Food and Chemical Toxicology 131 (2019) 110554

Contents lists available at ScienceDirect

Food and Chemical Toxicology

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

CLARITY-BPA Core Study: Analysis for non-monotonic dose-responses and biological relevance

Melissa A. Badding*, Leila Barraj, Amy L. Williams, Carolyn Scrafford, Richard Reiss

Exponent, Inc, United States

ARTICLE INFO

Keywords:
Bisphenol A
BPA
Non-monotonic dose response
Mammary gland
Endocrine disruption
Low dose

ABSTRACT

The results of a large 2-year bisphenol A (BPA) rat study conducted by the NTP, called the CLARITY-BPA Core Study, were recently released. This study addressed some of the toxicological issues associated with BPA, including endocrine disruption and non-monotonic dose responses (NMDR). The study involved oral gavage treatment of rats to BPA at doses of 2.5–25,000 μg/kg-bw/day. To address NMDR, the 81 statistically significant findings (based on the primary statistical tests) from the Core Study were evaluated using a recently published methodology that relies upon six checkpoints to determine if there is evidence for a NMDR. Failure to meet the majority of the checkpoints indicates limited evidence of NMDR. The analysis found that only 2 of the 81 findings met at least 5 of the checkpoints: an increase in percent basophils in stop-dose females and decreased total bile acids in stop-dose males. However, these findings are not concordant or consistent with those of other BPA data. Importantly, none of the endocrine-related or reproductive endpoints fulfilled at least 5 of the checkpoints. This analysis found limited evidence for NMDR associated with BPA treatment in the study. These results are consistent with the conclusions reached in the Core Study report.

1. Introduction

Bisphenol A (BPA) is a high production volume industrial chemical used as a monomer in the production of polycarbonate plastic and epoxy resins that have broad applications in consumer products, including medical devices and storage containers for foods and beverages (NTP, 2018). While the toxicity of BPA has been extensively studied, the conclusions derived from animal and human exposure studies have been scientifically debated. Safety assessments conducted by multiple regulatory agencies have concluded that current exposure levels of the general population via dietary exposure are low (< 0.5 μg/kg bw/day) and do not pose significant risks to human health (EFSA, 2015; FDA, 2014a; b; Health Canada, 2012). However, others have concluded that the overall weight of evidence from BPA studies indicates that BPA is likely a human health hazard (Gore et al., 2015; Richter et al., 2007; Rochester, 2013; Welshons et al., 2006; Zoeller et al., 2012); this conclusion is based on various effects observed in animal studies with BPA treatment (often related to endocrine disruption) and epidemiological associations of urinary BPA levels with a variety of adverse human health outcomes.

Some researchers have further suggested that BPA exhibits a non-monotonic dose-response (NMDR) (e.g., Vandenberg, 2014; vom Saal et al., 2007). For example, increased mammary gland adenocarcinomas following BPA treatment have been suggested to exhibit a NMDR (Acevedo et al., 2013; Jenkins et al., 2011). A NMDR is characterized by a change in the slope of the dose-response curve across the range of doses examined and is defined mathematically by a response where the slope of the curve changes sign from negative to positive or positive to negative somewhere between the low and high doses (Kohn and Melnick, 2002). In the case of certain essential vitamins, for example, toxicity is associated with both extremely low doses (where essential levels are not being met) as well as at extremely high doses. The potential for low dose effects is particularly important for risk assessment.

In an attempt to address and resolve some of the questions surrounding the toxicology of BPA, including its possible role in endocrine disruption and the potential to induce NMDRs, a two-year repeat-dose rat study of BPA (the Core Study) was conducted by the Consortium Linking Academic and Regulatory Insights on Bisphenol A Toxicity (CLARITY-BPA). The CLARITY-BPA Core Study was a collaborative effort involving the National Center for Toxicological Research (NCTR) of the U.S. Food and Drug Administration (FDA), the National Toxicology Program (NTP), and the National Institute for Environmental Health Sciences (NIEHS). It is part of a larger research program that includes NIEHS-sponsored research conducted by academic laboratories on animals or tissues received from the Core Study and in which researchers were blinded to the doses of BPA that the animals received. The results

* Corresponding author.
E-mail address: mbadding@exponent.com (M.A. Badding).

https://doi.org/10.1016/j.fct.2019.06.001
Received 20 February 2019; Received in revised form 23 May 2019; Accepted 1 June 2019
Available online 15 June 2019

of the Core Study were recently released (NTP, 2018); results from the academic investigations were released separately and are not subject to the analyses presented herein.

The CLARITY-BPA Core Study was conducted according to guideline-compliant research standards and included two arms. One involved oral gavage exposures to Sprague-Dawley rats from the NCTR breeding colony from gestation day 6 through the start of labor; the pups were then directly exposed from postnatal day (PND) 1 until termination at one or two years. This continuous-dose arm represents an in utero through lifetime exposure to BPA. The second arm (referred to as the stop-dose arm) was implemented to assess any effects that were due to early exposure only. It involved exposures from gestation day 6 through PND 21 with no further treatment until sacrifice at one or two years. The study included five BPA-treated groups with doses of 2.5, 25, 250, 2,500, and 25,000 μg/kg bw/day administered to 61–72 pregnant rats/dose group and then to 19–26 pups/sex/dose group for the one-year interim sacrifice groups and 46–50 pups/sex/dose group for the 2-year terminal sacrifice groups. As indicated by Rhomberg and Goodman (2012), to determine whether an endocrine-disrupting chemical has effects at low doses, it needs to be evaluated at doses relevant to human exposures and the lowest BPA dose in the CLARITY-BPA Core Study is in the estimated human exposure range (FDA, 2009). Because the CLARITY-BPA Core Study included 5 doses over a large (4 orders of magnitude) dose range, it provides a unique opportunity to evaluate whether BPA is associated with NMDR. The study also included two negative control groups (one per arm) as well as treatments with ethinyl estradiol (EE_2) at doses of 0.05 and 0.5 μg/kg bw/day (n = 26 rats/sex/dose group) as a positive control for estrogenic activity in the continuous-dose arm.

Many endpoints were evaluated, including body weights, litter parameters, age at vaginal opening, vaginal cytology, hematology and clinical pathology (interim sacrifice only), sperm parameters (interim sacrifice only), organ weights (interim sacrifice only), and histopathology (both interim and terminal sacrifices). For the purposes of assessing statistical significance, BPA-treatment groups were compared against the control group from the same arm only. As expected with a study involving the evaluation of many endpoints (i.e., more than 200) using multiple statistical tests, and because statistical significance was assessed using a p-value of 0.05, some statistically significant findings from the study are likely to be due to chance (Dahiru, 2008; Ogee et al., 2014, 2015). Thus, not all observed statistically significant findings may be treatment-related and some will be due to chance alone. Based on the observations in the study, the final study report concluded that "differences between BPA treatment groups, particularly below 25,000 μg BPA/kg bw/day [the high dose], and the vehicle control group detected by the low-stringency statistical tests applied to histopathology lesions, were not dose responsive, sometimes occurring in only one low or intermediate dose group, and did not demonstrate a clear pattern of consistent responses within or across organs within the stop- and continuous-dose arms and the interim and terminal sacrifices."

For assessing NMDRs, there are no established, agreed upon criteria. Some researchers have suggested using standard statistical methods in addition to visual inspections of dose-response data. For example, Vandenberg et al. (2014) recommends evaluating NMDRs using visual inspections and standard statistical tests for comparisons of group means and sets a criterion for a NMDR as the"[s]tatistical analysis indicates that lower doses produce an increase in response compared to untreated controls and high doses produce a decrease in response compared to untreated controls, or vice versa." However, due to sampling error (false positives), there will always be situations where this criterion is met simply due to chance. Vandenberg et al. (2019) analyzed the Core Study for NMDR by counting the number of statistically significant comparisons with the control group across the doses and found that there were some statistically significant comparisons at doses below the No Observed Adverse Effect Level (NOAEL). The Core Study authors used pairwise tests to compare each of the five dose groups to the vehicle controls for all endpoints. However, statistical significance alone does not necessarily imply the presence of a NMDR since unknown factors could distort the response in one of the dose groups, leading to the detection of a statistically significant difference at that dose. By the same note, a lack of statistical significance does not necessarily imply the lack of a NMDR; however, we assume that true NMDR are more likely to include a statistically significant finding (groupwise or trend) than not. For this reason, and because the Varret et al. (2018) methodology is time consuming to apply, we have limited our application of the Varret et al. (2018) methodology to endpoints with statistically significant findings.

A recent effort by Beausoleil et al. (2016) and Varret et al. (2018) endeavored to develop a systematic approach for evaluating non-monotonic effects (hereafter referred to as Varret et al. (2018)). The research was sponsored by the European Food Safety Authority (EFSA). The methodology described in Varret et al. (2018) for evaluating NMDR comprises assessment of six checkpoints, including some involving statistical analysis (comparing fits for monotonic and non-monotonic models, an approach that is more powerful than pairwise comparisons) and some qualitative evaluations (visual inspection of the dose-response curve and evaluation of the number of slope changes in the dose-response) for studies with at least five dose groups. Varret et al. (2018) suggest researchers consider a meta-analysis of the available data on a specific substance and endpoint when a dataset fulfills at least five of the six checkpoints.

The current paper applies the Varret et al. (2018) methodology to analyze all endpoints where the primary statistical analysis in the CLARITY Core Study indicated statistical significance. Endpoints that fulfilled at least five of the six checkpoints were also evaluated herein for biological concordance, consistency and plausibility, including a comparison of the findings with results from the different arms of the study, results from EE_2 treatment, and findings from another FDA-sponsored BPA study (Delclos et al., 2014).

The analysis of the CLARTIY-BPA Core Study data presented herein provides an integrated statistical and biological analysis of the findings in an effort to further elucidate the potential of BPA to induce adverse effects associated with NMDR from exposure during development and throughout a lifetime.

2. Materials and methods

2.1. Evaluating non-monotonicity in dose-response

The methodology proposed by Beausoleil et al. (2016) and Varret et al. (2018) requires at least five dose groups as a threshold for a NMDR analysis; the fitting of five dose-response models to the data, where two of the models allow for NMDR modeling; and a follow-up evaluation of the data and fitted models using a list of six checkpoints. The same five dose-response models that were used by Varret et al. (2018) to illustrate their methodology were fitted to the CLARITY-BPA Core Study data for the endpoints showing statistically significant findings (as listed in the CLARITY-BPA Core Study report (NTP, 2018)). The five dose-response models are listed in Table 1. Two of the models (NMDR1 and NMDR2 in Table 1) allow for non-monotonicity.

The models were fit in PROAST (Slob, 2002; EFSA, 2009), a software package developed by the National Institute for Public Health and the Environment (RIVM) and which runs in R (R Core Team, 2018). While Varret et al. (2018) used an earlier version (version 60.1) of the software, the current analysis uses version 65.11, which was obtained from the software developer. Both raw data at the animal level and summary data at the dose level can be used in the software. For the binary endpoints, these models were fitted as Latent Variable Models (LVM), which assume that the observed incidences originate from an underlying continuous response that is not directly observed.

The six checkpoints proposed by Varret et al. (2018) to address

Table 1
Statistical models fitted to the data.

Name	Equation	Model number in PROAST
Full model	y = group means	E11
No DR (null model)	y= a	E1
MDR: 4 parameter exponential	$y= a\left[c- (c-1)e^{-bx^d} \right]$	E5
NMDR1: 6 parameter double exponential	$y= a\left\{\left[c_1 - (c_1 - 1)e^{-(x/b_1)^d} \right]\left[c_2 - (c_2 - 1)e^{-(x/b_2)^d} \right]\right\}$	E21
NMDR2: 7 parameter double exponential	$y= a\left\{\left[c_1 - (c_1 - 1)e^{-(x/b_1)^{d_1}} \right]\left[c_2 - (c_2 - 1)e^{-(x/b_2)^{d_2}} \right]\right\}$	E33/E27 (continuous/quantal)

DR, dose-response; MDR, monotonic dose-response; NMDR, non-monotonic dose-response

Table 2
Checkpoint list.

Description	Method used to determine if the checkpoint is met
Checkpoint #1: Does at least one NMDR model fit significantly better than the null model?	Likelihood ratio (LR) test of models 21 and 33 (or 27 in the case on quantal data) vs. the null model. A p-value of 0.05 is used. If the LR test is non-significant, the non-monotonicity maybe due to random sampling error.
Checkpoint #2: Does at least one NMDR model fit significantly better than the MDR model?	Likelihood ratio test of models 21 and 33 (or 27 in the case of quantal data) vs. the MDR model (model E5). A p-value of 0.05 is used. If the LR test is non-significant, the non-monotonicity maybe due to random sampling error.
Checkpoint #3: Does the (visually) apparent NMDR depend on a single dose group (which might be an outlier)?	As described in Varret at al. (2018), this checkpoint is evaluated by drawing a smooth monotonic curve (or horizontal line) that passes through as many confidence intervals (CI) around the observed responses as possible. When that number was above one, then this checkpoint was considered fulfilled.
Checkpoint #4: Are the effect sizes in both directions of the NMDR higher than 5%?	Fig. 1 extracted from Beausoleil et al. (2016) illustrates this checkpoint. The curve of the fitted NMDR model had to have at least 2 directions for this checkpoint to be considered fulfilled.
Checkpoint #5: Is the steepness of the DR curve within the range of biologically plausible/realistic DR shapes?	The steepness of the curve is measured by the "d" parameters. "d" parameters that fall in the [0.25–4] range are considered biologically plausible and indicate that this checkpoint is fulfilled.
Checkpoint #6: Does the apparent NMDR consist of two directions?	This checkpoint was evaluated by drawing a smooth curve with the minimal number of directions needed to hit all the confidence intervals around the observed responses, and then counting the number of directions. If the number of directions was 2, the checkpoint is fulfilled.

possible non-monotonicity include.

- Checkpoint 1 - An evaluation of the goodness of fit of the two NMDR models versus the null model; if at least one of the NMDR models is significantly better than the null model then this checkpoint is fulfilled.
- Checkpoint 2 - An evaluation of the goodness of fit of the two NMDR models versus the monotonic dose response model; if at least one of the NMDR models is significantly better than the monotonic dose response model then this checkpoint is fulfilled.

Since lack of statistical significance in checkpoints 1 and 2 does not necessarily imply non-existence of a NMDR relationship and statistical significance could be due to non-random error in the data, Varret et al. (2018) proposed the following four additional checkpoints:

- Checkpoint 3 - An evaluation of the NMDR models via a visual inspection of the data to determine if an apparent NMDR is due to one or more dose groups; if it is due to more than a single dose, the checkpoint is fulfilled, since this may imply that the non-monotonicity is not due to an outlier.
- Checkpoint 4 - An evaluation of the effect sizes in both directions; if the effect sizes are both greater than 5%, the minimum benchmark dose response recommended by EFSA (2009), the checkpoint is fulfilled.
- Checkpoint 5 - An evaluation of the steepness parameters in the NMDR models; if these parameters are within the range of 0.25–4 that was shown to be biologically plausible by Slob and Setzer (2014), the checkpoint is fulfilled.
- Checkpoint 6 - A visual inspection of the fitted model; if the apparent NMDR has no more than two directions, the checkpoint is fulfilled, since a dose-response in more than two directions may be due to non-random errors.

2.2. Data used in the analysis

The summary data for the outcomes identified in Table 1 of the CLARITY-BPA Core Study report as showing statistically significant treatment effects for BPA relative to vehicle controls in the primary statistical analyses were extracted from the NTP report tables. Specifically, for continuous outcomes (i.e., organ weights, hematology, and clinical chemistry endpoints), the number of animals, mean, and standard error of the mean (SEM) for each dose group were extracted from the report tables. The exceptions were for albumin and % basophils, where the rounded SEM was reported as "0" in the report. In these cases, the unrounded SEM was calculated from the raw data. For the binary endpoints (i.e., for the neoplastic and non-neoplastic lesion outcomes), the number of animals and lesions per dose group were extracted from the report tables.

2.3. Implementation of the methodology

The five models were first fitted in PROAST using the default start values. The estimated parameters, the log-likelihood, Akaike information criterion (AIC), and a convergence indicator were saved for each outcome, together with graphical summaries of each model. The fitted models were then reviewed to confirm that they converged. Models that did not converge were re-run iteratively by changing the start values until they converged, and the models were then saved. The graphical summaries were reviewed to confirm via visual inspection that the models fit the data. When the visual inspection indicated that the curve could potentially be improved, the models were re-run until the curve provided a good visual fit to the data. This was necessary because the NMDR models being fitted to the data have more parameters (6 or 7)

Table 3
Checkpoints fulfilled for evaluated endpoints.

	Checkpoint 1: At least 1 NMDR fits better than null model	Checkpoint 2: At least 1 NMDR fits better than MDR	Checkpoint 3: The (visually) apparent NMDR does not depend on a single dose group	Checkpoint 4: Effect sizes in both directions of the NMDR > 5%	Checkpoint 5: Steepness is in the biologically plausible range	Checkpoint 6: Apparent NMDR consists of exactly 2 directions
Number of endpoints fulfilling the given checkpoint	10	15	3	42	8	51
Of these, number of endpoints reported in the CLARITY-BPA Core Study as showing a statistically significant pairwise difference or trend test:						
Pairwise difference	4	10	1	26	5	24
Trend test	8	6	3	22	4	30

than there are dose groups (5), thus potentially resulting in different curves with the same log-likelihood. The fit of the curves was independently reviewed by two authors (LB and CS) and any discrepancies resolved by mutual agreement.

The methodology used to determine if the six checkpoints were met for a given outcome is summarized in Table 2. Checkpoints 3, 4, and 6 were evaluated independently by two authors (LB and CS), and any inconsistencies were resolved by a further review by both authors. Additional information on the checkpoints and the associated rationale are available in Beausoleil et al. (2016) and Varret et al. (2018).

Based on the recommendations of Varret et al. (2018), any endpoints that met at least 5 of the 6 checkpoints were further examined for biological concordance and, consistency by examining these in comparison with outcomes for related endpoints and across different treatment arms, findings associated with EE2 treatment, and findings reported in another FDA/NCTR BPA study (Delclos et al., 2014).

3. Results

Eighty-one (81) endpoints were evaluated: 57 for which a statistically significant trend test had been reported in the CLARITY-BPA Core Study, 24 for which statistically significant differences from control had been reported for one or more BPA-treatment groups but without a significant trend test, and 16 for which both a statistically significant difference from controls and a significant trend had been reported.

Table 3 presents the number of outcomes fulfilling each of the six checkpoints; a listing of each of the 81 endpoints assessed and the specific checkpoints fulfilled for each can be found in Supplementary Table S5. Two checkpoints were more likely to be fulfilled than the other four checkpoints. These were Checkpoint 4, which evaluates whether effect size was greater than 5% and Checkpoint 6, which assesses whether the apparent NMDR consists of exactly two directions. Checkpoints 1 and 2, which evaluate the goodness of fit of the NMDR models versus the null model and monotonic dose response model, respectively, were fulfilled by a relatively small number of endpoints (10 for Checkpoint 1 and 15 for Checkpoint 2). Checkpoint 3, which evaluates whether more than one dose group is responsible for the apparent NMDR, was fulfilled by three of the endpoints only. Checkpoint 5, which evaluates the biological plausibility of the steepness of the DR curve, as measured by the "d" model parameters, was met by eight endpoints only.

As discussed by Beausoleil et al. (2016) and Varret et al. (2018), the evidence for a NMDR increases as the number of checkpoints fulfilled increases. Neither Beausoleil et al. (2016) or Varret et al. (2018) report the minimum number of checkpoints that should be met, although Varret et al. (2018) indicates that when all six checkpoints are fulfilled, the datasets "could be considered as providing evidence for NMDR". Varret et al. (2018) further state that "A future perspective could be to perform a meta-analysis of (additional) substance- and endpoint-specific data of DR datasets fulfilling six or five checkpoints." Hence, we decided *a priori* to use at least five of six checkpoints as a cut-off for further analysis.

Of the 81 evaluated endpoints, only two endpoints met at least five of the six checkpoints. These were % basophils at the 1-year time point in the stop-dose arm for females, which met all six checkpoints, and total bile acids at the 1-year time point in the stop-dose arm for males, which met all checkpoints except Checkpoint 3, which evaluates whether the apparent NMDR does not depend on a single dose group. Importantly, none of the endocrine-related or reproductive endpoints fulfilled at least 5 of the 6 checkpoints, including mammary gland adenocarcinoma, which has been discussed in recent assessments of potential NMDR effects for BPA (Acevedo et al., 2013; Jenkins et al., 2011). While significantly increased mammary gland adenocarcinomas were observed at the lowest dose in the stop-dose arm female rats at 2 years, this endpoint only met two of the six checkpoints, namely Checkpoint 4 which evaluates whether the effect size was > 5%, and

Table 4
Total and % basophils at the 1-year time point in stop-dose arm females (mean ± S.E.M.).

BPA (μg/kg bw/day)	0 (Vehicle)	2.5	25	250	2,500	25,000
Basophils, $10^3/mm^3$	0.01 ± 0	0.01 ± 0	0.02 ± 0	0.02 ± 0.01	0.01 ± 0	0.01 ± 0
% Basophils	0.1 ± 0[a]	0.2 ± 0	0.2 ± 0	0.3 ± 0.1	0.1 ± 0	0.1 ± 0

[a] Significant trend, $p < 0.05$.

Table 5
Total bile acids (μmol/L) at the 1-year time point in males (mean ± S.E.M).

BPA (μg/kg bw/day)	0 (Vehicle)	2.5	25	250	2,500	25,000
Stop-dose	36.4 ± 3.1[a]	34.6 ± 3.2	25.0 ± 2.7[a]	32.6 ± 6.5	33.7 ± 4.3	35.0 ± 2.5
Continuous-dose	32.8 ± 2.7[a]	33.3 ± 4.1	32.5 ± 2.4	35.8 ± 3.4	42.0 ± 4.7	28.1 ± 2.7

[a] Asterisks in the vehicle column indicate a significant trend ($p < 0.05$), while asterisks in BPA dose group columns indicate significant differences in pairwise comparisons to the vehicle group.

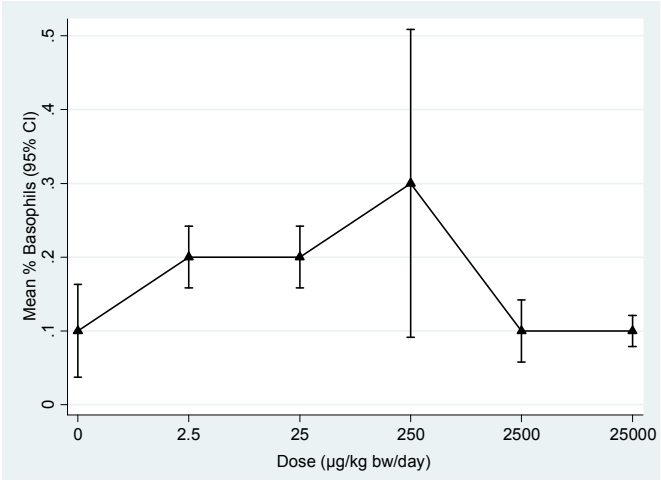

Fig. 1. Mean %basophils in stop-dose arm females with rounded values.

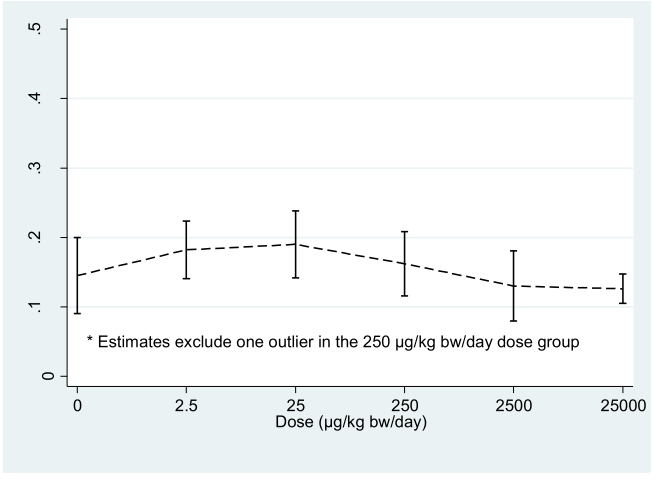

Fig. 2. Mean %basophils in stop-dose arm females with raw, unrounded values.

Checkpoint 6 indicates that the apparent NMDR consists of exactly two directions.

The % basophils at the 1-year time point in the stop-dose arm females met all six checkpoints. However, a corresponding change in total basophils or other white blood cell parameters was not observed (Table 4), and the increase in % basophils was very slight (from 0.1% in controls to a maximum of 0.3% in the 250 μg/kg bw/day dose group). A similar change in the % basophils also was not observed in males or in the other BPA-treatment arm (which experienced a longer duration of BPA exposure that extended beyond that of the stop-dose arm animals). Further, no change in total basophils or % basophils was found with the positive control, EE2 treatment (NTP, 2018).

Total bile acids at the 1-year time point in the stop-dose arm males met five of the six checkpoints. A decreasing trend was seen in males of both treatment arms (Table 5). This was the only clinical chemistry endpoint related to the liver that was altered in male rats; further, this endpoint was not significantly altered by BPA treatment for females in either treatment arm or with EE2 treatment.

4. Discussion

The CLARITY-BPA Core Study final report concludes that "differences between BPA treatment groups, particularly below 25,000 μg/kg bw/day, and the vehicle control group detected by the low-stringency statistical tests applied to histopathology lesions, were not dose responsive, sometimes occurring in only one low or intermediate dose group, and did not demonstrate a clear pattern of consistent responses within or across organs within the stop- and continuous-dose arms and the interim and terminal sacrifices." The evaluation presented herein provides additional analysis of these data, beyond that provided in the CLARITY-BPA Core Study report, using the methodology developed by Beausoleil et al. (2016) and Varret et al. (2018) to assess potential NMDR.

BPA is often cited as a substance linked to NMDR and low-dose responses (Vandenberg, 2014; vom Saal and Hughes, 2005). While the results have been mixed, recent reviews of the scientific literature by regulatory authorities concluded that oral exposure BPA studies in rodent models only rarely display endpoints with NMDR and the effects either lack reproducibility or are likely due to chance (EPA, 2013) and that the available data do not provide evidence that BPA exhibits a NMDR for critical endpoints (EFSA, 2015). The results of the CLARITY-BPA Core Study, because of its incorporation of five BPA-treatment groups, offer a unique opportunity to further evaluate the potential of BPA to induce NMDR.

Only two statistically significant endpoints from the CLARITY-BPA Core study met at least 5 of the 6 checkpoints, including % basophils at 1 year in the female stop-dose arm (met all 6 checkpoints) and total bile acids for males at 1 year in the stop-dose arm (met 5 of the 6 checkpoints). All other endpoints examined in this analysis met less than 5 of the checkpoints.

From a biological standpoint, elevated basophils could indicate a possible myeloproliferative disorder, chronic inflammation, or hypothyroidism (Healthline.com, 2018). However, there is no evidence to suggest any of those occurred in the stop-dose females. Further, as mentioned above, there was no corresponding change in the total basophil counts, no other white blood cell findings that were significantly

altered in the stop-dose females, and treatment with EE2 did not induce basophil count changes. Additionally, the statistical difference for this endpoint is an effect of the reporting precision of one decimal place. The individual animal data range from 0% to 2.2%. The 2.2% value is an outlier (beyond 1.5 times the interquartile range in a boxplot) for stop-dose females; the next highest value is 0.5%. The mean values for all groups were 0.1%, 0.2%, or 0.3%. The authors could have reported average values with two decimal places, but elected to keep the reporting at one decimal place consistent with the individual animal data. Thus, one should be cautious in interpreting differences between 0.1% and 0.2%, or 0.2% and 0.3%. The distortion resulting from the reporting precision of one decimal place is particularly evident in the stop-dose arm. The difference between the control group and the lowest dose group is 0.1% in the Core Study report. However, based on the individual animal data, the mean of the control group is 0.145% (rounded down to 0.1% in the report) and the mean of the lowest dose group is 0.182% (rounded up to 0.2%). Calculated in this manner, the difference is only 0.037% instead of 0.1% as tabulated in the report. Similarly, the two highest dose groups are reported at 0.1%. The actual values are 0.130% (2,500 μg/kg bw/day) and 0.126% (25,000 μg/kg bw/day). When the unrounded means and associated standard errors were used in PROAST and the resulting NMDR models were re-evaluated, only three of the six checkpoints in Varret et al. (2018) were satisfied, namely checkpoints 4, 5, and 6.

Further, if the 2.2% value in the 250 μg/kg bw/day dose group is omitted as an outlier, the mean value in the group would be reduced from 0.252% to 0.164% and the rounded values from 0.3% to 0.2%. When the means are calculated without rounding and the one clear outlier removed, the dose-response no longer appears to show any features indicative of non-monotonicity. Fig. 1 shows the dose-response with rounding and Fig. 2 shows the dose-response with raw unrounded values and the outlier in the 250 μg/kg bw/day group removed. The non-monotonic features are substantially dampened using the raw unrounded data such that the confidence intervals for the group mean values overlap for all dose groups. Thus, based on an evaluation of the individual animal data and the lack of other data to indicate that the finding is biologically relevant, it is highly likely that the increasing trend for % basophils in BPA stop-dose females was incidental and not treatment related or adverse to the animals.

For the endpoint of total bile acids in males at 1 year in the stop-dose arm, the NMDR fit appears to show a depression in total bile acids at the middle doses only. Total bile acids for males in the continuous-dose arm also showed a significant trend, but this endpoint only met two of the Varret et al. (2018) checkpoints. The relevance of the trend in decreased bile acids in males from both BPA treatment arms is unknown. A reduction in bile acids was also observed in males at the high-dose (300,000 μg/kg bw/day) in a previous 90-day FDA-sponsored BPA study (Delclos et al., 2014). However, it is important to note the large difference in the high dose levels between that study (300,000 μg/kg bw/day) and the Core Study (25,000 μg/kg bw/day). An increase, and not a decrease, in bile acids typically indicates liver toxicity (Chiang, 2013; Luo et al., 2014); therefore, the biological relevance of the reported reduction in bile acids, if any, is unclear. Further, this was the only clinical chemistry endpoint related to the liver that was altered in male rats in the CLARITY-BPA Core Study. This strongly suggests that the findings for bile acids were spurious.

Recent publications have reported increased mammary gland adenocarcinomas with BPA treatment (Acevedo et al., 2013; Jenkins et al., 2011). Further, this finding has been suggested to exhibit a NMDR in some publications. In our analysis, increased mammary gland adenocarcinomas observed at the lowest dose in the stop-dose arm female rats at 2 years met only two of the six Varret et al. (2018) checkpoints, indicating little evidence for a NMDR.

A potential limitation of the analysis presented herein is the use of NMDR models with more parameters than dose groups. As pointed out by Varret et al. (2018): “the evidence for NMDR in datasets with less than five doses can be expected to be low in general.” While the current analysis is of a dataset with five doses, over-parameterized models are used to estimate the steepness parameter “d” of the NMDR models. However, as discussed by Varret et al. (2018), evaluation of NMDR should not be based solely on one checkpoint but should be based on all six checkpoints. Further, the models were run iteratively to get the best visual fits and confirmation was determined independently by two of the authors.

Overall, our analysis found little evidence for NMDR in the endpoints evaluated in the CLARITY-BPA Core Study. The results of this analysis are consistent with and support the conclusions reached in the CLARITY-BPA Core Study report. Thus, the results of this large U.S. government-sponsored guideline-compliant study and the present analysis should provide some resolution of the controversy related to the toxicology of BPA.

5. Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was sponsored by the Polycarbonate/BPA Global Group at the American Chemistry Council. The analysis and conclusions presented herein are those of the authors and not of their employer or the sponsor.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fct.2019.06.001.

References

Acevedo, N., Davis, B., Schaeberle, C.M., Sonnenschein, C., Soto, A.M., 2013. Perinatally administered bisphenol A as a potential mammary gland carcinogen in rats. Environ. Health Perspect. 121, 1040–1045.

Beausoleil, C., Ormsby, J.N., Gies, A., Hass, U., Heindel, J.J., Holmer, M.L., et al., 2016. 5. Review of Non-monotonic Dose-Responses of Substances for Human Risk Assessment, vol. 13 EFSA Supporting Publication 2016 EN-1027.290pp. https://doi.org/10.2903/sp.efsa.2016.EN1027. https://ww.efsa.europa.eu/en/supporting/pub/1027e.

Chiang, J.Y.L., 2013. Bile acid metabolism and signaling. Comp. Physiol. 3 (3), 1191–1212.

Dahiru, T., 2008. P-value, a true test of statistical significance? A cautionary note. Ann. Ib. Postgrad. Med. 6 (1), 21–26.

Delclos, K.B., Camacho, L., Lewis, S.M., Vanlandingham, M.M., Latendresse, J.R., Olson, G.R., et al., 2014. Toxicity evaluation of bisphenol A administered by gavage to Sprague Dawley rats from gestation day 6 through postnatal day 90. Toxicol. Sci. 153, 174–197.

European Food Safety Authority (EFSA), 2009. Guidance of the scientific committee on use of the benchmark dose approach in risk assessment. EFSA J 1150, 1–72. http://www.efsa.europa.eu/sites/default/files/scientific_output/files/main_documents/1150.pdf.

European Food Safety Authority (EFSA), 2015. Report on the Two-phase Public Consultation on the Draft EFSA Scientific Opinion on Bisphenol A (BPA). EFSA Supporting Publication 2015:EN-740.

Gore AC, Chappell VA, Fenton SE, Flaws JA, Nadal A, Prins GS, Toppari J, Zoeller RT. EDC-2: the endocrine society's second scientific statement on endocrine-disrupting chemicals. Endocr. Rev. 36(6):E1-E150. 10.1210/er.2015-1010.

Health Canada, 2012. Health Canada's updated assessment of bisphenol A (BPA) exposure from food sources. In: F.D. Bureau of Chemical Safety, Health Products and Food Branch, Health Canada, . Available at: http://www.hc-sc.gc.ca/fn-an/securit/packag-emball/bpa/bpa hra-ers-2012-09-eng.php.

Healthlinecom, 2018. Available at: https://healthline.com/health/basophils#what-happens-if-yourlevel-is-high Accessed July, 2018.

Jenkins, S., Wang, J., Eltoum, I., Desmond, R., Lamartiniere, C.A., 2011. Chronic oral exposure to bisphenol A results in a nonmonotonic dose response in mammary carcinogenesis and metastasis in MMTV-erbB2 mice. Environ. Health Perspect. 119, 1604–1609.

Kohn, M.C., Melnick, R.L., 2002. Biochemical origins of the non-monotonic receptor-mediated dose-response. J. Mol. Endocrinol. 29, 113–123.

Luo, L., Schomaker, S., Houle, C., Aubrecht, J., Colangelo, J.L., 2014. Evaluation of serum

bile acid profiles as biomarkers of liver injury in rodents. Toxicol. Sci. 137 (1), 12–25.
National Toxicology Program (NTP), 2018. The CLARITY-BPA core study: a perinatal and chronic extended-dose-range study of bisphenol A in rats. In: Research Triangle Park, NC. National Toxicology Program, vol. 9 NTP RR Available at: https://ntp.niehs.nih.gov/ntp/results/pubs/rr/reports/rr09_508.pdf.
Ogee, A., Ellis, M., Stone, B.K., Scibilia, B., Pammer, C., Steele, C., 2014. How to correctly interpret P values. The minitab blog. April 17, 2014. Available at: https://blog.minitab.com/blog/adventures-in-statistics-2/how-to-correctly-interpret-p-values .
Ogee, A., Ellis, M., Stone, B.K., Scibilia, B., Pammer, C., Steele, C., 2015. Understanding hypothesis tests: significance levels (alpha) and P values in statistics. The minitab blog. March 19, 2015. Available at: https://blog.minitab.com/blog/adventures-in-statistics-2/understanding-hypothesis-tests-significance-levels-alpha-and-p-values-in-statistics .
R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria Available at: https://www.R-project.org/.
Rhomberg, L.R., Goodman, J.E., 2012. Low-dose effects and nonmonotonic dose-responses of endocrine disrupting chemicals: has the case been made? Regul Toxicol Pharmacol. Oct 64 (1), 130–133.
Richter, C.A., Birnbaum, L.S., Farabollini, F., Newbold, R.R., Rubin, B.S., Talsness, C.E., Vandenbergh, J.G., Walser-Kuntz, D.R., vom Saal, F.S., 2007. In vivo effects of bisphenol A in laboratory rodent studies. Reprod. Toxicol. 24 (2), 199–224. https://doi.org/10.1016/j.reprotox.2007.06.004.
Rochester, J.R., 2013. Bisphenol A and human health: a review of the literature. Reprod. Toxicol. 42, 132–155. https://doi.org/10.1016/j.reprotox.2013.08.008.
Slob, W., 2002. Dose-response modeling of continuous endpoints. Toxicol. Sci. 66 (2), 298–312.
Slob, W., Setzer, W., 2014. Shape and steepness of toxicological dose–response relationships of continuous endpoints. Crit. Rev. Toxicol. 44 (3), 270–297.
U.S. Environmental Protection Agency (EPA), 2013. State of the science evaluation: nonmonotonic dose responses as they apply to estrogen, androgen, and thyroid pathways and EPA testing and assessment procedures. June 2013. Available at: https://www.epa.gov/sites/production/files/2016-01/documents/nmdr.pdf.
U.S. Food and Drug Administration (FDA), 2009. Memorandum to M. Hepp (division of food contact notifications, regulatory group 1) Re: exposure to bisphenol a (BPA) for infants, toddlers and adults from the consumption of infant formula, toddler food and adult (canned) food, FDA-2010-N-0100- 0001. Available at: https://www.regulations.gov/document?D=FDA-2010-N-0100-0009 .
U.S. Food and Drug Administration (FDA), 2014a. Updated safety assessment of bisphenol A (BPA) for use in food contact applications. Available at: https://www.fda.gov/downloads/NewsEvents/PublicHealthFocus/UCM424266.pdf.
U.S. Food and Drug Administration (FDA), 2014b. Final report on the review of literature and data on BPA. Available at: https://www.fda.gov/downloads/Food/IngredientsPackagingLabeling/FoodAdditivesIngredients/UCM424011.pdf.
Vandenberg, L.N., 2014. Non-monotonic dose responses in studies of endocrine disrupting chemicals: bisphenol a as a case study. Dose Response 12 (2), 259–276.
Vandenberg, L.N., Hunt, P.A., Gore, A.C., 2019. Endocrine disruptors and the future of toxicology testing - lessons from CLARITY-BPA. Nat Rev Endocrinol. 2019 Jun 15 (6), 366–374. https://doi.org/10.1038/s41574-019-0173-y.
Varret, C., Beronuis, A., Bodin, L., Bokkers, B.G.H., Boon, P.E., Burger, M., et al., 2018. Evaluating the evidence for non-monotonic dose-response relationships: a systematic literature review and re-analysis of *in vivo* toxicity data in the area of food safety. Toxicol. Appl. Pharmacol. 339, 10–23.
vom Saal, F.S., Hughes, C., 2005. An extensive new literature concerning low-dose effects of bisphenol A shows the need for a new risk assessment. Environ Health Perspect. 2005 Aug 113 (8), 926–933.
vom Saal, F.S., Akingbemi, B.T., Belcher, S.M., Birnbaum, L.S., Crain, D.A., Eriksen, M., 2007. Chapel Hill bisphenol A expert panel consensus statement: integration of mechanisms, effects in animals and potential to impact human health at current levels of exposure. Reprod. Toxicol. 24, 131–138.
Welshons, W.V., Nagel, S.C., vom Saal, F.S., 2006. Large effects from small exposures. III. Endocrine mechanisms mediating effects of bisphenol A at levels of human exposure. Endocrinology (6 Suppl. l), S56–S69. https://doi.org/10.1210/en.2005-1159.
Zoeller, R.T., Brown, T.R., Doan, L.L., Gore, A.C., Skakkebaek, N.E., Soto, A.M., Woodruff, T.J., Vom Saal, F.S., 2012. Endocrine-disrupting chemicals and public health protection: a statement of principles from the Endocrine Society. Endocrinology 153 (9), 4097–4110. https://doi.org/10.1210/en.2012-1422.