Public Health Goals

Responses to Peer Review and Public Comments on Technical Support Document:

Public Health Goals for Haloacetic Acids in Drinking Water

December 2022

Pesticide and Environmental Toxicology Branch
Office of Environmental Health Hazard Assessment
California Environmental Protection Agency

Responses to Peer Review and Public Comments on Technical Support Document: Public Health Goals for Haloacetic Acids in Drinking Water

Prepared by

Pesticide and Environmental Toxicology Branch
Office of Environmental Health Hazard Assessment
California Environmental Protection Agency

December 2022

CONTENTS

INTRODUCTION

This document contains responses to public comments received by the Office of Environmental Health Hazard Assessment (OEHHA) on the Public Health Goal (PHG) technical support document for haloacetic acids (HAAs) during the first and second public comment periods, and to comments from the external scientific peer reviewers.

OEHHA released the first draft of this PHG document for public comment on January 31, 2020, and held a public workshop on April 28, 2020 via webcast. The public comment period closed on May 1, 2020. OEHHA received comments from the American Chemistry Council (ACC), Clean Water Action, the Environmental Working Group, the Southern California Water Coalition and Marc William Goff.

Pursuant to Health and Safety Code section 116365(c)(3)(D), OEHHA submitted the HAA PHG document for scientific peer review following the closure of the first comment period. Comments were received from the peer reviewers in October 2020.

The external scientific peer reviewers were:

Zhoumeng Lin, Ph.D.
Institute of Computational Comparative Medicine (ICCM)
Department of Anatomy and Physiology, College of Veterinary Medicine
Kansas State University
Manhattan, KS 66506

David H. Phillips, Ph.D., Professor
Department of Analytical, Environmental & Forensic Sciences
School of Population Health & Environmental Sciences
King's College London
Franklin-Wilkins Building
150 Stamfort Street
London SE1 9NH

Kan Shao, Ph.D.
School of Public Health
Indiana University Bloomington
PH 029
1025 E. 7th Street
Bloomington, IN 47405

Paul A. White, Ph.D.
Department of Biology
University of Ottawa
30 Marie Curie
Ottawa, Ontario, KIN 6N5 Canada

OEHHA made changes in response to the public and peer review comments as appropriate, and incorporated them into the second draft of the PHG technical support document posted on the OEHHA website on August 19, 2022 for a 30-day comment period. OEHHA received one comment from the Chlorine Institute.

The full public comments and peer review comment letters are posted on the OEHHA website along with this response document, and the final version of the PHG document.

In this document, comments appear in quotation marks where they are directly quoted from the submission. Note that for the public comments where the commenter included a footnote, OEHHA did not copy the footnote into the response document. Footnotes can be seen in the original public comment letters posted on the OEHHA website. Editorial comments resulting in non-substantive changes have been addressed and are not included in this document.

For further information about the PHG process or to obtain copies of PHG documents, visit the OEHHA web site at www.oehha.ca.gov.

OEHHA may also be contacted at:

Office of Environmental Health Hazard Assessment
P.O. Box 4010, MS-12B
Sacramento, California 95812-4010
Attention: PHG Program

PHG.Program@oehha.ca.gov

(916) 324-7572

RESPONSES TO EXTERNAL SCIENTIFIC PEER REVIEW COMMENTS

RESPONSES TO COMMENTS RECEIVED FROM DR. ZHOUMENG LIN

Comment 1: "It is mentioned that "While the benchmark dose (BMD) approach was applied to datasets amenable to modeling, the point of departure (POD) is determined from the no-observed-adverse-effect level (NOAEL) for systemic toxicity from an animal toxicology study." This decision may seem confusing initially. However, after reading the detailed description in the document (Pages 197-200), this decision is considered acceptable because the two candidate critical studies (DeAngelo et al., 1997; NTP, 1992) are of comparable quality, and the BMDL05 based on data from NTP (1992) is 3.4 mg/kg/day, which is similar to the NOAEL of 3.5 mg/kg/day based on the DeAngelo et al. (1997) study, which is not amenable to BMD analysis. Also, as the authors mentioned, the NOAEL of 3.5 mg/kg/day was chosen instead of BMDL05 of 3.4 mg/kg/day based on NTP (1992) because the observed mortality was "due to undetermined causes" and its relevance to human health was unclear in the NTP (1992) study."

Response 1: OEHHA acknowledges the peer reviewer's comment that the decision to use the NOAEL was acceptable.

Comment 2: "On Page 194, it is said that "The BMR is typically set at 5% above the background or the response of the control group for dichotomous data". … Based on my experience and my understanding of the US EPA guideline (US EPA, 2012), a response level of 10% extra risk is commonly used to define BMD for dichotomous data." The peer reviewer then goes on to quote from US EPA (2012) guidance; he also indicates that both EFSA (2017) and US EPA (2012) focus on the 10% response range in determining the BMR for dichotomous data. He suggests that "the authors re-consider their approach carefully, and either revise or at least provide a clear justification for each BMR selection by addressing the considerations outlined above from the US EPA guideline."

Response 2: OEHHA's guidance on BMR selection when evaluating dichotomous data is described in OEHHA (2008), *Technical Support Document for the Development of Noncancer Reference Exposure Levels*. The guidance went through public comment as well as scientific peer-review by the State's Scientific Review Panel on Toxic Air Contaminants (SRP)[1] (see response to Comment 3 below), and is consistent with the US EPA guidelines (US EPA, 2012). As noted in these peer-reviewed risk assessment guidelines (OEHHA, 2008),

> "[Reference concentration] determinations for various endpoints by the U.S. EPA have used either 5% or 10% as the benchmark response rate, depending on the statistical uncertainty in the data (U.S. EPA, 2002a; U.S. EPA, 2004). OEHHA has used the 5% response rate in several chronic [reference exposure levels], and showed that the lower 95% confidence bound on the BMC_{05} [benchmark concentration for a BMR of 5%] typically appears equivalent for risk assessment

[1] Approved by California's Scientific Review Panel on Toxic Air Contaminants June 18, 2008.

purposes to a NOAEL in well designed and conducted animal studies where a quantal measure of toxic response is reported…Therefore, OEHHA typically uses a 5% response rate as the default for determination of the BMC from quantal data (i.e. the effect is either present or it is not) in animals."

While the US EPA in its *Benchmark Dose Technical Guidance* (US EPA, 2012) notes, "For comparing potencies across chemicals or endpoints (e.g., for chemical rankings) for dichotomous data, a response level of 10% extra risk has been commonly used to define BMDs, also known as effective doses (i.e., ED_{10}s)," the guidelines also state, "[T]he BMD (BMDL) used as a POD may correspond to response levels below (or sometimes above) 10% extra risk. For standardization, rounded levels of 1%, 5%, or 10% have typically been used."

OEHHA added further details in the *Dose-Response Assessment* section of the technical support document regarding the choice of BMRs for BMD modeling.

Comment 3: "[S]ome of the default uncertainty factors for PHG derivation are different from those recommended by US EPA (US EPA, 2002; 2011; 2016a; 2016b). …[I]t is indicated that UF_H [intraspecies uncertainty factor] due to toxicokinetic component could be up to 10, and UF_H due to toxicodynamic component could be up to 10. … [A]ccording to US EPA guideline (US EPA, 2002; 2011; 2016a; 2016b), the UF_H of ≤10 is intended to account for potential variation in sensitivity among humans and is considered to include toxicokinetic/toxicodynamic processes. …I suggest the authors double check the UF_H for combined toxicokinetic/toxicodynamic processes, and revise or clarify. In addition, in Table 10.1, it is indicated that the database deficiency factor (UF_D) is $\sqrt{10}$. However, according to US EPA guideline (US EPA, 2002), the recommended value is ≤10 to address database deficiencies. Overall, I suggest the authors check their default certainty factors, provide a justification, and clarify why some of these factors are different from those recommended by US EPA."

Response 3: The draft PHG document notes that Table 10.1 (default uncertainty factors for PHG derivation) is adapted from and is consistent with OEHHA's *Technical Support Document for the Development of Noncancer Reference Exposure Levels* (OEHHA, 2008).[2] These guidelines were extensively peer reviewed and approved by the SRP. This guidance updated previous guidance to reflect scientific knowledge and techniques in response to California's Senate Bill 25 (Chapter 731, Statutes of 1999, Health and Safety Code sections 39669.5 et seq., Children's Environmental Health Protection Act), which mandates that infants, children and other sensitive subpopulations must be considered when evaluating health-based ambient air quality standards. The legislative mandate for PHG development similarly requires consideration of sensitive subgroups including infants, children, pregnant women and the elderly (Health and Safety Code section 116365, subsection (c)(1)(C)). OEHHA performed its own systematic, rigorous analyses of variability in susceptibility to toxicants and uncertainties in human health risk assessment, especially as they pertain

[2] https://oehha.ca.gov/media/downloads/crnr/noncancertsdfinal.pdf

to the increased susceptibility of sensitive individuals, including infants and children. The results of this systematic analysis and associated guidance can be found in OEHHA (2008). With regard to use of the UFD of √10, it is noted that this is within the US EPA range of 1 to 10 for this factor.

In response to this comment and request for clarification, a sentence has been added to the technical support document to refer readers to OEHHA's 2008 guidance document for a detailed discussion of its default uncertainty factors.

Comment 4: "On Page 199, the total UF of 1000 for MCA is based on the consideration of an UFH of 30 for variation in the human population, which includes √10 for toxicodynamics and 10 for toxicokinetics, which accounts for diversity, including infants and children, with no human kinetic data. As stated above, this UF_H of 30 needs to be justified, and the reason of why this factor is different from that recommended by US EPA should be clarified." The peer reviewer makes this comment regarding the UF_H of 30 for all 5 HAAs, thus other instances of this comment will not be shown in this document as they will be addressed by the same response.

Response 4: OEHHA's rationale for application of the UF_H of 30 is found in the *Dose-Response Assessment* section of the technical support document. A more detailed explanation of how OEHHA's default UFs were derived can be found in OEHHA (2008). Also, see Response 3.

Comment 5: "Regarding the BMD analyses, I have tried to reproduce the results of one representative noncancer dataset and one representative cancer dataset. I was able to obtain the same or similar results by using the same settings (BMR of 0.05 and extra risk type, etc.). However, if I change the BMR from 5% to 10%, the results will be quite different."

Response 5: OEHHA agrees that using a different BMR would result in a different BMDL value. Modeling parameters and outputs are included in the document in Appendices D and E for transparency, and so that modeling results can be replicated by interested parties.

Comment 6: "Similar to MCA and DCA, my concern on the scientific assumptions, findings, and conclusions for TCA is mainly on the use of 5% as BMR in the BMD modeling and the selection of the uncertainty factors, which are different from those recommended by US EPA. Please refer to my detailed description on these two points above. In addition, I have been trying to reproduce some of the calculated results presented in Table 10.14 entitled "TCA candidate cancer studies (BMDS analysis and CSF calculation)". The peer reviewer provides an example of how his CSF calculations differed from those presented in the document. "I understand that this may be due to adjustment for exposure duration. However, can the authors clarify how they calculated these results in the footnote, so that readers can understand this better."

Response 6: Regarding the use of BMR 5% and specific UFs, see Responses 2 through 4. OEHHA appreciates the effort by the commenter to double check the calculations. The formula for exposure duration adjustments is given in the *Adjusting for Experimental Duration* section. Additionally, the corresponding footnote to Table 10.14 was expanded to provide details on exposure duration adjustments for CSFs derived from the Bull et al. (2002) and DeAngelo et al. (2008) Study 1 subsets. The differences between their calculation and the values presented in Table 10.14 were due to the exposure duration adjustment.

Comment 7: "Similar to my comments on MCA, DCA, and TCA above, I have a concern on the uncertainty factors of MBA. Specifically, I am concerned about the total UF of 3000 for MBA and its associated description on Page 208. First, an UF of √10 for LOAEL-to-NOAEL extrapolation is used. However, Based on Table 10.8, the 5 mg/kg/day is a NOAEL, not LOAEL. In this case, why an UF for LOAEL-to-NOAEL extrapolation is needed here?"

Response 7: This is an error in the draft HAA document. The MBA assessment is based on a NOAEL identified in Dalgaard-Mikkelsen et al. (1955), and therefore, the assessment would not require the LOAEL-to-NOAEL UF. This study is of subchronic duration, and the UF of √10 was meant to address the uncertainty associated with the extrapolation from subchronic exposure to chronic exposure (Table 10.1). The MBA section in Chapter 10 of the draft PHG document was changed accordingly. The correction does not result in a change to the composite UF (3,000) or the PHG value for MBA.

Comment 8: "Second, based on the presented toxicity evidence, MBA is the most potent among all five regulated HAAs in many in vitro endpoints, and it is evident that MBA toxicity data are quite limited. As a result, I am wondering whether the authors have considered using an UF of 10 for database deficiency, instead of √10. Third, it is actually not clear which value was used to account for the uncertainty of database deficiency. On Page 228, in the paragraph starting with "Lack of cancer studies for MBA", it is indicated that "A database deficiency uncertainty factor of 10 was applied in the PHG calculation to account, in part, for the potential carcinogenicity of MBA." However, on Page 208, it is mentioned that an UF of √10 is used to account for database deficiency for MBA. Is the UF for database deficiency 10 or √10 for MBA?"

Response 8: While OEHHA typically applies a database deficiency UF of √10, OEHHA's guidelines also state, "In some cases, it may be appropriate to apply a database deficiency factor larger than three-fold. The need for the additional database deficiency UF will be evaluated on a chemical-by-chemical basis and justified in the individual [reference exposure level] summaries." Additionally, "Cumulative UF values are normally limited to between 1 and 3,000: if the latter value is exceeded it is generally taken to indicate that the source data are insufficient to support derivation of a [reference exposure level]." For MBA, OEHHA applied a database deficiency UF of √10, resulting in a cumulative UF of 3,000. In response to this comment, the typo has been corrected to √10, and the specific UFs applied are explicitly outlined.

Comment 9: "[I]f we assume the BMD analysis results [for DBA] are acceptable, then based on the justification provided on Pages 208-212, the selection of the LOAEL of 1 mg/kg/day from Veeramachaneni et al. (2007) appears to be reasonable."

Response 9: OEHHA acknowledges the peer reviewer's statement that the selection of the LOAEL appears to be reasonable.

Comment 10: "Page 3: "Both the California and federal MCLs of 60 ppb for total HAAs represent the highest allowable annual average sum of the concentrations of MCA, DCA, TCA, MBA, and DBA." The sum of the proposed PHGs of 53, 0.2, 0.1, 25, and 0.03 ppb for MCA, DCA, TCA, MBA, and DBA, respectively, is larger than the highest allowable annual average sum of the concentrations of MCA, DCA, TCA, MBA, and DBA. Can the authors provide a clarification as to how their recommended PHGs are related to the California and federal MCLs of 60 ppb for total HAAs? Are there sufficient scientific data or is it necessary to derive a PHG for the sum of the concentrations of MCA, DCA, TCA, MBA, and DBA since humans are usually exposed to mixtures of these five regulated HAAs?"

Response 10: In response to this comment, a sentence explaining why a group PHG was not developed was added to the *Summary* and the *Dose-Response Assessment* sections. California's current MCL for HAA5 was adopted from US EPA, prior to the development of any PHGs for these chemicals. State law requires OEHHA to develop a PHG for each drinking water contaminant that is regulated with an MCL. PHGs are entirely health-based whereas CA MCLs are to be set as close to the corresponding PHG as is economically and technologically feasible.

It is not possible to develop a group PHG for these HAAs due to the distinct critical effects of the individual HAAs. Three HAAs, i.e., DCA, TCA and DBA are carcinogens with low PHGs, while the two remaining HAAs, MCA and MBA, have higher PHGs based on noncancer effects. Since the five HAAs do not have the same critical adverse effect as the basis for their respective PHG, development of a group PHG would not be optimal.

Comment 11: "Page 13, regarding the statement "Mean values for MCA, DCA and TCA in samples of urban and rural air were 1.5-2.5 nanograms per cubic meter (ng/m^3), 0.66-1.1 ng/m^3 and 0.13-0.22 ng/m^3, respectively (Martin et al., 2003).", please clarify where the air samples were collected. If not from US or California, are there any studies that report air concentrations of these HAAs in the US or California?"

Response 11: In response to this comment, the text was amended to indicate that the Martin et al. (2003) air samples were collected in Ontario, Canada. Studies in US air are described later in the same section, *Air*.

Comment 12: "Page 24, Paragraph 2: The pharmacokinetics of HAAs is introduced in Sections of Absorption, Distribution, Metabolism, and Excretion. The second paragraph

on Page 2 is about plasma clearance. It is now placed in the Section of Absorption. I suggest moving this paragraph to the Section of Excretion."

Response 12: In response to this comment, this paragraph was moved to *Excretion.*

Comment 13: "Page 25: In the paragraph describing the study by Styles et al. (1991), it is stated that "Little of the radioactivity was covalently bound to plasma or liver proteins", whereas in the paragraph describing the studies by Schultz et al. (1999), Templin et al. (1993), and Yu et al. (2000), it is stated that "TCA appeared to bind significantly to plasma proteins after dosing by both routes". First of all, please clarify the extent of binding, i.e., what is the percentage of binding? Also, please clarify the discrepancy in the findings between the Styles et al. (1991) study and the studies by Schultz et al. (1999), Templin et al. (1993) and Yu et al. (2000)."

Response 13: In response to this comment, the text was modified. The percentage of TCA non-covalent binding to plasma proteins was added to the text. While Styles et al. (1991) analyzed covalent binding of TCA to plasma or liver proteins, in which an irreversible chemical bond is formed between TCA and the protein molecules, other reports analyzed reversible non-covalent binding. Thus, there is no discrepancy, as different types of binding are described.

Comment 14: "Page 27, Paragraph 3: Please clarify what the phrase "by a 2-2 to 23-times" means?"

Response 14: In response to this comment, this typo was corrected to "by 22- to 23-times."

Comment 15: "Page 27, Paragraph 4: The elimination half-life of MCA in humans can be calculated and described here based on the study by Kulling et al. (1992)."

Response 15: In response to this comment, the sentence was edited to include the phrase "with an approximate half-life of 2 hours."

Comment 16: "Pages 26-29: It is mentioned that plasma protein binding of TCA is species-dependent; DCA is minimally bound to plasma proteins; and DBA does not significantly bind to plasma proteins or accumulate in blood cells. What about the plasma protein binding properties of MCA and MBA? Plasma protein binding is an important property of a chemical as it is generally considered that the free fraction of a chemical is responsible for its pharmacological and toxicological effects. Therefore, I suggest that the authors create a small table listing the reported plasma protein binding percentages of each of the five regulated HAAs in different species from different studies."

Response 16: In response to this comment, a sentence was added to the *Distribution* sub-section of the *Pharmacokinetics* chapter, explaining that only plasma binding of TCA was investigated. Among the five HAAs, only plasma binding of TCA was investigated in some detail, in part due to its relative resistance to metabolism. The

results of these studies are covered in detail in the text. Minimal binding was found with DCA and DBA, and no studies were located with MCA and MBA. Thus, for some of the HAAs there are no plasma protein binding data to prepare a comparison table.

Comment 17: "Pages 23-34: The pharmacokinetic properties of the five HAAs are described only in the text. In this way, the differences in the PK of different HAAs are not easy to see. I suggest that the authors include a new table listing the key PK information of individual HAAs. As an example of this table, please refer to Table 3-2 of the NTP (2018) report entitled "Report on Carcinogens: Monograph on Haloacetic Acids Found as Water Disinfection By-Products". In the Absorption column, the authors could list oral bioavailabilities of different HAAs in different species. In the Metabolism column, the authors could list the metabolic rates (if available), metabolic enzymes, etc. In the Excretion column, it is important to provide information on the elimination half-lives of different HAAs in different species."

Response 17: The peer reviewer makes a good suggestion, and a table would be a good way to present these values. OEHHA will consider this suggestion for future PHGs.

Comment 18: "As stated in the text, the half-life of TCA is 5.4-6.4 hours in mice (Templin et al., 1993), 8 hours in rats (Schultz et al., 1999), and from 2.1 to 6.3 days in humans (Bader et al., 2004; Froese et al., 2002). This suggests that there may be a substantial interspecies difference in the toxicokinetics of TCA. These results have implications in terms of how to extrapolate animal toxicity data to interpret human risks. These results should be listed in the table, and the potential reasons underlying the observed species-difference should be explained in the text."

Response 18: As discussed below, text has been added to address the issue raised by the peer reviewer.

TCA appears to demonstrate species differences in toxicokinetics, particularly between mouse and human, which could have importance for the mouse-to-human extrapolation of the CSF or noncancer health-protective concentration derivation (i.e., dose conversion), and ultimately, the development of the PHG. Although OEHHA did spend a significant amount of effort on addressing this concern of TCA species differences, in the end, the toxicokinetic approaches (including PBPK models) were not found to be sufficiently reliable.

It is currently thought that the longer half-life ($T_{1/2}$) of TCA in humans could be due to a higher rate of binding to plasma proteins. This hypothesis complicates the comparison of mouse TCA $T_{1/2}$ values from Templin et al. (1993) with human data (Bader et al., 2004; Froese et al., 2002) in the following way. Mouse TCA $T_{1/2}$ was measured with a single exposure by gavage, and TCA plasma protein binding was 41%, 34% and 23% of total TCA at doses of 0.03, 0.12 or 0.61 mmol/kg, respectively (Templin et al., 1993). Consistent with the latter observations, the authors also stated that for lower levels of TCA in the serum, which would occur at lower TCA doses, 50-57% of total TCA was

bound to protein. Thus, it is likely that in the mouse, TCA $T_{1/2}$, as well as the TCA fraction bound to protein, would increase with lower TCA doses.

One big difference between mouse and human experimental systems used to estimate TCA $T_{1/2}$ is the dose. One can estimate the daily dose of TCA in humans in the Froese et al. (2002) study was 1 µg/kg-day, assuming a body weight of 50-70 kg, whereas the lowest of the three doses in the Templin et al. (1995) mouse study was 4.89 mg/kg, or 5,000 times higher numerically (although this was a single dose). Froese et al. (2002) also noted considerable inter- and intra-individual variability of data.

The mouse TCA $T_{1/2}$ of 5.4-6.4 hours and human $T_{1/2}$ of 2.1 to 6.3 days cannot be directly compared, given the dramatic differences in the experimental design of the Templin et al. (1993) and Froese et al. (2002) studies (dose levels more than three orders of magnitude different, repeated vs. single dose exposure) that would exacerbate the apparent species $T_{1/2}$ difference, and the hints in the mouse data that $T_{1/2}$ could be longer at lower doses. On the other hand, the interspecies allometric conversion in the cancer assessment and the UF_A (interspecies uncertainty factor) in the noncancer assessment would likely compensate for the observed difference in half-lives (although these factors would also be responsible for toxicodynamic differences between species). Therefore, given the high uncertainty of the available TCA toxicokinetics data, OEHHA used default species dose conversion methods, i.e., application of UF_A and allometric conversion using body weight scaling respectively, for noncancer and cancer dose-responses.

In response to this comment, a short paragraph was added to the *Excretion* section, and a sentence was added to the TCA subsection of the *Dose-Response Assessment* chapter to clarify OEHHA's position.

Comment 19: "Page 31, Figure 3.1: Please provide more detailed legend for Figure 3.1. For example, please clarify what "r, m >> h" and "h > r, m" mean. Both dashed arrows and solid arrows are used. Please clarify the differences between dashed arrows and solid arrows."

Response 19: In response to this comment, clarifications were added to the legend of Figure 3.1 (p.16). Expression of the style "r, m >> h" and "h > r, m" indicate relative importance of indicated pathway in species and are common in toxicokinetic literature.

Comment 20: "Pages 35-39: This section describes available PBPK models related the five regulated HAAs. The authors indicate that PBPK-based approaches were considered for use in the risk assessment and PHG derivation for TCA and DCA because human and mouse PBPK models for TCA and DCA were available. The authors decided not to use the existing TCA or DCA PBPK models for dose-response analyses due to some limitations of existing models. I agree that the available PBPK models for TCA and DCA are not sufficient for use in risk assessment."

Response 20: OEHHA acknowledges the peer reviewer's concurrence with OEHHA to not use the available PBPK models for TCA and DCA.

Comment 21: "However, this section should be updated to better reflect the current state of science in this field. Specifically, it appears that some related PBPK models are not discussed in this report. Also, all the cited PBPK studies were published at least 9 years ago. There are some related PBPK studies published in recent years that should also be discussed." The peer reviewer provides examples of updated PBPK models that are relevant to this assessment.

Response 21: It was suggested in this comment that OEHHA should expand the PBPK model section to include the PBPK models for TCA and DCA precursors, such as tetrachlorethylene (PCE) and trichlorethylene (TCE). Given the peer reviewer's agreement with OEHHA that ultimately, the available TCA and DCA PBPK models are not sufficient for use in risk assessment (see Comment 20), expanding the PBPK section to include related models would be primarily for information purposes. Although the PBPK models of HAA precursors take into consideration DCA and TCA kinetics and may provide some insights into their properties, such evidence would be indirect and compound the multiple uncertainties due to the presence of multiple other components in the models, such as precursor compounds and other metabolites, additional metabolic pathways, etc. The PHG draft document discussed the publications on the PCE model by Chiu and colleagues that included direct TCA validation studies (Chiu et al., 2009; Chiu and Ginsberg, 2011); however, more extensive discussion of this model, as well as of any other HAA precursor model lacking direct HAA validation studies, would not be helpful for HAA risk assessment due to the high uncertainty of such models.

Comment 22: "At the end of the PBPK section, the authors concluded that the existing PBPK models for TCA and DCA are not sufficient for use in risk assessment. I agree with this conclusion. Besides the limitations associated with existing PBPK models discussed in the report, another difficulty is that most of the discussed models were developed using a legacy software program, acslX, which was discontinued by the company in 2015. Also, many of the earlier PBPK studies did not report the entire model codes, nor did the authors share detailed instructions on how to use their models to do risk assessment. These factors also partially contribute to the inability or difficulty of using these earlier models in risk assessment. It is recommended that future PBPK studies of HAAs should use contemporary open-source programming tools (e.g., R program) and publish the entire model code along with a detailed instruction on how to use the model. This will increase the transparency and reproducibility of the model, and also make it straightforward for risk assessors to directly use the published PBPK models. The authors may want to mention this recommendation in their report."

Response 22: This comment states concurrence with OEHHA's analysis and decision to not use PBPK models. OEHHA agrees that increased transparency of PBPK models would contribute to their utility for risk assessors. However, OEHHA generally does not provide recommendations for researchers in PHG technical support documents.

Comment 23: "Here is another general suggestion on the section describing PBPK models. I suggest the authors create a new table listing the existing PBPK models/studies for each of the five HAAs. For MCA and MBA, if no models are available, just list None. For others, please list the basic features, strengths, and limitations of each model. This will give readers a clear understanding of what models are available, what data gaps are in this field, and what limitations that have prevented from using the existing models to do risk assessment."

Response 23: The peer reviewer makes a good suggestion, as a table of PBPK models is a good way to present the information. However, such an expanded comparison is beyond the scope of this document. The essential part of the PBPK section is the determination of whether the PBPK models can be used for HAA risk assessment, and the commenter has stated his agreement with OEHHA's conclusion.

Comment 24: "Page 40: Please consider providing a flowchart listing the literature search outcome. Specifically, how many references were retrieved in the initial search? How many references approximately that were excluded after screening the titles? How many references that were excluded after screening the abstracts, and how many were excluded after reading the full text, and how many were selected at the end? Also, for literature search, the end date was 4/26/2018. What was the start date of literature for each of the used databases?"

Response 24: In response to the comment, a sentence has been added to the *Reproductive and Developmental Effects* section of Chapter 4 directing the reader to the flowchart and discussion provided at the beginning of Appendix B.

Comment 25: "Page 48: Similar to the comment above, please consider providing a flowchart listing the outcome of the literature search on cancer effects."

Response 25: In response to the comment, a sentence has been added to the *Cancer Effects* section of Chapter 4 directing the reader to the flowchart and discussion provided at the beginning of Appendix C. The text further explains that the start date was 1/1/1985.

Comment 26: "Pages 52-193: A number of in vitro and in vivo toxicity studies for the five regulated HAAs have been discussed. However, one thing that is missing is the high-throughput in vitro toxicity assay data from ToxCast/Tox21 database (Kavlock et al., 2019; Krewski et al., 2020; Tox21, 2020). The ToxCast data have been considered in the risk assessment of other environmental contaminants, such as perfluorooctane sulfonate (US EPA, 2016a). I did a quick search and found that as of August 24, 2020, there are 2, 7, 18, 2, and 3 active assays for MCA, DCA, TCA, MBA, and DBA respectively. I am wondering if the authors would consider discussing these toxicity data in this document."

Response 26: OEHHA acknowledges the importance of the ToxCast/Tox21 database as a data source for in vitro toxicity. As noted in the comment, HAAs had several active

assays. However, these results were not discussed in the draft PHG document because overall they were found to be somewhat unreliable for these particular chemicals.

TCA appears to have a “U” grade (i.e., unknown/ inconclusive), likely indicating analytically failed quality control (QC). ToxCast/Tox21 assay results using TCA cannot be considered with satisfactory confidence, and TCA data were not used in the draft PHG.

With DCA, one of the four batches used had an “A” (passing) grade in the QC check, and three others had failing grades (“U”, “Ac” – caution, low concentration, “F” – wrong ID, “Fns” – no sample). It was not possible to identify which one of the 7 assays used specific QC-passing or failing batches of DCA, lowering the confidence in the results. Moreover, in all seven assays, just one concentration point was above baseline and fold induction was low, indicating a weak signal. Given QC doubts and a weak signal, DCA data were not used in the draft PHG.

MCA, had a QC grade of “A” for its two positive assays. The two positive assays each resulted from a single active concentration barely above the cutoff line, indicating a weak signal. These assays are transcription based, and weak activation of transcription is generally not considered relevant.

One batch of DBA passed the QC check (grade “A”) but the second failed (“F”, “Ac”), and it is not clear which assay used which batch, lowering the overall confidence in the results. The three active assays all demonstrated activation of ARE-dependent transcription which suggests increased activity of the Androgen receptor (AR). One assay had weak signal and two assays had 1.75-2.15-fold activation, which is what would be barely considered active (~2-fold activation). This is the strongest ToxCast/Tox21 data among the five HAAs. Still, given persisting QC doubts, OEHHA did not include this result in the draft PHG document. Multiple in vitro and in vivo studies were reviewed for DBA. While the ToxCast/Tox21 finding of possible androgen receptor activation would be consistent with the observed adverse effects and other in vitro biological observations for DBA, it would not significantly improve understanding of the underlying mechanisms.

MBA passed one QC check (grade “A”) but failed another (“U”). However, in the two active ToxCast/Tox21 tests activity was increased in just one lower concentration point (same point in both assays) but returned to background level for the next nine concentrations. These results appear to be an experimental artifact and the data were not used in the draft PHG.

Comment 27: “Page 62, Paragraph 2, First Sentence: “(53-60/dose group)”, please clarify if this is per sex. If so, please change it to “(53-60/sex/dose group)”.”

Response 27: In response to this comment, the text was clarified to indicate 53-60/sex/dose group.

Comment 28: "Page 63, Paragraph 1, "…identified the low dose (3.5 mg/kg-day) as a NOAEL based on increased relative liver weight (US EPA, 2006)". I am a little bit concerned about this conclusion and the use of this NOAEL in the risk assessment. Based on Table 5.5, I understand that the authors consider the endpoint of systemic toxicity as "decreased body weight and relative liver weight". At this low dose of 3.5 mg/kg-day, it did not decrease relative liver weight. Instead, it increased it. However, from the perspective of toxicology, this is a statistically significant increase in relative liver weight and there is no evidence that this increase is of no toxicity concern. Therefore, I would considered as LOAEL, not NOAEL. I understand this is somewhat an arbitrary selection, and depends on the inclusion criteria of sensitive endpoints for subsequent dose response analysis. However, I would like to raise this point for the authors' consideration."

Response 28: OEHHA considered the points raised. For the DeAngelo et al. (1997) study, the sensitive endpoints were decreased body weight and increased relative liver weight. The relative liver weight in the control and low dose (3.5 mg/kg-day) groups were 4.35±0.21% and 4.56±0.18% (mean ± standard error), respectively. The two-tailed p-value of unpaired t test equals 0.4504, and the difference between the groups is not statistically significant. Thus, OEHHA considered 3.5 mg/kg-day the NOAEL.

Comment 29: "Page 77: Last Paragraph: It is indicated that "OEHHA does not identify NOAELs and LOAELs for single-dose studies". This statement appears in many places throughout the document. Please provide a justification of this decision at least in one place."

Response 29: In response to this comment, OEHHA reconsidered this policy of excluding single-dose studies from NOAEL/LOAEL determination. As such, OEHHA concluded that there is value in presenting NOAELs and LOAELs from single dose studies. However, these values are marked with a footnote to indicate that these values are generally not considered as critical studies, as dose-response assessments cannot be adequately conducted. Furthermore, excluding single-dose studies from critical study consideration avoids creating the false impression that NOAELs/LOAELs identified from multi-dose and single-dose studies are of similar quality and comparable.OEHHA generally prefers using multi-dose studies for risk assessment since they contain additional information on the biological gradient of response (one of the Bradford-Hill criteria) and can be subjected to dose-response analysis.

However, when no multi-dose studies of acceptable quality are available, OEHHA may have no option but to use a single-dose study and identify its NOAEL/LOAEL as the POD. For example, the MBA PHG is based on the NOAEL from technically a single-dose study by Dalgaard-Mikkelsen et al. (1955).

Comment 30: "Page 121: It is said that "The biochemical effects, such as GSTz1 inhibition, observed at high DCA concentrations would likely be negligible at exposures to the relatively low environmental DCA concentrations found in drinking water". Can the

authors provide experimental evidence to support this statement? Is there a threshold level of the GSTz1 inhibition effect by DCA?"

Response 30: In response to this comment, a reference is provided to the study of quantitative evaluation of DCA kinetics in humans (Li et al., 2008), that concluded: "Apparent inhibition of GSTzeta mediated metabolism of DCA was minimal for low doses of DCA (µg/kg day)." Thus, at either the noncancer health-protective concentration (7.6 µg/kg-day) or PHG (0.03 µg/kg-day), the effect of GSTz1 inhibition would be negligible.

Comment 31: "Page 194: It is stated that "For PHG development, OEHHA uses the BMDL as the POD for the calculation of a health-protective drinking water concentration when the data are amenable to BMD modeling." However, it is apparent that there are exceptions. For example, for MCA, the NTP (1992) data were amenable for BMD modeling and resulted in a BMDL05 of 3.4 mg/kg/day. However, the NOAEL of 3.5 mg/kg/day from DeAngelo et al. (1997) was actually chosen as the POD. Please clarify that there are exceptions here."

Response 31: The excerpt cited applies to the determination of the POD (either using the NOAEL or BMDL) for a given study in general, and not to the process of selecting the critical study and critical POD for deriving the acceptable daily dose (ADD). For MCA, OEHHA chose DeAngelo et al. (1997) as the critical study and the NOAEL of 3.5 mg/kg-day as the critical POD because the DeAngelo study was considered to be a high-quality study (i.e. appropriate number of animals, administration of neutralized MCA in drinking water, and a comprehensive pathological examination and serum analysis). It is noted that the NTP (1992) study is of comparable quality to the DeAngelo et al. (1997) study. However, because the critical endpoint identified in the NTP (1992) study was increased mortality due to unidentified causes, and it is generally more appropriate to develop guidance values based on less severe effects than mortality, the DeAngelo et al. (1997) study was chosen over the NTP (1992) study.

Comment 32: "Page 211: It is stated that "BMD modeling was performed on the datasets presented in Table 10.9". However, Table 10.9 only lists the studies, and provides information on the Dose, Endpoint, and NOAEL/LOAEL/BMDL. It does not actually provide the response data that are needed in the dose-response analysis. Please either revise this sentence or update the table."

Response 32: In response to this comment, this sentence was changed to: "BMD modeling was performed on the datasets from the studies presented in Table 10.9."

Comment 33: "Page 220, Table 10.14: I have been trying to reproduce some of the calculated results." The peer reviewer provides an example of how his CSF calculation differs from what is presented in the document. "I understand that this may be due to adjustment for exposure duration. However, can the authors clarify how they calculated the results in the footnote, so that readers can understand this better."

Response 33: In response to this comment, the footnote was expanded to include the exposure duration adjustment factor as well as specific parameters for each adjusted study (Bull et al., 2002; DeAngelo et al., 2008). The differences in CSF calculation are indeed due to the adjustment for exposure duration.

Comment 34: "Page 228: In the paragraph starting with "Lack of cancer studies for MBA", it is indicated that "A database deficiency uncertainty factor of 10 was applied in the PHG calculation to account, in part, for the potential carcinogenicity of MBA." However, on Page 208, it is mentioned that an UF of √10 is used to account for database deficiency for MBA. Is the UF for database deficiency 10 or √10 for MBA? Please clarify."

Response 34: The UF is √10; the text was corrected accordingly.

Comment 35: "Appendix A, Pages 276-281: Appendix A contains a number of equations and parameters with values. I found one equation appears to be inconsistent with the textual description. Specifically, in the top paragraph of Page 279, it is said that "The total inhalation intake (Intakeinh) for a chemical in indoor air is obtained by summing the inhalation intakes in the active state, resting state, and in the shower/bath for each life-stage, as shown in the following equation". This description makes sense. However, the equation appears to have an extra item that is not necessary. I believe the equation should look like this: Intakeinh = Cair X (BRa X ETai + BRr X ETri) + Cbath_air X BRa X ETsb. Please correct the equation or clarify the textual description. Also, please check other equations and parameters carefully."

Response 35: The CalTOX model presented in Appendix A has been peer reviewed. The peer reviewer refers to the following equation:

$$\mathbf{Intake_{inh}} = \mathbf{C_{air}} \times (\mathbf{BR_a} \times \mathbf{ET_{ai}} + \mathbf{BR_r} \times \mathbf{ET_{ri}} - \mathbf{BR_a} \times \mathbf{ET_{sb}}) + \mathbf{C_{bath\_air}} \times \mathbf{BR_a} \times \mathbf{ET_{sb}}$$

In this equation, the term '**– $BR_a \times ET_{sb}$**' mentioned in the comment accounts for decreased exposure to ambient air due to the time spent in the shower or bath, where the subject is exposed to a different air concentration (**C_{bath_air}**).

Comment 36: "Page 276, Footnote 2: Is this multimedia total exposure model freely available for the public to use? Can the authors share the link to download this model or provide full citation information of this model?"

Response 36: The CalTOX description, documentation and sources are free and available for the public to download at: https://dtsc.ca.gov/caltox-download-instructions/. In response to this comment, this link has been added to the PHG document.

Comment 37: "Page 280: Can the authors provide a citation where the equation to calculate the steady-state skin permeability coefficient is from?"

Response 37: In response to this comment, a link has been added to the PHG document. Citations to specific parameters can be found in CalTOX documentation (see Response 36).

Comment 38: "Page 408: Appendix D provides raw results of BMD dose-response analyses of the noncancer datasets. I tried to reproduce the results of one representative dataset, i.e., the first dataset presented in Figure D1. I used the latest version of BMDS (Version 3.2) that was released on August 20th, 2020. Note that according to the latest BMDS User Guide, BMDS 3.2 contains the majority of commonly used models and features that were available in BMDS 2.7, so the results should be consistent between the two versions." The peer reviewer reports his modeling results and compares them to results in the document, stating that the BMD analysis of a representative dataset on noncancer endpoints from the draft PHG document was reproducible.

"It is also worth noting that I also tried to analyze this same dataset by setting the BMR as 0.1, which is the default specific effect per US EPA guideline and the default value in BMDS 3.2, the results of BMD and BMDL became 14.857 and 12.222 mg/kg/day, respectively."

Response 38: OEHHA acknowledges that the peer reviewer is able to reproduce the BMD analysis used by OEHHA. Thus, this comment is consistent with OEHHA's analysis. Regarding the selection of an alternative BMR, see response to comment 2 above.

Comment 39: "Page 448: Appendix E provides raw results of BMD dose-response analyses of the cancer datasets. I also tried to reproduce the results of one representative dataset, i.e., the first dataset on the effect of hepatic adenomas or carcinomas in male B6C3F1 mice at 52 weeks from DeAngelo et al. (1999) presented in Figure E1." The peer reviewer reports his modeling results and compares them to results in the document, stating that the BMD analysis of a representative dataset on cancer endpoints from the draft PHG document was reproducible.

"In addition, after I changed the BMR to 0.1, the BMD and BMDL values became 157.64 and 66.19 mg/kg/day, respectively. These results are quite different from the results based on a BMR of 0.05 as presented above."

Response 39: OEHHA acknowledges that the peer reviewer is able to reproduce the BMD analysis used by OEHHA. Thus, this comment is consistent with OEHHA's analysis. Regarding the selection of an alternative BMR, see response to comment 2 above.

RESPONSES TO COMMENTS RECEIVED FROM DR. DAVID H. PHILLIPS

Comment 1: "As the indications from thorough chronic administration to rodents are that it does not induce tumour formation, the conclusion is that MCA is not carcinogenic and thus deriving a public health goal (PHG) for cancer is neither appropriate nor possible."

Response 1: OEHHA acknowledges the peer reviewer's concurrence with OEHHA to not derive the PHG for MCA based on its carcinogenicity.

Comment 2: "Overall, there is evidence of genotoxicity of DCA. Although other modes of action of DCA carcinogenicity may also be involved, a genotoxic mode of action should be assumed and applied for purposes of quantitative risk assessment."

Response 2: OEHHA acknowledges the peer reviewer's comment that a genotoxic mode of action should be assumed for the purposes of quantitative risk assessment. This is consistent with the approach taken in OEHHA's analysis.

Comment 3: "Overall, there is evidence of genotoxicity of TCA in vivo. Lack of activity in vitro suggests that TCA requires metabolic activation, by pathways that are not adequately functional in the in vitro test systems, in order to exert its genotoxicity."

Response 3: OEHHA acknowledges the peer reviewer's comment and notes it is consistent with the information in the PHG technical support document.

Comment 4: "[DBA] can be considered to be a multi-organ carcinogen in two species of rodent and in both sexes. Activity profiles such as this are more often associated with genotoxic carcinogens than with non-genotoxic ones…Overall, the genotoxic activity of DBA both *in vitro* and *in vivo* indicates that the carcinogenicity of DBA should be considered to be by a genotoxic mode of action. IARC concluded that the evidence that DBA carcinogenicity involves a genotoxic mechanism is moderate."

Response 4: OEHHA acknowledges the peer reviewer's comment and notes it is consistent with the conclusions in the PHG technical support document.

Comment 5: "Benchmark dose (BMD) modelling has been applied to the animal tumour data of the 3 carcinogenic HAAs, DCA, TCA and DBA. To the best of my knowledge, the values for the PHGs of these compounds of 0.2, 0.1 and 0.03 ppb in drinking water, as representing a human cancer risk of 10^{-6} from daily lifetime (70 years) exposure, have been arrived at by due application of the modelling calculations."

Response 5: OEHHA acknowledges the peer reviewer's conclusion that the PHG values for the carcinogenic HAAs appear to be appropriately calculated.

Comment 6: "Although the data on MBA are, for the most part, limited to genetic toxicology studies in vitro, which are insufficient to conclude whether or not MBA is

carcinogenic, the fact that it is more potently active in several assays than DCA and TCA, both of which are carcinogens, should be viewed as a warning. The lack of evidence for MBA carcinogenicity should not be taken to assume that it is not carcinogenic. Treating it as such for the derivation of a PHG, based only on its noncancer toxic effects, in the same way that MCA (a non-carcinogen) has been treated, could be considered counter intuitive. While the lack of carcinogenicity data does not permit a cancer risk to be calculated formally, a precautionary approach would be to set a PHG for MBA at a similar level to that of carcinogenic HAAs DCA, TCA and DBA, rather than at the much higher level (at least two orders of magnitude) of the non-carcinogenic chemical MCA."

Response 6: OEHHA acknowledges the peer reviewer's comment that MBA demonstrates genotoxic potential in vitro, and may ultimately be carcinogenic but available studies are lacking. To account for the uncertainty regarding the potential carcinogenicity of this compound, an additional uncertainty factor of $\sqrt{10}$ is applied in the development of the acceptable daily dose (ADD). The resulting composite uncertainty factor (3,000) was the maximum recommended value based on recommendations of CalEPA Risk Assessment Advisory Committee (1996) and the US EPA (2002a).

While MBA appears to be more potent in vitro with genotoxicity and other endpoints compared to TCA, DCA and DBA, it also appears to be metabolized and/or excreted at a dramatically higher rate in the rat compared to DCA, TCA or DBA (Saghir and Schultz, 2005). The PHG document states, "Therefore, the higher potency of MBA in toxicological mechanisms of interest… may or may not be compensated for by increased metabolic clearance in the context of in vivo toxicity."

Comment 7: "Since these five HAAs have carcinogenic and/or toxicological activity, it is highly appropriate that risk assessment be carried out and Public Health Goals established that give guidance on the levels of exposure at which a defined cancer or toxicological risk is anticipated. Regardless of whether these exposure levels are advisory or are subject to regulatory enforcement, they should not be attained if in doing so the protective properties of water chlorination in preventing bacterial infection of the public are compromised. While it is desirable to keep exposure to HAAs to an acceptably low level, this must not be done to the detriment of controlling water borne pathogens."

Response 7: OEHHA acknowledges the importance of effective drinking water disinfection, as discussed in the draft PHG document.

In requiring OEHHA to prepare health risk assessments of drinking water contaminants, Health and Safety Code section 116365(c)(1) states:

> "The risk assessment shall contain an estimate of the level of the contaminant in drinking water that is not anticipated to cause or contribute to adverse health effects, or that does not pose any significant risk to health."

Thus, OEHHA's task is to review the toxicity of the five HAAs in drinking water and to assess direct risks from exposure to these chemicals. Evaluation of effects of disinfection on microbial contaminants and its correlation with residual HAAs, as well as health risk assessment of residual microbial contamination is outside the scope of this assessment, but falls within the mandates of the State Water Resources Control Board (SWRCB), a board that, like OEHHA, falls within CalEPA.

SWRCB establishes Maximum Contaminant Levels (MCLs) in consideration of multiple factors. The risk-benefit tradeoff between residual disinfection byproducts in drinking water and exposure to microbial contaminants in drinking water is one of the factors that SWRCB considers in developing California MCLs for disinfection byproducts. The PHG document does not provide an analysis of the health benefits of disinfection of drinking water, or the efficacy of different treatment technologies that result in different profiles of disinfection byproducts. Those considerations fall to the SWRCB.

Additional language has been added to the *Risk Characterization* section to indicate that a risk-benefit analysis comparing the risk of disinfection byproduct exposure vs. microorganism exposure falls outside the scope of the PHG evaluation.

RESPONSES TO COMMENTS RECEIVED FROM DR. KAN SHAO

Comments regarding MCA

Comment 1: "In DeAngelo et al 1997, 50/sex/dose Male F344/N rats was the number of animals at the beginning of the study, but the number of animals at final sacrifice (i.e., n=23, 24, 23, 25 for the four dose levels from low to high respectively) should be used in dose-response analysis and thus presented in Table 10.2."

Response 1: In response to this comment, a footnote was added to Table 10.2 to indicate the numbers of animals in the DeAngelo et al. (1997) study at terminal sacrifice, which were used in the dose-response analysis.

Comment 2: "Additionally, the number of animals per dose group in DeAngelo et al 1997 reported in Table 5.5 is not correct."

Response 2: In response to this comment, the number of animals per dose group in DeAngelo et al. (1997) in Table 5.5 was changed to 50, and the phrase "including interim sacrifices" was added after the duration of the study to indicate that each dose group was a combined number of animals sacrificed at different times.

Comment 3: "The relative liver weight was expressed as "mean ± standard error" in Table 5 in the original publication of DeAngelo et al 1997, and directly presented in the same format in Table 10.2 here. It is very important to clearly specify how the data are expressed using a footnote, because it is directly related to the results of the estimated NOAEL level and BMD/BMDL level."

Response 3: In response to this comment, standard error values were converted to standard deviation and a clarifying footnote was added to Table 10.2.

Comment 4: "Because the relative liver weight decrease in male F344/N rats reported in DeAngelo et al 1997 was selected as the basis for deriving POD, it would be appropriate to explain why the BMD method cannot be applied to the data set in a little more detail."

Response 4: In response to this comment, the PHG document was modified. Statistical analysis in BMDS (version 2.7) indicated that neither the assumption of constant nor modeled variance was appropriate for this dataset. This clarification was added to explain why the dataset was not amenable to modeling. In fact, BMDS 2.7 produced regression with several models; however, statistical analysis of the dataset indicated that these regressions would be unreliable.

Comment 5: "Actually, based on the BMD modeling results obtained from BMDS 2.7, the failure may be mainly caused by potential issues in the modeling algorithms implemented in BMDS 2.7. It is worth trying or considering using other modeling tools for BMD analysis, e.g., the web-based Bayesian benchmark dose (BBMD) modeling

system (Shao and Shapiro 2018) which can provide plausible BMD modeling results for this data set."

Response 5: In this instance and several others, as explained below, the peer reviewer proposes the web-based Bayesian benchmark dose (BBMD) modeling system (Shao and Shapiro, 2018) as an alternative tool for dose-response analyses.

BMD modeling with US EPA's Benchmark Dose Software (BMDS) is the approach adopted by OEHHA and the US EPA in the development of guidance values. While the BBMD modeling system, and other Bayesian modeling platforms, may have utility in dose-response analyses, they would require further theoretical development and validation before they can be used in the development of guidance values such as PHGs.

As an example, one area for further research and justification (compared to BMDS) is the goodness-of-fit analysis. BMDS provides statistical measures for the appropriateness and utility of the modelling approach, as well default statistical measures of the goodness-of-fit (p-value, based on χ^2 statistic) and model quality (Akaike information criterion, or AIC), which allow an informed choice among alternative models. Best practices have been developed for model selection using the BMDS package.

BBMD modeling is a Bayesian approach and does not include frequentist statistical tests. To estimate the goodness-of-fit of the model, the BBMD method defines the posterior predictive p-value, called the "PPP" as follows: "…likelihood values calculated using the predicted responses and the original data are computed and compared; … the probability that one type of likelihood is larger than the other … is estimated. The PPP can be approximated by counting the predicted responses that satisfy the inequality out of the entire posterior sample space. A large or small *p*-value means that a discrepancy in predicted data is very likely, further indicating a poor fit. Therefore, a PPP value within the range from 0.05 to 0.95 indicates an adequate fit" (Shao and Shapiro, 2018).

Without providing further detail, Shao and Shapiro (2018) indicate that the PPP was estimated as described by Dr. Andrew Gelman, currently at Columbia University, referring to the textbook Gelman et al. (2004). Gelman et al. (2013), the most recent edition of the same textbook, outlines the general approach to evaluate the fit of a Bayesian model. However, from the available information in Shao and Shapiro (2018), it is not clear how the approach operates in practice. It remains to be determined whether the method is an advance over the existing accepted BMD approach. In comparing the BMDS and BBMD runs of the same datasets, the criteria of acceptable regression fit are more lax in BBMD, and the question remains how and to what extent it represents an improvement.

The peer reviewer suggests that the situation of "poor model fit" in BMDS should be remediated by using BBMD, which can provide plausible BMD modeling results for such datasets. However, where modeling with the most recent software (BMDS v2.7) results

in failure to meet thresholds of acceptability on statistical tests for the model, the acceptance criteria in BBMD (such as PPP value) are not clear. More development and review is required, such as comparative analysis of BMDS and BBMD methods, before BBMD can be considered as an alternative to BMDS. OEHHA looks forward to exploring the use Bayesian modeling tools, such as BBMD, in future assessments. However, a detailed analysis and comparison of BMDS and BBMD is beyond the scope of this assessment, which is aimed at establishing PHGs for HAAs.

Additional language was added in the *Dose-Response Assessment* section to note OEHHA's consideration of BBMD.

Comment 6: "In Table 10.3, presenting "Goodness-of-fit p-value" as the criteria for model selection is insufficient, and the AIC value should be reported as well. Higher goodness-of-fit p-value typically indicates a better fit, however, for BMD analysis using BMDS, p-value is not used for model selection but an indicator demonstrating if the model can fit the data adequately. AIC value is used to compare different models by considering both goodness-of-fit and the complexity of the models. For example, when analyzing the increased mortality in female rats reported in NTP (1992) (i.e., 0/53, 4/53, 12/53) using BMDS 2.7, the LogLogistic model had a p-value of 1.0000 (higher than Quantal-linear model's 0.6395), but a higher AIC value than the Quantal-Linear model (89.0625 vs. 88.0049). AIC was the essential reason why the Quantal-linear model was picked."

Response 6: In response to this comment, a footnote has been added to Table 10.3, which is a summary table and does not present details of the BMD analysis. This table includes p-values as a measure of confidence that the presented models fit the data sufficiently well and a footnote has been added to Table 10.3 to indicate this. While AIC values are also important and are used to select the best of all models of acceptable fit for the same endpoint, presenting AIC values in this table would be out of context since only the best model for each endpoint is presented. In contrast, the model p-values do not require such comparison and even when presented by themselves, add to the understanding of the analysis.

.**Comment 7:** "In short, it is adequate to use systemic toxicity (i.e., decreased relative liver weight) reported in DeAngelo et al 1997 as the critical endpoint to derive the POD and ADD. However, because of the importance of this data set, it is worth putting extra effort (e.g., using other BMD modeling tools to derive BMDL estimate) to justify the plausibility of using 3.5 mg/kg-day (i.e., the NOAEL) as the POD even though the BMD analysis may provide a similar result."

Response 7: OEHHA acknowledges the peer reviewer's comment that the use of decreased liver weight in the DeAngelo et al. (1997) study is an adequate basis for the derivation of the MCA POD, as was done in OEHHA's analysis. To determine the POD for this dataset, OEHHA followed the BMDS and US EPA guidelines (Davis et al., 2011; USEPA, 2012), and considers the NOAEL as the appropriate choice. Also, as stated in

response 5, BBMD needs to be further evaluated before it can be used as an alternative to BMDS in OEHHA's risk assessments.

Comments regarding DCA

Comment 8: "The selection of 7.6 mg/kg-day as the POD [for increased relative liver weight] using the NOAEL method is not very well supported by the evidence for a few reasons:

a. The conclusion of "poor model fit" presented in Table 10.5 for this data set is mainly based on the goodness-of-fit p-values reported in EPA's BMDS for various dose-response models. It has been questioned by a few experts that whether the goodness-of-fit p-values for continuous data are correctly calculated in BMDS. Based on the fitted dose-response models visualized in BMDS, the model fits (including the Hill model, and several models in the Exponential model family) are reasonably well. In addition, all eight models can be appropriately fitted in the BBMD modeling system.
b. The study design (especially dose placement) makes the estimated NOAEL relatively conservative. The p-value of dose group 77 mg/kg-day is 0.023, just a little smaller than the cut-off line 0.05. Therefore, a no-effect-level should be a little below 77, but not an order of magnitude smaller than 77.
c. Using the BBMD system to analyze the data presented in Table 1 with a BMR = 5% (i.e., 5% increase in central tendency of response comparing to the control), the estimated BMD and BMDL are 20.6 and 19.0 mg/kg-day given by the best fitted model (i.e., Exponential 2), and 92.2 and 63.7 mg/kg-day using the model averaging technique. These results are more consistent with the outcomes reported in DeAngelo et al (1999)."

Response 8: Table 10.5 is a summary table and does not present details of the BMD analysis (detailed model outputs can be found in Appendix D). This table includes p-values as a measure of confidence that the presented models fit the data sufficiently well and a footnote has been added to indicate this. While adequate visual fit is essential for model choice in BMDS, it is not the only criterion for choosing a model (Davis et al., 2011; USEPA, 2012). OEHHA's BMD model selection criteria are listed at the beginning of Appendix D of the PHG document.

As detailed in Response 5, the Bayesian-based BMD (BBMD) method for data fitting is in an earlier phase of development and acceptance than BMDS. In this dose-response analysis, OEHHA followed the US EPA and BMDS guidelines (Davis et al., 2011; USEPA, 2012) and considers the NOAEL as the appropriate POD. While the BBMD shows promise as a new tool for data fitting, additional validation steps would be required to characterize the utility of this method for BMDL derivation suitable for PHG development.

Footnote c in Table 1 in DeAngelo et al. (1991) indicates statistical significance of $p<0.02$ for the 77 and 486 mg/kg-day dose groups (shown as 0.5 and 5 g/L, respectively) but not for the 7.6 mg/kg-day dose group (shown as 0.05 g/L). OEHHA

conducted a pairwise statistical comparison of mean values using the Student's t-test, which generally produces comparable results to more involved methods, such as Tukey's test. Using the t-test, the 7.6 mg/kg-day dose was not significantly different from control (p=0.4057), while the next dose (77 mg/kg-day) was significantly different from control (p=0.0127). These conclusions are similar to those reported in the DeAngelo et al. (1991) report. Therefore, 7.6 mg/kg-day was identified as a NOAEL for this endpoint in this study.

The "true" no effect level of the study lays somewhere between 7.6 and 77 mg/kg-day doses. One consideration in using the p-values for the respective pairwise comparisons as indicators of where this value may be located is that the standard deviation (relative liver weight = 6.83±1.92, mean ± standard deviation) at the 77 mg/kg-day dose is notably larger than the control (5.01±0.32) or the low dose (5.25±0.78) group. Thus, statistical comparisons of relative liver weight between the low dose and control would clearly result in lower p values. The effect of dose group spacing on POD determination using the NOAEL/LOAEL method is a disadvantage in comparison to a BMD model-driven dose-response, but it is standard practice when no acceptable BMD model can be generated.

As detailed in Response 5, OEHHA has determined it is premature to use the BBMD method as an alternative to the BMD method that is currently used by OEHHA and US EPA as a standard approach of establishing PODs at this time.

Comment 9: "[I]t seems that the input data used in this BMD analysis were not correct. The relative liver weight of male dogs reported in Table 6 in the original publication of Cicmanec et al (1991) is most likely expressed as "mean and SD" instead of "mean and SE" (given the reporting style used in Tables 1 and 2 in that paper). Therefore, the data can be directly used in BMDS software without conversion. Consequently, the results reported in Table 10.5 for relative liver weigh increase in male dogs in Cicmanecet al 1991 should be updated."

Response 9: In response to this comment, OEHHA double checked the paper to confirm that the numbers used for BMD modeling were correct. The data from Table 6 of the Cicmanec et al. (1991) were used as is for BMD modeling, thus it is not necessary to update Table 10.5, which presents the results of the dose-response analysis based on reported mean ± standard deviation values.

Comment 10: "The last row in Table 10.5, i.e., the BMD estimates based on testicular degeneration data in Cicmanec et al 1991, should be noted that the estimated BMD/BMDL were not based on BMR=1SD."

Response 10: In response to this comment, a footnote was added to Table 10.5 to clarify that the BMD and BMDL in the last row were determined with a BMR of 0.05.

Comment 11: "It is not clear why only 1st degree multistage (i.e., LMS) model was used to model the DCA candidate cancer data sets. As described in the "Dose-

Response Model" section on Page 213, the Multistage-Cancer model, which can have as many parameters as the number of dose groups, is used for modeling cancer endpoints. However, no justification was provided to explain why only LMS was applied for the data sets presented in Table 10.12. The LMS is not always the best-fit model in the Multistage-Cancer model family. For example, for hepatic adenomas or carcinomas in male B6C3F1 mice at 52 weeks reported in DeAngelo et al. (1999), the 3rd degree multistage model (parameters q1 and q2 were reduced) has better fitting performance (P-value: 0.5510, AIC: 41.3154) than the LMS (P-value: 0.3673, AIC: 41.789)." Similar comments regarding the use of the 1st degree multistage model were also made for TCA, and DBA.

Response 11: The term linearized multi-stage (LMS) model applies to all models of the form '$p(d) = \beta + (1- \beta) \times \exp[-(q_1 d + q_2 d^2 + ... + q_i d^i)]$', and not just the 1st degree model.

In modeling cancer datasets, OEHHA uses all available LMS models (at least 1st degree LMS and 2nd degree LMS for a 3-dose study). As noted in the *Cancer Dose-Response Analyses and Cancer Potency Derivation* section of the PHG document, among the LMS models with acceptable fit ($p>0.05$), it is OEHHA's policy to choose the model with the fewest parameters based on the scientific principle of parsimony, which is consistent with BMDS guidance. In this case, the 1st degree LMS provided acceptable fit to the DCA cancer dataset, and therefore was chosen as the preferred model.

Comment 12: "The cancer slope factor was estimated based on the male mice hepatic tumor data (52-100 weeks) reported in DeAngelo et al 1999. However, unlike other data sets listed in Table 10.12, the Multistage Weibull (MSW) model was applied to analyze this data set and the estimated BMDL was used to derive the cancer slope factor. Although the report provided explanation on using the MSW model instead of Multistage model (i.e., adjust tumor rates for possible underestimates due to early treatment-dependent mortality), the method is still not well justified. The main reason is that no statistics were provided to evaluate how well the MSW model fit the data (no p-value or dose-response plot), and adequate fit is very important when applying the BMD method."

Response 12: In response to this comment, an additional statement was added to the cancer dose-response indicating that US EPA's MSW model does not provide a p-value or scaled residuals. In training sessions given by BMDS technical staff, "[u]sers are advised to choose the simplest adequate model (i.e., the model with the lowest AIC value that still affords a reasonable fit to the data)."[3]

Comment 13: "Additionally, it seems that the MSW model has been removed from the BMDS 2.7, the reason to adopt the results generated from an earlier version of BMDS should be discussed."

[3] https://clu-in.org/conf/tio/bmds/slides/BMDS_Cancer_Models.pdf

Response 13: The text of the technical support document was modified to address this comment. US EPA's MSW model is not included in the BMDS suite of models. It is a separate executable file that can be downloaded, along with its user manual and technical documentation files, from US EPA at https://cfpub.epa.gov/ncea/bmds/recordisplay.cfm?deid=217055. A footnote with the URL for this model can be found in the *Cancer Dose-Response Analyses and Cancer Potency Derivation* section of the PHG document.

Comment 14: "In short, the selection of critical studies and endpoints for deriving the POD and cancer slope factor are appropriate…"

Response 14: OEHHA acknowledges the peer reviewer's concurrence with the selection of studies and endpoints for the analyses leading to the PHG for DCA.

Comments regarding TCA

Comment 15: "The second paragraph on Page 206 is the first time explaining why a BMR of 5% was chosen for noncancer dichotomous endpoints (5% BMR has also been used for cancer endpoints in this report). This justification should be moved to the section of "Point of Departure" on Page 194 where "5% BMR" was first mentioned in Chapter 10."

Response 15: In response to this comment, the justification for use of BMR 5% has been moved to the *Point of Departure* section.

Comment 16: "Actually, using 5% BMR for dichotomous data BMD modeling is relatively conservative. "When data were expressed as counts of dichotomous endpoints, the NOAEL was approximately 2–3 times higher than the BMDL for a 10% probability of response above control values and 4–6 times higher than the BMDL for a 5% excess probability of response." (US EPA, BMD Technical Guidance, 2012). Moreover, EFSA pointed out "the size of the estimated effect at the NOAEL is, on average over a number of studies, close to 10% (quantal responses) or 5% (continuous responses)" (EFSA, 2017). Therefore, the report should discuss the rationale to use 5% BMR as default choice for dichotomous data of both non-cancer and cancer endpoints, which may result in relatively conservative POD estimates."

Response 16: OEHHA's guidance on BMR selection is described in OEHHA (2008). In response to this and other comments, OEHHA added further details in the *Dose-Response Assessment* section of the technical support document regarding the choice of BMRs for the BMD modeling. Also, see Responses 2 and 3 to comments from Dr. Zhoumeng Lin above.

Comment 17: "To keep the format consistent in this Chapter, the column name of the last two columns in Table 10.7 should be named as "BMD05 (mg/kg-day)" and "BMDL05 (mg/kg-day)."

Response 17: In response to this comment, columns in Table 10.7 were renamed as suggested.

Comment 18: "When relative liver weight is expressed as "5.3 ± 1.0" in Table 10.6, a footnote should be used to clarify that it is expressed as "Mean ± Standard Deviation."

Response 18: In response to this comment, the clarifying footnote to Table 10.6 was added.

Comment 19: "Additionally, like what have been mentioned in the comments above, the "poor model fit" was primarily determined by the "p-value" reported in BMDS 2.7, which was sometimes contradicted by the dose-response plot produced by BMDS itself and modeling results provided in some other software."

Response 19: As noted in Response 6, p-values are shown in the BMD modeling results tables as a measure of confidence that the presented models fit the data sufficiently well. When determining acceptable model fit, other criteria are taken into account, including visual inspection of the output plot, whether variances are adequately modeled, if scaled residuals are less than the absolute value of 2, and whether a p-value can be calculated (e.g., when degrees of freedom = 0). Regarding results provided by "some other software," it is current OEHHA policy to perform dose-response analyses with US EPA's Benchmark Dose Software.

Comment 20: "In short, the selection of critical studies and endpoints for deriving the POD and cancer slope factor appears to be appropriate…"

Response 20: OEHHA acknowledges the peer reviewer's concurrence with the selection of studies and endpoints for the analyses leading to the PHG for TCA.

Comments regarding MBA

Comment 21: "Given the very limited data that are available for dose-response assessment for MBA, the NOAEL identified in the report and used as the POD for deriving ADD is plausible."

Response 21: OEHHA acknowledges the peer reviewer's concurrence with the selection of studies and endpoints for the analyses leading to the PHG for MBA.

Comments regarding DBA

Comment 22: "The dose-response analyses for both non-cancer and cancer endpoints of DBA are adequate. The report provided sufficient arguments to justify why the LOAEL value from Veeramachaneni et al (2007) was chosen as the POD over the BMD estimates of four non-cancer endpoints reported in Table 10.10."

Response 22: OEHHA acknowledges the peer reviewer's concurrence with the selection of studies and endpoints for the analyses leading to the PHG for DBA.

Comment 23: "A column of AIC values should be added in Tables 10.3, 10.5, 10.7, and 10.10. For dichotomous data of non-cancer endpoints, no model is considered as a default. Thus, all dichotomous dose-response models should fit the data, and then the AIC value is used to compare different models and select the most appropriate one."

Response 23: OEHHA agrees that the AIC value should be used as one criterion for comparing and choosing an appropriate model from different models for the same dataset. However, Tables 10.3, 10.5, 10.7 and 10.10 present candidate critical endpoints, each with its own distinct dataset and its corresponding best fit model, if one exists. The purpose of these tables is not to show alternative models for each dataset and that is why AIC values are not included.

Comment 24: "To keep the format of Tables 10.12, 10.14, and 10.15 consistent, a p-value for the fitted model should be reported in these tables."

Response 24: In response to this comment, Tables 10.12, 10.14 and 10.15 were edited to contain p-values, where applicable. MSW and multi-site BMDS results do not have p-values reported.

General comment

Comment 25: "To sum up, the draft PHGs for Haloacetic Acids in drinking water were derived based on comprehensive literature review and sophisticated analytics using scientifically solid methods and practices. Properly addressing issues mentioned in comments above can improve the quality of the report. No other scientific subjects need to be discussed or described."

Response 25: OEHHA acknowledges the peer reviewer's comments stating that no other scientific studies need to be discussed or described, that OEHHA's literature review was comprehensive, and that OEHHA used scientifically valid methods.

RESPONSES TO COMMENTS RECEIVED FROM DR. PAUL WHITE

Comment 1: "Based on my expertise and experience, I can say with confidence that the document comprises an impressive, scholarly review of the scientific literature, suitably augmented with rigorous analyses of available dose-response data, followed by judicious interpretation of PoD (Point of Departure) values (i.e., BMD or Benchmark Dose) for determination of the "level of a contaminant in drinking water at which adverse health effects are not expected to occur from a lifetime of exposure…". The health protective drinking water concentrations, which constitute the PHGs, are calculated using suitably-adjusted non-cancer ADD values and suitably-adjusted carcinogenicity CSF (cancer potency) values. There is no question that the overall analysis and interpretation is rigorous, particularly with respect to the determination of cancer health-protective values, for which determination of robust BMDL and CSFhuman values can be complex and challenging."

Response 1: OEHHA acknowledges the peer reviewer's comment, including his finding that the PHGs, are calculated using suitably-adjusted non-cancer ADD values and suitably-adjusted carcinogenicity CSF (cancer potency) values.

Comment 2: "Nevertheless, despite the overall quality of the document, there are some noteworthy shortcomings, and consequent room for improvement. As outlined in the comments below, I am primarily concerned about (1) careful and judicious evaluation of available genotoxicity test results to determine the strength of the evidence to support a genotoxic MOA underlying the carcinogenicity of DCA, TCA and DBA, and (2) the criteria used to evaluate the regulatory suitability/utility of available studies; moreover, the utility of PoD values (i.e., BMDs) determined via quantitative dose-response analyses."

Response 2: The peer reviewer's comments regarding these two main issues are addressed in detail below (Responses 3, 5-48, 53).

Comment 3: "With respect to #1, although it is not necessarily unreasonable to conduct risk assessments for DCA, TCA and MBA that assume a genotoxic MOA for carcinogenicity, doubts specified in the literature should be outlined and itemised. This is particularly necessary for DCA, since the literature contains explicit statements regarding a lack of sufficient evidence to justify a genotoxic MOA for exposures to levels present in finished drinking water, i.e., "….not considered to play a primary role in its carcinogenicity" (see below)."

"…both DCA and TCA elicit positive responses on the Mouse Lymphoma Forward Mutation Assay, i.e., both yield responses that exceed what is referred to as the GEF or Global Evaluation Factor [14]. However, although positive, suitably elevated responses are only observed at very high doses. … '[I]t seems unlikely that [DCA] would be mutagenic (or possibly carcinogenic) at the levels seen in finished drinking water.' … Thus, "the authors need to carefully consider whether a cancer risk assessment for DCA that is based on a mutagenic MOA can be robustly justified [1]."

Response 3: OEHHA agrees that the hypothesis of DCA exerting genotoxic action only at high concentrations is a legitimate concern that may affect the MOA determination for DCA at concentrations in drinking water.

In response to this comment, text was added to the *Genetic Toxicity* section for DCA to address the plausibility of DCA genotoxicity at lower doses, as part of the general reworking of this section to address this and related comments. Briefly, some earlier reviews, such as Richardson et al. (2007) (indicated as reference [1] in the comment) postulated the hypothesis that DCA genotoxicity was observed in vivo only at high doses and would not be mutagenic at levels seen in drinking water. However, multiple studies (published after 2007) reported DCA genotoxicity at low concentrations in vitro and at low doses in vivo (e.g., micronucleus induction in human peripheral lymphocytes at 25-100 µg/ml in Varshney et al. (2013) and DNA single strand breaks in hepatic tissue in mice treated with 12.5-50 mg/kg-day sodium dichloroacetate for 13 weeks in Hassoun et al. (2014)). Given this additional evidence, it is not possible to exclude a genotoxic MOA for DCA in the low dose range.

This comment highlights the results of the mouse lymphoma cell forward mutation assay, as indicative of the lack of genotoxic response at low doses. This experimental system was used in Harrington-Brock et al. (1998). While this report designates DCA as a "weak direct-acting mutagen in mouse lymphoma cells", it also notes that "a dose-related [ ] mutagenic effect [ ] was observed at concentrations between 100-800 µg/ml" (Harrington-Brock et al., 1998), which is the full range of applied concentrations. Thus, this report does not appear to provide evidence of the lack of genotoxic response at lower concentrations, but does lend support to a genotoxic MOA.

Comment 4: "Regarding the dermal exposure discussion on p. 21, the criteria underlying the statements, and the level of uncertainty, could be mentioned. Estimations using EPA-recommended methods indicated that dermal dose is negligible, but what are the criteria underlying this determination, and what is the degree of uncertainty? Obviously related to vapour pressure, Henry's law constant and pKa. Not requesting much additional detail, perhaps another sentence or two to elaborate. Later on p. 23, the document mentions skin permeability coefficient Kp; perhaps this should be mentioned on p. 21?"

Response 4: In response to this comment, the text has been edited to clarify that OEHHA did not estimate dermal dose but was citing US EPA's conclusion regarding the potential for dermal exposure.

Comment 5: "In numerous places, the authors refer to data that "were amenable to dose-response modelling"; moreover, the suitability of BMD values for determination of ADDs and CSFs. This is not necessarily problematic; however, the reader is not provided with explicit statements regarding the criteria employed to determine dataset and/or BMD utility/suitability."

Response 5: In response to this and other comments, the *Point of Departure* section of the document was expanded to include specific criteria for dose-response modeling and choosing the best model. In the draft PHG document, the term 'amenable to dose-response modelling' is used to indicate the fact that BMDS analysis is appropriate for the given dataset and that BMD modeling should be performed for POD determination. In contrast, for datasets described as "not amenable to dose-response modeling," it means the data were not appropriate for BMD modeling, for example, if the dose-response was non-monotonic or there was an unusually high response rate in all dose groups, and thus modeling was not done. Similarly, "poor model fit" indicates that BMD modeling was done on a given dataset, but the results did not meet acceptability criteria, as outlined in the PHG document. This has been clarified in the document.

Comment 6: "I assume that the authors followed the criteria outlined in Section 2.1.5 of the USEPA's 2012 Benchmark Dose Technical Guidance Document [2], e.g., a statistically significant response with dose-related trend, a dataset containing information on the dose-response relationship between the control and maximum dose, etc. With respect to the latter, for example, that the analysed dataset includes responses between the control level and the level associated with the maximum dose, that the non-control doses do not elicit responses that are all essentially the same, etc. Did the authors follow the flowchart provided in USEPA (2012), i.e., Fig 2A on p. 16? If yes, this should be explicitly mentioned. Perhaps tables summarizing the results of BMD analyses should include a column indicating the suitability of the dataset, and the feasibility of determining a reliable BMD, i.e., according to criteria specified by the EPA? For example, insufficient dose-groups, no evidence of dose-related trend, etc."

Response 6: In response to this comment, the document has been modified to be clear for each applicable case why modeling was not done. OEHHA has established criteria for determining the suitability of data for BMD modeling that are comparable to US EPA guidelines, so although the flow chart may not be followed exactly, the general principles are similar. While factors such as number of dose groups, evidence of dose-related trend, etc. are considered in deciding whether a dataset should be modeled. The text now clarifies that these criteria have been met for the modeled datasets displayed in the tables summarizing BMD modeling results and that datasets not modeled did not meet the requisite criteria.

Comment 7: "In my opinion, the text on pp. 194-195 pertaining to PoD determination needs to be expanded, i.e., need to provide the reader with much more information on the BMD analyses approach and methodology (e.g., model selection criteria, goodness-of-fit evaluation, BMD suitability for regulatory decision-making, criteria for dataset exclusion, etc.). Some information is provided at the top of p. 408, but it is inadequate. Readers will almost certainly want more detail, presumably with explicit reference to the data analysis and interpretation criteria outlined in the aforementioned EPA Guidance Document (2012)."

"As noted, the authors' rationale for selecting the PoD for ADD determination needs to be clearly delineated, presumably in summary tables like Table 10.5. Perhaps the

authors could insert a comment column within which PoD suitability could be indicated and suitably justified."

Response 7: The text on pp. 194-195 was expanded to include requested details on dose-response modeling and choosing the best model (see also Response 5 to Dr. White).

Comment 8: "The authors do not comment on the precision of the BMD values employed to determine ADD and CSF values. Although not specified in the aforementioned technical guidance document, numerous researchers have employed the BMDU-to-BMDL ratio, or BMD-to-BMDL ratio, as an indicator of BMD precision; moreover, as an indicator of a BMDs suitability for regulatory decision-making. Although the EPA and EFSA guidance documents do not explicitly address the utility of the BMD:BMDL ratio [2, 3], the BMDS Wizard software employs BMD:BMDL for model choice decisions and BMD uncertainty evaluation [4]. …The authors are strongly encouraged to examine the BMDU:BMDL ratio, or the BMD:BMDL ratio, as an indicator of BMD precision and utility. I would even recommend including the metric in all tables summarizing BMD values."

"The authors need to take care with respect to the criteria employed to indicate that BMDs are unsuitable. I suggest using BMD precision to guide statements about regulatory suitability/utility. Of course, there will be instances where the model fit is unacceptable due to, for example, all non-control doses eliciting the same response [2]."

"Again, the authors should be using a metric like BMDU:BMDL ratio or BMD:BMDL ratio to evaluate BMDL precision and its suitability for regulatory evaluations and decision-making."

Response 8: In response to this comment, OEHHA's model selection criteria are now outlined in the *Dose-Response Assessment* section. That section notes that OEHHA takes into account the BMD:BMDL ratio as one of the criteria in considering acceptable model fit. It is also described in the text when it is part of the rationale for rejecting a BMDL for use as a POD. Other considerations used in choosing the best model and data set to be used for POD determination are described as they are applied to particular cases in the technical support document.

Comment 9: "Similarly, the authors note that BMD values are not suitable when the BMD is below the lowest tested dose. What is this criterion based on? To my knowledge, there is no explicit statistical or theoretical reason to deem such BMDs unsuitable. Granted, the model fit may be unacceptable, and the BMD precision so low that the value is deemed unsuitable for regulatory evaluations and decision-making. Even EPA (2012) indicates on p.15 "…. dose spacing and the proximity of the BMR to the observed response level will influence the uncertainty in the BMD estimate". It does not say that the BMD is necessarily unsuitable, it only says that BMD precision will likely be low."

“An example that illustrates my concerns is the DCA testicular degeneration study of Cicmanec et al. (1991) (see pp. 202-203 and pp. 432-433). On p. 203 the authors indicate that the BMD is “not useful as a PoD”; the rationale is “BMD05 and BMDL05 are much lower than the low dose…., indicating very large uncertainty in the model prediction and extrapolation to the low end of the dose response”. In actuality, the proximity of the BMD/BMDL to the lowest dose does not necessarily have any bearing on PoD uncertainty. What matters is the BMD:BMDL ratio or BMDU: BMDL ratio. As noted earlier, Haber et al (2018) noted that “the default settings for BMDS Wizard indicate that a BMD:BMDL ratio of >20 results in a model being placed in the questionable bin, and a ratio of >5, results in a caution flag”. In this case, the BMD:BMDL ratio is 2.26, indicating that the PoD can likely be regarded as suitable for regulatory decision-making.”

Response 9: In response to this comment, the text has been changed to read, “In this case, the BMD values are not suitable because the BMDL is far below the lowest tested dose, 70-fold lower. Moreover, the response levels all non-control doses had plateaued.”

Comment 10: “In general, as indicated above, the authors need to judiciously outline the criteria used to assess the overall suitability of BMDs for regulatory decision-making; moreover, explicitly indicate the precision of BMD values associated with acceptable datasets and acceptable analyses (e.g., model fits).”

Response 10: In response to this and other comments, criteria used to determine if the model provides an acceptable fit for the dataset and specific model selection criteria have been added to the *Dose-Response Assessment* chapter of the PHG document.

Comment 11: “With respect to the discussion of the DeAngelo et al (1996) testicular weight data on p. 203, in this case the preclusion of BMD modelling is also acceptably justified, i.e., non-monotonic dose-response. But suitable justification often not provided.”

Response 11: OEHHA acknowledges the agreement with the justification provided for not modeling the DeAngelo et al. (1996) testicular weight data. In response to this and other comments, OEHHA provided explanations as to why particular datasets were not modeled.

Comment 12: “With respect to extrapolation below the lowest experimental dose, the authors may be interested to know that Slob et al (2005) noted that high dose effects can actually be helpful for estimating the doses associated with small effects (i.e., the BMD) [7].”

Response 12: OEHHA generally uses all dose levels in its models and appreciates the peer reviewer bringing this paper to OEHHA’s attention.

Comment 13: "I have similar concerns regarding the authors' statements about the BMDL1SD for relative liver weight based on the male Beagle dog dose-response data from Cicmanec et al (1991). Again, the authors note that the value cannot be used as a PoD because of the "high uncertainty of extrapolation outside the range of experimental observations". Not only is there no evaluation of BMD uncertainty, but there is no statistical or theoretical reason why a BMDL below the lowest experimental dose cannot be used as a PoD. In my experience, this happens very frequently, and in most cases the BMDL precision is entirely acceptable."

Response 13: In response to this comment, explanatory text was added. See also Responses 9 in this section.

Comment 14: "What is more concerning about this particular dataset is the similarity of the responses at all non-control doses."

Response 14: OEHHA agrees. Responses for all non-control doses were elevated above control values at similar levels and this is one reason why OEHHA deemed the BMDL questionable and did not use it for a POD.

Comment 15: "In general, as already stated, the authors need to do a better job convincing the reader that the selection of PoDs for ADD calculation is sound and justifiable. In numerous cases, I am not convinced. There are several reasons; a primary reason is the authors repeated referral to BMD uncertainty without any attempt to calculate a metric indicative of precision and uncertainty."

Response 15: In response to this and other comments, further explanation on PoD selection was added to the document. See also responses to Comments 8-10 in this section

Comment 16: "Similarly, I am not entirely convinced that the need for an additional UF is a sound basis for PoD ranking and selection."

Response 16: The need to include an additional UF is one of several factors considered and is not a stand-alone criterion for POD ranking and selection. The need for an additional UF indicates higher overall uncertainty. All else being equal, if one POD requires such an operation, and the other one does not, the latter case would have lower overall uncertainty for ADD (and eventual) PHG determination.

A brief statement outlining this position was added to the *Study and Endpoint Selection* section of the PHG document.

Comment 17: "All this being said, in some instances, the criteria used to select a PoD for regulatory decision-making are in fact clearly delineated. For example, on the top of p. 205, the authors clearly outline model selection criteria for the results presented in Table 10.7. This type of statement about criteria used to select models, and evaluate PoDs, should be explicitly provided. I suggest a clear delineation at the start of Section 10 (p. 194)."

Response 17: In response to this comment, text in the *Point of Departure* section was expanded to include the details suggested by the peer reviewer (see also Response 5).

Comment 18: "The authors should take care regarding use of language such as "better study" (i.e., p. 203). How is better defined? Are the authors trying to say that the route of exposure employed in the Cicmanec et al (1991) study is not suitably aligned with calculation of an ADD that can convincingly be employed to determine a drinking water PHG?"

Response 18: In response to this comment, the text has been expanded to better explain why DeAngelo et al. (1991) is the preferred study.

The draft PHG document stated: "Compared to the Mather et al. (1990) and Cicmanec et al. (1991) studies, the DeAngelo et al. (1991) study was chronic in duration, employed a greater number of animals per dose and animals were exposed to drinking water rather than gelatin capsules as in Cicmanec et al. (1991), making it a better study." This language has been further developed.

The intent was not to suggest that the exposure via gelatin capsule is not suitably aligned with the determination of a drinking water PHG. When taken into context, DeAngelo et al. (1991) was better because it was chronic in exposure duration (and other studies were not), it employed a greater number of animals per dose than other studies, and the route of exposure (via drinking water) was closer to human exposure from drinking water. Route of exposure (e.g., capsules vs. drinking water) would affect absorption, exposure duration during the day, and plasma concentration profile for DCA, which is readily metabolized. In this instance, internal metrics of exposure would be more consistent between the test animals and humans with the similar route of exposure, drinking water.

Comment 19: "…on p. 204, the authors note that some TCA non-cancer studies "provide non-cancer datasets of acceptable quality for dose-response analysis". Again, the basis for the adjective "acceptable" is not clearly specified."

Response 19: The text has been modified to address this comment. The *Chronic Toxicity in Animals* section provides study descriptions and limitations. Studies considered unsuitable for dose-response analysis were indicated, and the reasons for this determination were described. The term "acceptable" is also further described.

Comment 20: "Similar concerns about the use of the term "poor" on p. 205. In this case, in accordance with the USEPA Technical Guidance document (i.e., p. 33), the statement presumably indicates chi-squared $p<0.1$. If yes, this should be clearly indicated.**"**

"Similar to the concerns noted above for DCA and TCA, I am also concerned about statements on p. 211 (i.e., section on MBA dose-response analyses) such as "amenable to BMD modeling" and "ideal for producing reliable BMDL estimates.

Response 20: In response to this and other comments, the meaning of these terms has been further explained in the text. The document explains that there are multiple reasons why a BMDS model would be considered unacceptable ("poor model fit"), of which the chi-squared p-value (e.g., $p < 0.1$) is one. The document further explains that the phrase "amenable to BMD modeling" and similar statements mean that the dataset is appropriate for BMD modeling (as determined by the criteria outlined in the *Point of Departure* section of the PHG document), and generally conforms to US EPA's BMD guidelines (US EPA, 2012). Also, see response to Comment 5 in this section.

Comment 21: "Nevertheless, even in circumstances where $p<0.1$, the BMD Technical Guidance document states "Some of these less adequate fits may be satisfactory when other criteria are taken into account (including the nature of the variability of the endpoint, visual fit, and residuals in the most relevant region of the data range); expert judgment is useful in these cases". Was expert judgement employed, or are the entries in Table 10.7 merely based on the chi-squared test for goodness of fit?"

Response 21: In response to this comment, further explanation was added. As outlined in Appendix D of the PHG document, "Model selection criteria when comparing outputs of different models for the same endpoint/dataset were: the lowest Akaike's information criterion (AIC), goodness of fit p-value ≥0.05, scaled residual ≤ the absolute value of 2, and visual inspection of the dose-response curve." A paragraph including these criteria was also added to the *Point of Departure* section of the PHG document to ensure that OEHHA's model selection criteria are clear. Though not explicitly stated, expert judgment is also employed on a case-by-case basis.

Comment 22: "With respect to the interpretation of the [D]BA dose-response data more specifically, I cannot understand how high incidence rate in control animals can be used to discount the female rat nephropathy results from NTP (2007). In this case, the BMD:BMDL ratio is <2 (p. 441); thus one could argue that the BMDL can be used for calculation of an ADD. Granted, the authors may still be able to justify the use of the male reproductive toxicity LOAEL from Veeramachaneni et al (2007), particularly since the NTP (2007) neuropathy BMDL would yield an ADD of 0.0021 mg/kg-day (i.e., 0.62/300)."

Response 22: In response to this comment, further evaluation of the female rat nephropathy data was conducted and the document was modified.

The rat nephropathy reported in the NTP (2007) study likely represents chronic progressive nephropathy (CPN) based on the similarity of the described symptoms (thickened basement membrane, glomerular thickening, etc). CPN is commonly observed in rats, is well studied and has complex etiology and multiple factors driving its spontaneous occurrence. It has been argued that rat CPN does not represent a relevant adverse effect for human health (Hard et al., 2009). One of the proposed criteria for assessment of CPN in rats is treatment-dependent severity; however, in the female rat dataset (NTP, 2007) lesions were ranked as minimum to mild, and no dose-dependent severity was observed.

Rat strain differences in CPN are well-established, with Fisher 344 and Sprague-Dawley rats reported as possessing the highest rates of CPN (Hard et al., 2009, and references therein). CPN is commonly viewed as an 'old rat' disease, yet Fisher 344 and Sprague-Dawley rats start developing CPN at 2-3 months of age (Hard et al. 2009, and references therein). However, among the three available DBA studies in rats, only one study (the NTP 2-year study) found a treatment-dependent increase in CPN. Conversely, the >90 day Christian et al. (2002) and the 90-day segment of the NTP (2007) studies found no significant association with increased CPN. Based on several considerations - lack of consistency among studies, mild severity, lack of dose-dependent increase in severity and high background in the only positive study – OEHHA decided not to use CPN in female rats as a candidate critical endpoint. This endpoint has been removed from consideration and the justification is found in the *Dose-Response Assessment* section.

Comment 23: "BMD analysis for continuous responses (e.g., body weight or organ weight) employed a BMR of 1SD above control. Although this is commonly used, it is fraught with problems, and some guidance documents recommend 5% increase above control [3, 8]. As noted in White et al. (2020), "The 1SD approach has been criticized, particularly for endpoints with low response variability whereby it is unlikely that a 1SD change from control (i.e., background) could be deemed adverse (Haber et al. 2018). Conversely, for endpoints with high control variability, the 1SD approach will yield larger CES [BMR] values, that is, the percentage increase corresponding to a 1SD increase above control will be relatively large. Larger CES [BMR] values will yield larger BMD and BMDL values, which may be less desirable from a regulatory point of view (i.e., less restrictive)" [4, 5]. I recommend that the authors note these issues; moreover, that other approaches may have merit."

Response 23: OEHHA acknowledges the peer reviewer's concern about applicability of guidelines, such as specific BMR values for dose-response analysis. In choosing the BMR, OEHHA follows its own (OEHHA, 2008) and US EPA's (US EPA 2012) BMDS guidelines. These suggest using 1SD as a default BMR for continuous datasets, with additional considerations in special types of studies. Following the OEHHA and US EPA guidelines, a BMR of 1SD (one standard deviation) is appropriate when there is not a clear consensus on the degree of change that is adverse. In cases of extremely high variation (given as a hypothetic example in the peer review comment), OEHHA would consider alternative BMRs, as specified in the guidelines.

In contrast, the peer reviewer cites the recommendations of the scientific committee for the European Food Safety Authority that recommends use of 5% BMR as default for continuous datasets. To some extent, the difference in recommendations reflects differences in scientific opinion between the agencies and their advising bodies. Haber et al. (2018) (as cited by the peer reviewer) discusses the distinct US EPA and EFSA approaches in choosing BMR, and concludes that "rationales of the two groups reflect different approaches in priorities (i.e., what the groups are trying to estimate) and differences in areas of comfort with uncertainty," and furthermore, "[c]learly, the definition of the BMR includes judgement and elements of science policy." OEHHA

concurs with this assessment. An in-depth analysis of the differences between the guidelines and their supporting technical documents would go beyond the scope of the PHG assessment. However, OEHHA appreciates the peer reviewer's comment and provided resources, and will consider these in future updates of its guidelines.

Comment 24: "The authors repeatedly note that data presented in graphs could not be used. Obtaining data from graphs is very simple; indeed, it can be effectively done using free software that is accessible via a browser. See https://automeris.io/WebPlotDigitizer/. I strongly recommend extracting data from graphs, analysing the extracted data, and using the results for the determination of PHGs."

"Inability to use data displayed graphically was specifically noted on p.200 (i.e., Pereira et al., 1996 DCA liver toxicity data). Cannot see any reason why these data could not be extracted and analysed."

"Same would apply to the DeAngelo et al. (1997) body weight data mentioned in Table 10.7 (p. 205)." "Oddly, the issue with the DeAngelo et al. (1997) TCA data is shown in a table, the issue with the Pereira et al (1996) DCA data is not shown in an analogous table (i.e., Tables 10.4 and/or 10.5). This is inconsistent."

Response 24: OEHHA did in fact extract data from graphs several times in the PHG document, using GetData Graph Digitizer. While data could be extracted from graphs in this way, they were not always useful for POD consideration due to a lack of statistical analysis (e.g., standard deviations were not available). Descriptions of these instances, where statistical data were unobtainable, were added to the document.

The reason that the graphical data from Pereira (1996) could not be used for quantitative analysis is that large-sized symbols for data points overlapped and masked smaller error bars, making it impossible to extract statistical information necessary for BMD modeling. This rationale has been added to the PHG document.

Graphically presented data in DeAngelo et al. (1997) did not include any measure of variance, and therefore it could not be modeled with BMDS. The corresponding line in Table 10.7 was amended to indicate the lack of a measure of variance. In contrast, the NOAEL for DeAngelo et al. (1997) body weight data was available and statistical significance was indicated for the highest dose in the original report. A footnote has been added to Table 10.6 to clarify this.

Comment 25: "The organization of the genotoxicity information is OK, but it could certainly be better. Generally preferable to organize by endpoint, then separate the summary into in vitro studies and in vivo studies. That would allow readers to easily see the results of studies that examined frank genotoxicity endpoints, i.e., mutations and chromosome abnormalities including breaks, translocation, whole chromosome loss/gain and changes in ploidy. This type of organization differentiates between the

frank effects, which are severe and irreversible, from entirely reversible effects such as strand breaks and DNA damage reporter signals, e.g., SOS response in E. coli."

Response 25: In response to this comment, presentation of genotoxicity data in the tables was changed along the lines of the suggestions. The discussion of genotoxicity was also modified to follow the suggested order and hierarchy of experimental systems and effects.

Comment 26: "The authors are referred to recent IARC monographs within which the data are organized as (1) mutation, (2) chromosome damage, and (3) other DNA damage endpoints (e.g., DNA damage reporter signals). In each category the reviewed information starts with human, then animal in vivo, then mammalian cells in vitro, then other eukaryotic cells in vitro, then bacterial cells in vitro. Basically, IARC monograph sections pertaining to mechanistic support for human carcinogenic hazard start with mutagenicity, then within that (i) human in vivo, (ii) animal in vivo (plants generally listed last), (iii) human cells in vitro, (iv) animal cells in vitro, (v) other eukaryotes in vitro, (vi) bacterial cells in vitro. Then same order for cytogenetic effects. This is followed by review of assessments of reversible effects such as DNA damage as bulky and oxidative lesions, DNA damage as strand breaks (e.g., alkaline unwinding assay, SCGE or comet assay), and lastly, acellular in vitro induction of DNA damage. DNA damage reporter assays are on the bottom of the list since signals can be turned off when the stimulus is removed (e.g., prophage induction assay, SOS Chromotest, umuC assay, etc)."

"Results presented in the document need to be organized and interpreted in the same hierarchical fashion. In particular, it is critically important to differentiate between genotoxicity (e.g., strand breaks and bulky or oxidative lesions), mutagenicity, and clastogenicity (i.e., chromosomal abnormalities)."

"As noted earlier, all genotoxicity results must be interpreted in a hierarchical fashion."

Response 26: OEHHA agrees with the peer reviewer's suggestions for the organization of genotoxicity data and this is something that could be considered for future PHGs. In response to this comment, tables and discussions in the Genetic Toxicity sections were modified using a hierarchy of experimental systems and observed effects that approximates the genotoxicity data presentation in IARC monographs.

Comment 27: "I definitely have reservations regarding the interpretation of the in vitro genotoxicity information. First, with respect to the quality of the studies, the authors should to refer to the relevant OECD Test Guidelines, e.g., 471 for bacterial reverse mutation assays and 476/490 for mammalian cell in vitro mutagenicity [9-11]."

Response 27: OEHHA recognizes the importance of standardized guidelines, such as the OECD Test Guidelines for the purpose of regulatory safety testing However, for risk assessment purposes, OEHHA evaluates genotoxicity studies using an extensive set of criteria which includes publication in a peer-reviewed journal, adequate experimental

conditions (reported compound purity, use of adequate strains and exposure conditions, presence of controls, among others) and adequate reporting of results. In these evaluations, OEHHA did not adhere to the OECD Test Guidelines because doing so would potentially be creating a bias against higher quality studies that do not meet every requirement of an existing test guideline.

The purpose of OECD Guidelines is to promote generation of the experimental data of the highest quality to enable the OECD Mutual Acceptance of Data. The PHG risk assessment process has a different focus and collects all available data of adequate quality for the weight-of-evidence analysis in which studies of higher quality are given higher consideration. In this framework, studies with non-fatal flaws can still be considered in the overall analysis, which could compensate for scarce experimental data for a given chemical. For example, many bacterial reverse mutation reports included in genotoxicity sections in the PHG draft pre-date and would not meet the OECD guideline 471 requirements to use five different strains of bacteria *and* include S9 metabolic fraction *with* an appropriate control for S9 activation. Nonetheless, the data reported in these studies are still valuable in overall considerations regarding whether a substance is genotoxic.

Comment 28: "With respect to the bacterial reverse mutation assays more generally, i.e., the assays collectively referred to as the Ames Test, it is critically important to recognize that these are assays plural and not a single assay. For example, results obtained with TA98 and TA100 are not necessarily redundant. Rather, because these strains are reverted by different types of mutations, i.e., frameshift and base-pair substitution, respectively, responses on the two strains is generally complementary."

"Bottom line is that a "weight of evidence" approach cannot be employed when interpreting Ames test results across different bacterial strains. In essence, responses on different strains can essentially be viewed as responses for different assays (e.g., Salmonella assay with TA98 versus Salmonella assay with TA100). Granted, many mutagens elicit responses on both the base-pair and frameshift strains; however, some mutagens only elicit responses on the base-pair stains or the frameshift strains. Moreover, some agents only elicit responses on strains such as TA102 and TA104, base-pair strains that respond to DNA cross-linking agents such as Mitomycin C. Note that the bacterial reverse mutation assay based on E. coli WP2 can also detect DNA cross-linking agents; it was recently defined as redundant with TA102 and TA104 [12]."

"Bottom line -it is absolutely critical that bacterial reverse mutation assay results are interpreted with strain differences in mind. For information about strain genotypes and the types of mutations they detect, the authors are referred to Maron and Ames (1983) [13]. In essence, TA98, TA97, TA1538 and TA1537 detect frameshift mutagens; strains TA100 and TA1535 detect base-pair mutagens. As noted, some compounds (e.g., polycyclic aromatic hydrocarbons) elicit both types of responses; some compounds exclusively elicit responses on one or other type of strain. For example, N-ethyl-N¬nitrosourea is a base-pair mutagen that, to my knowledge, can only be detected with a strain that detects base-pair substitution mutations (e.g., TA100 or TA1535)."

Response 28: OEHHA recognizes the importance of different bacterial strains, and reports and interprets them separately. In response to this comment, additional language was added to differentiate observed genotoxicity result by bacterial strains where relevant.

Comment 29: "The complementarity of Salmonella strain responses is particularly important with respect to statements such as that on the bottom of p. 84, i.e., "Several studies employing reverse mutation assays in S. typhimurium did not observe genotoxicity of DCA, while other studies employing similar strains and methods reported weak or moderate genotoxicity". What do the authors mean by "similar strains"? By the way S. typhimurium is now called Salmonella enterica Serotype Typhimurium."

Response 29: 'Similar' was incorrectly used and has been changed to 'the same.' OEHHA uses *S. typhimurium* for simplicity.

Comment 30: "Granted, in terms of the bacterial reverse mutation assay results for DCA, the results are truly mixed, e.g., mixed results for both the base-pair (TA100 and TA1535) and the frameshift (TA98, TA1538 and TA1537) strains. That being said, it was impossible to be sure because of the way that results are summarised in Table 6.3, e.g., mix of base pair (TA100 and TA1535) and frameshift (TA1537) results indicated in a single table row. This needs to be changed if the base-pair and frameshift responses are different."

Response 30: OEHHA agrees that the statement, "Evidence of in vitro genetic toxicity of DCA is mixed (Table 6.3)" is unclear in the context of how the data are presented in Table 6.4 (formerly Table 6.3). That sentence has been changed to, "Evidence of in vitro genetic toxicity of DCA is inconsistent (Table 6.4)."

When two different strains are presented in a single row, this means that both strains have the same result. Strains are combined in one row solely to save space. Consistent with the peer reviewer's recommendation, when the results among strains differ, an additional row is added with the different result, as is the case with TA1535 and TA1537 strains in Herbert et al. (1980) shown in Table 6.3.

Comment 31: "Overall, with respect to DCA, there is convincing evidence that the substance is a base-pair mutagen, i.e., positive responses on TA100 except for the older Herbert et al study that examined very low concentrations. Importantly, OECD Test Guideline 471 indicates that for a definitive test, the maximum tested concentration should be 5mg/plate, or limit of solubility, or limit of cytotoxicity."

Response 31: This comment is consistent with OEHHA's analysis, and the overall conclusion is consistent with OEHHA's conclusion on the genotoxicity of DCA.

Comment 32: "Granted, with respect to DCA genotoxicity, there seems to be fairly convincing in vivo evidence, e.g., positive for mouse peripheral blood MN assay. … Bottom line – despite some disagreement in the literature, the evidence of a genotoxic

MOA does seem to be convincing. Especially given the transgenic rodent results (i.e., lacI mutation) of Leavitt et al (1997). Additionally, the other possible MOAs (e.g., peroxisome proliferation and regenerative proliferation) do not seem to be supported by the available evidence."

"Nevertheless, the reservations stated in the literature should be acknowledged, and the results should be interpreted with caution."

"As a final comment, it may be useful to evaluate the Leavitt et al. study in relation to OECD Test Guideline 488 [18]. However, the study was conducted long before the TG was published, and the positive responses for 2 doses at 60 weeks seems quite convincing."

Response 32: This comment is consistent with OEHHA's analysis. A discussion of reservations regarding the genotoxicity of DCA has been added to the *Risk Characterization* section of the PHG document. OEHHA agrees that the positive results of the Leavitt et al. (1997) study are quite convincing and do not require further evaluation in relation to OECD Test Guideline 488 (OECD, 2013).

Comment 33: "With respect to the calculation of CSF values (i.e., pp. 214-215), the analyses are convincing; nevertheless, as noted above, the authors need to judiciously mention uncertainty related to the assertion that DCA, TCA and DBA are genotoxic carcinogens."

Response 33: Additional text has been added to the *Carcinogenicity* and *Risk Characterization* sections of the PHG document to discuss uncertainty related to the genotoxicity of DCA, TCA and DBA. The underlying mechanisms of carcinogenicity appear complex. Given the presence of positive genotoxic evidence, a genotoxic MOA cannot be ruled out for DCA, TCA or DBA. In this case, a default linear extrapolation is used according to OEHHA guidelines.

Comment 34: "Important to note that the use of terms like "weak +" can be misleading. For example, Table 6.3 on p.86 notes that MLA mutagenicity of DCA (i.e., Harrington-Brock et al., 1998) is a weak positive. According to the paper, weak refers to potency, and not to the dichotomous call. Readers may be under the impression that the response was not a clear +. In fact, in this type of table the result should just be indicated as + since the response is undoubtedly +. Low potency, but + nonetheless."

Response 34: In response to this comment, OEHHA eliminated the designation 'weak' from the tables.

Comment 35: "With respect to the text pertaining to S9 metabolic activation (e.g., bottom of p. 84), it is important to note that S9 is not derived from hepatic cytosol. S9 is what is known as a PMS or post-mitochondrial supernatant. As such, it contains microsomes and cytosol. It is the microsomes that are critically important since they

contain Cytochrome P450 isozymes that are often essential for conversion of seemingly benign substances into DNA-reactive genotoxicants, e.g., aromatic amines."

"Interestingly, the text on p. 133 refers to "liver extract containing metabolizing enzymes". This is quite different from the earlier mention that S9 is derived from cytosol, i.e., only contains cytosolic enzymes. The authors need to correct these sections. The inconsistency in the text likely reflects the fact that sections were written by different authors."

Response 35: The description of S9 fraction was changed to "a mixture of unfractionated microsomes and cytosol containing a wide variety of metabolic enzymes."

Comment 36: "I also have concerns about the summary of TCA genotoxicity data. On p. 133, the authors comment on results for "the majority of the reverse mutation assays". As noted above, bacterial reverse mutation data must be interpreted in the context of the strain genotype and type pf mutation required for reversion to histidine prototrophy."

Response 36 In response to this comment, text was added to clarify that assays for detecting frame shift mutations (TA98, TA97, TA1538 and TA1537) or base-pair changes (TA100 and TA1535) produced negative results in most studies, and positive results in some studies. Positive results were also described in the text, and specific study results are presented in the tables.

Comment 37: "It is clear from the results presented in Table 7.4 that the bacterial reverse mutation assay results for TCA are very mixed, e.g., for Salmonella TA100 six negatives and two positives. That being said, few of the assessments meet the maximum test concentration requirements specified in OECD TG 471. The studies that tested at or near the recommended level of 5mg/plate (i.e., Nestmann et al., 1980 and Moriya et al., 1983) are primarily negative on both the base-pair and frameshift strains. Nevertheless, the mammalian cell mutagenicity data of Harrington-Brock et al (1988) do indicate that TCA is a weak mutagen. Importantly, Harrington-Brock et al note "The weight-of-evidence for TCA suggest that it is less likely to be a mutagenic carcinogen"."

Response 37: In response to this comment, text was added to clarify the classification of *S. typhimurium* strains according to the type of mutation, and to mention the maximum recommended concentration.

OEHHA agrees that the evidence from bacterial reverse mutation assays for TCA is inconsistent. In evaluation of in vitro studies, OEHHA did not directly use OECD guidelines, as outlined in the Response 27 above. Briefly, OECD guidelines set a high standard for the study quality, and strict application of these guidelines would likely severely limit the pool of studies available for evaluation. OEHHA uses its own set of criteria that are related to OECD guidelines regarding quality of the studies. However,

the overall approach is more flexible, and lower quality studies can be still considered in the weigh-of-evidence analysis while given less weight.

To illustrate the practical limitations of the proposed use of the OECD guidelines, one can consider Morya et al. (1983), which is a study cited in the comment as supportive evidence within the OECD TG 471 regarding the maximum applied concentration in the assay. Although this study met the guideline requirements for maximum concentrations, it also used TCA of unknown provenance and unknown purity, tested it with only 2 bacterial strains (instead of 5 required by the guideline), did not test with S9 metabolic fraction (as required by the guideline) and does not appear to include any positive or negative controls. While this study does not meet all OECD TG 471 requirements (similar to most genotoxicity studies for HAAs), it is included in the risk assessment but given less consideration due to quality limitations. Thus, the negative finding in this study would not be considered as definitive by OEHHA in the overall weigh-of-evidence analysis of TCA genotoxicity.

OEHHA notes that Harrington-Brock et al. (1988) formulated their conclusions 23 years ago, based on a significantly smaller dataset. Many more positive in vitro and in vivo genotoxicity studies on TCA (see Tables 7.4 and 7.5) have been published since 1988. The observation that TCA can metabolize to DCA in vivo (Larson and Bull (1992)) suggests that DCA MOA considerations for carcinogenesis would be applicable to TCA (See Comment 39 in this section).

Comment 38: "Thus, for both DCA and TCA, the evidence for a genotoxic MOA at the levels present in treated drinking water seems rather shaky. Granted, there does appear to be some evidence that TCA is genotoxic in vivo. That being said, it is not at all clear that the positive chromosomal aberration and MN studies listed in Table 7.5 were properly conducted, thus yielding usable data. As noted earlier, OECD TG 474 is very clear about appropriate timing for collection of bone marrow and peripheral blood. The authors need to provide information in the table about tissue examined (i.e., bone marrow or peripheral blood) and post-exposure sample collection time; moreover, judiciously interpret the data in the context of the OECD Test Guideline(s)." The peer reviewer describes instances where studies would not meet the OECD Test Guidelines, and questions whether the studies are acceptable for evaluation.

"Nevertheless, as noted earlier, these results should be reviewed and further scrutinized in relation to the requirements outlined in the relevant OECD Test Guidelines. It should be noted, that the OECD TG pertaining to the bone marrow chromosomal aberration assay is #475."

Response 38: In response to the comment, the timing of post-treatment sacrifice of animals for bone marrow assays was added to the third column of Table 7.4 of the final PHG document. In all exposure scenarios, these collection times were 24 hours and sometimes, 48 hours. The OECD Guideline 471 recommends 24 or 48 hours as collection times and is consistent with the reported experiments.

As explained in responses to Comments 27 and 37 above, OEHHA does not use OECD testing guidelines for study evaluation but rather employs its own vigorous set of criteria that are not inconsistent with the guidelines. Not every study evaluation detail is reported in the PHG to keep the overall document concise.

The peer reviewer's comment that the genotoxicity evidence for DCA and TCA is mixed concurs with OEHHA's evaluation. Given the positive evidence for genotoxicity of DCA and TCA, a genotoxic MOA for carcinogenicity cannot be ruled out. See Response 3 in this section. Finally, the peer reviewer's overall conclusions regarding DCA and TCA carcinogenicity MOAs also concur with those in the draft PHG document (see Comments 32, 40 and 41 in this section).

Comment 39: "Again, as noted earlier, since the effect is reversible, results pertaining to strand breaks (e.g., SCGE assay and alkaline unwinding assay) should not be used, in my opinion, as the sole support for assertion of a genotoxic MOA. This would also apply to assays that examined the frequency of DNA damage as oxidative 8-OHdG lesions."

Response 39: The discussion of these endpoints is consistent with OEHHA's analysis. The results pertaining to strand breaks and oxidative 8-OHdG are not the sole line of evidence used in reaching conclusions about the genotoxicity of TCA.

Comment 40: "Importantly, with respect to both DCA and TCA carcinogenesis, the authors' consideration of the available evidence is reasonably judicious and balanced. For example, the statements on p. 157 pertaining to the carcinogenic MOA of TCA indicate that it likely involves mixed MOAs. This is consistent with EPA's 2011 evaluation. Moreover, that the available evidence is sufficient to justify an assumption that TCA is a genotoxic carcinogen. In particular, the evidence pertaining to in vivo genotoxicity is reasonably compelling."

Response 40: OEHHA acknowledges the peer reviewer's concurrence with OEHHA's assessment and conclusions regarding the mode of action and carcinogenicity of DCA and TCA.

Comment 41: "On page 157, the authors present concluding remarks regarding the carcinogenic MOA for TCA. The same is warranted for DCA on p. 122, i.e., an overall summary for DCA. With respect to DBA there is no effective summary of information pertaining to carcinogenic MOA. Here and elsewhere, there are inconsistencies in the document. They likely reflect composition by numerous authors."

"With respect to the carcinogenic MOA of DBA more generally, there is reasonable evidence to support a genotoxic MOA. Unless I missed it, this was never explicitly stated."

Response 41: In response to this comment, remarks were added to the *Carcinogenicity* sections for DCA and DBA. OEHHA acknowledges the peer reviewer's concurrence

with OEHHA's approach and conclusions regarding the mode of action and carcinogenicity of DBA.

Comment 42: "Granted, the in vivo genotoxicity data for DBA are weak, but the NTP (2007a) study is convincing; moreover, there is strong evidence to support in vitro genotoxicity. With respect to the latter, DBA elicits a positive mutagenic response in both bacterial and mammalian cells. Collectively, this is important information that should be summarised. Presumably, on or about p. 193."

Response 42: In response to this comment, a remark on genotoxicity has been added to the DBA *Carcinogenicity* section.

Comment 43: "Should be noted that IARC evaluations of human carcinogenic hazard always include any human biomonitoring data pertaining to genotoxic effects, e.g., chromosome damage as micronuclei in peripheral blood lymphocytes. Granted, there does not seem to be any data for DCA and TCA in Monograph 84, and no data for DBA in Monograph 101. I would have expected some published data; such data are useful to support the supposition that carcinogenic effects are exerted via a genotoxic MOA. I am wondering if there is any way to use published information on urinary mutagenicity. Perhaps not, since the identity of putative urinary mutagens are not known, and most of the published studies seemed to have focussed on THMs and alkylnitrosamines. The authors may wish to carefully review this literature (i.e., urinary mutagenicity literature) to see if there is anything there that could support a genotoxic MOA for the HAAs examined."

Response 43: In response to this comment, OEHHA checked the literature and did not locate this type of data for any of the HAAs covered.

Comment 44: "With respect to carcinogenic MOA more generally, the authors should organise the presented information in a manner that is aligned with IARC's Key Characteristics of Carcinogens, i.e., the systematic approach now being used by IARC to organize mechanistic information pertaining to carcinogenesis. The authors are referred to recent works by Smith et al and Guyton et al [19-22]. Granted, the organization of the carcinogenic MOA information (e.g., pp. 118-119) is somewhat aligned with the Key Characteristic framework. The authors should note this fact, i.e., that their summary of MOA considerations is aligned with the IARC approach for evaluating mechanistic evidence related to human carcinogenic hazard. For example, on the bottom of p. 121, the author could note that the available evidence points towards two Key Characteristics (i.e., characteristics 2 and 4 in Smith et al., 2016), with genotoxicity being the most strongly supported by the available evidence."

Response 44: OEHHA acknowledges the emerging importance of the Key Characteristics of carcinogens framework to organize mechanist information in carcinogenesis discussions. The current DCA and TCA MOA sections in the document already somewhat follow the logic of this approach, and new assessments that OEHHA will begin in the future will be more closely aligned with the organization of Key

Characteristics used by IARC. As suggested, a statement about characteristics 2 and 4 was added.

Comment 45: "With respect to a mutagenic mode of action, the presented information need not be restricted to oncogenes like ras (e.g., H-ras). The authors should probably indicate that currently-available information highlights a wide range of cancer driver mutations. Granted, details would be beyond the scope of the current document, but the existence of a wide range of driver mutations could be mentioned. The authors are referred to the work of Stratton et al. (2009) and Bailey et al. (2018) [23, 24]."

Response 45: The document presents the currently available evidence from oncogene studies conducted with HAAs, and OEHHA did not limit its search to ras. Additional general information on cancer driver mutations is beyond the scope of this assessment, but the references are appreciated.

Comment 46: "Overall, I recommend that the authors (1) summarise the genotoxicity information in a hierarchical fashion, and (2) summarise the available information in a manner that is aligned with IARC's Key Characteristics of Carcinogens."

Response 46: In response to this comment, genotoxicity tables have been reorganized to present data in a hierarchical fashion, and summaries have been added to the *Genetic Toxicity* sections. Summarizing available information to align with IARC's Key Characteristics of Carcinogens is a good recommendation that will guide future assessments.

Comment 47: "With respect to analysis and interpretation of carcinogenicity dose-response data (e.g., hepatic adenoma and carcinoma data from DeAngelo et al., 1999), the rationale underlying combining adenomas and carcinomas is not clear. It looks like the authors requested the per-animal data specifically for the purposes of combining lesion incidence values for the later time points. Not necessarily unreasonable, but should be justified."

Response 47: The document was modified to provide further justification. According to McConnell et al. (1986), it is appropriate to combine hepatocellular adenomas and carcinomas in B6C3F1 mice "to obtain a better understanding of carcinogenicity." This guideline notes that "morphological studies of spontaneous hepatocellular lesions in mice… have indicated that adenomas… may represent early stages in the formation of carcinomas." Referring to McConnell et al. (1986), the US EPA Guidelines for Carcinogen Risk Assessment (2005) state: "The incidence of benign and malignant lesions of the same cell type, usually within a single tissue or organ, are considered separately but may be combined when scientifically defensible." It is general practice in NTP cancer bioassay reports to present hepatocellular adenomas and carcinomas separately and combined.

OEHHA added a sentence to the *Cancer Dose-Response Analyses and Cancer Potency Derivation* section justifying the combining of adenoma and carcinoma data on a per animal basis for dose-response analysis.

Comment 48: "I am confused about the daily water consumption values used to interpret the DCA carcinogenicity data (i.e., p. 117). I checked the EPA Exposure factor Handbook, and there does not appear to be any information for B6C3F1 mice."

Response 48: The daily water consumption rate in B6C3F1 mice used to convert doses in Bull et al. (2002) study is from Gold and Zeiger (1996), which is a different document from US EPA's Biological Values document (US EPA, 1988).

Comment 49: "Interestingly, the summary presented by tera.org indicates daily drinking water intake values for B6C3F1 mice in the range of 8.5-8.8 mL (tera.org/Tools/ratmousevalues.pdf). Obviously, if the daily water consumption value is reduced from approximately 8.5mL per day for B6C3F1 mice to the 5mL value used by the authors, this will affect the PHG value. If the dose calculations for B6C3F1 mice use, for example, 5mL per day instead of 8.5mL per day, the result would be an approximately 40% decrease in dose, given a constant BW. If doses are shifted downward by using a lower DW daily intake value, the BMD will be shifted to a lower value, and PHG will be lower by extension, i.e., calculated dose required to elicit the BMR will be lower and the calculated PHG will be lower. The authors need to make sure the daily DW intake values used to determine dose per unit BW per day are as accurate as possible, i.e., strain matched wherever possible."

Response 49: OEHHA agrees that it is important to use as specific an estimate of biological parameters as possible, but notes that both DW daily intake values mentioned in the comment are estimates and not experimentally measured values. While US EPA (1988) lists 8.3-8.5 ml/day as a suggested reference DW intake value for B6C3F1 mice, it does not list a single B6C3F1 study among multiple mouse studies that reported drinking water consumption (Table 5-1, US EPA (1988)). On the other hand, the 5 ml/day recommendation (Gold and Zeiger, 1996) is more recent and was reported based on the review of specifically cancer studies. Thus, each method has its advantages but also its uncertainties. Ultimately, the choice of the dose conversion method for a given study did not affect the proposed PHG values since none of the critical studies in the PHG draft required dose conversion (doses were reported either in the study or the associated US EPA document), and application of an alternative dose conversion method to other candidate critical studies would not have affected the choice of the critical study.

Comment 50: "The methodology employed for calculation of DBA dose (i.e., p. 172) is useful and interesting for the reader. Why wasn't this information provided for DCA and TCA?"

Response 50: Multiple DBA studies presented in Table 9.1 required dose conversion and to simplify the presentation of results, the generic conversion formulas were

described separately in the text preceding the table, and the required parameters were provided for each individual study in the footnotes to the Table 9.1. In contrast, the vast majority of DCA and TCA studies had estimated oral doses, expressed as mg/kg-day, in the original report. An occasional study, requiring a dose conversion (e.g., Bull et al. (2002)) was best addressed with a specific footnote.

Comment 51: "With respect to the sections on DCA and TCA, were water consumption values body weight corrected? Perhaps this type of information, which pertains to evaluation of all the HAA toxicological data, should be presented elsewhere, i.e., in a section pertaining to dose determination. Or more generally, in a section pertaining to data manipulation and interpretation, i.e., Section 10."

Response 51: Doses in the drinking water studies are presented in units of milligrams per kilogram of body weight per day (mg/kg-day), and have been corrected for body weight.

Comment 52: "p. 177 notes that "IARC concludes that there is moderate evidence of a genotoxic mechanism". IARC likely phrased as something like "strength of the evidence to support a genotoxic mode of action is moderate". Minor difference, but important, i.e., the way strength of evidence is denoted. From Monograph 84 – "The strength of the evidence that any carcinogenic effect observed is due to a particular mechanism is assessed, using terms such as weak, moderate or strong."

Response 52: In response to this comment, the sentence has been changed to "Although the mechanism of carcinogenicity of DBA is unknown, IARC notes, 'Several comparative genotoxicity and mutagenicity studies … have demonstrated that dibromoacetic acid is more potent than its chlorinated analogue, dichloroacetic acid, and that they have several molecular and biochemical activities in common.'"

Comment 53: "The text on p. 194 indicates that the BMR for dichotomous response endpoints was set at 5%. Additionally, 5% extra risk for carcinogenicity dose-response analysis is specified on p. 214. Yet EPA's 2012 Technical Guidance Document indicates on p. 20, "for dichotomous data, a response level of 10% extra risk has been commonly used to define BMDs". … "Moreover, Hardy et al (2017) note, "…..the BMDL10 may be an appropriate default…..a BMR of 10% appears preferable for quantal data because the BMDL can become substantially dependent on the choice of dose-response model at lower BMRs". In light of the guidance from the EPA and the EFSA, why did the authors choose a BMR of 5% extra risk for interpretation of carcinogenicity dose-response data?"

Response 53: See response to Comment 16 from Dr. Shao.

Comment 54: "I found the section pertaining to Uncertainty Factors very confusing. In large part because it differs from recommendations of the WHO/IPCS, the ICH, the USFDA, and the ECHA [25-28]. Granted, OEHHA likely has its own rationale, traditions

and guidelines, and discussions and considerations regarding alternatives may be outside the scope of this document."

"I was particularly curious about the inter-species UF, and the contrast with FDA recommendations for use of surface area to BW ratios. The ICH uses the same paradigm (i.e., ICH Q3C(R6)) [28]. More specifically, the ICH states that the inter-species UF "takes into account the comparative surface area:body weight ratios for the species concerned and for man….. Surface area (S) is calculated as S = kM0.67….in which M = body mass, and the constant k has been taken to be 10" [28]. This is aligned with the FDA approach that recommends "Conversion of Animal Doses to Human Equivalent Doses Based on Body Surface Area" [25]. Thus, the UF for mouse would be approximately 12; for rat it is in the 5-6 range. Alternatively, depending on how the calculation is conducted, approximately 0.08 for mouse and 0.2 for rat. Looks like OEHHA is recommending 3.16 for all cases where toxicity assessment data is from tests with any non-primate species. This seems quite odd. Granted, this value would only compensate for interspecies toxicokinetic considerations. WHO/IPCS recommends an additional UF to compensate for interspecies toxicodynamic differences, i.e., $10^{0.4}$ = 2.5 [26]. Seems like OEHHA is recommending 3.16 for non-primate studies with no data on interspecies toxicodynamic differences."

Response 54: The text has been added to provide more explanation that OEHHA uses its own peer-reviewed guidelines on uncertainty factors (UFs) for non-cancer dose response analyses (OEHHA, 2008) The guidance went through public comment as well as scientific peer-review (see Responses 2 and 3 to Dr. Lin above). The OEHHA guidance on interspecies extrapolation accounts for both pharmacokinetic and pharmacodynamic differences – general defaults are shown in Table 10.1 - a factor of 3.16 (i.e., √10) for pharmacokinetics and 3.16 for pharmacodynamics. These are the factors commonly used in the absence of chemical specific data. When they are both applied in composite, they result in a factor of 10 interspecies adjustment.

As noted in OEHHA (2008): "Schmidt et al. (1997) evaluated interspecies variation between human and five other animal species. Sixty compounds had human data that could be matched to one or more animal species. The animal to human ratio of 10 represented approximately the 85th percentile. … Where both chemical- and species-specific data are unavailable, and therefore a [human equivalent concentration] cannot be estimated, a 10-fold [interspecies uncertainty factor] is normally used."

The references provided in the comment for dose conversion methods would give higher values for the toxicokinetic component of the interspecies UF.

Historically, the interspecies UFs were not based on separate estimations of toxicokinetic and toxicodynamic components. In the last 15-20 years, along with the increasing use of pharmacokinetic modeling, the UF of 10 was proposed to comprise separate toxicodynamic and toxicokinetic components, and that perspective is reflected in the OEHHA (2008) guidance.

A sentence referring the reader to the OEHHA (2008) *Technical Support Document for the Derivation of Noncancer Reference Exposure Levels* was added to the PHG document.

Comment 55: "I was similarly perplexed by the study durations UFs employed by the authors. ECHA recommends 2 for rodent 90-day studies (i.e., approximately 12% of lifespan), 6 for rodent 30-day studies (i.e., < 8% of lifespan) [27]. With respect to study duration, ICH recommends 2 for a rodent 6-month study, 5 for a rodent 90-day study, and 10 for shorter duration studies [28]. The authors may not have the latitude to consider other UFs; moreover, considerations of other UFs may push the work outside the scope of determination of PHGs for the state of California. Nevertheless, it would be helpful for the reader if the authors elaborate a bit regarding where the UFs come from. Just 2 or 3 lines should suffice."

Response 55: ECHA and ICH differ from each other for adjustments for study durations. Defaults used in California assessments are as follows:

1	study duration >12% of estimated lifetime
√10	study duration 8-12% of estimated lifetime
10	study duration <8% of estimated lifetime

These values are in Table 10.1 in the document. Thus, where ECHA would use a factor of 2 to adjust for a 90 day study, and ICH would use a factor of 5, OEHHA uses a value that falls between these two at 3.16. For a short duration study, OEHHA and ICH would use 10, whereas ECHA would use 6. OEHHA's values are consistent with guidance established by US EPA (2002a).

Regarding the table of default uncertainty factors for PHG derivation, the PHG document states, "Table 10.1 below is adapted from OEHHA's *Technical Support Document for the Development of Noncancer Reference Exposure Levels* (OEHHA, 2008). A statement was added referring readers to OEHHA (2008) for the detailed rationale supporting these default UFs.

Comment 56: "As noted earlier, there are numerous instances where the authors note that collected dose-response data are not amenable to dose-response analysis (e.g., p. 200 bottom), or that the resultant BMD cannot be used for regulatory purposes. In section 10, the authors should outline the criteria used to evaluate (1) the suitability of reported/collected data for dose-response analyses, (2) the ability to reliably determine a BMD, and (3) the utility of the BMD. The analyses presumably followed the USEPA (2012) guidance. #3 will likely require calculation of BMD:BMDL ratio, and inclusion of the ratio in Tables of BMD results, e.g., Table 10.5. Personally, I prefer BMDU-to-BMDL ratio as an indicator of BMD precision and utility."

Response 56: In response to this comment, the *Point of Departure* section of the draft PHG document was expanded to include the suggested information. These edits

address points (1) the suitability of reported/collected data for dose-response analyses, (2) the ability to reliably determine a BMD.

Regarding point (3), the utility of BMD, the BMD:BMDL ratio (a strictly mathematical characteristic of the dose response) is one consideration OEHHA uses in addition to other factors. See Response 8 above.

Comment 57: "In some cases the authors have analysed dose-response data with only 2 non-control doses. Although this is the minimum required, most researchers conducting BMD analyses would likely say that the minimum required for effective BMD analysis is 3 non-control doses. That being said, low BMD precision that can occur when there are few dose group will be reflected by the aforementioned ratio metrics. The authors may be interested to know that works such as Kuljus et al (2006) note that to avoid the risk of dose placements that do not favour precise determination of a BMD, a minimum of 4 doses (i.e., 3 plus control) is recommended [29]."

Response 57: While OEHHA generally prefers to analyze dose-response studies with a higher number of dose groups, other criteria contribute to the selection of critical studies. Removing from consideration a high-quality study with two dose groups (plus control) that reports adverse effects at low doses (sensitivity) would potentially discard valuable toxicological information.

Comment 58: "Interesting that the authors are using allometric animal-to- human scaling for calculation of human CSF values. Why wasn't this strategy used for animal-to-human adjustments for non-cancer endpoints? I believe the FDA recommends calculation of a conversion factor as $(W_{animal}/W_{human})^{(1-b)}$, where b = 0.67 (i.e., rather than 0.75) [25]. The FDA method would yield a smaller conversion factor, particularly for a small species like mouse; consequently, a smaller CSF_{human} and a larger cancer health-protective drinking water concentration. By the way, the calculation method employed needs a citation. It's presumably USEPA (1992) [30]."

Response 58 In response to this comment, OEHHA added a citation (OEHHA, 2009) to the *Cancer Dose Response and Cancer Potency Derivation* section. The use of 0.75 power for animal-to-human adjustments for cancer dose response assessment is used across OEHHA programs and is adopted in regulation (California Code of Regulations, Title 27, Section 25703(6)) used in implementing a related program. OEHHA follows its risk assessment guidelines, which establish distinct approaches for cancer and noncancer animal-to-human extrapolations. OEHHA and US EPA both use scaling to the three-quarters power in interspecies adjustments for cancer. Originally both agencies scaled based on surface area adjustment, namely b=0.67.

Comment 59: "On p.215 I'm not sure I understand how the authors can so readily dispense with rat data. Are rats uniformly less sensitive? Please provide a citation."

Response 59: OEHHA agrees that the DCA rat studies require a better justification in consideration for the choice of critical studies. In response to this comment, this section

has been revised. In the process, it was found that statistical significance for tumor incidences in Study 1 for DeAngelo et al. (1996) in Table 6.10 was indicated incorrectly. While it was previously indicated that in Study 1, either hepatic adenomas or combined hepatic adenomas and carcinomas (5/29 and 7/29, respectively) were significantly different from the controls (1/23 and 1/23, respectively), this determination was incorrect. In fact, there are no significant differences in either case using the exact Fisher text and the table has been corrected.

The following paragraph was included in the *Cancer Dose-Response Analysis* section for DCA: "The available male F344 rat studies (Richmond et al., 1995; DeAngelo et al., 1996) observed significant toxicity at the highest dose, resulting in early sacrifice and/or progressively decreased dose. While the multidose study of Richmond et al. (1995) observed significantly increased hepatic adenomas at the highest dose (296 mg/kg-day), and the single dose (plus control) study of DeAngelo et al. (1996) observed significantly increased hepatic carcinomas, and adenomas and carcinomas at the highest dose (139 mg/kg-day), neither is considered as a candidate critical study for DCA carcinogenesis due to increased toxicity at the doses where tumors were observed, and due to lower sensitivity of the studies."

Comment 60: "With respect to the cancer dose-response analysis and potency determination for TCA, where is the calculation of the CSF_{human}? On p. 218, the reader is walked through the calculation for DCA, why isn't there a similar overview for TCA? Similarly, why are the table entries/format for 10.14 different from 10.12? Please keep consistent. Similar inconsistencies with respect to Table 10.15 for DBA. Even the titles are inconsistent. Where is the study citations for Table 10.15?"

Response 60: In response to this comment, the sections describing the selection of the critical study and CSF_{human} calculation for TCA and DBA were changed to match that of DCA.

Tables 10.12, 10.14 and 10.15 were re-formatted to be consistent among each other and to contain all necessary information.

Comment 61: Regarding Table 10.15, "Why is there no p value for the 3rd row?"

Response 61: The BMDS multi-site cancer model does not provide a p value – see Appendix E. A footnote was added to Table 10.15 regarding the multisite p-value.

Comment 62: "With respect to TCA, I am wondering why the authors did not mention regenerative proliferation in Section 12 (i.e., p. 231). Didn't DeAngelo et al (2008) report regenerative proliferation in murine hepatocytes, i.e., hepatic proliferation in B6C3F1 mice?"

Response 62: TCA-dependent compensatory cell proliferation is discussed in detail in the section *Cytotoxicity and Cell Proliferation*.

Comment 64: "I also have doubts about the strength of the evidence for MBA genotoxicity. Yes, the in vitro results presented in Table 8.2 are quite convincing, but the scant in vivo data shown in Table 8.3 is certainly cause for concern. One might even say less than adequate for assertion of a genotoxic MOA for MBA carcinogenicity."

Response 64: Both DBA and MBA were negative in genotoxicity assays in nematode and newt. However, DBA was active in all mammalian genotoxicity assays (Tables 8.3, 9.4). Thus, the lack of positive in vivo genotoxicity findings for MBA may be due to the absence of mammalian studies. Yet, there is not enough evidence to determine whether MBA is a carcinogen and the draft PHG document states: "MBA is not assessed for carcinogenicity in this document." However, given the strength of in vitro genotoxicity evidence, there is certainly a concern that MBA may have carcinogenic potential in vivo. This concern, together with the lack of reproductive and developmental studies, was addressed with a database deficiency uncertainty factor √10.

Comment 64: "In section 12, I suggest summarizing the PHG values in a table."

Response 64: Per this suggestion, a summary table has been added after the separate subsections for each HAA in Chapter 12.

Comment 65: "As a final comment, I would be curious to know what the SWRBC will do with the DCA, TCA and MBA PHGs, which are all far less than 1ppb. The values are clearly very low relative to the regulatory standards for HAA5 summarized in Table 11.3."

Response 65: Per the SWRCB website[4], "Health & Safety Code §116365(a) requires a contaminant's MCL to be established at a level as close to its PHG as is technologically and economically feasible, placing primary emphasis on the protection of public health."

Comment 67: "Actually, shouldn't this be Table 12.1?"

Response 67: In response to this comment, Table 11.3 has been changed to Table 12.1.

[4] https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/MCLsandPHGs.html

RESPONSES TO COMMENTS MADE DURING THE FIRST PUBLIC COMMENT PERIOD

RESPONSES TO COMMENTS RECEIVED FROM AMERICAN CHEMISTRY COUNCIL (May, 2020)

General Comments:

Comment 1: "We appreciate that the HAA TSD recognizes findings from the WHO and the International Agency for Research on Cancer (IARC) regarding the essentiality of drinking water disinfection, from centralized treatment facilities to individual taps, relative to incremental reductions in DBP concentrations. These definitive statements call for a quantitative analysis of the potential public health risks that may result from further efforts to reduce DBP concentrations in drinking water—particularly those associated with individual MCLs that are an order of magnitude lower than the current group MCL for HAAs. Yet such analysis does not exist in this draft TSD; as it did in the TSD for THMs, OEHHA is deferring this analysis to the SWRCB."

Response 1: Such an analysis is beyond the scope of the HAA PHG assessment. The statute for PHGs (California Health and Safety Code Section 116365) states that OEHHA "shall prepare and publish an assessment of the risks to public health posed by each contaminant for which the state board proposes a primary drinking water standard…The risk assessment shall contain an estimate of the level of the contaminant in drinking water that is not anticipated to cause or contribute to adverse effects, or that does not pose any significant risk to health." Thus, a PHG is specific to its respective contaminant only.

Comment 2: "Currently, California regulates HAAs under the MCL for total HAAs of 60 ppb as the sum of the concentrations of MCA, DCA, TCA, MBA, and DBA. In the draft HAA TSD, OEHHA has replaced the single group PHG with separate PHGs for each of the five regulated HAAs."

Response 2: OEHHA did not previously develop a single PHG for these 5 HAAs. The current MCL (Maximum Contaminant Level) for total HAAs is the same MCL adopted by US EPA and was not based on a PHG.

Comment 3: "In light of the significant disparity between OEHHA's estimates of cancer risk presented by DBPs and the epidemiological evidence, OEHHA should consider the proposed PHGs for DCA, TCA, and DBA in the larger context of the long history of chlorine disinfection in the state, declining concentrations of HAAs and other DBPs in finished drinking water, and the overall trends in liver cancer incidence."

Response 3: PHGs are health-based determinations that do not depend on the environmental levels of the pollutant in question. In drafting these PHGs, OEHHA did not identify any published epidemiological studies exploring correlations of liver cancer and HAAs, and therefore no conclusion can be made in this respect. Multiple causes can contribute to carcinogenesis of any organ or system, including liver cancer.

Comment 4: "The Draft PHG for DCA is based on reports of liver tumors in studies conducted in male mice. The evidence in female mice is less consistent, however, and studies in rats suggest lower sensitivity than in mice."

Response 4: It is correct that male mice are more sensitive, and it is OEHHA's policy to base the PHG on the most sensitive study of sufficient quality.

Comment 5: "Moreover, the mice in the key study selected by OEHHA for the DCA risk assessment (DeAngelo et al., 1999) exhibited a high rate of spontaneous liver tumors and significant mortality and body weight decreases at the two highest doses. As a result, it does not appear that this study is appropriate for deriving a cancer slope factor (CSF)."

Response 5: The control group incidence of hepatic adenomas and carcinomas in the DeAngelo et al. (1999) study does not diminish the significance of the findings. When compared to concurrent controls, the top two dose groups were highly significant, with 100% of treated animals displaying adenomas and/or carcinomas.

In order to account for significant mortality in the experiment, a multistage-in-dose Weibull-in-time (MSW) model was used for dose-response analysis. This model accounts for early deaths in the treated animals, as well as the reduced lifetime of animals sacrificed before the end of the study. The authors noted that "the early mortality … was due almost wholly to liver cancer" (DeAngelo et al., 1999). The MSW model accounts for this effect as well.

Regarding the body weight decreases in the two highest dose groups compared to the controls, the study authors noted: "Based upon the water consumption and body weight gain through 78 wk of treatment, 2 g/L DCA did not exceed the maximum tolerated dose (highest DCA concentration not resulting in a >10% body weight reduction). The lower body weight seen in this group (and for 3.5 g/L DCA) at the termination of the study was due almost entirely to the cachexia that accompanied a significant liver tumor burden. It was therefore possible to generate a reliable dose-response curve for the development of hepatocellular cancer in male B6C3F_1 mice exposed to DCA in the drinking water" (DeAngelo et al., 1999). OEHHA concurs with this determination.

Comment 6: "The OEHHA analysis, in fact, notes limitations for all of the cancer studies considered as candidates for deriving the proposed PHG. Given these limitations, it is not clear why OEHHA did not derive the geometric mean of the CSFs for the most relevant studies (i.e., 0.027 per mg/kg per day)—rather than selecting the highest CSF among the male mouse studies."

Response 6: The three studies considered as candidate critical studies for DCA were DeAngelo et al. (1999), Bull et al. (2002) and Wood et al. (2015). Of these, DeAngelo et al. (1999) was considered to be of higher quality than the other two studies, in part because it exposed mice to DCA over a lifetime (100 weeks). While this study demonstrated some pre-term mortality in the two high dose groups and had several

interim sacrifice groups, these issues were addressed by application of the MSW model. Thus, in this study, the majority of the animals in the 100 week dose groups lived to term, and cancer rates in these animals were measured directly. In contrast, Bull et al. (2002) exposed mice to DCA for only 52 weeks. While the resulting cancer slope factor (CSF) was adjusted from this shorter duration to 104 weeks using a different method (poly-3 adjustment), none of the animals in this study lived past their half-lifetime mark (52 weeks) and therefore, there is added uncertainty in adjusting for one additional year of hypothetical exposure. Wood et al. (2015) exposed mice to DCA only over the first 10 weeks of the 94-week study, requiring the Armitage-Doll adjustment, which would significantly increase uncertainty in comparison to a study that does not require this dose adjustment. Therefore, OEHHA chose the DeAngelo et al. (1999) study over the Bull et al. (2002) and Wood et al. (2015) studies.

In this case, DeAngelo et al. (1999) was determined to be of better quality, i.e., the study providing a higher level of confidence, compared to the other candidate studies. Deriving a geometric mean of the CSFs of the three studies would only lower the confidence of the result. This discussion is presented in the *Cancer Dose-Response Analyses and Cancer Potency Derivation* section.

Comment 7: "Moreover, although DCA appears to be weakly genotoxic, and only at higher doses, OEHHA assumes that the liver tumors result from a genotoxic mechanism. As noted by USEPA, there is little basis for judging whether genotoxic effects—including alterations in the genetic messages for various proto-oncogenes—are important in the carcinogenic response, and if so, whether the dose-response curve for genotoxic effects is linear or nonlinear."

Response 7: In its IRIS report, the US EPA states that the genotoxicity data for DCA are inconsistent and seem to indicate that DCA is a weak mutagen. Furthermore, "Nevertheless, in the absence of causal data, EPA considers it prudent to assume that DCA might be genotoxic, at least under in vivo exposure levels that are associated with detectable increases in tumor incidence (particularly at the higher doses). Whether DCA is genotoxic at lower doses (which would suggest a linear dose-response curve for cancer risk) is not known" (US EPA, 2003). US EPA goes on to state, "Because the mode of action by which DCA increases cancer risk is not understood, extrapolation to low dose was performed by assuming a no-threshold linear dose-response curve between the origin and the POD" (US EPA, 2003).

After conducting its own assessment of the mechanistic data, OEHHA reached a similar conclusion regarding the mode of action (MOA) of DCA carcinogenesis and the dose-response method. As explained in the draft PHG document, DCA carcinogenesis could be the result of multiple MOAs, including genotoxic and non-genotoxic mechanisms. In this case, OEHHA applies the health-protective approach of non-threshold linear extrapolation.

In addition, OEHHA carefully considered available mechanistic evidence regarding liver cytotoxicity and cell regeneration, and concluded, "These effects occurred at higher

doses than those required to induce tumors in rodents. This indicates that cytotoxicity with regenerative proliferation is not likely a key event in the MOA for DCA tumorigenesis." Thus, in its analysis, OEHHA did not find strong evidence for the threshold MOA. The Commenter did not submit additional studies in support of the threshold mechanism of DCA carcinogenesis.

Comment 8: "TCA Is Not a Genotoxic Carcinogen."

Response 8: The available studies of TCA genotoxicity are discussed in detail in the draft PHG, and the in vitro and in vivo studies are summarized in Tables 7.4, and 7.5, respectively. Specifically, out of 21 in vivo studies of genotoxicity, 13 studies reported positive results. While most *S. typhimurium*-based in vitro studies found limited or no genotoxicity (Table 7.5), this experimental system tests for direct carcinogenic action and lacks the full metabolic capacity of in vivo test systems, in which 13 studies showed positive results. One reported effect of in vivo metabolism of TCA is conversion to DCA, which has genotoxic potential. Based on the evidence, OEHHA cannot exclude the possibility that TCA acts through a genotoxic mechanism.

Comment 9: "As the Draft PHG indicates, while there is consistent evidence of liver tumors in male mice exposed to TCA, the evidence for tumors is less consistent in female mice and tumors have not been reported in rat studies."

Response 9: As reported in the draft PHG document, one study employing a relatively high dose of TCA (Bull et al., 1990) did not find a significant increase in hepatic tumors in female B6C3F1 mice, whereas two other studies (Pereira, 1996; Pereira and Phelps, 1996) found significantly increased tumors at high doses. While this outcome is consistent with lower sensitivity of female B6C3F1 mice to TCA carcinogenesis compared to male mice, the overall evidence of TCA carcinogenesis in female mice is clear. OEHHA acknowledges the lack of evidence of TCA carcinogenesis in rats.

Comment 10: "As with DCA, the key [TCA] study selected by OEHHA (DeAngelo et al., 2008) reported a high incidence of tumors in the control group which diminishes the significance of the findings in the dose groups."

Response 10: The high control group incidence of hepatic adenomas and carcinomas in the DeAngelo et al. (2008) study does not diminish the significance of the findings. Hepatocellular adenomas and carcinomas are commonly observed in male B6C3F1 mice. In the historical control data, high background levels of combined hepatocellular adenomas or carcinomas (247/339 or 72.9%) were observed in drinking water studies conducted by NTP from 1984 to 1994 (NTP, 1999), the period of time overlapping with experimental work (1991-1993) in the DeAngelo et al. (2008) study, as indicated in the data files obtained from US EPA for these studies. Thus, an incidence of 31/56 (55.4%) for combined hepatocellular adenoma and carcinoma in the control group (DeAngelo et al., 2008) is not unusual.

Moreover, there was a significant increase in combined hepatocellular adenoma and/or carcinoma in DeAngelo et al. (2008), as determined by the exact trend test. Other studies (DeAngelo et al. (2008) at 60 weeks and Bull et al. (2002), Table 10.13) also demonstrated a significant increase in tumor incidence as demonstrated by the exact trend test or pairwise comparison with controls. Three additional studies in male B6C3F1 mice (Bull et al., 1990; Ferreira-Gonzalez et al., 1995; Herren-Freund et al., 1987) also found increased hepatic tumors in a single dose study design. Thus, the observation of increased hepatic tumors in male B6C3F1 mice with TCA treatment in six independent studies is unlikely due to chance.

Comment 11: "Although OEHHA considered and rejected two other studies with male mice, it is not clear why they did not include the study by Pereira (1996) which reported liver tumors in female mice exposed to TCA for up to 576 days (82 weeks). Benchmark dose (BMD) modeling of the results of the Pereira study produces a 95% lower confidence limit on the BMD for a 10% response (BMDL10) of 4.67 mg/kg per day compared to a BMDL10 of 1.50 mg/kg per day for the study by DeAngelo et al. (1999)."

Response 11: Regarding TCA dose-response in Pereira (1996) and other studies in female mice, the PHG draft states, "Since it appears that the endpoint of hepatocellular tumors in female B6C3F1 mice was less sensitive in comparison to male mice, only studies of hepatocellular tumors in male mice were considered for dose-response assessment and PHG derivation." The reason the study by Pereira (1996) was not considered as a critical study is the lower sensitivity of the study, as stated.

OEHHA was not able to reproduce the BMDS results provided in this comment, and no further documentation was submitted to inform this calculation. In OEHHA's calculations, the cancer incidence data from the Pereira (1996) study produced $BMDL_{05}$ values about an order of magnitude higher than the $BMDL_{05}$ values obtained from the chosen critical studies, consistent with the description of the Pereira (1996) study in the PHG document and the stated reason not to consider this study as a candidate critical study.

Comment 12: "Peroxisome proliferation has also been demonstrated in a number of short- and long-term TCA exposure studies in both rats and mice. Considering the very limited evidence for the genotoxicity of TCA, it is likely that the mouse liver tumors result from a non-genotoxic mechanism defined by an exposure threshold below which the cancer risk would be zero."

Response 12: The issue of PPARα activation as a possible MOA of TCA carcinogenesis is discussed in detail in the draft PHG document. OEHHA concludes, "Although the available evidence does not exclude the possibility that at least some TCA tumors could originate with PPARα activation, there is evidence suggesting it is not the only MOA for TCA carcinogenesis."

The available studies of TCA genotoxicity are discussed in detail in the PHG draft document, and the in vivo and in vitro studies are summarized in Tables 7.4, and 7.5,

respectively. Specifically, out of 21 in vivo studies of genotoxicity, thirteen studies reported positive findings of genotoxicity. OEHHA does not consider this evidence very limited and cannot rule out a genotoxic MOA for TCA carcinogenesis. In this case, OEHHA utilizes the default linear approach, as further explained in the document.

To base the CSF on a threshold MOA of carcinogenesis, OEHHA would need to rule out MOAs associated with low-dose linearity. OEHHA was not able to locate such evidence. In contrast, the existing mechanistic studies point at several likely underlying mechanisms, including genotoxicity, as explained above.

Comment 13: "The cancer evidence for DBA is limited to a National Toxicology Program (NTP) study which reported liver tumors in male and female mice and an increase in lung tumors in male mice. Liver and lung tumors were not observed in rats in the NTP study. The control groups for both the male and female mice exhibited a high rate of spontaneous liver tumors, however, and the incidence of lung tumors was increased in the control group of the male mice. In addition, the lung tumors did not show a clear dose-response in the male mice. Tumors were significantly increased at a mid-dose of 500 mg/L (ppb) [sic], but not at the highest dose of 1000 mg/L (ppb) [sic]."

Response 13: The NTP (2007) report concluded that "there was some evidence of carcinogenic activity of dibromoacetic acid in male rats based on an increased incidence of malignant mesothelioma" and "there was some evidence of carcinogenic activity ... in female rats based on an increased incidence and positive trend of mononuclear cell leukemia." Furthermore, there was "clear evidence of carcinogenic activity … in male and female mice based on increased incidences of hepatocellular neoplasms and hepatoblastoma (males only)." Thus, positive findings of DBA carcinogenicity were reported in four different rodent studies of high quality, and OEHHA considers this a sufficient basis for the carcinogenicity determination of DBA.

Regarding the comment that liver and lung tumors were not observed in rats in the NTP study, US EPA (2005) states, "[S]ite concordance is not always assumed between animals and humans. … [C]ertain modes of action with consequences for particular tissue sites (e.g., disruption of thyroid function) may lead to an anticipation of site concordance." Similar reasoning would apply to site concordance between different species of laboratory animals, such as rats and mice. Since neither the MOA of DBA carcinogenesis nor organ-specific effects are known, there is no mechanistic basis to expect site concordance for DBA carcinogenesis.

The incidence of liver tumors in the control groups of male and female mice, and incidence of lung tumors in the control group of male mice (NTP, 2007a) does not diminish the significance of the carcinogenesis finding. NTP (2007) concluded that "[i]ncreased instances of lung neoplasm in male mice were … considered to be exposure related."

As indicated in the Table 9.15 of the draft PHG document, alveolar/bronchiolar adenomas in male mice demonstrated a significant trend ($p<0.05$). Similarly, NTP

(2007) concluded that increased incidences of lung neoplasms in male mice were exposure related. The observation that tumor instances in some dose groups do not reach significantly different levels does not mean that there is no overall trend or that the effect is not treatment related.

Comment 14: "Given the limited cancer data available for DBA, and the conflicting results reported in mice and rats, the mouse cancer data should not be used as the basis for the PHG. Moreover, any estimate of cancer risk should not include the lung tumors in male mice as a result of the high spontaneous incidence in the control animals and the lack of a clear dose-response in the male mice."

Response 14: As detailed in Response 13, the evidence of DBA carcinogenesis includes findings in four rodent studies, including mice and rats of both sexes. Thus, the available evidence is not limited. The lack of site concordance between mice and rats is common for many carcinogens, and does not preclude the dose-response analysis of the most sensitive study, i.e., the study in male mice. NTP (2007) also concluded that increased incidences of lung neoplasms in male mice were exposure related, and thus OEHHA determined they should be considered in the PHG.

Comment 15: The public comments on the first public review draft of the THM TSD were submitted well ahead of the external scientific peer review reports on that document, yet there is no indication in the peer review reports that the reviewers considered those comments. OEHHA's passive approach to notifying peer reviewers about the availability of public comments, rather than specifically including those comments in the materials submitted to the peer reviewers, tends to produce peer reviews that focus only on the studies and the OEHHA analysis provided in the TSD. This approach appears inconsistent with the applicable peer review statute, which requires OEHHA to submit "the scientific portions of the proposed rule, along with a statement of the scientific findings, conclusions and assumptions on which the scientific portions of the proposed rule are based and the supporting scientific data, studies, and other appropriate materials, to the external scientific peer review entity for its evaluation" (emphasis added).The public comments constitute "other appropriate materials" because they provide important supplemental information that either was not included or not properly analyzed in OEHHA's TSD."

Response 15: The peer reviewers are asked to evaluate OEHHA's scientific findings, conclusions and assumptions as detailed in the draft PHG document as specified in Health and Safety Code § 57004(a)(2). Since the draft PHG document was completed prior to OEHHA's receipt of public comments, the comments are not "empirical data or other scientific findings, conclusions, or assumptions" upon which OEHHA's scientific findings or conclusions in the draft PHG document are based. Nevertheless, the public comments on the draft PHG document for HAAs were posted on OEHHA's website and the peer reviewers were provided a link to the comments in the peer review request letter.

Comment 16: "The peer review process is not transparent. Stakeholders have no visibility into how OEHHA develops charge questions or how it coordinates with the University of California to identify or select peer reviewers. The timeframe for peer review reports is unclear, and OEHHA does not post peer review reports for public inspection as they are submitted. Subsequent public review draft TSDs typically provide no indication of how OEHHA addressed peer reviewer comments in its proposed changes. OEHHA should correct these procedural deficiencies in future PHG peer reviews, starting with this one."

Response 16: OEHHA follows the required procedures for the review process set out in Health & Safety Code § 57004. OEHHA prepares a peer review package that contains a draft technical support document, along with a statement of the scientific findings, conclusions, and assumptions on which the scientific portions of the proposed PHGs are based. Reviewers are asked to determine whether the scientific work product is "based upon sound scientific knowledge, methods, and practices." The State Water Resources Control Board (SWRCB) coordinates with the University of California to identify and select peer reviewers. Once a peer review request/package is submitted to SWRCB, OEHHA has no knowledge of the peer reviewers' identities until the completed peer review package, with a report that contains an evaluation of the draft PHG, is returned. The timeframe for peer review reports depends on factors outside of OEHHA's control: identifying appropriate reviewers and waiting for their reviews. Comments from the external peer review of the proposed PHGs for HAAs were posted on OEHHA's website shortly after they were received. Furthermore, when the document is finalized, OEHHA posts detailed responses to the external scientific review comments, as well as public comments, in a separate document. The statute does not require OEHHA to identify changes to the draft document occasioned by the peer reviewers' comments. Therefore, OEHHA is following the requirements for peer review in statute..

RESPONSES TO COMMENTS RECEIVED FROM SOUTHERN CALIFORNIA WATER COALTION (May 2020)

Comment 1: "[O]ur principal concerns with this TSD are the threats it presents to effective drinking water disinfection and related risk communication."

Response 1: Effective drinking water disinfection and related risk communication are not within the scope of this assessment.

Comment 2: "These proposed PHGs have the potential to drive enforceable regulatory limits below levels that can be readily and affordably achieved by the chlorine-based disinfection technologies employed by most drinking water purveyors in California—especially if they lead to development of MCLs for individual HAAs."

Response 2: PHGs, based solely on health effects, are non-regulatory and are not supposed to consider affordability or the technologies required to achieve such goals. California law mandates that the SWRCB set MCLs that are economically and technologically feasible.

Comment 3: "Absent a rigorous quantitative risk balancing analysis, it is unclear how a future regulation that considers only long-term, theoretical cancer risks from exposure to DBPs will protect public health from potentially severe acute risks posed by microbiological contaminants that may occur both in source water and in the drinking water distribution system."

Response 3: Future regulations will consider more than cancer risks, including waterborne illnesses caused by microorganisms. California Health and Safety Code (HSC) §116350 requires the SWRCB to take into account technological and economical feasibility when setting a primary drinking standard. Beyond that, the California Safe Drinking Water Act requires SWRCB to administer other provisions related to the regulation of drinking water to protect public health. Water sanitation and microbial contamination are governed by different statutes (such as the Revised Total Coliform Rule[5]), procedures, and monitoring. Additionally, HSC §116360 states, "The department shall take all reasonable measures it determines necessary to reduce the risk to public health from waterborne illnesses in drinking water caused by cryptosporidium and giardia, to the extent those micro-organisms are not yet able to be adequately controlled through existing drinking water treatment and other management practices."

Comment 4: The Commenter expresses concerns about how the HAA MCL would affect other aspects of drinking water regulation including exposure to viruses (particularly, the novel coronavirus that causes COVID-19), aging infrastructure and drinking water rates.

[5] https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/rtcr.html

Response 4: OEHHA agrees that risk assessment of drinking water disinfection is an important health safety concern, as discussed in the PHG document. The SWRCB takes many factors into account, as explained in Response 3.

Comment 5: "Three of the five HAAs addressed in this TSD - trichloroacetic acid (TCA), dichloroacetic acid (DCA) and dibromoacetic acid (DBA) - have been added to the Proposition 65 list based on evidence that these substances cause tumors in studies of laboratory mice…In each case, there are open questions in the scientific literature about the relevance of the mouse data upon which the listings were based to human health risk assessment."

Response 5: Although the scientific literature may contain differing scientific views, the Proposition 65 listings were based on conclusions by expert authorities after comprehensive reviews of the entire scientific literature. TCA was listed as a carcinogen under Proposition 65 via the Labor Code mechanism, which mandates the listing of chemicals identified by the World Health Organization's International Agency for Research on Cancer (IARC) as causing cancer in humans or laboratory animals. DCA and DBA were listed via the Authoritative Bodies mechanism, which specifies that a chemical will be added to the Proposition 65 list when an authoritative body (e.g., US EPA for DCA and NTP for DBA) designated by the State's Qualified Experts committee formally identifies it as causing cancer.

Comment 6: "For TCA, there is consistent evidence of liver tumors in male mice but evidence for tumors is less consistent in female mice, and tumors have not been reported in rat studies. In addition, the mouse tumors appear to result from a non-genotoxic mechanism that can be defined as a threshold mechanism (i.e., no cancer risk below a threshold exposure level). Separate evaluations by the National Toxicology Program and U.S. EPA indicate that the PHG for TCA should not be based on carcinogenic effects."

Response 6: The available studies of TCA genotoxicity are discussed in detail in the PHG draft, and the in vivo and in vitro studies are summarized in Tables 7.4, and 7.5, respectively. Specifically, 13 out of 21 in vivo studies of genotoxicity reported positive findings of genotoxicity. Thus, OEHHA utilized the default linear approach, as further explained in the document.

In order to consider the threshold mechanism of carcinogenesis, OEHHA would require clear evidence that this is the sole MOA for TCA carcinogenesis. OEHHA was not able to locate such evidence. In contrast, the existing mechanistic studies point at several likely underlying mechanisms, including genotoxicity.

Furthermore, no NTP listing supports this comment. In its *Report on Carcinogens (RoC) Monograph on Haloacetic Acids Found as Water Disinfection By-Products*, the NTP did not designate TCA as a carcinogen (hazard identification) because it did not meet its criteria of reported cancer in two different species of laboratory animals (NTP, 2021). OEHHA utilizes different criteria for hazard identification of carcinogens and determined

TCA to be a likely human carcinogen based on the observation of cancer in male and female mice.

The US EPA determined that a dose-response analysis for TCA carcinogenicity is appropriate (US EPA, 2013). US EPA's dose-response analysis was based on the same critical study, DeAngelo et al. (2008), used by OEHHA.

While there are some differences in classification and MOA considerations between OEHHA and these other agencies, all assessments cover similar studies and acknowledge the observation of liver tumors in DeAngelo et al. (2008). OEHHA addresses differences in MOA interpretation in the PHG draft document, as described above.

Comment 7: "DCA appears to be weakly genotoxic and only at higher doses, which may indicate a threshold cancer mechanism."

Response 7: See response to Comment 7 from the American Chemistry Council.

Comment 8: "It should be noted that DCA has been used therapeutically in humans at doses as high as 25 mg/kg-day."

Response 8: Human studies of DCA, conducted in the clinical setting, are described in detail in the PHG document. Studies involved small numbers of subjects (less than 50) and lasted for a limited time, with the longest follow-up of 4-5 years (Barshop et al., 2004; Mori et al., 2004). As detailed in the PHG draft document, most studies were conducted in severely ill patients and focused on specific endpoints, which did not include cancer. The relatively short follow-up period would be insufficient for tumor development, and the small number of participants would further decrease the statistical power of the studies. It is worth noting that Mori et al. (2004) reported liver enlargement in all patients with the average dose of 30 mg/kg-day, although the nature of the underlying pathology was not further investigated.

DCA is approved for therapeutic use in humans, as a cauterizing agent and medical disinfectant, and DCA and its salts have been used in the treatment of congenital lactic acidosis. However, the approval of a drug for use as medication does not imply that it cannot cause long-term adverse effects, such as cancer. In fact, many drugs used acutely or for severe life-threatening conditions can increase the risk of cancer. For example, chloramphenicol sodium succinate is used as a broad-spectrum antibiotic for serious infections, but has been classified as a carcinogen by the Food and Drug Administration, and is listed under California's Proposition 65 as a carcinogen.

Comment 9: "Moreover, the mice in the key study selected by OEHHA for the DCA risk assessment exhibited a high rate of spontaneous liver tumors, which complicates interpretation of the study results. This study does not appear to be an appropriate foundation for a quantitative health risk assessment."

Response 9: Please see Response 13 in the responses to comments submitted by the American Chemistry Council above.

Comment 10: "Although there is more evidence of the genotoxicity of DBA (liver tumors in male mice, rare spontaneous tumors in rats), the mechanism for tumor induction has not been clearly identified and may involve precursor events that are non-genotoxic."

Response 10: In the presence of positive findings of genotoxicity, and when the mechanism of tumor induction has not been clearly defined, OEHHA cannot rule out a genotoxic MOA. In this case, the default choice is the linear extrapolation of cancer risk to lower dose, as was performed for DBA dose-response and PHG calculation in the PHG draft document.

RESPONSES TO COMMENTS RECEIVED FROM THE ENVIRONMENTAL WORKING GROUP (March 2020)

Comment 1: "EWG applauds OEHHA's approach of using Age Sensitivity Factors for different life stages for the cancer risk assessment of haloacetic acids and other contaminants. OEHHA's pioneering 2009 analysis[6] convincingly demonstrated the need for age-specific susceptibility factors for the assessment of carcinogens' impact on human health. This approach is also supported by the peer reviewed research literature,[7] which demonstrates that, at a minimum, a susceptibility factor of 10 should be applied to account for infants' and the developing fetus' greater vulnerability to toxic chemicals."

Response 1: OEHHA acknowledges the comment.

Comment 2: "In the table below, we summarize cancer slope factors for trihalomethanes and haloacetic acids published by OEHHA and the EPA. We note the overall similarity of the cancer slope factors, which supports OEHHA's proposed approach on both haloacetic acids and trihalomethanes, and we support OEHHA's decision to use the 5 percent increased risk benchmark for calculating the cancer slope factor. Further, cancer-based public health goals for haloacetic acids are supported by the findings from human epidemiological studies. EWG agrees with the references that OEHHA cites in the draft public health goal document that link the ingestion of drinking water containing disinfection byproducts to an increased risk of bladder cancer."

Response 2: OEHHA acknowledges the comment.

Comment 3: "In conclusion, EWG agrees with the methodology OEHHA used to derive the cancer and noncancer risk values for these chemicals, and we support OEHHA's approach to making the proposed public health goals protective for everyone, including those in vulnerable life stages, such as young children and the developing fetus. EWG urges OEHHA to finalize these proposed values as the final public health goals for the state of California."

Response 3: OEHHA acknowledges the comment.

RESPONSES TO COMMENTS RECEIVED FROM CLEAN WATER ACTION (May 2020)

Comment 1: "Clean Water Action, on behalf of its members throughout California, is pleased to support OEHHA's proposed public health goals (PHGs) of 0.2 parts per billion (ppb) for dichloroacetic acid, 0.1 ppb for trichloroacetic acid, and 0.03 ppb for dibromoacetic acid in order to correspond to a one in a million cancer risk in keeping with state policy. We also support the proposed PHGs for monochloroacetic acid (53 ppb) and monobromoacetic acid (25 ppb), which are based on non-cancer effects. While the need to disinfect drinking water is essential in protecting public health, we believe that these health goals will allow California to effectively offset unintended health consequences from these disinfection by-products. We particularly applaud that concentrations of the by-products consider sensitive populations such as infants and children who consume greater amounts of water by body weight and incur a greater cancer risk due to early exposures compared to adult exposures. OEHHA has thus taken a proper approach and we encourage them to finalize these PHGs with all expediency."

Response 1: OEHHA acknowledges the comment.

RESPONSES TO COMMENTS RECEIVED FROM MARC WILLIAM GOFF SR (March 2020)

Comment 1: "Data from research studies indicate that several HAAs, e.g., dichloroacetic acid and trichloroacetic acid, may be carcinogenic in laboratory animals. Exposure to other HAAs has also been associated with reproductive and developmental effects in laboratory animals. The current Maximum Contaminant Level (MCL) set for HAA5 is because of concern that exposure to HAAs over many years may increase the risk of cancer."

Response 1: OEHHA acknowledges the comment.

RESPONSES TO COMMENTS MADE DURING THE SECOND PUBLIC COMMENT PERIOD

RESPONSES TO COMMENTS RECEIVED FROM THE CHLORINE INSTITUTE (September 2022)

Comment 1: The Chlorine Institute comments on the role of chlorinating drinking water for public health and expresses concern that the proposed HAA PHG would "increase public health risk if the result is a move away from proven chlorine-based treatment technology." The Chlorine institute comments, "As the development of a PHG is a risk-based process, the public health risk of decreasing levels of chlorine disinfectants and the concomitant increase in microbial risk should be a necessary component of this process. If the proposed PHGs were to become final, water systems using chlorine-based disinfection could not readily achieve them, potentially jeopardizing the public health of Californians. The TSD should include a risk-benefit comparison of tradeoffs between the proposed PHG's and the loss of the ability to utilize chlorine-based disinfection methods." Through the comments, the stakeholder reiterates the need to incorporate a risk-benefit analysis in PHG development.

Response 1: The PHG development is not, in fact, a risk-based process that considers microbial risks. Rather, a PHG is derived solely based on health effects of the chemical in question. The concerns of balancing risks in modifying the chlorination of drinking water is addressed later by the Water Board at MCL development.

Comment 2: The Chlorine Institute states that their comments "are intended to reinforce those previously made by the American Chemistry Council (ACC) on May 1, 2020, regarding the First Public Review Draft." Regarding basing the PHGs on cancer risks, "It should be recognized that the studies that point in the direction of a relationship between cancer risk and DBP exposure are not consistent."

Response 2: The comments made by the American Chemistry Council are addressed earlier in this document.

Comment 3: "We are concerned about the manner in which this document was developed and reviewed. Although the announcement of the second draft indicates revisions have been made in light of previous comments, no substantive changes have been made to the conclusions of the second draft relative to the first as reported within the summaries."

Response 3: OEHHA carefully considered all peer review and public comments and made changes as appropriate. The rationale for specific changes (or lack thereof) can be found in this document.

REFERENCES CITED IN RESPONSES (ALL SECTIONS)

Bader E, Hrudey S, Froese K (2004). Urinary excretion half life of trichloroacetic acid as a biomarker of exposure to chlorinated drinking water disinfection by-products. Occupational and environmental medicine 61(8): 715-716.

Barshop BA, Naviaux RK, McGowan KA, et al. (2004). Chronic treatment of mitochondrial disease patients with dichloroacetate. Mol Genet Metab 83(1-2): 138-149.

Bailey MH, Tokheim C, Porta-Pardo E, et al. (2018) Comprehensive characterization of cancer driver genes and mutations. Cell 173: 371-385.

Bhunya S, Das P (1987). Bone marrow chromosome aberration and sperm abnormality in mice in vivo induced by monochloroacetic acid (MCA). Chromosome Inform Serv 42: 28-30.

Bokkers BG, Slob W (2007). Deriving a data-based interspecies assessment factor using the NOAEL and the benchmark dose approach. Critical Reviews In Toxicology 37(5): 355-373.

Bull RJ, Sanchez IM, Nelson MA, Larson JL, Lansing AJ (1990). Liver tumor induction in B6C3F1 mice by dichloroacetate and trichloroacetate. Toxicology 63(3): 341-359.

Bull RJ, Orner GA, Cheng RS, et al. (2002). Contribution of dichloroacetate and trichloroacetate to liver tumor induction in mice by trichloroethylene. Toxicol Appl Pharmacol 182(1): 55-65.

CalEPA Risk Assessment Advisory Committee (1996). A Review of the California Environmental Protection Agency's Risk Assessment Practices, Policies, and Guidelines.

Chiu WA, Okino MS, Evans MV (2009). Characterizing uncertainty and population variability in the toxicokinetics of trichloroethylene and metabolites in mice, rats, and humans using an updated database, physiologically based pharmacokinetic (PBPK) model, and Bayesian approach. Toxicol Appl Pharmacol 241(1): 36-60.

Chiu WA, Ginsberg GL (2011). Development and evaluation of a harmonized physiologically based pharmacokinetic (PBPK) model for perchloroethylene toxicokinetics in mice, rats, and humans. Toxicol Appl Pharmacol 253(3): 203-234.

Christian MS, York RG, Hoberman AM, et al. (2002). Oral (drinking water) two-generation reproductive toxicity study of dibromoacetic acid (DBA) in rats. Int J Toxicol 21(4): 237-276.

Cicmanec JL, Condie LW, Olson GR, Wang SR (1991). 90-Day toxicity study of dichloroacetate in dogs. Fundam Appl Toxicol 17(2): 376-389.

Crump KS (1995). Calculation of benchmark doses from continuous data. Risk Analysis 15(1): 79-89.

Dalgaard-Mikkelsen S, Kvorning SA, Moller KO (1955). Toxic effects of monobromacetic acid on pigs. Acta Pharmacol Toxicol (Copenh) 11(1): 13-32.

Davis JA, Gift JS, Zhao QJ (2011). Introduction to benchmark dose methods and US EPA's benchmark dose software (BMDS) version 2.1. 1. Toxicol Appl Pharmacol 254(2): 181-191.

DeAngelo AB, Daniel FB, Stober JA, Olson GR (1991). The carcinogenicity of dichloroacetic acid in the male B6C3F1 mouse. Fundam Appl Toxicol 16(2): 337-347.

DeAngelo AB, Daniel FB, Most BM, Olson GR (1996). The carcinogenicity of dichloroacetic acid in the male Fischer 344 rat. Toxicology 114(3): 207-221.

DeAngelo AB, Daniel FB, Most BM, Olson GR (1997). Failure of monochloroacetic acid and trichloroacetic acid administered in the drinking water to produce liver cancer in male F344/N rats. J Toxicol Environ Health 52(5): 425-445.

DeAngelo AB, George MH, House DE (1999). Hepatocarcinogenicity in the male B6C3F1 mouse following a lifetime exposure to dichloroacetic acid in the drinking water: dose-response determination and modes of action. J Toxicol Environ Health A 58(8): 485-507.

DeAngelo AB, Daniel FB, Wong DM, George MH (2008). The induction of hepatocellular neoplasia by trichloroacetic acid administered in the drinking water of the male B6C3F1 mouse. J Toxicol Environ Health A 71(16): 1056-1068.

ECHA (European Chemicals Agency). (2012) Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.8: Charaterization of dose[concentration]-response for human health. ECHA-2010-G-19-EN. Helsinki, Finland, European Chemical Agency.

EFSA (2017). Update: guidance on the use of the benchmark dose approach in risk assessment. EFSA Journal 15(1): 4658.

Ferreira-Gonzalez A, DeAngelo AB, Nasim S, Garrett CT (1995). Ras oncogene activation during hepatocarcinogenesis in B6C3F1 male mice by dichloroacetic and trichloroacetic acids. Carcinogenesis 16(3): 495-500.

Froese KL, Sinclair MI, Hrudey SE (2002). Trichloroacetic acid as a biomarker of exposure to disinfection by-products in drinking water: a human exposure trial in Adelaide, Australia. Environ Health Perspect 110(7): 679.

Gelman A, Carlin JB, Stern HS, et al. (2013). Bayesian data analysis, CRC press.

Gold LS, Zeiger E (1996). Handbook of carcinogenic potency and genotoxicity databases, CRC Press.

Guyton KZ, Rusyn I, Chiu WA, et al. (2018) Application of the key characteristics of carcinogens in cancer hazard identification. Carcinogenesis 39: 614-622.

Guyton KZ, Rieswijk L, Wang A, et al. (2018) Key characteristics approach to carcinogenic hazard identification. Chem. Res. Toxicol 31: 1290-1292.

Haber LT, Dourson ML, Allen BC, et al. (2018). Benchmark dose (BMD) modeling: current practice, issues, and challenges. Critical Reviews In Toxicology 48(5): 387-415.

Hard GC, Johnson KJ, Cohen SM (2009). A comparison of rat chronic progressive nephropathy with human renal disease—implications for human risk assessment. Critical Reviews In Toxicology 39(4): 332-346.

Harrington-Brock K, Doerr CL, Moore MM (1998). Mutagenicity of three disinfection by-products: di-and trichloroacetic acid and chloral hydrate in L5178Y/TK+/−− 3.7. 2C mouse lymphoma cells. Mutation Research/Genetic Toxicology and Environmental Mutagenesis 413(3): 265-276.

Herbert V, Gardner A, Colman N (1980). Mutagenicity of dichloroacetate, an ingredient of some formulations of pangamic acid (trade-named "vitamin B15"). Am J Clin Nutr 33(6): 1179-1182.

Herren-Freund SL, Pereira MA, Khoury MD, Olson G (1987). The carcinogenicity of trichloroethylene and its metabolites, trichloroacetic acid and dichloroacetic acid, in mouse liver. Toxicol Appl Pharmacol 90(2): 183-189.

ICH (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use). (2016) Impurities: Guidelines for Residual Solvents Q3C(R6). Geneva, Switzerland: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use.

Leavitt SA, DeAngelo AB, George MH, Ross JA (1997). Assessment of the mutagenicity of dichloroacetic acid in lacI transgenic B6C3F1 mouse liver. Carcinogenesis 18(11): 2101-2106.

Li T, Schultz I, Keys DA, Campbell JL, Fisher JW (2008). Quantitative evaluation of dichloroacetic acid kinetics in human—a physiologically based pharmacokinetic modeling investigation. Toxicology 245(1-2): 35-48.

Kavlock RJ, Austin CP, Tice RR. 2019. Chapter 2.13 - us vision for toxicity testing in the 21st century. In: The history of alternative test methods in toxicology. Elsevier, inc. Pp. 129–136.

Krewski D, Andersen ME, Tyshenko MG, Krishnan K, Hartung T, Boekelheide K, Wambaugh JF, Jones D, Whelan M, Thomas R et al. (2020). Toxicity testing in the 21st century: Progress in the past decade and future perspectives. Arch Toxicol. 94(1):1-58.

Kuljus K, von Rosen D, Sand S and Victorin, K. (2006) Comparing experimental designs for benchmark dose calculations for continuous endpoints. Risk Anal 26: 1031-1043.

Kurinnyĭ A (1984). Cytogenetic activity of the herbicide sodium trichloroacetate. TSitologiia i genetika 18(4): 318-319.

Larson JL, Bull RJ (1992). Metabolism and lipoperoxidative activity of trichloroacetate and dichloroacetate in rats and mice. Toxicol Appl Pharmacol 115(2): 268-277.

Mackay JM, Fox V, Griffiths K, et al. (1995). Trichloroacetic acid: investigation into the mechanism of chromosomal damage in the in vitro human lymphocyte cytogenetic assay and the mouse bone marrow micronucleus test. Carcinogenesis 16(5): 1127-1133.

Maron DM, Ames BN (1983). Revised methods for the Salmonella mutagenicity test. Mutation Research/Environmental Mutagenesis and Related Subjects 113(3-4): 173-215.

Mather GG, Exon JH, Koller LD (1990). Subchronic 90 day toxicity of dichloroacetic and trichloroacetic acid in rats. Toxicology 64(1): 71-80.

McConnell EE, Solleveld HA, Swenberg JA, Boorman GA (1986). Guidelines for combining neoplasms for evaluation of rodent carcinogenesis studies. J Natl Cancer Inst 76(2): 283-289.

Mori M, Yamagata T, Goto T, Saito S, Momoi MY (2004). Dichloroacetate treatment for mitochondrial cytopathy: long-term effects in MELAS. Brain Dev 26(7): 453-458.

Moriya M, Ohta T, Watanabe K, et al. (1983). Further mutagenicity studies on pesticides in bacterial reversion assay systems. Mutat Res 116(3-4): 185-216.

Nestmann ER, Chu I, Kowbel DJ, Matula TI (1980). Short-lived mutagen in Salmonella produced by reaction of trichloroacetic acid and dimethyl sulphoxide. Can J Genet Cytol 22(1): 35-40.

NTP (1992). NTP Toxicology and Carcinogenesis Studies of Monochloroacetic Acid (CAS No. 79-11-8) in F344/N Rats and B6C3F1 Mice (Gavage Studies). Natl Toxicol Program Tech Rep Ser. 396: 1-245.

NTP (1999). Toxicology data management system. Historical control tumor incidence summary. Oral water studies. National Toxicology Program, US Department of Health and Human Services.

NTP (2007). Toxicology and Carcinogenesis Studies of Dibromoacetic Acid (CAS No. 631-64-1) in F344/N Rats and B6C3F1 Mice (Drinking Water Studies). Research Triangle Park, NC, National Toxicology Program, US Department of Health and Human Services.

NTP (2009). NTP Historical Controls Report. National Toxicology Program, US Department of Health and Human Services.

NTP (2018). Report on carcinogens monograph on haloacetic acids found as water disinfection by-products. Peer review draft. March 30, 2018. Office of the Report on Carcinogens, Division of the National Toxicology Program, National Institute of Environmental Health Sciences, United States Department of Health and Human Services.

NTP (2021). Report on Carcinogens, Fifteenth Edition.; Research Triangle Park, NC: Division of the National Toxicology Program, National Institute of Environmental Health Sciences, US Department of Health and Human Services, Public Health Service. https://ntp.niehs.nih.gov/go/roc15

OECD (1997). OECD Guideline for the Testing of Chemicals. Bacterial Reverse Mutation Test 471. Paris, France, Organization for Economic Cooperation and Development.

OECD (2013). OECD Guideline for the Testing of Chemicals. Transgenic Rodent Somatic and Germ Cell Mutation Assays Test 488. Paris France, Organization for Economic Cooperation and Development.

OECD (2014). OECD Guideline for the Testing of Chemicals. Mammalian Erythrocyte Micronucleus Test 474. Paris, France, Organization for Economic Cooperation and Development.

OECD (2016). OECD Guideline for the Testing of Chemicals, Section 4. Mammalian Bone Marrow Chromosomal Aberration Test 475. Paris, France, Organization for Economic Cooperation and Development.

OEHHA (2008). Air Toxics Hot Spots Risk Assessment Guidelines: Technical Support Document for the Derivation of Noncancer Reference Exposure Levels. Office of Environmental Health Hazard Assessment, California Environmental Protection Agency, Sacramento, CA.

OEHHA (2009). Technical Support Document for Cancer Potency Factors: Methodologies for Derivation, Listing of Available Values, and Adjustments to Allow for Early Life Stage Exposures. Office of Environmental Health Hazard Assessment, California Environmental Protection Agency, Sacramento, CA.

Pereira MA (1996). Carcinogenic activity of dichloroacetic acid and trichloroacetic acid in the liver of female B6C3F1 mice. Fundam Appl Toxicol 31(2): 192-199.

Pereira MA, Phelps JB (1996). Promotion by dichloroacetic acid and trichloroacetic acid of N-methyl-N-nitrosourea-initiated cancer in the liver of female B6C3F1 mice. Cancer Lett 102(1-2): 133-141.

Saghir SA, Schultz IR (2005). Toxicokinetics and oral bioavailability of halogenated acetic acids mixtures in naive and GSTzeta-depleted rats. Toxicol Sci 84(2): 214-224.

Schmidt CW, Gillis CA, Keenan RE and Price PS (1997). Characterizing inter-chemical variation in the interspecies uncertainty factor (UFA) [Abstract #1057]. The Toxicologist 36(1 Pt 2): 208.

Schultz IR, Merdink JL, Gonzalez-Leon A, Bull RJ (1999). Comparative toxicokinetics of chlorinated and brominated haloacetates in F344 rats. Toxicol Appl Pharmacol 158(2): 103-114.

Shao K, Shapiro AJ (2018). A web-based system for Bayesian benchmark dose estimation. Environ Health Perspect 126(1): 017002.

Slob W, Moerbeek M, Rauniomaa E, Piersma AH (2005). A statistical evaluation of toxicity study designs for the estimation of the benchmark dose in continuous endpoints. Toxicological Sciences 84(1): 167-185.

Smith MT, Guyton KZ, Gibbons CF, et al. (2016) Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. Environ Health Perspect 124: 713-721.

Stratton MR, Campbell PJ and Futreal PA (2009) The cancer genome. Nature 458: 719-724

Styles JA, Wyatt I, Coutts C (1991). Trichloroacetic acid: studies on uptake and effects on hepatic DNA and liver growth in mouse. Carcinogenesis 12(9): 1715-1719.

Templin MV, Parker JC, Bull RJ (1993). Relative formation of dichloroacetate and trichloroacetate from trichloroethylene in male B6C3F1 mice. Toxicol Appl Pharmacol 123(1): 1-8. [Erratum 133(131):177, July, 1995].

Templin MV, Stevens DK, Stenner RD, et al. (1995). Factors affecting species differences in the kinetics of metabolites of trichloroethylene. J Toxicol Environ Health 44(4): 435-447.

US EPA (1988). Recommendations for and Documentation of Biological Values for Use in Risk Assessment. EPA-600/6-87/008. NTIS PB88-179874/AS. Environmental Criteria and Assessment Office, Office of Research and Development. Cincinnati, Ohio, United States Environmental Protection Agency.

US EPA (1992) A cross-species scaling factor for carcinogen risk assessment based on equivalence of Mg/Kg0.75/Day. Federal Register 57: 24152-24173.

US EPA (2002a). A review of the reference dose and reference concentration processes. EPA/630/P-02/002F. Risk Assessment Forum. Washington, DC., United States Environmental Protection Agency.

US EPA (2002b). Consideration of the FQPA Safety Factor and Other Uncertainty Factors in Cumulative Risk Assessment of Chemicals Sharing a Common Mechanism of Toxicity. U.S. Environmental Protection Agency, Washington, DC. Available at: https://www.epa.gov/sites/production/files/2015-07/documents/apps-10x-sf-forcra.pdf.

US EPA (2004). An Examination of EPA Risk Assessment Principles and Practices. EPA/100/B-04/001. Washington (DC): Office of the Science Advisor, United States Environmental Protection Agency.

US EPA (2005). Guidelines for carcinogen risk assessment. EPA/630/P-03/001F. Washington, DC., United States Environmental Protection Agency.

US EPA (2006). National Primary Drinking Water Regulations: Stage 2 Disinfectants and Disinfection Byproducts Rule; Final Rule. Code of Federal Register (CFR) 40 CFR Parts 9, 141, and 142. Fed Reg 71(2): 388-493.

US EPA (2011) Toxicological Review of Trichloroacetic Acid (CAS No. 76-03-9). In Support of Summary Information on the Integrated Risk Information System (IRIS). U.S. Environmental Protection Agency, Washington, DC. Available at: https://cfpub.epa.gov/ncea/iris/iris_documents/documents/toxreviews/0655tr.pdf

US EPA (2012). Benchmark Dose Technical Guidance. EPA/100/R-12/001. Washington, D.C., United States Environmental Protection Agency.

US EPA (2013). Trichloroacetic acid. In: IRIS, updated 9/30/2011. U.S. EPA, Washington, DC, accessed at: http://www.epa.gov/iris/subst/0655htm.

US EPA (2016a). Drinking Water Health Advisory for Perfluorooctanoic Acid (PFOA). EPA 822-R-16-005. Washington DC, United States Environmental Protection Agency.

US EPA (2016b). Drinking Water Health Advisory for Perfluorooctane Sulfonate (PFOS). EPA 822-R-16-002. Washington DC, United States Environmental Protection Agency

White PA, Long AS, Johnson GE (2020). Quantitative interpretation of genetic toxicity dose-response data for risk assessment and regulatory decision-making: current status and emerging priorities. Environmental and molecular mutagenesis 61(1): 66-83.

Wood CE, Hester SD, Chorley BN, et al. (2015). Latent carcinogenicity of early-life exposure to dichloroacetic acid in mice. Carcinogenesis 36(7): 782-791.