Kan Shao, PhD
Associate Professor
Department of Environmental and Occupational Health
Indiana University School of Public Health - Bloomington
1025 E Seventh Street
Bloomington, IN 47405

September 7, 2020

Re: Request for an External Peer Review of the Office of Environmental Health Hazard Assessment Draft Public Health Goals for Haloacetic Acids in Drinking Water

Based on my expertise and experience, I am reviewing the findings, assumptions, or conclusions I agreed I could review with confidence, including:

1. Conclusion 1: After reviewing the published literature on monochloroacetic acid (MCA), OEHHA concludes that systemic toxicity is the most sensitive adverse health effect associated with exposure to this chemical. OEHHA is basing the PHG on this critical endpoint and its supporting studies. Available cancer bioassays for MCA showed no evidence of carcinogenic activity. 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.
2. Conclusion 2: After reviewing the published literature on dichloroacetic acid (DCA), OEHHA concludes that cancer is the primary adverse health effect associated with human exposure to this chemical. OEHHA is basing the PHG on this critical endpoint and its supporting studies. For the PHG, OEHHA is using the BMD approach for cancer slope factor (CSF) determination from animal toxicology studies. While the BMD approach was applied to noncancer datasets amenable to modeling, the POD for DCA is determined from the NOAEL for liver toxicity from an animal toxicology study.
3. Conclusion 3: After reviewing the published literature on trichloroacetic acid (TCA), OEHHA concludes that cancer is the primary adverse health effect associated with human exposure to this chemical. OEHHA is basing the PHG on this critical endpoint and its supporting studies. For both the PHG and noncancer health-protective concentration, OEHHA is using the BMD approach for CSF and POD determination, respectively, from animal toxicology studies.
4. Conclusion 4: Studies of monobromoacetic acid (MBA) are very limited. After reviewing the available studies, OEHHA concludes that muscular degeneration is the most sensitive adverse health effect associated with exposure to this chemical. OEHHA is basing the PHG on this critical endpoint. While many studies on the genotoxicity of MBA were positive, OEHHA was not able to locate

any carcinogenicity studies for this chemical. The POD for MBA is determined from the NOAEL for muscular degeneration from an animal toxicology study.

5. Conclusion 5: After reviewing the published literature on dibromoacetic acid (DBA), OEHHA concludes that cancer is the primary adverse health effect associated with human exposure to this chemical. OEHHA is basing the PHG on this critical endpoint and its supporting studies. For the PHG, OEHHA is using the BMD approach for CSF determination from animal toxicology studies. While the BMD approach was applied to noncancer datasets amenable to modeling, the POD for DBA is determined from the lowest-observed-adverse-effect level (LOAEL) for male reproductive toxicity from an animal toxicology study.

The comments provided in this document mainly focus on dose-response modeling and assessment (including both the BMD method and NOAEL method) for cancer and/or noncancer endpoints in each of the five conclusions listed above.

Conclusion 1: After reviewing the published literature on monochloroacetic acid (MCA), OEHHA concludes that systemic toxicity is the most sensitive adverse health effect associated with exposure to this chemical. OEHHA is basing the PHG on this critical endpoint and its supporting studies. Available cancer bioassays for MCA showed no evidence of carcinogenic activity. 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.

The systemic toxicity (i.e., decreased body weight and relative liver weight) reported in DeAngelo et al (1997) was selected as the critical effect and used to derived the POD for MCA. A number of issues in Chapter 10 from Page 197 to Page 199 should be addressed or clarified:

(1) To facilitate the presentation of dose-response assessment and its results, it is better to list data that are ready for dose-response analysis in Table 10.2, rather than showing the data directly collected from the cited studies which is little redundant (much information has been presented in Chapter 5).
   a. In DeAngelo et al 1997, 50/sex/dose Male F344/N rats was the number of animal 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. (Additionally, the number of animals per dose group in DeAngelo et al 1997 reported in Table 5.5 is not correct.)
   b. 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.

(2) 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. 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.

(3) 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.

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.

<u>Conclusion 2</u>: After reviewing the published literature on dichloroacetic acid (DCA), OEHHA concludes that cancer is the primary adverse health effect associated with human exposure to this chemical. OEHHA is basing the PHG on this critical endpoint and its supporting studies. For the PHG, OEHHA is using the BMD approach for cancer slope factor (CSF) determination from animal toxicology studies. While the BMD approach was applied to noncancer datasets amenable to modeling, the POD for DCA is determined from the NOAEL for liver toxicity from an animal toxicology study.

For the dose-response assessment of non-cancer effects, a few important issues should be addressed or clarified.

(1) Relative liver weight increase in male B6C3F1 mice reported in DeAngelo et al (1991) was selected as the basis for deriving the POD. The data set is shown in Table 1 below. The selection of 7.6 mg/kg-day as the POD using the NOAEL method is not very well supported by the evidence for a few reasons:

Table 1. Relative liver weight of male B6C3F1 mice after exposure to DCA for 60 weeks

Dose (mg/kg-day)	N	Mean	SD

0	9	5.01	0.32
7.6	9	5.25	0.78
77	9	6.83	1.92
410	12	11.63	1.13
486	30	17.57	4.37

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).

(2) Based on Figure D7 presented on Page 424, it 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 Cicmanec et al 1991 should be updated. In addition, as mentioned in (1)a the data of female dogs might be fit adequately if not just relying on the goodness-of-fit p-value or if using another BMD modeling tool.

(3) 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.

For the dose-response assessment of cancer effects, two issues should be addressed.

(1) 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 q_1 and q_2 were reduced) has better fitting performance (P-value: 0.5510, AIC: 41.3154) than the LMS (P-value: 0.3673, AIC: 41.789).

(2) 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. 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.

In short, the selection of critical studies and endpoints for deriving the POD and cancer slope factor are appropriate, but the dose-response modeling approaches applied should be replaced by a more appropriate model/approach or better justified.

<u>Conclusion 3</u>: After reviewing the published literature on trichloroacetic acid (TCA), OEHHA concludes that cancer is the primary adverse health effect associated with human exposure to this chemical. OEHHA is basing the PHG on this critical endpoint and its supporting studies. For both the PHG and noncancer health-protective concentration, OEHHA is using the BMD approach for CSF and POD determination, respectively, from animal toxicology studies.

For the dose-response assessment of non-cancer effects, a few issues should be addressed or clarified.

(1) 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. 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.

(2) To keep the format consistent in this Chapter, the column name of the last two columns in Table 10.7 should be named as "BMD_{05} (mg/kg-day)" and "$BMDL_{05}$ (mg/kg-day)"

(3) 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". 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.

For the dose-response assessment of cancer effects, the cancer slope factor was estimated based on hepatocellular adenoma or carcinoma in B6C3F1 male mice after 104 weeks exposure to TCA as reported in DeAngelo et al (2008). In Table 10.14, the BMD and BMDL value were estimated using the 1st degree multistage model. However, no justification was provided in the report explaining why the 1st degree MS model was chosen over the 2nd degree MS model. There are three dose levels in this data set, so 2nd degree MS model can be used. As reported in BMDS 2.7, the goodness-of-fit p-value and AIC value for 2nd degree MS model are 0.2307 and 210.01 respectively, and the corresponding values for 1st degree MS model are 0.1608 and 210.539. Therefore, based on these statistics, 2nd degree MS is a more plausible model to derive the cancer slope factor.

In short, the selection of critical studies and endpoints for deriving the POD and cancer slope factor appears to be appropriate, but the BMD modeling approach and results should be better clarified and justified.

<u>Conclusion 4</u>: Studies of monobromoacetic acid (MBA) are very limited. After reviewing the available studies, OEHHA concludes that muscular degeneration is the most sensitive adverse health effect associated with exposure to this chemical. OEHHA is basing the PHG on this critical endpoint. While many studies on the genotoxicity of MBA were positive, OEHHA was not able to locate any carcinogenicity studies for this chemical. The POD for MBA is determined from the NOAEL for muscular degeneration from an animal toxicology study.

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.

<u>Conclusion 5</u>: After reviewing the published literature on dibromoacetic acid (DBA), OEHHA concludes that cancer is the primary adverse health effect associated with human exposure to this chemical. OEHHA is basing the PHG on this critical endpoint and its supporting studies. For the PHG, OEHHA is using the BMD approach for CSF determination from animal toxicology studies. While the

BMD approach was applied to noncancer datasets amenable to modeling, the POD for DBA is determined from the lowest-observed-adverse-effect level (LOAEL) for male reproductive toxicity from an animal toxicology study.

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. For cancer endpoints, although the higher degree Multistage models were eventually reduced to the first degree multistage model (i.e., LMS), explanation should be provided in the report.

In addition to the specific comments on each of the five conclusions provided above, two comments regarding the format of tables are provided below to make the presentation of the report more consistent and clear:

(1) 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.

(2) 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.

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.

References:

EFSA Scientific Committee, Hardy A, Benford D, Halldorsson T, Jeger MJ, Knutsen KH, More S, Mortensen A, Naegeli H, Noteborn H, Ockleford C, Ricci A, Rychen G, Silano V, Solecki R, Turck D, Aerts M, Bodin L, Davis A, Edler L, Gundert-Remy U, Sand S, Slob W, Bottex B, Abrahantes JC, Marques DC, Kass G and Schlatter JR, 2017. Update: Guidance on the use of the benchmark dose approach in risk assessment. EFSA Journal 2017;15(1): 4658, 41 pp.

Shao K; Shapiro A. A Web Based System for Bayesian Benchmark Dose Estimation. Environmental Health Perspectives. 2018. 126 (1): 017002.

US EPA (US Environmental Protection Agency). Benchmark dose technical guidance document. EPA/100/R-12/001. 2012. Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum.

Review report on the draft document entitled “Public Health Goals – Haloacetic Acids in Drinking Water: Monochloroacetic Acid, Dichloroacetic Acid, Trichloroacetic Acid, Monobromoacetic Acid, and Dibromoacetic Acid”

Zhoumeng Lin, BMed, PhD, DABT, CPH
Assistant Professor of Toxicology
Institute of Computational Comparative Medicine (ICCM)
Department of Anatomy and Physiology, College of Veterinary Medicine
Kansas State University, Manhattan, KS
September 10, 2020

Based on my expertise and experience, I am reviewing the findings, assumptions, and conclusions I agreed I could review with confidence. Based on Page Two of the OEHHA letter of review request sent to me, there are five expertise types required for this review, and these expertise types are cross-referenced to specific five haloacetic acid (HAA) proposed public health goals. My expertise is in pharmacokinetic modeling, dose-response analysis, and exposure assessment. This expertise type is needed to assess noncancer risk assessment for all five HAAs (expertise type #1), and the cancer risk assessment for DCA, TCA, and DBA (expertise type #3). As discussed with Dr. Gerald Bowes, my review below will focus on the findings, assumptions, and conclusions of all 5 HAAs that are related to my expertise in pharmacokinetic modeling, dose-response analysis, and exposure assessment. These related findings, assumptions, and conclusions, along with my comments, are listed below by number, as they are referred to in Attachment 2 of the review request. In addition, after reading through the entire document, I have a number of specific comments that are numerically listed below according to the page number.

1. Scientific Assumptions, Findings, and Conclusions for Monochloroacetic Acid (MCA) to be Commented:

After reviewing the published literature on monochloroacetic acid (MCA), OEHHA concludes that systemic toxicity is the most sensitive adverse health effect associated with exposure to this chemical. OEHHA is basing the PHG on this critical endpoint and its supporting studies. Available cancer bioassays for MCA showed no evidence of carcinogenic activity. 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.

PHG: Carcinogenicity studies of MCA are limited. The two available two-year studies of orally administered MCA (NTP, 1992; DeAngelo et al., 1997) did not show evidence of carcinogenic activity. MCA has not been listed as a carcinogen by the US Environmental Protection Agency (US EPA) or the National Toxicology Program (NTP). The International Agency for Research on Cancer (IARC) has not classified the carcinogenicity of MCA.

Systemic (NTP, 1992; DeAngelo et al., 1997) and cardiovascular (Daniel et al., 1991; NTP, 1992; DeAngelo et al., 1997) toxicities appear to be the most sensitive endpoints for MCA in animal studies. Among the candidate critical studies, NTP (1992) and DeAngelo et al. (1997) produced the lowest candidate PODs.

NTP (1992) reported increased mortality, which produced a $BMDL_{05}$ (lower 95% confidence limit of the benchmark dose associated with 5% extra risk) of 3.4 mg/kg-day for female rats. However, this value was not selected as the POD because the observed mortality was described as “due to undetermined causes,” and its relevance to human health was unclear.

The statistically significant decreases in relative liver weight and final body weight (collectively referred to as systemic toxicity) reported in DeAngelo et al. (1997) had the lowest NOAEL, 3.5 mg/kg-day, which was selected as the POD. After applying a total uncertainty factor (UF) of 1,000, a relative source contribution (RSC) of 80% and a

lifetime-weighted average drinking water intake rate of 0.053 L/kg-day, the resulting health-protective concentration for noncancer effects is 53 µg/L or parts per billion (ppb).

Comment:

(1.1) 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 $BMDL_{05}$ 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 $BMDL_{05}$ 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.

(1.2) I have a concern on the BMD analysis method. 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”. I am concerned about this statement and this criterion that was used in the benchmark dose-response analysis. 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. Specifically, the criteria of setting the BMR from the US EPA guideline (US EPA, 2012) cited by the authors are quoted here:

- An extra risk of 10% is recommended as a standard reporting level for quantal data, for the purposes of making comparisons across chemicals or endpoints. The 10% response level has customarily been used for comparisons because it is at or near the limit of sensitivity in most cancer bioassays and in noncancer bioassays of comparable size. Note that this level is not a default BMR for developing PODs or for other purposes.
- Biological considerations may warrant the use of a BMR of 5% or lower for some types of effects (e.g., frank effects), or a BMR greater than 10% (e.g., for early precursor effects) as the basis of a POD for a reference value.
- Sometimes, a BMR lower than 10% (based on biological considerations), falls within the observable range. From a statistical standpoint, most reproductive and developmental studies with nested study designs easily support a BMR of 5%. Similarly, a BMR of 1% has typically been used for quantal human data from epidemiology studies. In other cases, if one models below the observable range, one needs to be mindful that the degree of uncertainty in the estimates increases. In such cases, the BMD and BMDL can be compared for excessive divergence. In addition, model uncertainty increases below the range of data.

Also, according to a recent critical review on benchmark dose-response modeling by a group of prominent risk assessors and toxicologists (Haber et al., 2018), it is mentioned that “Both EFSA (2017) and US EPA (2012) focus on the 10% response range in determining the BMR for dichotomous data, for similar reasons, although the implementation specifics differ”.

Based on US EPA guideline (US EPA, 2012), the review by Haber et al. (2018), and my experience, I suggest 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.

(1.3) I also have a concern on the uncertainty factors. As listed in Table 10.1 on Page 195, some of the default uncertainty factors for PHG derivation are different from those recommended by US EPA (US EPA, 2002; 2011; 2016a; 2016b). For example, regarding the intraspecies uncertainty factor (UF_H), it is indicated that UF_H due to toxicokinetic component could be up to 10, and UF_H due to toxicodynamic component could be up to 10. What about the UF_H due to combined toxicokinetic and toxicodynamic components? This is not mentioned in the table, but does it mean that it could be up to 100 (10 X 10 = 100)? Please note that according 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. Also, based on the toxicological review document for TCA by US EPA Integrated Risk Information System (US EPA, 2011), the UF to account for human variation is also 10. Therefore, 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.

(1.4) On Page 199, the total UF of 1000 for MCA is based on the consideration of an UF_H of 30 for variation in the human population, which includes $\sqrt{10}$ 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.

2. Scientific Assumptions, Findings, and Conclusions for Dichloroacetic Acid (DCA) to be Commented:

After reviewing the published literature on dichloroacetic acid (DCA), OEHHA concludes that cancer is the primary adverse health effect associated with human exposure to this chemical. OEHHA is basing the PHG on this critical endpoint and its supporting studies. For the PHG, OEHHA is using the BMD approach for cancer slope factor (CSF) determination from animal toxicology studies. While the BMD approach was applied to noncancer datasets amenable to modeling, the

POD for DCA is determined from the NOAEL for liver toxicity from an animal toxicology study.

PHG: In NTP's 2018 Report on Carcinogens Monograph on Haloacetic Acids Found as Water Disinfection By-Products, DCA was classified as reasonably anticipated to be a human carcinogen based on sufficient evidence in experimental animals and supporting mechanistic data. IARC (2014) classified DCA as Group 2B, possibly carcinogenic to humans, based on sufficient evidence in experimental animals and US EPA classified DCA as a "likely human carcinogen" in 1998. DCA has been listed as a carcinogen under California's Proposition 65 since 1996. Numerous studies have reported liver cancer in animals orally exposed to DCA. Although the overall data are mixed, positive results have been reported for a number of mutagenicity and genotoxicity assays in bacterial and mammalian test systems. Thus, the weight of the evidence indicates DCA is mutagenic and genotoxic.

To determine the PHG, OEHHA first derived a cancer slope factor (CSF) of 0.041 $(mg/kg\text{-}day)^{-1}$ from the data for hepatic adenoma and carcinoma in male B6C3F1 mice reported by DeAngelo et al. (1999). Using this CSF, along with age sensitivity factors (ASFs) to account for the increased susceptibility of infants and children to carcinogens, and a cancer risk of 10^{-6}, the proposed PHG is 0.2 ppb.

Noncancer public health-protective concentration: Due to its use as a proposed therapy for some severe human pathologies, such as lactic acidosis, there are numerous reports of neuropathy in humans exposed subchronically to DCA. In animal experiments, DCA exposure induced several noncancer adverse effects, including liver, developmental and reproductive toxicity, and neurotoxicity. DCA is listed as a developmental toxicant under Proposition 65.

To determine a health-protective concentration for noncancer health effects, OEHHA analyzed the dose-response data from 5 studies reporting liver toxicity (Toth et al., 1992; Mather et al., 1990; Cicmanec et al., 1991; DeAngelo et al., 1991; DeAngelo et al., 1999) and from 2 studies reporting reproductive toxicity (Toth et al., 1992; Cicmanec et al., 1991). BMD modeling was performed on all the datasets; however, the NOAEL of 7.6 mg/kg-day for increased liver weight from DeAngelo et al. (1991) was selected as the POD. After applying a total UF of 1,000, an RSC of 80% and a lifetime-weighted average drinking water intake rate of 0.053 L/kg-day, the resulting health-protective concentration for noncancer effects is 115 ppb.

Comment:

(2.1) It is indicated that "While the BMD approach was applied to noncancer datasets amenable to modeling, the POD for DCA is determined from the NOAEL for liver toxicity from an animal toxicology study.". A justification of the selection of the critical study and the endpoint for noncancer toxicity of DCA is provided on Page 203. Based on this justification and based on the values of various NOAELs and $BMDL_{05}$ from different studies, I consider this selection as justifiable. However, similar to my comment on MCA above, I also have a concern on the use of 5% as the BMR in the BMD modeling and the selection of the uncertainty factors described on Page 204. Please provide a clarification on the use of 30 to account for human variability on Page 204. Please also refer to my Comment 1.3 and 1.4 above.

(2.2) 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. Please refer to my specific comments on Appendix D and Appendix E as listed below (Comments 6.44 and 6.46, respectively).

3. Scientific Assumptions, Findings, and Conclusions for Trichloroacetic Acid (TCA) to be Commented:

After reviewing the published literature on trichloroacetic acid (TCA), OEHHA concludes that cancer is the primary adverse health effect associated with human exposure to this chemical. OEHHA is basing the PHG on this critical endpoint and its supporting studies. For both the PHG and noncancer health-protective concentration, OEHHA is using the BMD approach for CSF and POD determination, respectively, from animal toxicology studies.

PHG: US EPA (2011, 2013) concluded that "there is suggestive evidence of carcinogenic potential for TCA based on significantly increased incidences of liver tumor," and IARC classified TCA as group 2B, possibly carcinogenic to humans, based on sufficient evidence for carcinogenicity in animals, and inadequate evidence in humans (Guha et al, 2012; IARC, 2014). TCA was listed as a carcinogen under Proposition 65 in 2013, based on the IARC identification of TCA as a carcinogen. However, NTP (2018) considered the existing evidence of TCA carcinogenicity as not sufficient for listing in its Report on Carcinogens.

There are several chronic studies in mice of both sexes that demonstrate a statistically significant increase in liver tumors. While results from in vitro genotoxicity studies mostly negative, in vivo administration of neutralized TCA appears to be genotoxic in rodents (OEHHA, 1999; US EPA, 2005).

To determine the PHG, OEHHA first derived a CSF of 0.071 $(mg/kg\text{-}day)^{-1}$ from the data for hepaticellular adenoma or carcinoma in male B6C3F1 mice reported by DeAngelo et al. (2008). Using this CSF, along with ASFs to account for the increased susceptibility of infants and children to carcinogens, and a cancer risk of 10^{-6}, the proposed PHG is 0.1 ppb.

Noncancer public health-protective concentration: In animal studies, subchronic and chronic oral exposure to TCA appears to primarily affect liver size and weight, collagen deposition, lipid and carbohydrate metabolism, peroxisome proliferation, and body weight. To determine a health-protective concentration for noncancer health effects, OEHHA analyzed the dose-response data from 2 studies reporting liver toxicity (DeAngelo et al., 2008; Bull et al., 2002) and 1 study reporting decreased body weight (DeAngelo et al., 1997). BMD modeling was performed on 3 datasets; the data for decreased body weight were presented in a graph and could not be modeled. The $BMDL_{05}$ of 8.45 mg/kg-day for hepatocellular necrosis was selected as the POD. After applying a total UF of 1,000, an RSC of 80% and a lifetime-weighted average drinking water intake rate of 0.053 L/kg-day, the resulting health-protective concentration for noncancer effects is 128 ppb.

Comment: 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)". For example, in Table 10.14, if the animal CSF is 0.011 mg/kg/day, by using the equation of $0.011 * (70/0.04)^{0.25}$, I got 0.071 mg/kg/day for human CSF, which is exactly the same as presented in Table 10.14. However, for the other two datasets, if the animal CSF is 0.0018, by using the equation of $0.0018 * (70/0.04)^{0.25}$, I got 0.0116 mg/kg/day for human CSF, which is different from 0.061 presented in Table 10.14. Also, I got a different human CSF value if the animal CSF is 0.0045. 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.

4. Scientific Assumptions, Findings, and Conclusions for Monobromoacetic acid (MBA) to be Commented:

Studies of monobromoacetic acid (MBA) are very limited. After reviewing the available studies, OEHHA concludes that muscular degeneration is the most sensitive adverse health effect associated with exposure to this chemical. OEHHA is basing the PHG on this critical endpoint. While many studies on the genotoxicity of MBA were positive, OEHHA was not able to locate any carcinogenicity studies for this chemical. The POD for MBA is determined from the NOAEL for muscular degeneration from an animal toxicology study.

PHG: No chronic studies with MBA are available. There are two oral MBA studies of subacute to subchronic duration (Linder et al., 1994; Dalgaard-Mikkelsen et al., 1955); however, the Linder et al. (1994) study had limitations, such as its 2-week exposure duration and single dose exposure, that precluded its use as a critical study.

While the Dalgaard-Mikkelsen et al. (1955) also had several limitations, it was selected as the critical study for PHG derivation. This is a multigenerational study in pigs, in which first generation (F0) animals showed no adverse effects following a 15-month exposure to an estimated average dose of 5 mg/kg-day MBA. Animals in the second generation (F1) demonstrated an array of toxic effects, including skeletal muscle degeneration. OEHHA determined a NOAEL of 5 mg/kg-day, based on the lack of adverse effects in the F0 generation, as the POD. After applying a total UF of 3,000, an RSC of 80% and a lifetime-weighted average drinking water intake rate of 0.053 L/kg-day, the resulting health-protective concentration for noncancer effects is 25 ppb.

Comment: 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 $\sqrt{10}$ 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? 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? Fourth, as stated above, the UF of 30 for intraspecies variability needs clarification, too.

5. Scientific Assumptions, Findings, and Conclusions for Dibromoacetic acid (DBA) to be Commented:

After reviewing the published literature on dibromoacetic acid (DBA), OEHHA concludes that cancer is the primary adverse health effect associated with human exposure to this chemical. OEHHA is basing the PHG on this critical endpoint and its supporting studies. For the PHG, OEHHA is using the BMD approach for CSF determination from animal toxicology studies. While the BMD approach was applied to noncancer datasets amenable to modeling, the POD for DBA is determined from the lowest-observed-adverse-effect level (LOAEL) for male reproductive toxicity from an animal toxicology study.

PHG: DBA was listed as a carcinogen under Proposition 65 in 2008. NTP (2018) classified DBA as reasonably anticipated to be a human carcinogen based on sufficient evidence in experimental animals and supporting mechanistic data. Several studies have shown that DBA is genotoxic in vitro and in vivo, and IARC (2013) concluded there is moderate evidence of a genotoxic mechanism for its carcinogenicity.

NTP (2007) examined the carcinogenicity of DBA in male and female rats and mice exposed via drinking water. An increase in malignant mesothelioma in male rats and mononuclear cell leukemia in male and female rats was observed. DBA treatment also resulted in a statistically significant dose-related increase in hepatocellular adenomas and carcinomas in both sexes of mice, and hepatoblastomas and alveolar/bronchiolar adenomas in male mice. To determine the PHG, OEHHA first derived a CSF of 0.25 $(mg/kg\text{-}day)^{-1}$ from a multisite analysis of the data for hepaticellular adenoma, carcinoma and/or hepatoblastoma, and alveolar/bronchiolar adenoma in male B6C3F1 mice reported by NTP (2007). Using this CSF, along with ASFs to account for the increased susceptibility of infants and children to carcinogens, and a cancer risk of 10^{-6}, the proposed PHG is 0.03 ppb.

Noncancer public health-protective concentration: Noncancer adverse effects considered for POD determination include hepatotoxicity, nephropathy, alveolar epithelial hyperplasia, developmental, and reproductive toxicity. Datasets amenable to BMD modeling were those for hepatic cystic degeneration, nephropathy, alveolar epithelial hyperplasia, and relative liver weight. However, none of the resulting BMDLs was selected as the POD due to low sensitivity compared to the NOAEL/LOAEL from other studies or due to uncertainties in the modeling results. The LOAEL of 1 mg/kg/day for male reproductive toxicity, specifically testicular lesions and increased morphologically abnormal sperm (Veeramachaneni et al., 2007) is selected as the POD.

After applying a total UF of 3,000, an RSC of 80% and a lifetime-weighted average drinking water intake rate of 0.053 L/kg-day, the resulting health-protective concentration for noncancer effects is 5 ppb.

Comment: Similar to my comments on other HAA5, my concern on the scientific assumptions, findings, and conclusions for DBA is mainly on the use of BMR of 5% in the BMD analysis and the default uncertainty factors. Please refer to the comments above for detailed description. Nevertheless, if we assume the BMD analysis results 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.

6. Specific Comments.

After reading through the document, I have the following specific comments numerically listed according to the page number for easy referencing.

(6.1) Page 3: Table S1, there is a typo in the row of "Dichloroacetic acid". Please change "Liver toxcity" to "Liver toxicity".

(6.2) 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?

(6.3) Page 6, Paragraph 1: Please include "RoC" in the list of commonly used abbreviations.

(6.4) Page 8, Paragraph 1: Please cite a few representative study to support this statement: "In addition, adverse noncancer effects such as liver toxicity, developmental and reproductive toxicity, and kidney toxicity have been linked to DBPs.".

(6.5) 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?

(6.6) Page 13, regarding the statement "In the air discharged from municipal waste incinerators, MCA concentrations were 3.2-7.8 µg/m^3 (WHO, 2004d)", please cite the primary reference that reported this result, which is Mowrer and Nordin (1987) according to the WHO (2004d) report.

(6.7) Page 22, Paragraph 2: "…a biomarker to DPB exposure…". Did the authors intend to mention "DBP exposure"?

(6.8) 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.

(6.9) 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).

(6.10) Page 27, Paragraph 3: Please clarify what the phrase “by a 2-2 to 23- times” means?

(6.11) 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).

(6.12) 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.

(6.13) 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. 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.

(6.14) 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.

(6.15) 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. 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. For example, regarding TCA, the authors describe the TCA submodel as part of the PCE PBPK model published by Dr. Weihsueh A. Chiu and his colleagues (Chiu and Ginsberg, 2011; Chiu et al., 2009). Recently, Dr. Chiu’s group published an updated PBPK model for PCE in mice that incorporated the glutathione conjugation pathway. This model better describes the toxicokinetics of PCE, and thus its metabolite TCA. Since it is in mice only, it may not be directly used in human health risk assessment, but this study still deserves mentioning in this report to better reflect the latest science on this subject. Additionally, besides being a submodel for the PCE PBPK model, there are a few PBPK studies that reported TCA or DCA submodel from the PBPK model for TCE (Chiu et al., 2014; Evans et al., 2009; Fisher et al., 1998). The strengths, limitations, and relevance to TCA or DCA risk assessment of these studies should be discussed in this report, too. In addition, Dr. Jeffrey W. Fisher published a review article on selected compartmental and PBPK models for TCE and its metabolites (e.g., TCA) (Fisher, 2000). This study may be worth to be included in this report, too.

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.

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.

(6.16) 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?

(6.17) Page 48: Similar to the comment above, please consider providing a flowchart listing the outcome of the literature search on cancer effects.

(6.18) Page 52, Paragraph 3, Last Sentence: Please change “poisoning cases cases” to “poisoning cases”.

(6.19) 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.

(6.20) 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)”.

(6.21) 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.

(6.22) 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.

(6.23) 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?

(6.24) Page 123, Last Paragraph: Please check and revise this sentence if needed "OEHHA did not identify NOAELs and LOAELs are not identified for single dose studies". I believe these words, "are not identified", can be deleted.

(6.25) Page 194: I have a concern on the use of BMR of 5% in the BMD modeling analysis, which is described in detail in Comment 1.2 above.

(6.26) 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 $BMDL_{05}$ 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.

(6.27) Page 195, Table 10.1: Please refer to my detailed Comment 1.3, and provide a justification of the uncertainty factors and explanation why some of these default uncertainty factors are different from those recommended in US EPA guideline (US EPA, 2002; 2011; 2016a; 2016b).

(6.28) Page 199: As stated in Comment 1.4, 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.

(6.29) Page 204: Similar to MCA, an UF_H of 30 is used for DCA. Please provide a justification and clarification.

(6.30) Page 206: Please provide a clarification of the use of 30 for UF_H for TCA, and why not using 10 as recommended by US EPA.

(6.31) Page 207, Last Paragraph: Please change "reported by by" to "reported by".

(6.32) Page 208: I am also concerned about the UF for MBA. Please refer to my Comment 4 as described above.

(6.33) 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.

(6.34) Page 212: Similar to other chemicals, please provide a clarification on the use of 30 for UF_H for DBA.

(6.35) Page 214: Please provide a justification of why BMR of 5% is used, rather than 10% as recommended by US EPA. Please refer to the detailed description presented above.

(6.36) Page 214, Bottom Paragraph: Please provide a reference to support the use of the equation in the calculation of adjusted animal cancer slope factor.

(6.37) Page 220, Table 10.14: I have been trying to reproduce some of the calculated results. For example, in Table 10.14, if the animal CSF is 0.011 mg/kg/day, by using the equation of $0.011 * (70/0.04)^{0.25}$, I got 0.071 mg/kg/day for human CSF, which is exactly the same as presented in Table 10.14. However, for the other two datasets, if the animal CSF is 0.0018, by using the equation of $0.0018 * (70/0.04)^{0.25}$, I got 0.0116 mg/kg/day for human CSF, which is different from 0.061 presented in Table 10.14. Also, I got a different human CSF value if the animal CSF is 0.0045. 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.

(6.38) 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 $\sqrt{10}$ is used to account for database deficiency for MBA. Is the UF for database deficiency 10 or $\sqrt{10}$ for MBA? Please clarify.

(6.39) 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 ($Intake_{inh}$) 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: $Intake_{inh} = C_{air} \times (BR_a \times ET_{ai} + BR_r \times ET_{ri}) + C_{bath\_air} \times BR_a \times ET_{sb}$. Please correct the equation or clarify the textual description. Also, please check other equations and parameters carefully.

(6.40) Page 276, Last Paragraph: I believe “Table 1” and “Tables 2 and 3” should be changed to “Table A1” and “Tables A2 and A3”.

(6.41) 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?

(6.42) Page 280: Can the authors provide a citation where the equation to calculate the steady-state skin permeability coefficient is from?

(6.43) Appendix B, Page 282: Please change the definition of HAA5 to “sum of MCA, DCA, TCA, MBA, and DBA”

(6.44) 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. Here is what I found. I used the dataset of increased mortality in male rats from NTP (1992), the same specified effect of 0.05, risk type of “Extra Risk”, and confidence level of 0.95, and my results showed that the best model was the logistic model based on the lowest AIC value, and the BMD and BMDL were 10.236 and 7.848 mg/kg/day. These results are consistent with the results presented on Pages 408-410, where the authors also used the logistic model and reported BMD and BMDL as 10.236 and 7.84716 mg/kg/day, respectively. Therefore, I am able to reproduce the authors’ BMD dose analysis results of a representative dataset on noncancer endpoints.

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.

(6.45) Page 448: Please delete the extra “in male B6C3F1 mice” in the title of Figure E1.

(6.46) 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.

I used the same specified effect of 0.05, risk type of “Extra Risk”, and confidence level of 0.95, and my results showed that the best model was the multistage model (Degree 1), and the BMD and BMDL were 36.56062 and 22.96986 mg/kg/day, respectively. These results are consistent with the results presented on Pages 448-450, where the authors also used the multistage model and reported BMD and BMDL as 36.5606 and

22.9711 mg/kg/day, respectively. Therefore, I am able to reproduce the authors' BMD dose analysis results of a representative dataset on cancer endpoints.

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.

(6.47) Page 482, Paragraph 1: When mentioning hutchon.net, please cite the reference by Bland and Altman, The odds ratio. BMJ, 2000;320:1468, as requested on the hutchon.net website.

References:

Bader EL, Hrudey SE, Froese KL. 2004. Urinary excretion half life of trichloroacetic acid as a biomarker of exposure to chlorinated drinking water disinfection by-products. Occup Environ Med. 61(8):715-716.

Chiu WA, Campbell JL, Jr., Clewell HJ, 3rd, Zhou YH, Wright FA, Guyton KZ, Rusyn I. 2014. Physiologically based pharmacokinetic (pbpk) modeling of interstrain variability in trichloroethylene metabolism in the mouse. Environ Health Perspect. 122(5):456-463.

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.

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.

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.

EFSA. 2017. Update: Guidance on the use of the benchmark dose approach in risk assessment. EFSA J. 15:4658.

Evans MV, Chiu WA, Okino MS, Caldwell JC. 2009. Development of an updated pbpk model for trichloroethylene and metabolites in mice, and its application to discern the role of oxidative metabolism in tce-induced hepatomegaly. Toxicol Appl Pharmacol. 236(3):329-340.

Fisher JW. 2000. Physiologically based pharmacokinetic models for trichloroethylene and its oxidative metabolites. Environ Health Perspect. 108 Suppl 2:265-273.

Fisher JW, Mahle D, Abbas R. 1998. A human physiologically based pharmacokinetic model for trichloroethylene and its metabolites, trichloroacetic acid and free trichloroethanol. Toxicol Appl Pharmacol. 152(2):339-359.

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):679-687.
Haber LT, Dourson ML, Allen BC, Hertzberg RC, Parker A, Vincent MJ, Maier A, Boobis AR. 2018. Benchmark dose (bmd) modeling: Current practice, issues, and challenges. Crit Rev Toxicol. 48(5):387-415.
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.
Kulling P, Andersson H, Bostrom K, Johansson LA, Lindstrom B, Nystrom B. 1992. Fatal systemic poisoning after skin exposure to monochloroacetic acid. J Toxicol Clin Toxicol. 30(4):643-652.
Martin JW, Mabury SA, Wong CS, Noventa F, Solomon KR, Alaee M, Muir DC. 2003. Airborne haloacetic acids. Environ Sci Technol. 37(13):2889-2897.
Mowrer J, Nordin J. 1987. Characterization of halogenated organic acids in flue gases from municipal waste incinerators. Chemosphere. 16(6):1181-1192.
NTP. 1992. Toxicology and carcinogenesis studies of monochloroacetic acid (cas no. 79-11-8) in f344/n rats and b6c3f1 mice (gavage studies). National toxicology program, nih publication no. 92-2851.
NTP. 2018. Report on carcinogens monograph on haloacetic acids found as water disinfection byproducts. Peer review draft. March 30, 2018. Office of the report on carcinogens, divison of the national toxicology program, national institute of environmental health sciences, u.S. Department of health and human services.
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.
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.
Tox21. 2020. United states federal government tox21 collaboration advancing toxicology to improve environmental health and pharmaceutical safety. Tox21 fact sheet. Available at: Https://tox21.Gov/wp-content/uploads/2019/04/tox21_factsheet_apr2019.Pdf.
US EPA. 2002. 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-for-cra.Pdf.

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. Risk assessment forum, u.S. Environmental protection agency, wahington dc. Epa/100/r-12/001.
US EPA. 2016a. Drinking water health advisory for perfluorooctane sulfonate (pfos). U.S. Environmental protection agency, washington, dc. Available at: Https://www.Epa.Gov/sites/production/files/2016-05/documents/pfos_health_advisory_final-plain.Pdf.
US EPA. 2016b. Drinking water health advisory for perfluorooctanoic acid (pfoa). U.S. Environmental protection agency, washington, dc. Available at: Https://www.Epa.Gov/sites/production/files/2016-05/documents/pfoa_health_advisory_final-plain.Pdf.
Veeramachaneni DN, Palmer JS, Klinefelter GR. 2007. Chronic exposure to low levels of dibromoacetic acid, a water disinfection by-product, adversely affects reproductive function in male rabbits. J Androl. 28(4):565-577.
WHO. 2004d. Monochloroacetic acid in drinking-water, background document for development of who guidelines for drinking-water quality. Geneva, switzerland, world health organization.
Yu KO, Barton HA, Mahle DA, Frazier JM. 2000. In vivo kinetics of trichloroacetate in male fischer 344 rats. Toxicol Sci. 54(2):302-311.

Review of "Proposed Public Health Goals for Haloacetic Acids in Drinking Water", First Public Review Draft, January 2020. Office of Environmental Health Hazard Assessment, California Environmental Protection Agency.

David H. Phillips BA PhD DSc FRCPath
Professor of Environmental Carcinogenesis

King's College London
Department of Analytical, Environmental and Forensic Sciences
School of Population Health and Environmental Sciences
Faculty of Life Sciences and Medicine
Franklin-Wilkins Building
150 Stamford Street
London SE1 9NH
UK

Preamble

As my research interests and specialist knowledge are in carcinogenesis, with particular emphasis on chemicals, the following remarks are confined for the most part to the carcinogenicity and genotoxicity of haloacetic acids (HAAs). I have not considered in any detail the evidence for the reproductive toxicity or other noncancer effects of the chemicals, as I consider that these lie outside my area of expertise.

Use and occurrence of HAAs

A number of HAAs are manufactured commercially and have industrial uses. They have been variously used for organic synthesis, as fungicides and herbicides, and for medical treatments. More importantly from a public health point of view, they are formed as by-products of water disinfection by chlorination, due to the presence of organic molecules and bromide ions in water.

Human exposure to HAAs

HAAs are present at low concentrations in air, soil and natural waters, but the major source of human exposure is ingestion of drinking water. Inhalation and dermal exposures from water (e.g. from washing and showering) also occur, but these are minor routes of exposure relative to ingestion.

Five HAAs are considered in the current First Public Review from the Office of Environmental Health Hazard Assessment, California Environmental Protection Agency. The evidence on their carcinogenicity and genotoxicity will be considered individually in turn.

1. Monochloroacetic acid (MCA)

Carcinogenicity of MCA

MCA has been tested in two-year animal bioassays involving male and female rats and mice, administered by gavage (up to 21 mg/kg/day for rats and up to 71 mg/kg/day for mice) and, in the case of male rats, in drinking water (up to 60 mg/kg/day). In none of these studies did MCA show evidence of carcinogenicity. MCA has also been administered to mice dermally, by subcutaneous injection and by i.p. injection, none of which resulted in increased tumour formation in the treated animals.

Genotoxicity of MCA

There is some evidence for weak genotoxicity of MCA in short-term assays for mutagenicity or genotoxicity. MCA is mostly inactive in bacterial mutagenicity assays (the Ames test) with just one study reporting positive results in the mM range of test concentrations. Positive results have been reported in the chromosome aberration assay in human lymphocytes and in a mammalian (mouse) cell mutagenicity assay when tested at high concentrations. MCA also caused DNA damage in cells in culture, but not in mouse tissues *in vivo*.

These activities indicate that such genotoxic activity that MCA has is, at most, weak. 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.

2. Dichloroacetic acid (DCA)

Carcinogenicity of DCA

DCA has shown clear evidence of carcinogenicity when tested in male and female mice and in male rats, inducing tumours at high frequency at the highest doses and with some evidence of dose response when tested at multiple doses. The International Agency for Research on Cancer (IARC) concluded that the evidence for DCA carcinogenicity in experimental animals is sufficient and classified it as Group 2B, possibly carcinogenic to humans. In the US, several agencies (NTP, EPA) have classified DCA as a carcinogen.

Hepatic adenomas and carcinomas were induced in both sexes of mice and in male rats, and in mice bronchiolar/alveolar adenomas and squamous cell papillomas were also induced.

Overall, the carcinogenic profile of DCA is of a multi-species, multi-organ carcinogen, active in both sexes. These are characteristic more associated with carcinogens acting through a genotoxic mechanism that with non-genotoxic ones.

Genotoxicity of DCA

There is some evidence of mutagenicity of DCA in bacteria, but results from different studies are not consistent. DCA showed some mutagenic activity, in some cases weak, in mammalian cells (mouse lymphoma assay, chromosomal aberrations, micronucleus induction). In *in vivo* studies, there is evidence of DCA mutagenicity (in Big Blue transgenic mice) and DNA damage (DNA strand breaks in DNA unwinding and comet assays) in livers of mice, but not rats.

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.

3. Trichloroacetic acid (TCA)

Carcinogenicity of TCA

Chronic treatment of mice with TCA has resulted in significant increases in hepatic adenomas and carcinomas in both males and females in several studies. In a single study in male rats, no increase in tumour formation was seen. TCA also increased hepatic tumours when administered as a tumour promoter following initiating doses of N-methyl-N-nitrosourea (MNU).

IARC evaluated the evidence on TCA and considered that there is sufficient evidence of its carcinogenicity in experimental animals, leading to the classification as Group 2B, possibly carcinogenic to humans. Several US agencies, although not all, also list TCA as a carcinogen.

Genotoxicity of TCA

In common with the HAAs already considered, TCA generally shows weak, if any, evidence of mutagenic activity in bacterial test systems. Evidence for mutagenicity in mammalian cells *in vitro* is also weak. In contrast, TCA demonstrates activity in several *in vivo* assays, with various endpoints. These include chromosomal aberrations in mice and chickens, and micronucleus formation in mice. TCA also caused DNA damage (strand breaks detected by the DNA alkaline unwinding assay and alkaline elution assay) in hepatic tissue of mice and rats after oral administration.

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.

4. Monobromoacetic acid (MBA)

Carcinogenicity of MBA

As MBA has not been tested for carcinogenicity in experimental animals, there are no data on which to base a judgement as to whether MBA is a carcinogen or non-carcinogen.

Genotoxicity of MBA

MBA has shown some activity in bacterial mutation (Ames) assays, as well as inducing mutations, chromosomal aberrations and micronucleus formation in mammalian cells. It also caused DNA damage (comet assay) in cultured mammalian cells *in vitro*. Furthermore, there is evidence that MBA causes oxidative stress, which can result in DNA damage and which is considered a characteristic common to many carcinogens.

Investigations of activity *in vivo* are very limited, amounting to only two studies, one a micronucleus test in *P. walti* (newt larvae) and the other a test for long amplicon quantitative PCR in *C. Elegans* (nematodes); both gave negative results.

Overall, MBA has greater mutagenic and genotoxic activity *in vitro* than the analogous chloro compound MCA (a non-carcinogen), and MBA is also more active than DCA and TCA, which are both considered to be carcinogens.

5. Dibromoacetic acid (DBA)

Carcinogenicity of DBA

DBA induced a significant increase in lung neoplasms (alveolar/bronchiolar adenomas and carcinomas) in male mice but not females. Hepatocellular adenomas were significantly increased at several doses in both sexes of mice. Combined incidences of hepatic tumours (adenomas and carcinomas) were significantly increased at all doses in males and in 2/3 doses in females. In male rats DBA induced malignant mesothelioma at the highest dose tested, but at marginal statistical significance (p=0.07). In females the highest dose tested caused a significant increase in mononuclear cell leukaemia; in males it was the lowest dose that increased this malignancy significantly.

Thus, 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.

IARC has concluded that there is sufficient evidence for carcinogenicity of DBA in experimental animals, leading to the classification of Group 2B, possibly carcinogenic to humans.

Genotoxicity of DBA

DBA is generally active in bacterial tests for mutagenicity. It also induced gene mutation (Hprt locus) and DNA damage (comet assay) in mammalian cells. It also induced micronucleus formation in male and female mice (peripheral blood erythrocytes) and induced oxidative damage to DNA (8-oxoguanine formation) in the livers of mice following oral administration.

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.

Risk assessment of HAAs

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.

For MCA, a non-carcinogen, risk assessment based on the systemic toxicity of the compound results in the much higher PHG value of 53 ppb. Similarly for MBA, a value of 25 ppb is arrived at related to the effect of muscular degeneration.

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.

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.

David H. Phillips

30 September 2020

External Scientific Peer Review of OEHHA's Draft PHGs for Haloacetic Acids in Drinking Water

Prepared by:
Paul A. White, PhD, Department of Biology, University of Ottawa, Ottawa, Ontario, Canada
Submitted: October 1, 2020.

The document submitted for external peer review contains the OEHHA's "comprehensive review of the pertinent and relevant scientific literature on the regulated HAAs" (haloacetic acids), with the goal of establishing Public Health Guidelines (PHGs). More specifically, establishing "health-protective drinking water concentrations" derived from quantitative analysis of available toxicological dose-response data for 5 targeted HAAs, i.e., health-protective drinking water concentrations for MCA, DCA, TCA, MBA and DBA calculated using suitably-adjusted non-cancer ADD (Acceptable Daily Dose) and human CSF (Cancer Slope Factor) values. With respect to the ADDs and CSFs, values were determined via rigorous collection and analysis of dose-response data from the peer-reviewed scientific literature.

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 CSF_{human} values can be complex and challenging.

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. 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) [1] .

In some cases, the comments and recommendations outlined herein stem from a difference of opinion regarding the approach and methodology employed to analyse the collected dose-response data; moreover, the approach underlying the use of BMD values for determination of ADD and CSF values. Thus, some of the comments and recommendations might be viewed as suggestions for consideration.

Apologies if the comments below seems repetitive; moreover, that they do not progressively follow the sequential document sections.

General Comments Pertaining to the Entire Document, and Specific Comments Pertaining to the

Evaluations of DCA, TCA and DBA –

A few spelling errors in the document: e.g., *International* on p. 4, *proposition* on p. 5.

"-1" on p. 6 should be superscripted.

Don't think text ever refers to Table 10.8.

I appreciated the table of physical-chemical properties on p. 10. Very useful for the reader. For example, some published papers that commented on the genotoxicity of HAAs suggest the necessity to conduct *in vitro* assays in a sealed container (e.g., see [1]). Although this would clearly be necessary for THMs; clearly not required for the HAAs examined by the authors since the vapour pressure values are so low.

Minor comment – the authors might consider using sub-headings, e.g., "Production and Use" on p. 11 would be section 2.1. Would make it easier for readers to follow the document.

Drinking water concentration information on pp. 15-22 is very interesting and appreciated. Similar comment for the judicious review of human epidemiologic studies, which includes detailed considerations of study strengths and weaknesses. Similarly appreciated the rigorous assignment of quality scores to the reviewed cancer epidemiology studies. Instills confidence regarding the human epi concluding statements on p. 51.

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?

Information on metabolism (i.e., pp. 29-34) is very important and useful for the reader. Helpful for interpretation of the toxicological data presented later. For example, interpretation of TCA and DCA data in light of the fact that TCA can generate DCA, and DCA can generate glyoxylic acid. Additionally, that fecal excretion is insignificant, and hepatic first pass is a main contributor to elimination of an orally-administered dose.

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. 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. 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).

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]. As explicitly noted by Haber et al (2018), "the default settings for BMDS Wizard are that a BMD:BMDL ratio of >20 results in a model being placed in the 'questionable' bin, and a ratio of >5, while not altering the bin in which a model is placed, does result in a 'caution' flag" [4]. More recent works even go so far as to state that a BMDU:BMDL ratio >100 indicates that the BMD is unsuitable for regulatory decision-making [5, 6]. 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.

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. 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].

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). 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 "BMD_{05} and $BMDL_{05}$ 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. The assertion that the "$BMDL_{05}$ from this dataset is not useful as a PoD" is not supported by the BMD:BMDL ratio, and is not consistent with the most recent BMD technical guidance. The point being that the authors need to be much more cautious with respect to the rationale underlying the denotation that a PoD is "not useful". 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. Granted, with respect to the Cicmanec et al (1991) study, the fact that the non-control response levels are similar is cause for concern; the concern would be consistent with statements in the USEPA's 2012 Technical Guidance.

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.

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].

I have similar concerns regarding the authors' statements about the $BMDL_{1SD}$ 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. What is more concerning about this particular dataset is the similarity of the responses at all non-control doses. 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. Similarly, I am not entirely convinced that the need for an additional UF is a sound basis for PoD ranking and selection. 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).

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? Here and throughout, please choose words carefully, and provide a convincing argument based on judicious study evaluations. Similarly, 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. 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. 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?

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". 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.

With respect to the interpretation of the MBA 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).

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.

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.

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. 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).

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]. 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 assay**s** 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). 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

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. 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.

With respect to the genotoxicity of DCA in particular, there seems to be some disagreements about whether the observed responses *in vitro* can be used as the basis to claim a genotoxic MOA underlying carcinogenicity. Importantly, 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. Moore and Harrington-Brock (2000) note that "weight of the evidence argues that chemically-induced mutation is unlikely to be a key event in the induction of human tumors that might be caused by TCE itself (as the parent compound) and its metabolites, CH, DCA, and TCA. This conclusion derives primarily from the fact that these chemicals require very high doses to be genotoxic" [15]. This type of inconsistency with the OEHHA report calls into question the PHG determined for carcinogenic effects, i.e., may not be a mutagenic MOA. Interestingly, Harrington-Brock (1998) noted that "it seems reasonable to postulate that mutational events are involved in the etiology of the observed mouse liver tumors induced by DCA at drinking water doses of 0.5 to 3.5 g/l". However, they also noted that "it seems unlikely that it would be mutagenic (or possibly carcinogenic) at the levels seen in finished drinking water" [16]. Doubts about the genotoxic MOA of DCA were reiterated by Richardson et al (2007) who noted that "the weak genotoxicity of DCA, which is exhibited only at high concentrations, is not considered to play a primary role in its carcinogenicity." 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].

Granted, with respect to DCA genotoxicity, there seems to be fairly convincing *in vivo* evidence, e.g., positive for mouse peripheral blood MN assay. That being said, the NTP (2007b) study did not yield a significant positive response. What is critically important here is that the results of any/all *in vivo* MN studies should be interpreted in the context of OECD Test Guideline 474. The TG is very specific about the post-exposure timing for collection of bone marrow or peripheral blood. More specifically, "Samples of bone marrow are taken at least twice …. starting not earlier than 24 hours after treatment, but not extending beyond 48 hours after treatment with appropriate interval(s) between samples,…..Samples of peripheral blood are taken at least twice ….. starting not earlier than 36 hours after treatment, with appropriate intervals following the first sample, but not extending beyond 72 hours" [17]. All that being said, I suspect that the Fuscoe et al. and NTP studies were conducted in accordance with OECD TG 474. 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.

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.

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.

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.

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. 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.

As noted earlier, all genotoxicity results must be interpreted in a hierarchical fashion. 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". 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). More specifically, with respect to the Bhunya and Das (1987) study, sample collection was too early. OECD TG 474 indicates "If 2 daily treatments are used (e.g. two treatments at 24 hour intervals), samples should be collected once between 18 and 24 hours following the final treatment for the bone marrow or once between 36 and 48 hours following the final treatment for peripheral blood". Indeed, it is not clear whether the results from several of the listed studies are acceptable, i.e., Bhunya and Das (1987), MacKay et al (1995), and Kurinnyi et al (1984). More specifically, for the MacKay et al study is that 2x daily doses with 6- and 24-hr post-exposure sampling time? If 2x daily doses, then the 24-hr sampling time would be acceptable. For the Kurinnyi et al study, is it a single acute dose followed by sample collection at 20-hrs post-exposure? If yes, it is not an acceptable study since sampling is too early. 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.

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. 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.

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. 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.

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.

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. 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].

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.

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.

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. 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.

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? 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.

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".

Several parts of Section 10 (i.e., dose-response assessment) are confusing. For example, 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". The summary presented on p. 21 of EPA (2012) notes, "An extra risk of 10% is recommended as a standard reporting level for quantal data….. The 10% response level has customarily been used for comparisons because it is at or near the limit of sensitivity in most cancer bioassays ….". Moreover, Hardy et al (2017) note, "…..the $BMDL_{10}$ 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?

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 $S = kM^{0.67}$….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. 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.

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. With respect to tables such as Table 10.5 more specifically, the authors need to outline criteria underlying statements such as "poor model fit".

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].

On p. 214, the author note the assumption that "lifetime incidence of cancer increases with the 3rd power of age". A citation should be added. 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].

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.

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? Why is there no p value for the 3rd row?

On p.225 please provide a citation for the age adjustment. Guessing its EPA (2005) [31].

With respect to Sections 11 and 12, I have already outlined my reservations regarding study and PoD selection criteria; moreover, the interpretation of mixed genotoxicity data for DCA and TCA. 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? 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. In section 12, I suggest summarizing the PHG values in a table.

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. Actually, shouldn't this be Table 12.1?

Citations

1. Richardson, S.D., Plewa, M.J., Wagner, E.D., Schoeny, R. and Demarini, D.M. (2007) Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: a review and roadmap for research. *Mutat. Res.,* **636**, 178-242

2. USEPA (United States Environmental Protection Agency). (2012) *Benchmark Dose Technical Guidance* **EPA/100/R-12/001**. Washington, DC: United States Environmental Protection Agency.

3. Hardy, A., Benford, D., Halldorsson, T., Jeger, M., Knutsen, K., More, S., Mortensen, A., Naegeli, H., Noteburn, H., Ockleford, C., Ricci, A., Rychen, G., Silano, V., Solecki, R., Turck, D., Aerts, M., Bodin, L., Davis, A., Edler, L., Gundert-Remy, U., Sand, S., Slob, W., Bottex, B., Abrahantes, J., Marques, D., Kass, G. and Schlatter, J. (2017) Update: Use of the benchmark dose approach in risk assessment. *EFSA Journal,* **15**, 4658

4. Haber, L.T., Dourson, M.L., Allen, B.C., Hertzberg, R.C., Parker, A., Vincent, M.J., Maier, A. and Boobis, A.R. (2018) Benchmark dose (BMD) modeling: current practice, issues, and challenges. *Crit. Rev. Toxicol.,* **48**, 387-415

5. White, P.A., Long, A.S. and Johnson, G.E. (2020) Quantitative interpretation of genetic toxicity dose-response data for risk Assessment and regulatory decision-making: current status and emerging priorities. *Environ. Mol. Mutagen.,* **61**, 66-83

6. White, P.A., Zeller, A., Pfuhler, S. and Johnson, G.E. (2019) Re: Gi et al. 2018, In vivo positive mutagenicity of 1,4-dioxane and quantitative analysis of its mutagenicity and carcinogenicity in rats, Archives of Toxicology 92:3207-3221. *Arch. Toxicol.,* **93**, 211-212

7. Slob, W., Moerbeek, M., Rauniomaa, E. and Piersma, A.H. (2005) A statistical evaluation of toxicity study designs for the estimation of the benchmark dose in continuous endpoints. *Toxicol. Sci.,* **84**, 167-185

8. European Food Safety Authority (EFSA). (2009) Use of benchmark dose approach in risk assessment: guidance of the scientific committee. *EFSA Journal,* **1150**, 1-72

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

10. OECD. (2014) *OECD Guideline for the Testing of Chemicals. In Vitro Mammalian Cell Gene Mutation Test 476*. Paris France: Organization for Economic Cooperation and Development.

11. OECD. (2015) *OECD Guideline for the Testing of Chemicals. In Vitro Mammalian Cell Gene Mutation Tests Using the Thymidine Kinase Gene 490*. Paris France: Organization for Economic Cooperation and Development.

12. Williams, R.V., DeMarini, D.M., Stankowski, L.F.,Jr, Escobar, P.A., Zeiger, E., Howe, J., Elespuru, R. and Cross, K.P. (2019) Are all bacterial strains required by OECD mutagenicity test guideline TG471 needed?. *Mutat. Res.,* **848**, 503081

13. Maron, D.M. and Ames, B.N. (1983) Revised methods for the Salmonella mutagenicity test. *Mutat. Res.,* **113**, 173-215

14. Moore, M.M., Honma, M., Clements, J., Bolcsfoldi, G., Burlinson, B., Cifone, M., Clarke, J., Delongchamp, R., Durward, R., Fellows, M., Gollapudi, B., Hou, S., Jenkinson, P., Lloyd, M., Majeska, J., Myhr, B., O'Donovan, M., Omori, T., Riach, C., San, R., Stankowski, L.F., J., Thakur, A.K., Van Goethem, F., Wakuri, S. and Yoshimura, I. (2006) Mouse lymphoma thymidine kinase gene mutation

assay: follow-up meeting of the International Workshop on Genotoxicity Testing--Aberdeen, Scotland, 2003--Assay acceptance criteria, positive controls, and data evaluation. *Environ. Mol. Mutagen.,* **47**, 1-5

15. Moore, M.M. and Harrington-Brock, K. (2000) Mutagenicity of trichloroethylene and its metabolites: implications for the risk assessment of trichloroethylene. *Environ. Health Perspect.,* **108 Suppl 2**, 215-223

16. Harrington-Brock, K., Doerr, C.L. and Moore, M.M. (1998) Mutagenicity of three disinfection by-products: di- and trichloroacetic acid and chloral hydrate in L5178Y/TK +/- (-)3.7.2C mouse lymphoma cells. *Mutat. Res.,* **413**, 265-276

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

18. 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.

19. Guyton, K.Z., Rusyn, I., Chiu, W.A., Corpet, D.E., van den Berg, M., Ross, M.K., Christiani, D.C., Beland, F.A. and Smith, M.T. (2018) Application of the key characteristics of carcinogens in cancer hazard identification. *Carcinogenesis,* **39**, 614-622

20. Guyton, K.Z., Rieswijk, L., Wang, A., Chiu, W.A. and Smith, M.T. (2018) Key characteristics approach to carcinogenic hazard identification. *Chem. Res. Toxicol.,* **31**, 1290-1292

21. Smith, M.T., Guyton, K.Z., Gibbons, C.F., Fritz, J.M., Portier, C.J., Rusyn, I., DeMarini, D.M., Caldwell, J.C., Kavlock, R.J., Lambert, P.F., Hecht, S.S., Bucher, J.R., Stewart, B.W., Baan, R.A., Cogliano, V.J. and Straif, K. (2016) Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. *Environ. Health Perspect.,* **124**, 713-721

22. Smith, M.T., Guyton, K.Z., Kleinstreuer, N., Borrel, A., Cardenas, A., Chiu, W.A., Felsher, D.W., Gibbons, C.F., Goodson, W.H., Houck, K.A., Kane, A., La Merrill, M.A., Lebrec, H., Lowe, L., McHale, C.M., Minocherhomji, S., Rieswijk, L., Sandy, M.S., Sone, H., Wang, A., Zhang, L., Zeise, L. and Fielden, M. (2020) The key characteristics of carcinogens: relationship to the hallmarks of cancer, relevant biomarkers, and assays to measure them. *Cancer Epidemiol. Biomarkers Prev.,* **Mar 9**, cebp.1346.2019

23. Stratton, M.R., Campbell, P.J. and Futreal, P.A. (2009) The cancer genome. *Nature,* **458**, 719-724

24. Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., Ng, P.K., Jeong, K.J., Cao, S., Wang, Z., Gao, J., Gao, Q., Wang, F., Liu, E.M., Mularoni, L., Rubio-Perez, C., Nagarajan, N., Cortes-Ciriano, I., Zhou, D.C., Liang, W.W., Hess, J.M., Yellapantula, V.D., Tamborero, D., Gonzalez-Perez, A., Suphavilai, C., Ko, J.Y., Khurana, E., Park, P.J., Van Allen, E.M., Liang, H., MC3 Working Group, Cancer Genome Atlas Research Network, Lawrence, M.S., Godzik, A., Lopez-Bigas, N., Stuart, J., Wheeler, D., Getz, G., Chen, K., Lazar, A.J., Mills, G.B., Karchin, R. and Ding, L. (2018) Comprehensive characterization of cancer driver genes and mutations. *Cell,* **173**, 371-385.e18

25. USFDA (United States Food and Drug Administration). (2005) *Guidance for Industry. Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers*. Washington, DC: United States Food and Drug Administration.

26. WHO (World Health Organization). (2001) *Guidance Document for the Use of Data in the Development of Chemical-specific Adjustment Factors (CSAFs) for Interspecies Differences and Human*

Variability in Dose/Concentration-response Assessment **WHO/PCS/01.4**. Geneva, Switzerland: World Health Organization.

27. 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.

28. 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.

29. 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

30. USEPA (United Sates Environmental Protection Agency). (1992) A cross-species scaling factor for carcinogen risk assessment based on equivalence of Mg/Kg0.75/Day*. Federal Register,* **57**, 24152-24173

31. USEPA (United Sates Environmental Protection Agency). (2005) *Supplemental Guidance for Assessing Cancer Susceptibility From Early-life Exposure to Carcinogens* **EPA/630/R-03/003F**. Washington, DC: United Sates Environmental Protection Agency