw vote to accept the statement as written, which yielded 1 member in favor (Dr. Conner), with the other 4 voting no. Dr. Dorman asked the panel members to elaborate on their concerns.

Dr. Bunton said her concern was with the wording of the interpretation, particularly "distinguishable only statistically from background." She said she believed important effects had occurred but were not shown statistically.

Dr. Treinen also had an issue with "only," as it implied a lack of biological plausibility. She felt that "only" was too definitive. Dr. Tooze agreed and added concern about the word "minimal." She was not convinced that the report had adequately described what was considered a minimal effect versus a non-minimal effect in terms of biological significance. She also said that some effects were distinguishable statistically and the word "only" implies that they were not biologically meaningful. Dr. Rosol said that he agreed in spirit, but also could not accept the language in the statement. He noted that he interpreted "minimal" to mean effects had occurred, but they were small. He was not willing to go that far, he said, with only some equivocal findings below 25,000. Ultimately, he felt the statement was not accurate.

Dr. Conner said that despite his yes vote in the straw poll, the statement could go away without a major impact on interpretation of the study. He shared reservations about the word "only," but not to the point where he would vote no.

After clarifying that moving to *not* accept the interpretation as written would be acceptable, Dr. Dorman called for a motion. Dr. Rosol moved to not accept the interpretation as written. Dr. Bunton seconded the motion. The panel voted 5 yes, 0 no, to **recommend not accepting Interpretation #1 as written**.

IV.F.2 Draft Scientific Interpretation #2

Many of the statistically significant BPA effects were not dose-responsive or occurred only in one dose group, typically not the highest dose group, and there was not a clear pattern of consistent responses in the stop-dose and continuous-dose study arms.

Dr. Dorman conducted a straw vote to accept the statement as written, which yielded 4 yes votes and 1 no vote, by Dr. Treinen. She explained she voted no because she struggled with what some of the conclusions did not say. Regarding the lack of consistent responses in the stop-dose and continuous-dose study arms, she asked why they would have been expected to be consistent. Dr. Walker explained that within each of the two study arms, no clear dose-response patterns were detected; they were not comparing the two arms. Dr. Treinen said that the report contained comments implying that the stop-dose was almost used for comparison, which made interpretation complicated. She noted that within discrete groups, she thought the statement was true, but given the write-up in the report, that did not seem to be the intent. Dr. Tooze agreed, and said she had the same trouble interpreting the statement. Dr. Walker noted that in the fourth statement, comparison between the study arms was made.

Dr. Dorman called for a motion to accept the statement as written. Dr. Bunton so moved, Dr. Conner seconded. The panel voted 5 yes, 0 no, to **recommend accepting Interpretation #2 as written**.

IV.F.3 Draft Scientific Interpretation #3

Background BPA exposure, from the diet or other unintended sources, did not affect the interpretation of the study.

Dr. Dorman conducted a straw vote to accept the statement as written, which yielded 3 yes votes and 2 no votes.

Dr. Treinen explained that she voted no because there was no control group that did not have some exposure to BPA or estrogenic compounds. She said there was no way of knowing whether the statement was accurate, and that stating unequivocally no effect occurred would be difficult. Dr. Tooze was concerned with the very strong phrasing, "did not affect the interpretation of the study." Although the BPA levels from the diet were apparently minimal, she noted, she was still concerned about the potential contamination, with the significant differences observed in the sensitivity analysis.

Dr. Dorman clarified that the sensitivity analysis Dr. Tooze mentioned was associated with the cohort 1/cohort 0 sensitivity analysis performed due to concerns for cross-contamination with animals being cohabitated in the same room as the cohort 0 academic animals.

Dr. Rosol said that he read the statement quite literally, in that it did not affect the *interpretation* of the study, not that the background exposures did not affect the animals' responses. He said those levels were clearly explained in the study, and the reader should decide whether they had an effect. The levels were stated, and people should make their own interpretations, he noted. Thus, he did not have a problem with the statement, because readers are free to interpret it their own ways.

Dr. Walker noted that the goal is always to interpret relative to control, and that the BPA background levels had been measured, such as the 10-fold lower measurement in the diet, which yielded confidence that an actual exposure differential existed between the 2.5 µg BPA/kg bw/day dose group and the control cohort.

Dr. Treinen asked for clarification from Dr. Rosol that he interpreted the statement as saying no effect occurred. Dr. Rosol said his impression was the data were there, but did not affect the interpretation of the study. He also stated achieving a control group with zero BPA exposure would be virtually impossible. He concluded by saying the wording of the interpretation was not perfect.

Dr. Dorman asked the panel whether they had concerns that the low levels confounded the study. No panel members responded. Dr. Dorman asked for a motion to accept the interpretation as written. Dr. Conner so moved and Dr. Rosol seconded. The panel voted 3 yes, 2 no, to **recommend accepting Interpretation #3 as written**.

Dr. Treinen and Dr. Tooze were the "no" votes and cited their earlier comments as their reasons for voting against the motion.

Dr. Rosol asked for clarification regarding how the panelists' comments were going to be taken into account. Drs. Wolfe and Walker explained the process for how NTP and NCTR would

consider the panel's advice when making revisions and how final decisions on accepting or rejecting recommendations would be made.

IV.F.4 Draft Scientific Interpretation #4

Although animals in the stop-dose and continuous-dose arms were handled differently and were not statistically compared, in several cases statistically significant increases in lesion incidences in BPA treatment groups relative to the vehicle control in one study arm were similar to the vehicle control group in the other study arm.

Dr. Dorman conducted a straw vote to accept the statement as written. There were 0 yes votes, 5 no votes.

Dr. Conner explained he voted no because the statement was too universal, too broad. He said that his point comes under the previous discussion regarding adding appropriate things to the weight of evidence and does not apply to all of the findings being interpreted. Dr. Dorman stated how critical this discussion is. Dr. Bunton felt that the statement was not necessary and went a step too far. She also questioned the statement's validity. Dr. Treinen said that she found the statement to be true, and looked at what it did not say. She said the controls were different between the two arms and she struggled with the intent of the statement and how it was presented in the report.

Dr. Delclos said that he understood the impetus for the comments and appreciated them. Dr. Walker said that they were trying to "thread the needle" between providing biological context and making appropriate statistical comparisons. He agreed with the comments that the statement is too broad and asked if including specific cases important to biological interpretation would improve the statement. Dr. Conner said that would help and clarified that voting no would not affect the validity of NTP's interpretations.

Dr. Rosol explained his no vote as stemming from his belief that the statement did not belong as a general conclusion. He said the statement sounded rather apologetic and opened the door for criticism. Dr. Tooze agreed that the statement was too broad and unnecessary, and said that the results needed to be presented with balance.

Dr. Dorman asked for a motion to not accept the interpretation as written. Dr. Conner so moved and Dr. Treinen seconded. The panel voted 5 yes, 0 no, to **recommend not accepting Interpretation #4 as written**.

IV.F.5 Draft Scientific Interpretation #5

Some of the effects observed at 25,000 µg BPA/kg bw/day (e.g., pituitary and epididymis in males, uterus and vagina in females) may be treatment-related.

Dr. Dorman conducted a straw vote to accept the statement as written, which was 5 yes, 0 no.

Therefore, the motion was to accept the statement as written. Dr. Treinen so moved and Dr. Bunton seconded. The panel voted 5 yes, 0 no, to **recommend accepting Interpretation #5 as written**.

Dr. Treinen indicated that the word "may" really helped in her decision to vote yes.

IV.F.6 Draft Scientific Interpretation #6

The high EE_2 dose (0.5 µg EE_2/kg bw/day) elicited effects in females and males.

Dr. Dorman conducted a straw vote to accept the statement as written, which was 5 yes, 0 no.

Therefore, the motion was to accept the statement as written. Dr. Rosol so moved and Dr. Conner seconded. The panel voted 5 yes, 0 no, to **recommend accepting Interpretation #6 as written**.

Dr. Tooze said that she agreed with the statement, although it was not very specific, and that more detail on the systems with effects would be helpful. Dr. Conner said that he had interpreted the EE_2 group as showing sensitivity to estrogenic effects, which was the purpose of the group in the study design.

IV.F.7 Draft Scientific Interpretation #7

The points above suggest that the observed variations in endpoints (e.g., body weights, lesion incidences), at least those below the highest BPA dose, within a given study arm were within the range of normal biological variation.

Dr. Dorman conducted a straw vote to accept the statement as written. The result was 0 yes, 5 no.

Dr. Conner said the statement was too dismissive of the other arguments that had been brought up that could support or detract from a given finding as treatment related. He said the statement sounded apologetic and too defensive.

Dr. Tooze cited the distinction made between biological plausibility and biological variability and said she would feel more comfortable with discussions of plausibility of effects, rather than variability. She noted that in a very large study, statistical methods take variability into account, and findings could be statistically significant when they are not clinically or biologically meaningful. She called for more interpretation in the report in terms of biological plausibility.

Dr. Bunton agreed that the statement was too broad and dismissive and was trying too hard to fit everything into biological variability.

Dr. Treinen noted that “the incidence of adenocarcinomas” were outside the stated variability. The statement was very broad and not necessarily true for every single finding.

Dr. Rosol agreed that the statement was too all-encompassing for a general conclusion. He agreed with using variability as a specific argument for specific data sets.

Dr. Conner said building the argument for or against biological plausibility would be important. He said that removing the statement, as a broad statement, would actually strengthen the report by focusing on each finding with all of the available data.

Dr. Berridge asked the panel to comment on the basic questions of whether BPA is bioactive, and if so, whether it is endocrine active. Also, he asked for comment on what the activity is and at what doses it occurs. He asked the panel whether members believed that the study being considered was answering any of those questions.

Dr. Conner said that Dr. Berridge’s questions might best be answered by the colleagues at NTP and NCTR. The purpose, in his perception, was to design a study to allow some, but not all, of

those questions to be answered, which would perhaps emerge from the academic elaborations. He felt that the study, as conducted, answers more limited questions, and perhaps not the broader questions expressed by Dr. Berridge. Dr. Conner noted that in the context of the endpoints analyzed in this study, except for the highest dose of BPA, possibly no morphological effects were observed.

Dr. Berridge said that one of the study's aims was to understand whether a typical guideline study allows clarity, and that he perceived from Dr. Conner's comments that the clarity is not present within the context of the study. Dr. Conner replied that the study contributes to the clarity, but one challenge is the expectation that it would be the be-all and end-all study, which is not what it was designed nor intended to be. The conclusions of the study were limited to what was measured, he observed, and the ancillary studies will help answer the broader questions.

Dr. Dorman asked for a motion to not accept the statement as written. Dr. Tooze so moved and Dr. Conner seconded. The panel voted 5 yes, 0 no, to **recommend not accepting Interpretation #7 as written**.

V. Adjournment

Dr. Berridge thanked Dr. Dorman for chairing the peer review and offered his congratulations to everyone at NCTR and NTP who contributed to the complex study. He thanked Dr. Wolfe and her associates for their efforts in organizing the meeting, and thanked the peer reviewers for their time and effort. Dr. Wolfe also expressed her appreciation to the panel.

Dr. Dorman adjourned the meeting at 3:00 pm, April 26, 2018.

VI. Approval of the Peer-Review Report by the Chair of the Peer-Review Panel

This peer-review report has been read and approved by the chair of the April 26, 2018 NTP Research Report Peer-Review Panel.

David Dorman, DVM, PhD, DABVT, DABT

Peer-Review Panel Chair

Date: 6 JUL 18