Re-evaluation of the risks to public health from bisphenol A (BPA) in foodstuffs – Overview, methods and weight of evidence

This is a paper for discussion.

This does not represent the views of the Committee and should not be cited.

Introduction

1.　　In December 2021 the EFSA Panel on Food Contact Materials, Enzymes and Processing Aids (CEP) published a draft opinion re-evaluating the health risks arising from the presence of Bisphenol A (BPA) in food. The panel propose a significant reduction in the current temporary Tolerable Daily Intake (TDI) of 4 μg/kg body weight (bw)/day to 4 ng/kg bw. This reduction would mean that both mean and high level consumers for all age groups would exceed the new TDI by 2-4 orders of magnitude.

Background

2.　　BPA is a monomer used in the manufacture of polycarbonates, epoxy resins and other polymeric materials, as well as in thermal printing in certain paper products. Polycarbonates are used in food contact materials such as reusable beverage bottles, infant feeding bottles, tableware and storage containers. Epoxy resins are used in the protective linings of food and beverage cans and vats (EFSA, 2021).

3.　　BPA is authorised for use as a monomer in plastic food contact materials in accordance with Commission Regulation (EU) No 10/2011/EU1 on plastic materials and articles intended to come into contact with foodstuffs and

retained UK legislation. The specific migration limit for BPA is 0.05 mg/kg, reduced from 3 mg/kg following the EFSA 2015 evaluation of BPA.

2015 EFSA evaluation of BPA

4.　　　In 2015, the EFSA Panel on Food Contact Materials, Enzymes, Flavourings and Processing Aids (CEF) established a temporary TDI (tTDI) of 4 μg/kg body weight (bw)/day (EFSA, 2015). The toxicity of BPA was evaluated using a weight of evidence approach. "Likely" adverse effects reported in animal studies were considered to be in the kidney and mammary glands. These underwent benchmark dose (BMDL10) response modelling. A BMDL10 of 8,960 μg/kg bw per day was calculated for changes in mean relative kidney weight in a two generation toxicity study in mice. No BMDL10 could be calculated for mammary gland effects. Using data on toxicokinetics, the BMDL10 was converted to a Human Equivalent Dose (HED) of 609 μg/kg bw per day. The CEF Panel applied a total uncertainty factor of 150 (for inter- and intra-species differences and uncertainty in mammary gland, reproductive, neurobehavioural, immune and metabolic system effects) to establish a temporary TDI (t-TDI) of 4 μg/kg bw per day. The CEF panel compared this t-TDI with exposure estimates and concluded that there was no health concern for any age group from dietary exposure and low health concern from aggregated exposure. The CEF Panel noted considerable uncertainty in the exposure estimates for non-dietary sources, whilst the uncertainty around dietary estimates was relatively low.

2021 Re-evaluation of BPA

5.　　　In 2016, the CEP Panel received a new mandate from the European Commission which stated 'In accordance with Article 29(1)(a) of Regulation (EC) No 178/20024, the European Commission asks EFSA to:

- establish a protocol detailing the criteria for new study inclusion and for toxicological evidence appraisal for the re-evaluation of BPA, to ensure an efficient and transparent re-assessment of BPA.
- re-evaluate the risks to public health related to the presence of BPA in foodstuffs.

In particular, the re-evaluation should take into consideration new data available from the results of the US National Toxicology Program (NTP)/Food and Drug Administration (FDA) study due in 2017 as well as all other new available information not previously evaluated by EFSA and which fulfil the criteria laid down in an established protocol. This re-evaluation should seek to clarify 410 the

remaining uncertainties concerning the toxicological endpoints of BPA, especially those concerning the mammary gland, reproductive, metabolic, neurobehavioural and immune systems and to establish a full tolerable daily intake (TDI on the basis of the new information available.'

6.	In 2017, a BPA hazard assessment protocol was published following public consultation; this was not commented on by the Committee. It was stated that the new methodology would be tested on a selection of papers assessed during the 2015 review. This testing phase would ensure that the methodology used for the 2015 BPA opinion and 2016 statement on immunotoxicity (EFSA CEF Panel, 2016) was robust, even though not as structured as the new one.

7.	The second part of the mandate was the re-evaluation of new scientific evidence dating from 2013-October 2019 and whether this supported the tTDI. The evaluation covered:

1)	The adverse effects in humans associated with the exposure to BPA via any route;

2)	the adverse effects in animals after:

- oral exposure to BPA at doses equal or below the cut-off of 10 mg/kg bw per day (based on the benchmark dose lower confidence interval (BMDL10) used by the EFSA CEF Panel to set the t-TDI in 2015)
- other exposure routes [subcutaneous (s.c.), intraperitoneal (i.p.), intravenous (i.v.), inhalation and intratesticular] at doses equal or below the cut-off of 10 mg/kg bw per day, when converted to an oral dose, taking into account the interspecies kinetics differences (see toxicokinetics- Chapter 3.1.1.5 of the opinion). No cut-off was applied for dermal studies.

For point 2, when all the doses in one study converted from other routes to oral will result in a dose above the oral cut-off of 10 mg/kg bw per day, the study will be excluded from every step of the assessment.

3)	the human and animal toxicokinetics of BPA.

Methods used in the re-evaluation

Population

8.	The target population of the hazard assessment was the general EU population, including specific vulnerable groups (embryos, fetuses and infants).

The target chemical substance was BPA; BPA derivatives were not included.

Health Outcome Categories

9.　　　Any endpoint was considered potentially relevant for the assessment and a similar categorisation system of Health Outcome Categories, as used in the EFSA opinion of 2015, was used in the new review with the categories being as follows:

- General toxicity (e.g. liver and kidney),
- Reproductive and developmental,
- Neurotoxicity and neurodevelopmental toxicity,
- Immunotoxicity,
- Metabolic effects,
- Cardiotoxicity,
- Carcinogenicity and mammary gland proliferative effects, and
- Genotoxicity.

In addition, toxicokinetic aspects of BPA were examined.

10.　　　Within each Health Outcome Category (HOC), clusters were identified that included several toxicologically relevant endpoints that are physiologically or toxicologically related, and that together shed light on the likelihood of an effect of BPA exposure in that cluster. The endpoints are measures of an individual parameter that is adverse in itself, i.e. an apical endpoint, or that might be involved in the development of an adverse condition, i.e. an intermediate endpoint. For example, the HOC General Toxicity would have a cluster Liver toxicity and within that the endpoints ALT, AST and gamma-glutamyl transpeptidase (γ-GTP).

11.　　　Where the endpoints were a distinct disease entity such as obesity, these were identified as a core cluster element. Additional relevant endpoints were identified within this such as BMI, leptin or adiponectin levels. The chapter in the International Classification of Diseases to which this particular disease belonged (e.g. Endocrine, nutritional and metabolic diseases) was identified, and a relevant HOC with the same name was created. These HOCs were linked with a Toxicological Effect Category whenever appropriate. Within this group all the identified disease entities were included as separate clusters (e.g. Diabetes Mellitus; Thyroid Gland Disorders; etc. These were then merged with the HOC tree.

12.　　The clusters were formed taking the timing of exposure into account (e.g during pregnancy, childhood or adulthood.

13.　　It was noted that there were problems with exposure assessment, such that there were no studies which measured BPA exposure in a way appropriate to the assessment and that the assessment was focussed on identifying adverse effects associated with BPA exposure. It was therefore decided to conduct Weight of Evidence (WoE) analysis only on those clusters/exposure periods for which at least two studies were available and of these at least one study reported a statistically significant effect.

Search and study selection

14.　　The evaluation was based on studies published between 1 January 2013 to 15 October 2018. For genotoxicity evidence, however, this was extended to 21 July 2021. The methods used for data collection (including the call for data) are described in chapter 3.2 of Annex A of the opinion and in Appendix A to that Annex. Following the literature search the records were uploaded into DistillerSR.

15.　　Study selection (chapter 4, Annex A). Screening of titles and abstracts for relevance to humans, animals or Mode of action was performed by two researchers working independently. The full text of the relevant studies then underwent a second screening by two researchers working independently. This was also the first step in categorising the studies into separate HOCs. The inclusion and exclusion criteria are set out in Tables 2,3 and 4 of Annex A.

16.　　The outcome of the literature searches is given in the Table 2 of the main opinion and reproduced below:

Records identified	Through database searching, n = 13,636
	Through call for data, n = 7
Title and abstract screening	n = 13,643
Full text screening	n = 3,231

Appraisal	Animal General toxicity, n = 54
	Animal Immunotoxicity, n = 42
	Animal Carcinogenicity and mammary gland proliferative effects, n = 46
	Animal Metabolic effects, n = 82
	Animal Neurotoxicity and developmental neurotoxicity, n = 94
	Animal Reproductive and developmental toxicity, n = 153
	Animal Cardiotoxicity, n = 22
	Human Case–control, n = 26
	Human Cohort, n = 99
Extraction	Animal studies, n = 298
	Human case–control and cohort studies, n = 105
Narrative review	Animal Mode of Action studies, n = 288 In vitro Mode of Action studies, n = 310 Human Mode of Action studies, n = 33 Human Mode of Action studies, n = 33 Human Cross-sectional studies, n = 177

17.   Data extraction is described in chapter 5 of Annex A to the opinion.

Identification of endpoints relevant to the hazard assessment.

18.   Functionally interrelated endpoints from human and animal studies were grouped in clusters for assessing in WoE analysis. The identification of the relevant endpoints for the weight of the evidence was identified based on the following criteria:

1)   Endpoints identified as key in the 2015 EFSA Scientific opinion:

- Human studies: endpoints assessed at least as As Likely as Not (ALAN) in the WoE), belonging to the HOCs developmental and reproductive effects, neurological, neurodevelopmental and neuroendocrine effects, immune effects, cardiovascular effects and metabolic effects.
- Animal studies: endpoints from Section 3.2.5 of the 2015 opinion (not included in the uncertainty analysis tables) for the HOC general toxicity, and from Section 4.3.2 of the 2015 EFSA Scientific opinion (included in the uncertainty analysis tables) for the HOCs mammary gland proliferation, carcinogenicity, reproductive toxicity, neurotoxicity, immunotoxicity, metabolic effects.

2) Endpoints identified in the current assessment:

- Human studies: endpoints belonging to relevant clusters, i.e. clusters composed of at least two studies, one of them showing a statistically significant effect for one of the endpoints measured;
- Animal studies: endpoints identified as statistically significant in at least one Tier 1 or Tier 2 study.

19.   In order to be considered and assessed in the WoE analysis, the relevant endpoints identified from the animal studies also needed to be expressed quantitatively and in a relevant animal model. Moreover, Tier 3 studies (see below) containing relevant endpoints but with less than three doses (control + 3 BPA doses) were excluded from the WoE. All endpoints were considered adverse unless otherwise stated. Further details are given in chapter 2.5, Annex A.

Internal validity

20.   The internal validity of the studies was assessed and a Tier rating determined to indicate the quality of the study. It is explained that internal validity relates to whether a study answers its research question "correctly", that is, in a manner free from bias. The BPA protocol considered two aspects of this,

(i)   those that introduced a systematic difference between the control and the exposed group only (e.g. non-randomised allocation of animals to study groups); and

(ii)   (ii) those potentially affected to the same extent the control and exposed study groups (e.g. the reliability of the method used to test the outcome).

21.   A structured approach was used to appraise the internal validity of human epidemiological and experimental animal studies, whereas for Mode of Action (MoA) studies a narrative approach was used. The internal validity of human and animal studies was evaluated by study design and by endpoint according to step 4 of the NTP Handbook for Conducting a Literature-Based Health Assessment using the Oral Health Assessment Tool (OHAT) Approach for Systematic Review and Evidence Integration (NTP-OHAT, 2015).

22.   Each of the studies was scored from ++, +,- , -/not reported or -- covering the range definitely low to definitely high risk of bias; this was done by two independent researchers in series, discussing any disagreement to reach consensus. The questions asked to assess internal validity were are set out in Tables 8 and 9 for human and animal studies respectively (chapters 6.1 and 6.2 of Annex A to the opinion) and reproduced below.

Table 8: Internal validity appraisal tool for human data (case–control and cohort study design) (adapted from NTP-OHAT, 2015)

#	Key Q	Question	Domain	Rating (++, +, -, --)
1	A	Did selection of study participants result in appropriate comparison groups?	Selection	
2		Were outcome data completely reported without attrition or exclusion of experimental units from analysis?	Attrition	
3	B	Can we be confident in the exposure characterisation?	Detection	
4	C	Can we be confident in the outcome assessment?	Detection	
5	D	Did the study design or analysis account for important confounding and modifying variables?	Confounding	
6		Were all measured outcomes reported?	Selective reporting	
7	E	Do the statistical methods seem appropriate?	Other sources of bias	

Table 9: Internal validity tool for experimental animal studies (adapted from NTP-OHAT, 2015)

#	Key Q	Question	Domain	Rating (++, +, -, --)
1		Was the administered dose or exposure level adequately randomised?	Selection	
2		Was the allocation to study group adequately concealed	Selection	
3		Were the experimental conditions identical across study groups?	Performance	
4		Were outcome data completely reported without attrition or exclusion from analysis?	Attrition	
5		Can we be confident in the exposure characterisation?	Detection	
	A	Sub-question: Did the test compound contain any impurities?		
6		Can we be confident in the outcome assessment?	Detection	
	B	Sub-question: Were the outcome assessors adequately blinded to the study group?		
7		Were all measured outcomes reported?	Selective reporting	
8		Were the statistical methods and the number of animals per dose group appropriate?	Other sources of bias	
	C	Sub-question: Was the number of animals per dose group appropriate?		

23.   This allowed the studies to be graded as Tier 1, 2 or 3 as follows.

24.   For human data:

- Tier 1: all key questions scored ++/+ and no more than one non key question scored - and no non key questions scored --.
- Tier 2: any combination not Tier 1 or 3.
- Tier 3 : any key questions scored --/- or any non-key question scored –.

More detail is given in appendices B1 and B2 of Annex A to the opinion.

25.   For animal data;

- Tier 1: all key sub questions scored ++/+ and no more than one question scored -.
- Tier 2: any combination not Tier 1 or 3.
- Tier 3 : any key sub questions scored --/- or more than 4 questions scored --.

26.   A tiered approach was taken starting with question 5 (see Table 9 above). If it was considered that there was a high risk of bias, the study was classed as Tier 3 and the evaluation stopped. If not, the evaluation moved on to question 6; again, if it was considered that there was a high risk of bias, the study was classed as Tier 3 and the evaluation stopped and so on to question 8. The

evaluation then moved on to the rest of the questions.

27.   For genotoxicity a specific internal validity approach was used.

External validity

28.   In the BPA protocol, external validity is the relevance for human health of measuring a given endpoint in a given animal model. The assessors were asked to consider whether the specific endpoints measured in a specific animal model would be relevant to humans. Thus, animal models differing from humans in terms of target anatomical or pathophysiological features for the chemical under investigation would not be considered relevant.

Weight of Evidence analysis

29.   The WoE analysis is described in Chapter 8 of Annex A.

30.   Following the appraisal of the individual human and experimental animal studies for internal and external validity (only for the animal studies), the experts evaluated the confidence in the overall body of evidence by clusters of endpoints for each HOC.

31.   Within the WoE it was taken into account whether an endpoint could be considered as apical (e.g. breast cancer) or as intermediate (e.g. mammary gland proliferation/hyperplasia). An apical endpoint is defines as an observable outcome in a whole organism, such as a clinical sign or pathologic state, which is indicative of a disease state that could result from exposure to a toxicant. Intermediate endpoints were considered to be events occurring at a step between the molecular initiating event and the apical outcome: they were toxicologically relevant to the apical outcome (a necessary element of the MoA or a biomarker of effect and were experimentally quantifiable.

32.   There were two options for synthesising the data – meta analysis on a particular endpoint or, where the conditions for meta-analysis were not met, data were extracted in summary tables containing the appropriate information (set out in Tables 5 and 6 of the annex) for all the studies containing the relevant endpoint and grouped by HOC and cluster. This allowed easy visualisation of the data. A specific approach was taken to the genotoxicity HOC.

33.   Confidence ratings in the overall body of evidence were reached by assessing the weight of relevant clusters (in human) or of clusters of relevant endpoints (in animals) per different exposure categories. For the epidemiological

studies, these categories were:

- 'Exposure during pregnancy',
- 'Exposure during childhood'
- 'Exposure during adulthood'.

For the animal studies:

- 'Developmental exposure (pre-natal and/or post-natal until weaning)',
- 'Developmental and adult exposure (pre-natal and post-natal in pups until adulthood)',
- 'Growth phase/young age',
- 'Adult exposure (after puberty)',
- Indirect (germline) exposure'

34.   Where exposure was via the dams or sires, effects on the F1, F2 and F3 generations were considered separately, as their exposure period was different.

35.   The protocol states that biological plausibility is a fundamental concept, with concordance increasing confidence in the body of evidence. Where there was no concordance, priority was given to apical endpoints. Since in all study types, the apical endpoints were generally considered to be the most direct, or applicable, to the assessment of the health outcome (e.g. incidence of cancer of the mammary gland). However, in some cases, intermediate endpoints may be as decisive as apical endpoints.

36.   *In vitro* and mechanistic data were used to support the evidence for intermediate endpoints in qualitative terms but were used in deriving the conclusions on hazard identification but not the WoE.

37.   The likelihood of a health effect in the overall body of evidence was evaluated using a modified version of step 5 of the NTP Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration (NTP-OHAT, 2015) and of the VKM risk assessment of energy drinks and caffeine (VKM, 2019).

38.   The studies on a specific cluster were grouped according to study design features. As detailed in the NTP-OHAT (2015) an initial confidence rating of human and animal studies was assigned on the basis of the study design and its intrinsic ability to potentially set up an association between exposure to a substance and a subsequent effect. The following four descriptors were used to determine this initial level of confidence:

- Controlled exposure conditions.
- Exposure preceding the effect onset.
- Outcome being assessed at individual level.
- Presence of an appropriate comparison group.

Fulfilment of all features would receive an initial rating of high confidence (++++). Lower ratings, i.e. moderate (+++), low (++) or very low (+), correspond to the number of features fulfilled.

39.   For experimental animal studies, the initial confidence was rated high (++++), as this design ensured the fulfilment of all the four key study features listed above. For cohort and case–control human studies the initial confidence was rated moderate (+++) as the 'controlled exposure conditions' feature was not fulfilled. Considerations on whether the exposure preceded the outcome were carried out at internal validity level of the process, thus resulting in considering this aspect as fulfilled.

40.   The grouped studies were then assessed for elements which would up or downgrade the likelihood of a health effect at study level and then overall at the body of evidence level. The considerations were similar to those described in the NTP-OHAT tool (NTP-OHAT, 2015) and in Balshem et al. (2011) and were applied in deciding whether to downgrade or upgrade the likelihood of a health effect.

41.   In brief, on a cluster basis the following elements were considered to downgrade the initial confidence in a body of evidence.

- internal validity;
- external validity (for animal studies only);
- unexplained inconsistency;
- imprecision (for human studies only).

42.   Conversely the following elements were considered for upgrading the confidence in the body of evidence.

- effect size (in human studies only);
- dose–response;
- consistency across study design type/dissimilar populations/animal models or species (at body of evidence level only);
- residual confounding (this applies mainly to human observational studies. If a study reports an effect or association despite the presence of residual confounding, confidence in the association was increased).

43.   Downgrading confidence only occurred when there were serious limitations in a study or body of evidence.

44.   The likelihood of a health effect was assessed per exposure period taking into account the internal and external validity (for animal studies only) of the different studies. This is set out in Table 11 of Annex A, which was adapted from the NTP-OHAT, 2015 evidence profile table. After potential upgrading or downgrading, five judgements on the likelihood of a health effect were possible:

- Very Likely: There is very high confidence in the body of evidence for an association between exposure to the substance and health effect/s (e.g. there is much evidence showing consistent effect/s).
- Likely: There is high confidence in the body of evidence for an association between exposure to the substance and health effect/s (e.g. there is evidence showing consistent effects).
- As Likely As Not (ALAN): There is low confidence in the body of evidence for an association between exposure to the substance and health effect/s (e.g. there is evidence showing inconsistent effects).
- Not Likely: There is very low confidence in the body of evidence for an association between exposure to the substance and health effect/s (e.g. there is evidence showing consistent no effects).
- Inadequate evidence: There is insufficient evidence available to assess if the exposure to the substance is associated with and health effect/s or data are missing.

45.   The likelihood levels were scored by the experts taking into account the elements above.

46.   A number of scenarios are described (lines 790- 809 of Annex A) setting out how the Tier rating of different studies were taken into account in determining the likelihood levels.

47.   The final likelihood level for a cluster was determined by the highest likelihood level in the exposure periods in the cluster.

48.   Table 12 (p27) of the Annex is the template used to grade confidence in the body of evidence and is reproduced below:

Table 12 Template for grading confidence in the body of evidence on the likelihood of a health effect per cluster in animal studies (adapted from NTP-OHAT, 2015 and VKM Risk assessment of energy drinks and caffeine (VKM, 2019))

Cluster /Exposure time	RefID	Species	Endpoint measured	Internal validity	External validity (animal model)	External validity (Endpoints)	Is there an effect?	Is there a MDR or a NMDR or is a single-dose study?	Unexplained inconsistency of the results between different studies (per endpoint)	Likelihood of effect (per endpoint) rationale	Unexplained inconsistency of the results between different studies (per cluster)	Likelihood of effect (per cluster), rationale
RefID n – Citation 1												
RefID n – Citation 2												

49.   To complete this, the assessors followed the indications set out on page 28 of the Annex.

50.   The integration of the human and animal evidence was done at a cluster level, the overview is given in figure 1 of Annex A, as below

Figure 1. Integration of human and animal evidence for the final assessment of a health effect.

Human evidence						
	Very likely	VL	VL	VL	VL	VL
	Likely	L	L	L	L	VL
	ALAN	ALAN	ALAN	ALAN	L	VL
	Not Likely	Not Likely	Not Likely	ALAN	L	VL
	Inadequate evidence	Non-classifiable	Not Likely	ALAN	L	VL
		Inadequate evidence	**Not Likely**	**ALAN**	**Likely**	**Very Likely**
		Animal evidence				

51.   The highest level of evidence for a health effect among the different exposure periods within a cluster was considered as the likelihood of a health effect for the whole cluster. The level of evidence for a health effect at cluster level resulting from the human evidence stream was combined with that deriving from the animal evidence stream to reach a single hazard identification conclusion (using a process adapted from step 7 of the NTP-OHAT Handbook, NTP-OHAT, 2015).

52.   The highest level of evidence for a health effect among the different exposure periods within a cluster was considered as the likelihood of a health effect for the whole cluster. The level of evidence for a health effect at cluster level resulting from the human evidence stream was combined with that deriving from the animal evidence stream to reach a single hazard identification conclusion (using a process adapted from step 7 of the NTP-OHAT Handbook, NTP-OHAT, 2015).

53.   For hazard characterisation, BMD analysis was then performed on all "likely" or "very likely" effects using human and/or experimental animal studies.

Low Dose Effects

54.   One area of interest in the assessment was the possibility of low dose effects. This is not clearly defined but was takes as < 5 mg/kg bw/day in line with the 2015 EFSA review. This level was consistent with the No Observed Adverse Effect Level NOAEL established by EFSA in 2007, based on changes in body and organ weights in a 2 generation reproductive toxicity study. This dose range was covered in a number of studies in the assessment including the CLARITY study.

55.   The opinion also noted the interests in non-monotonic dose responses. The Panel followed the recommendations of the 2018 EFSA Working Group on the implications of non-monotonic response on risk assessment (DN. COT comments) and applied their criteria to establish whether there was evidence non monotonic dose responses (NMDR) in individual studies or the overall body of evidence. These are set out in section 2.3.3 of the opinion. The studies where either the authors or the CEP panel considered that there were indications of NMDR are given in Table 3 of the opinion with the study by Montévil et al., 2020 discussed in more detail in Appendix B. However, overall the evidence for NMDR in this study was considered weak and inconclusive.

The CLARITY study

56.   Part of the mandate received by EFSA from the European Commission for the re-evaluation of BPA included an evaluation of the data coming from the US National Toxicology Program (NTP) research programme, called Consortium Linking Academic and Regulatory Insights on BPA toxicity (CLARITY-BPA). In particular, it was stated that that the re-evaluation should take into consideration new data available from the results of the US NTP/FDA study (The CLARITY-BPA Core Study: A Perinatal and Chronic Extended-Dose-Range Study of Bisphenol A in Rats), published in October 2018.

57.		The CLARITY-BPA program has two components:

1)	Core Study: A two-year guideline-compliant study of potential BPA toxicity in rats.

2)	Grantee Studies: Investigational studies conducted by university researchers testing a range of additional endpoints.

The Core Study tested potential BPA toxicity in rodents; findings were published in Camacho et al. (2019).

58.		The Clarity study was a 2 year study NTP based study in which NCTR Sprague-Dawley rats were given doses of BPA (0, 2.5, 25, 250, 2,500,and 25,000 µg/kg body weight (bw)/day) by gavage in a 0.3% carboxymethylcellulose vehicle (Camacho *et al*, 2019). The rats were dosed from gestation day (GD) 6 through to the start of parturition and then directly to pups from the day after birth until either postnatal day 21 (the stop-dose arm) or continuously until termination at one (interim sacrifice) or two years. The stop-dose arm was included to assess the potential for any BPA effects that were due to developmental exposure.

59.		It was reported that no BPA-related effects were evident in the in-life and non-histopathology data. Neoplastic and non-neoplastic lesions were observed in both females and males but were common age-associated lesions that were variable across control and BPA-treated groups. The lack of consistent responses within the continuous- and stop-dose arms within and across tissues suggested that these might not be plausibly relationship to BPA treatment. There was a possible relationship between the increased incidences of lesions in the female reproductive tract and the male pituitary and exposure in the 25,000 µg BPA/kg bw/day dose group.

HOC Genotoxicity

60.		A specific approach was taken to genotoxicity. In the assessment of genotoxicity, the CEP panel examined whether new data from the published literature could provide new evidence on the potential genotoxicity of BPA. A literature search was performed as described in Annex A to the opinion. The references from the previous (2015) CEF Panel opinion were included in the assessment using the same appraisal criteria applied to the newly published data and considering the EFSA Scientific Committee guidance documents on genotoxicity published after 2015 (EFSA, 2017, EFSA, 2021b).

61.		The genotoxicity studies considered for this assessment were:

- in vitro and in vivo studies (88 publications) retrieved from the literature search.
- in vitro and in vivo studies (15 publications) considered in the 2015 opinion.

62.   In vitro and in vivo studies were grouped based on the genotoxicity endpoint investigated:

- gene mutations (e.g. bacterial reverse mutation assay);
- chromosomal damage (CA and micronucleus assays);
- DNA damage (comet assay).

63.   The studies were summarized into synoptic tables which are presented in Annex L to the opinion.

64.   The studies were assessed for reliability using a scoring system based on criteria published by Klimisch et al. (1997) as explained in Chapter 2.3.5. In assigning the reliability score, compliance with the Organization for European Economic Cooperation and Development (OECD) Test Guidelines (TGs) or standardized methodology and the completeness of the reporting as detailed below were considered. The reliability scores were:

1) reliable without restriction

2) reliable with restrictions

3) insufficient reliability

4) reliability cannot be evaluated

5) reliability not evaluated, since the study is not relevant and/or not required for the risk assessment (in case the study is reported for reasons of transparency only).

Fuller explanations of these scores are given in pages 32 -33 of the opinion. The summary Tables in Annex L contain a box giving the reliability score and com ments justifying it. In a second step, the relevance (high, limited, low) of the study results was assessed based mainly but not exclusively on the following criteria (taken from Chapter 2.3.5).

- Genetic endpoint (high relevance for gene mutations, structural and numerical chromosomal alterations as well as results obtained in an in vivo comet assay, which belongs to the assays recommended by the EFSA Scientific Committee (2011) for the follow-up of a positive in vitro result;

lower relevance for other genotoxic effects). Other test systems although potentially considered of limited or low relevance may provide useful supporting information.

- Route of administration (e.g. oral vs. intravenous, intraperitoneal injection, subcutaneous injection, inhalation exposure) in case of in vivo studies.
- Status of validation (e.g. for which an OECD TG exists or is in the course of development, internationally recommended protocol, validation at national level only, no validation).
- Reliability and relevance of the test system/test design irrespectively of whether a study has been conducted in compliance with GLP or not.
- Information on BPA purity grade and/or the supplier. If only the supplier was available, the company's website was consulted to retrieve the purity grade, or the authors were contacted to ask for it. If none of the two information were reported or obtained, the relevance was considered low and the study was excluded from the WoE assessment.

65.   Genotoxicity studies evaluated as of low relevance were not considered in the assessment.

66.   Studies not investigating classical genotoxicity endpoints (e.g. γH2AX, oxidative DNA damage, DNA binding, ROS generation) and studies in humans are considered in the MoA and as supportive evidence. All the studies evaluated were summarised in a narrative form (Appendix E).

67.   In the summary tables, the studies are grouped based on genetic endpoints or test systems and arranged chronologically. The results were evaluated and presented as positive, negative, equivocal or inconclusive. If considered relevant for the interpretation of the genotoxicity endpoints, non-genotoxicity endpoints (e.g. reactive oxygen species (ROS) production) were reported in a narrative way only, but the results were not classified as "positive" or "negative." Descriptive summaries of the genotoxicity studies are given in Appendix E to the opinion.

Weight of evidence- genotoxicity

68.   The WoE approach taken by the panel for the evaluation and interpretation of the genotoxicity data took into account not only the quality and availability of the data on genotoxicity, but other relevant data; these include data on MoA and on toxicokinetics when available. The main steps of the WoE approach used were:

- Assembling of the evidence into lines of evidence of similar type. In a first step, the CEP Panel evaluated all available in vitro and in vivo studies addressing the three main endpoints of genotoxicity: gene mutations, structural and numerical chromosomal aberrations in addition to DNA damage endpoint (evaluated by Comet assay). Only the studies of high and limited relevance were included.

Studies investigating the BPA MoA were considered, e.g. DNA oxidation, ROS production (when genotoxicity was also investigated in the same study), DNA binding, interference with proteins involved in chromosome segregation during cell division, modulation of expression of genes involved in DNA repair or in chromosome segregation and markers of DNA double strand breaks (DSBs) (e.g. γH2AX). It was considered that evidence from the mechanistic studies could support the lines of evidence for the genotoxicity endpoints.

- Weighting of the evidence. The CEP panel considered that a quantitative method to weight the evidence was not appropriate due to the quantity and heterogeneity of the evidence to be integrated. A qualitative method based on expert judgment was therefore applied. All studies evaluated for reliability and relevance (as described above) were listed in the synoptic tables in Annex L. The evaluation of the studies of high and limited relevance was described in the opinion, including the conclusion for each line of evidence. The consistency of the evidence was assessed and presented in the opinion.
- Integrating all the evidence. The lines of evidence of the above genotoxicity endpoints were assessed separately. To elucidate the MoA of BPA, mechanistic studies were considered. Integrating evidence from the MoA with lines of evidence from genotoxicity endpoints allows a reduction in the uncertainty on the potential genotoxicity. In case genotoxic effects were observed, evidence from the MoA may allow clarification if the genotoxicity is due to a direct or indirect mechanism.

Conclusions

69.   The conclusions drawn by the EFSA panel are set out in section 4.1 of the opinion and summarised below.

Toxicokinetics

70.   The studies in mice and rats did not contribute to a better understanding of the toxicokinetic aspects of BPA. The studies in ewes showed that the absolute bioavailability was lower when BPA was given by nasogastric tubing compared with BPA administration via pellets. It was considered that this finding was most probably explained by the buccal absorption of BPA.

71.   The human data showed that nearly 100% of BPA is absorbed and undergoes significant pre-systemic metabolism to glucuronide and sulfate conjugates. The concentration in the systemic circulation is low. However, the dose-corrected Area Under the Curve (AUC)s were clearly different in the two studies. The most probable explanation for this was that in the study with the higher AUC values, the contact time with the buccal mucosa could be longer as the BPA was administered in cookies versus BPA in soup). The CEP Panel decided to use the median value of the AUCs from both studies for the calculation of the HEDF, because both modes of administration were realistic for humans. The median value was 15.7 nM × h, which is 4-fold higher than the modelled AUC value used for calculating the HEDF in the 2015 EFSA opinion.

72.   To calculate the HEDF, the AUC data were used from the 2015 EFSA opinion for mice, rats, monkeys and dogs. For ewes, the data reported in the current opinion were used. The following HEDFs were obtained: 0.0115 for mice, 0.165 for rats, 0.095 for monkeys, 0.1395 12156 for dogs, 0.1197 for ewes (gavage) and 0.4357 for ewes (diet). Specific factors were applied to convert the doses from studies in which BPA was given by routes other than oral to allow to the doses to be compared.

General toxicity

73.   The newly available literature data indicate that in the General toxicity HOC several organs are potential targets for BPA toxicity and that haematological parameters can be affected.

74.   No human studies were available in this cluster, while 10 clusters with relevant endpoints were identified in animal studies. These were: body weight, liver effects, kidney effects, lung effects, thyroid effects, parathyroid effects, pituitary gland effects, adrenal gland effects, bone marrow effects and effects on haematological parameters.

75.   Overall, none of the evaluated clusters' effects was considered Very Likely or Likely. In each of the evaluated clusters, effects were noted at least in one exposure period, but there were less consistent results among the available

studies and, therefore, these effects were judged as ALAN in all the clusters.

76.　　MoA studies suggested oxidative stress as a potential pathogenetic mechanism for kidney damage. Similarly, oxidative stress in liver cells may be related to impaired mitochondrial function and liver toxicity. MoA studies also suggested that epigenetic changes via DNA methylation may affect different signalling pathways related to lipid and carbohydrate metabolism. MoA studies in the lungs suggested that BPA can delay fetal lung maturation as indicated by reduced alveolar airspace and thickened septa. Both these findings may be related to an increase in the lung weight. MoA studies on thyroid cells suggested mechanisms responsible for an increase in proliferation, supporting the limited evidence of hyperplastic changes observed in the animal studies. Moreover, it was suggested that BPA could enhance the susceptibility to thyroid carcinoma in combination with other endogenous or external factors.

77.　　No studies taken forward for BMD analysis

Immunotoxicity

78.　　The newly available data from the literature indicate that the immune system is a target for BPA toxicity.

79.　　Within the Immunotoxicity HOC, one relevant cluster of endpoints was identified in the human studies. This was asthma/allergy, and included data from the exposure periods pregnancy and childhood.

80.　　In the animal studies, five clusters of relevant endpoints were identified: innate immunity, cellular immunity, humoral immunity, inflammation and allergic lung inflammation.

81.　　Based on the human data, a positive association between BPA exposure and asthma/allergy was judged as ALAN.

82.　　Based on the animal data, the clusters cellular immunity and allergic lung inflammation showed effects that were judged as Likely.

83.　　In the other clusters, effects were also noted, but there were fewer consistent results, and these effects were judged as ALAN. In the allergic lung inflammation cluster, the effect noted was the production of specific IgE in response to an allergen. This was deemed to be adverse as it is a crucial parameter in inducing allergic reactions in the respiratory tract. Other effects in that cluster supported the likelihood of this effect. The likely effect in the cluster

cellular immunity was supported by the consistency of the different endpoints within that cluster.

84.   The most sensitive parameter affected by BPA was the increased number of Th17 cells. Although Th17 cells are T cells, and therefore were put in the cluster cellular immunity, they play a role in allergic responses, and therefore the effect on Th17 cells is consistent with the effect on specific IgE noted above. In vivo evidence was supported by MoA studies. In vitro studies indicated the ability of BPA to induce immune deregulation, increasing susceptibility to develop inflammatory diseases. Th17 cells are a specific subset of CD4+ T helper cells, which participate in various immune diseases, including asthma and autoimmune diseases. Potential mechanisms by which BPA may contribute to immune-mediated disorders included modulation of ERK1/2 phosphorylation, NF-kB activation, modulation of estrogen receptors, Glucocorticoid Receptors and androgen receptors as well as cytokine/chemokine secretion, and oxidative stress. Effects may be on non-specific cells belonging to the immune system or influencing the immune system (such as APCs and epithelial cells). This will, through presentation of antigens to T-lymphocytes or release of mediators, influence the regulatory homeostasis of the immune system, suppressing T regulatory cells and stimulating Th-17 cells. Thus, BPA appeared to promote multiple interwoven pathways involved in immune deregulation that may play a role in immune related disorders.

85.   This was taken forward for BMD analysis.

Metabolic effects

86.   The newly available literature data indicate that BPA may induce adverse metabolic effects.

87.   Within the HOC Metabolic effects, five clusters of endpoints were identified in the human studies: These were obesity, cardiometabolic effects, thyroid effects, Type 2 Diabetes mellitus and gestational diabetes mellitus, including data from one or more of the exposure periods pregnancy, childhood and adulthood.

88.   In the animal studies, eight clusters of relevant endpoints were identified: obesity, fat deposition in the liver, glucose regulation, blood lipids, uric acid, Type 1 Diabetes mellitus, other metabolic hormones and thyroid hormones. The clusters included data from one or more of the exposure periods developmental until weaning, developmental until adulthood, growth phase, adult

exposure and indirect (germline) exposure.

89.	Based on the human data, none of the metabolic clusters showed effects that were considered Likely or Very Likely. A positive association between BPA exposure and obesity and Type 2 Diabetes mellitus was judged as ALAN, while a positive association between BPA exposure and cardiometabolic effects, thyroid effects and gestational diabetes mellitus was judged as Not Likely.

90.	Based on the animal data, no metabolic clusters were considered Very Likely. However, the cluster uric acid was considered Likely (in the adult exposure period), as increased levels of uric acid were observed in the liver of mice and in the serum of mice and rats after BPA exposure. The other metabolic endpoints were considered either ALAN (obesity, fat deposition in the liver, glucose regulation, blood lipids and Type 1 Diabetes Mellitus) or Not Likely (other metabolic hormones and thyroid hormones), in one or more exposure periods.

91.	There are substantial amounts of supporting evidence for plausible MoAs of BPA were available on obesity, fat deposition in the liver and glucose regulation, mostly from animal and in vitro studies. The MoA data in animals showed that BPA could increase the formation of hepatic uric acid by increasing the activity of the enzyme xanthinoxidase, which catalyses the conversion of the purines hypoxanthine and xanthine into uric acid. The MoA data on Type 1 Diabetes Mellitus were very limited and the results depended on the animal model used.

92.	Urate was taken forward for BMD analysis.

Neurotoxicity and Developmental Neurotoxicity

93.	 The newly available literature data indicate that the central nervous system is a target for BPA toxicity.

94.	Within the HOC Neurotoxicity and developmental neurotoxicity, the evaluation of the human data considered endpoints from the cluster neurodevelopment. In the animal studies, three clusters of endpoints were identified: neuromorphology, nervous system functionality and behaviour.

95.	Based on the human data, it was concluded that the evidence for an association between BPA exposure and impaired neurodevelopment was Not Likely.

96.   Based on the animal data, all three neurotoxicity clusters showed effects that were judged as Likely:

- In the neuromorphology cluster, Likely effects were found for the endpoints dendritic spine density of pyramidal cells in hippocampus (CA1 and dentate gyrus areas) after developmental exposure and for the endpoints number of neurons in hippocampus (CA1 and CA3 areas), and dendritic spine density in pyramidal cells in the medial part of the PFC after exposure during the growth phase/young age.
- In the nervous system functionality cluster, a Likely effect on the endpoint AChE activity during the adult exposure period was identified.
- In the behaviour cluster, Likely effects were noted for the endpoint anxiety/emotionality during all exposure periods (developmental, growth phase/young age, adult and exposure through the male germline). Furthermore, the endpoint learning/memory showed a Likely influence of BPA from developmental and growth phase/young age exposure, and effects on sensory-motor coordination and salt preference were considered Likely in adults.

97.   It was considered that the mechanisms of action that link the identified effects of BPA on various endpoints of brain structure, function and development have not been sufficiently explored in the literature to draw conclusions. There is evidence for the involvement of steroid-hormone-dependent pathways (oestrogen, androgens, corticosterone); oxidative stress, mitochondrial function and calcium regulation; gene expression changes through DNA methylation and other signalling pathways (canonical and non-canonical Wnt pathways, kinases).

98.   This was taken forward for BMD analysis.

Reproductive and Developmental Toxicity

99.   The Panel considered that the newly available literature data indicate that the reproductive system is a target for BPA toxicity.

100.   Within the Reproductive and Developmental toxicity HOC, five relevant clusters of endpoints were identified in the human studies. These were fetal and post-natal growth, prematurity, pre-eclampsia, male fertility and female fertility, including data from one or more of the exposure periods pregnancy, childhood and adulthood.

101.   In the animal studies, three clusters of relevant endpoints were identified: developmental toxicity, female reproductive toxicity and male reproductive toxicity. The clusters included data from one or more of the exposure periods developmental until weaning, developmental until adulthood, growth phase, adult exposure and indirect (germline) exposure.

102.   Based on the human data, none of the clusters showed effects that were judged as Likely or Very Likely. An association between maternal BPA exposure and impaired pre- and post-natal growth, shorter duration of gestation or preterm delivery, reduced male fertility and pubertal development when exposed during childhood, was judged as Not Likely. An association between BPA exposure and reduced female fertility and pre-eclampsia during adulthood and pubertal development when exposed during pregnancy was judged as ALAN.

103.   Based on the animal data, both female and male reproductive toxicity clusters showed effects that were judged as Likely:

- In the female reproductive toxicity cluster, there were Likely effects on ovary weight and histology and uterus histology after developmental exposure, on ovary histology after developmental and adult exposure, on implantation rate after growth phase/young age exposure and on ovary histology (follicle counts) after adult exposure.
- In the male reproductive toxicity cluster, there were Likely effects on epididymis (exfoliated germ cells and inflammation) after developmental exposure (pre-natal and/or post-natal until adult), on testis histology (decreased seminiferous tubule diameter) after growth phase/young age exposure and on sperm (motility, viability and acrosome reaction) after adult exposure.

104.   In the developmental toxicity cluster, effects were also noted, but the results were less consistent, and were judged as ALAN for the endpoints bone development, mammary gland histology, body weight (in the developmental exposure period), mammary gland weight and mammary gland histology (in the developmental and adult exposure) as well as body weight and age at first oestrus (in the growth phase/young age exposure).

105.   Supporting evidence for plausible MoAs of BPA on reproductive toxicity effects was available. This included estrogen and androgen receptor interactions and associated downstream and cross-stream effects, including epigenetic changes. Other possible mechanisms, including BPA-induced generation of oxidative stress, have been less explored.

106.   This was taken forward for BMD analysis.

Cardiotoxicity

107.   The newly available literature data investigated the cardiovascular system as a target of toxicity for BPA.

108.   Within the human HOC Cardiotoxicity, no case-control or cohort studies were available. Therefore, the evidence for a positive association between BPA exposure and cardiotoxicity in human was considered Inadequate.

109.   In the animal studies, five clusters of relevant endpoints were identified: absolute and relative heart weight, incidence of cardiac lesions, cardiac structural changes (as measured by echocardiography), effects on cardiac function (as measured by echocardiography), blood pressure and atherosclerotic lesions.

110.   Based on the animal studies, the evidence of BPA effects was judged as Not Likely in the majority of the cardiotoxicity clusters, and in few clusters as Inadequate, in one or more exposure periods. Given the functional relationship between the endpoints, the outcome of the WoE was considered biologically plausible.

111.   This was not taken forward for BMD analysis.

Carcinogenicity and Mammary gland proliferative effects

112.   The newly available literature data indicate that, in the HOC Carcinogenicity and mammary gland proliferative effects, the following organs are targets of BPA-induced toxicity: mammary gland, prostate and uterus.

113.   Within the HOC Carcinogenicity and mammary gland proliferative effects, no human studies were available, while five clusters with relevant endpoints were identified in animal studies. These were: mammary gland weight, mammary gland histology, prostate histology, uterus weight and uterus histology. For histology, four subclusters were considered, if available: non-neoplastic changes, pre-neoplastic lesions, neoplastic lesions, proliferation and apoptosis as evaluated by quantitative immunohistochemistry.

114.   The cluster mammary gland weight was judged Not Likely. The clusters mammary gland histology, prostate histology and uterus weight showed effects that were not consistently reported in the available studies and, therefore, these effects were judged as ALAN.

115.   Also, regarding the subclusters linked to lesions in the mammary gland, inconsistencies were noted: in the developmental until weaning exposure period no increase in pre-neoplastic lesions (Not Likely), but a higher incidence in neoplastic lesions (Likely) was observed. In the developmental to adult exposure period an increase in pre-neoplastic lesions (ALAN) was reported but no increase in neoplastic lesions was detected (Not Likely). Therefore, these effects contributed to the overall judgement ALAN in the cluster mammary gland histology.

116.   In the cluster uterus histology, the non-neoplastic changes gland cellular anomalies, squamous metaplasia and cystic endometrial hyperplasia were considered adverse and judged as Likely based on studies with developmental exposure (pre-natal and/or post-natal until weaning) to BPA.

117.   MoA studies in mammary gland addressing epigenetic effects, changes in gene expression and changes in hormone receptor levels suggested various MoAs of BPA possibly involved in the induction of proliferative/morphological changes. Some in vivo studies indicated that stromal-epithelial interactions may play a crucial role in the BPA-induced developmental mammary gland. In vitro studies provided some support for the hypothesis that BPA contributes to a higher susceptibility to mammary gland carcinogenesis. MoA studies on prostate cancer indicated that BPA can enhance the susceptibility to tumorigenesis in rodents co-treated with very high levels of oestradiol and testosterone, while developmental and chronic exposure to BPA without additional sex hormones did not demonstrate a direct tumorigenic effect. In vitro MoA studies on uterine cells indicated that BPA increases the proliferative rate. Data from other in vitro studies suggested that BPA modulates various mechanisms underlying the onset, growth and invasion of uterine tumours. However, the results of rodent studies did not demonstrate a tumorigenic activity of BPA.

Genotoxicity

118.   The analysis of the available literature data indicate that BPA does not induce gene mutations in bacteria. However, BPA induces DNA strand breaks, clastogenic and aneugenic effects in mammalian cells in vitro. Oxidative stress-related mechanism(s) are likely to be involved in this DNA damaging and clastogenic activity.

119.   In contrast with consistent positive in vitro findings, the in vivo findings in several studies with high/limited reliability were inconsistent. The CEP Panel concluded that the evidence does not support an in vivo genotoxic hazard posed

by BPA through direct interaction with DNA.

120.	The CEP Panel concluded that it is unlikely to very unlikely that BPA presents a genotoxic hazard, the causes of which include a direct mechanism, and that the balance of evidence allows a HBGV to be established.

Hazard Characterisation

121.	The Panel used BMD analysis on any endpoints judged Likely or Very Likely.

122.	After conversion of the doses to HED, the CEP Panel selected the lowest BMDL value of 0.93 ng/kg bw per day for the effect of BPA on Th17 cells in mice to be used as a Reference Point (RP) for the risk assessment of BPA.

123.	Based on an assessment of other endpoints that could require an additional UF, the WG's overall probability that no additional UF was needed was in the range 85-87%. The CEP Panel concluded that no additional UF was needed and that a HBGV based on the identified RP is justified.

124.	The CEP Panel applied the UFs for inter-species toxicodynamic difference (2.5) and intra-human variability in toxicokinetics and toxicodynamics (10) and established a TDI of 0.04 ng/kg bw per day.

Risk Characterisation

125.	The comparison of the dietary exposure estimates from the 2015 EFSA opinion with the new TDI showed that both the mean and the 95th percentile dietary exposures in all age groups (including all infant and toddler groups) exceeded the TDI by two to four orders of magnitude.

126.	The CEP Panel is aware that the exposure assessment presented in the 2015 opinion may not fully represent the current dietary exposure. Even considering this uncertainty, since the exceedance was so large, the CEP Panel concluded that there is a health concern from dietary BPA exposure for all age groups of the general population.

Summary and Discussion

127.	The EFSA CEP panel have undertaken a review of the extensive data base on BPA. This was done using a very structured approach to the data. In the absence of appropriate exposure data, a hazard assessment was conducted

based on endpoints grouped into HOCs and clusters, which then underwent WoE analysis. Endpoints deemed likely on Very Likely then underwent BMD analysis.

128.   The approach taken to the data differed from that of the SETE subgroup who did not recommend the use of any prescriptive, generic checklist or numerical scoring approach for quality ranking of studies, as such an approach is likely to be limiting and inflexible. Instead, the document(s) developed by SETE aim to provide guidance for experts and Committees to assess all information and apply good judgment transparently in a weight of evidence approach. Especially for epidemiological data, scoring methods are difficult to replicate, are not transparent to the final user of the risk assessment and do not reflect the usefulness of an individual study. EFSA dismisses studies in their assessment once a question in their WoE approach/checklist is considered to have a high risk of biases. SETE stressed, that even a study that 'scores low' may provide valuable evidence in the context of assessing a particular form of bias. The synthesis of evidence thus requires a broader approach than simply evaluating the quality of each individual study and weighting studies according to this assessment. Instead, it should use the classical considerations for judging causality. Evidence synthesis should thereby consider the entire body of evidence available and not just individual studies in isolation.

129.   In line with EFSA's assessment, SETE considered the main aspects in these considerations are whether or not the data indicate robust evidence of an effect in animals and whether the same effect has been reported in human/epidemiological studies. If the same effect has been reported in both animal and human studies, considerations should be given as to how the effect levels compare.

130.   However, rather than following a checklist and scoring system, if a predominantly positive answer can be given to the main considerations, then the weight of evidence strongly supports causality. However, it is important to establish the strength and robustness of the evidence for each line of evidence and reflect on how the uncertainties may influence the weight of evidence. Taken together these should provide information on how the various lines of evidence influence the overall conclusion, increasing or decreasing the likelihood of a conclusion of causality.

Considerations should be given to whether or not a line of evidence is considered sufficient by itself or provides a significant contribution to the overall weight of evidence. In this, the relative impacts of epidemiological and toxicological evidence are plotted against each other.

131.   Following weight of evidence analysis and BMD modelling, the endpoint that was the basis of the Reference Point was an increase in Th17 cells in mice. This is an intermediate rather than apical endpoint, however, was supported by related evidence in the same cluster of effects. These aspects are considered in accompanying papers.

Questions for the Committee

1.   Do Members have any comments on the methods used by EFSA for a) study selection, b) weight of evidence analysis.

2.   Does the Approach taken fairly reflect the data – are negative findings properly weighed?

3.   Do Members have any comments on the MOC/clusters approach?

4.   Do Members have any comment on the approach taken with regard to exposure assessment ?

5.   Do Members have any comments in general on the use of intermediate endpoints?

References

Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, Falck-Ytter Y, Meerpohl J, Norris S and Guyatt GH, 2011. GRADE guidelines: 3. Rating the quality of evidence. Journal of Clinical Epidemiology, 64(4), 401–406. doi: 10.1016/j.jclinepi.2010.07.015.

Camacho L, Lewis SM, Vanlandingham MM, Olson GR, Davis KJ, Patton R, Twaddle NC, Doerge DR, Churchwell MI, Bryant MS, Mclellen FM, Woodling K, Felton RP, Maisha MP, Juliar BE, Gamboa da Costa G and Delclos KB, (2019). NTP CLARITY-BPA report (2018). A two-year toxicology study of bisphenol A (BPA) in Sprague-Dawley rats: CLARITY-BPA core study results. Food and Chemical Toxicology, 132. 110728.

EFSA 2011. EFSA Scientific Committee. Scientific opinion on genotoxicity testing strategies applicable to food and feed safety assessment. EFSA Journal 2011;9(9):2379, 69 pp. doi:10.2903/j.efsa.2011.2379.

EFSA 2015. EFSA CEF Panel (EFSA Panel on Food Contact Materials, Enzymes, Flavourings and Processing Aids), 2015. Scientific Opinion on the risks to public

health related to the presence of bisphenol A (BPA) in foodstuffs: Executive summary. EFSA Journal 2015;13(1):3978, 23 pp. doi:10.2903/j.efsa.2015.3978.

EFSA 2017. EFSA Scientific Committee, Hardy A, Benford D, Halldorsson T, Jeger M, Knutsen HK, More S, Naegeli H, Noteborn H, Ockleford C, Ricci A, Rychen G, Silano V, Solecki R, Turck D, Younes M, Aquilina G, Crebelli R, Gürtler R, Hirsch-Ernst K, Mosesso P, Nielsen E, van Benthem J, Carfì M, Georgiadis N, Maurici D, Parra Morte J and Schlatter J, 2017c. Clarification of some aspects related to genotoxicity assessment. EFSA Journal 2017;15(12):5113, 25 pp. doi:10.2903/j.efsa.2017.5113.

EFSA (2021a) Re-evaluation of the risks to public health related to the presence of bisphenol A (BPA) in foodstuffs. PC-1019.

Public Consultation: (europa.eu)

EFSA (2012b) EFSA Scientific Committee, More SJ, Bampidis V, Bragard C, Halldorsson TI, Hernández-Jerez AF, Hougaard Bennekou S, Koutsoumanis K, Lambré C, Machera K, Naegeli H, Nielsen SS, Schlatter J, Schrenk D, Turck D, Younes M, Aquilina G, Bignami M, Bolognesi C, Crebelli R, Gurtler R, Marcon F, Nielsen E, Vleminckx C, Carfì M, Martino C, Maurici D, Parra Morte J, Rossi A and Benford D, 2021a. Guidance on aneugenicity assessment. EFSA Journal 2021;19(8):6770, 27 pp.

Klimisch HJ, Andreae M and Tillmann U, 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regulatory Toxicology and Pharmacology, 25(1), 1—5. doi:10.1006/rtph.1996.1076.

Montévil M, Acevedo N, Schaeberle CM, Bharadwaj M, Fenton SE and Soto AM, 2020. A combined morphometric and statistical approach to assess nonmonotonicity in the developing mammary gland of rats in the CLARITY-BPA study. Environmental Health Perspectives, 128(5), 57001. doi:10.1289/EHP6301. 32438830.

NTP-OHAT (National Toxicology Program – Oral Health Assessment Tool), 2015. Handbook for conducting a literature-based health assessment using OHAT approach for systematic review and evidence integration. Available online:

Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration (nih.gov)

VKM (Norwegian Scientific Committee for Food and Environment), 2019. Protocol for the risk assessment of energy drinks and caffeine. Available online:

[VKM rapportmal](#).