TASK ORDER 68HERH21F0090 UNDER
CONTRACT EP-C-17-017

EXTERNAL PEER REVIEW OF EPA'S DRAFT AQUATIC LIFE AMBIENT WATER QUALITY CRITERION FOR PERFLUOROOCTANE SULFONATE (PFOS)

FINAL PEER REVIEW REPORT

August 26, 2021

Submitted to:
U.S. Environmental Protection Agency
Office of Water, Office of Science and Technology
1200 Pennsylvania Avenue, NW
Washington, DC 20460
Attn: James Justice
Justice.JamesR@epa.gov

Submitted by:
Eastern Research Group, Inc.
110 Hartwell Avenue
Lexington, MA 02421

ERG
www.erg.com

CONTENTS

1.0 INTRODUCTION

The U.S. Environmental Protective Agency (EPA) Office of Water (OW) is charged with protecting ecological integrity and human health from adverse anthropogenic, water-mediated effects, under the purview of the Clean Water Act (CWA). In support of this mission, EPA has developed draft water quality criteria to protect aquatic life and aquatic-dependent wildlife from the presence of Perfluorooctanoic Acid (PFOA) and Perfluorooctane Sulfonate (PFOS) in freshwater and saltwater environments. Derivation of these criteria is described in two draft documents: *Aquatic Life Ambient Water Quality Criteria for Perfluorooctanoic Acid (PFOA)* and *Aquatic Life Ambient Water Quality Criteria for Perfluorooctane Sulfonate (PFOS)*.

This report documents the results of an independent letter peer review of the EPA's draft *Aquatic Life Ambient Water Quality Criteria for Perfluorooctane Sulfonate (PFOS)*. Eastern Research Group, Inc. (ERG), a contractor to EPA, organized this external peer review for EPA OW and developed this report. Independent peer review of the draft *Aquatic Life Ambient Water Quality Criteria for Perfluorooctanoic Acid (PFOA)* document is covered in a separate report.

Section 2.0 of this report presents the individual reviewer comments organized by charge question. Appendix A provides EPA's charge to reviewers and Appendix B presents the reviewer comments organized by reviewer.

1.1 Development of the Draft Documents

Toxicity studies used to derive the PFOA and PFOS criteria were carefully evaluated and thoroughly reviewed to ensure studies were of sufficient data quality to use in criteria derivation. Scientists from EPA OW and Office of Research and Development (ORD) conducted an extensive review of the PFOA and PFOS toxicity studies. Additionally, EPA obtained replicate-level (or treatment-level, when replicates were unavailable) concentration-response (C-R) data from publications, supplemental materials, or via contacting authors so that EPA could independently fit C-R models to estimate acute LC_{50} and chronic EC_{10} values that were used to derive the criteria to ensure endpoints used were statistically sound. Individual C-R models and resultant point estimates were also reviewed and discussed between OW and ORD to ensure the most statistically robust models informed the derivation of the PFOA and PFOS criteria. In addition to contacting study authors for C-R data (when not reported in the open literature), EPA also consulted primary authors for methods clarifications in many instances during the data quality review phase to ensure that the studies used to derive criteria were of high quality.

Overall, due to the paucity of measured freshwater toxicity data, EPA included a number of tests with unmeasured treatments to derive criteria to ensure the dataset was representative of a range of taxa and there were sufficient data to develop criteria. EPA also conducted meta-analyses to evaluate the relationship between nominal and measured test concentrations using tests with measured treatment concentrations. These meta-analyses (described in detail as Appendix L of the PFOA criteria document and Appendix O of the PFOS criteria document) suggested measured concentrations were similar to nominal concentrations and that the use of unmeasured tests, in light of data limitations, was appropriate. Additionally, estuarine/marine toxicity data limitations did not allow for the direct derivation of acute or chronic estuarine/marine criteria for PFOA or PFOS. Therefore, to develop recommendations that states and tribes could use in adopting protective values for estuarine/marine waters, EPA developed acute PFOA and PFOS protective benchmarks using a New Approach Methodology (detailed in Appendix K of the PFOA criteria document and Appendix L of the PFOS criteria document).

Addressing data limitations to derive robust criteria/benchmarks, extensively reviewing studies, and calculating point estimates meant that the derivation of the PFOA and PFOS aquatic life criteria were developed via comprehensive, rigorous process that included collaborations across EPA scientists in OW and

ORD. Beyond detailed discussions between OW and ORD, the PFOA and PFOS drafts also underwent two rounds of review with the EPA Scoping Workgroup (consisting of additional scientists from both OW and ORD) and one round of review with a group of internal EPA reviewers that included representatives from the OW, ORD, other EPA Program Offices, and EPA Regions.

Subsequently, EPA contracted with ERG to organize an independent external peer review of both draft documents. Results of the PFOS review are described in this report. Results of the PFOA review are documented in a separate report.

1.2 Peer Reviewers

ERG identified, screened, and selected the following five experts who met technical selection criteria provided by EPA and had no conflict of interest in performing this review:

- **Jason Conder, Ph.D**.; Principal, Geosyntec Consultants
- **Anu Kumar, Ph.D.**; Principal Research Scientist, Environment Protection and Technologies, Commonwealth Scientific and Industrial Research Organization (CSIRO)
- **Ryan Prosser, Ph.D.**; Associate Professor, University of Guelph
- **Christopher J. Salice, Ph.D.**; Director, Environmental Science and Studies Program, Towson University
- **Jamie G. Suski, Ph.D.**; Senior Scientist, EA Engineering, Science, and Technology, Inc.

ERG provided reviewers with instructions, the draft *Aquatic Life Ambient Water Quality Criteria for Perfluorooctane Sulfonate (PFOS)*, and the charge to reviewers (Appendix A of this report) prepared by EPA. Reviewers worked individually to develop written comments in response to the charge questions. After receiving reviewer comments, ERG compiled responses by charge question (see Section 2.0) and included the responses organized by reviewer (Appendix B of this report).

2.0 SUMMARY OF REVIEWER COMMENTS ORGANIZED BY CHARGE QUESTION

This section organizes reviewer comments by charge question (see Appendix B for reviewer comments organized by reviewer).

2.1 Please comment on the overall clarity of the document as it relates to the derivation of each criterion.

2.1. Clarity of Document as it Relates to the Derivation of Each Criterion	
Reviewer	**Comments**
Reviewer 1	Overall, the document is clear and the reader can follow the logic of criteria derivation, and track the values used back to the cited research articles or values calculated by EPA.
Reviewer 2	I thought that the document was well written and laid out. I thought that the document clearly laid out the approach that the EPA used to derive each criterion. I thought it clearly outlined the approach that the EPA chose in deciding which data to use in their derivation and how these data would be used in derivation.

2.1. Clarity of Document as it Relates to the Derivation of Each Criterion	
Reviewer	**Comments**
	The appendices are very useful in providing added detail and the data that were used in the derivation of the criteria. The appendices allow for a high level of transparency around how the criteria were generated. In Table 3-1, the acronym "GMAV" was used as a heading in the table, but I could not locate where this acronym was defined earlier in the document. The captions of figures and tables are not sufficiently detailed. Figures and tables should be able to stand on their own. Also, the use of acronyms in the caption and headings of tables and figures decreases clarity, e.g., Figs 3-1, Tables 3-1, 3-2, 3-3, 3-4, 3-5. The use of acronyms in the figure or table is valid to save space, as long as they are defined in the caption of the figure or table.
Reviewer 3	Main Question: Perhaps this was missed in the draft document, but, is there guidance when one criteria is exceeded and the other is not? For example, if the tissue based criteria are exceeded yet, the CCC is not; perhaps this is an unlikely scenario as those receptors have been accumulating PFOS for a duration likely under higher water concentrations (> 0.014mg/L). Although if sediment concentrations remain elevated (but not water column concentrations) this may also be a likely route of PFOS exposure to fish with sediment dwelling prey. There are additional domestic criteria missing from the previously published criteria section; please review those for Texas, Florida and California. Are there two Sharpe et al.'s, I believe this is only one publication but flipping between Sharpe et al. 2010, Sharpe et al. 2010a and Sharpe et al. 2010b throughout the document. If the goal is to distinguish between supplemental vs the manuscript proper I suggest just clarifying in the text instead of the reader looking for two pubs by Sharpe et al. 2010. Page 238 – error: The study authors reported a 96-hour LC_{50} of 58.47 mg/L PFOS, based on the results of the range finding test. The independently-calculated toxicity value was x.xx mg/L. Page 296 – error: The independently-calculated toxicity value was x.xx mg/L. Table 3-6 is not referenced/described in the text. Additionally, the title reads "Six" most sensitive and lists "Seven". Overall comment: ranking of sensitive genera flips back and forth between most and least sensitive among tables, consistency would help the reader.
Reviewer 4	EPA has drafted the PFOS aquatic life criteria to be consistent with methods described in EPA's "*Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*" (U.S. EPA 1985). I congratulate the EPA Team for a very thorough, comprehensive analysis of the toxicological data to derive each criterion. • The report is technically sound and is very clearly written.

2.1. Clarity of Document as it Relates to the Derivation of Each Criterion	
Reviewer	**Comments**
	• The criteria have been derived using strong science-based evidence. • Sub-sections on overview of PFAS, PFAS nomenclature, problem formulation, exposure pathways, transformation and degradation of PFOS precursors in the aquatic environment • sources, concentration reported in environment and existing criteria (both national and international) help to set the scene before toxicological data is presented and assessed for developing various criterion. • The freshwater acute water column-based criterion, the chronic water column-based chronic criterion, the chronic fish whole-body tissue criterion, the chronic fish muscle tissue criterion and the chronic invertebrate whole-body tissue criterion have been developed and documented in this report are based on comprehensive assessment of the toxicological data and consistent with the *Guidelines*. • Acute and chronic MDRs for PFOS estuarine/marine criteria derivation were not met due to fewer empirical PFOS toxicity data. To address this gap, the EPA Team developed an acute aquatic life benchmark for estuarine/marine environments based on Interspecies Correlation Estimation (ICE) model. Such predictive models should be used when there is limited toxicity data. • EPA Team has provided extensive background information on toxicity data assessment and collated this information in various Appendices as additional line of evidence. • Tables and Figures are very well laid out throughout the document and provide additional information of the toxicity data used in developing Water Quality Criteria for PFOS.
Reviewer 5	Overall, the document is clearly written and generally free of grammatical errors. I applaud the scientists and EPA for compiling an impressive amount of work and communicating it in a reasonably clean and coherent way. That said, there are a few instances of redundancy – literally, the same sentences repeated. I have made note of these in the actual report and will include that along with this document, if requested. Although these are easy enough to see with careful review. The document is VERY LONG and very detailed so any efforts to shorten or make more concise would be welcomed. As for the clarity of technical elements of the document, I feel that many of the decisions to use or not use data or endpoints could be more consistent and/or communicated better. For example, in some cases the geometric mean of endpoints for a certain taxa is used for the chronic value (pimephales) but for other taxa, this is not the case (e.g. daphnia; zebrafish). In other cases, the decision to not use certain data seems as if it could be communicated more clearly. So while the language of the document is pretty clear, the actual technical aspects are less so.

2.1. Clarity of Document as it Relates to the Derivation of Each Criterion	
Reviewer	**Comments**
	One major concern I have is with the overall approach of using the 4 most sensitive toxicity values to then derive the final acute and final chronic values. Using this approach it would seem the AWQC are then very sensitive to changes in any 1 of the 4 toxicity values. For example, when EPA explored the impact of different toxicity values on the chronic freshwater water column criteria, using higher toxicity values (e.g., pimephales in place of fatmucket Table 4-3) resulted in a lower chronic criteria. This is nonintuitive and suggests a possible flaw in the approach. It's possible this is not the case and it makes sense both mathematically (steeper slope) and perhaps even from a protection standpoint. Either way, EPA should explain why this happens and what it means for the overall approach. I suspect the EPA is somewhat constrained by the 1985 guidelines in developing the AWQC but I also see that New Approach Methods (WEB-ICE) were used to derive criteria with limited data. I wonder if using a species sensitivity distribution approach in which all the chronic or acute (freshwater)data are used would result in more defensible criteria that are less impacted by changes in any one toxicity value? At the very least, I think including a full SSD would be useful for comparison as part of the characterization piece. In the PFOA document I mention revisiting and publishing and updated 1985 guidelines…this is warranted when EPA has the bandwidth to do so. Having said all this, I am aware that EPA likely has justification for their approach of using the 4 most sensitive tox values but it would perhaps be good to mention this again as I suspect a lot of people that may read the AWQC will not also read the 1985 guidelines. Lastly, I had a hard time keeping track of all the decisions to use or not use data for each of the tox values that supported the criteria. I think a more detailed table with all the tox values considered (data shown in Figure 3-3) and including whether the data used were author-reported, re-calculated by EPA, along with the lowest reported/calculated value that wasn't used and why. This may be asking a lot and this information is throughout the document but not in a single, easy to locate and read location.

2.2 Please comment on the approach used to derive the draft criterion for PFOS. Please provide detailed comments.

- **Is the technical approach used to derive the criterion elements logical?**
- **Does the science support the conclusions?**
- **Is it consistent with the protection of freshwater aquatic life from acute, chronic, and bioaccumulative effects?**

2.2. The Technical Approach Used to Derive the Draft Criterion for PFOS	
Reviewer	**Comments**
Reviewer 1	• **Is the technical approach used to derive the criterion elements logical?**

2.2. The Technical Approach Used to Derive the Draft Criterion for PFOS	
Reviewer	**Comments**
	Yes, the technical approach used to derive the criteria elements is generally logical. I disagree with some of the elements of the analyses, as noted in my detailed comments (see below, responses to charge question 8) • **Does the science support the conclusions?** In general, the science is supportive of the general conclusions. As noted in my below detailed responses to other charge questions, I believe the science is not supportive of the work in a few key instances including: I believe the Criterion Continuous Concentration (CCC) should be potentially re-calculated considering my comments provided in response to charge question 5a. The science does not support the assumption of a 10-year recovery time for PFOS in aquatic systems. The generation of tissue criteria is weakly supported, and the uncertainty associated with these criteria should be emphasized. The NAM-generated marine Final Acute Value (FAV) and FAV/2 values (Appendix L) are highly uncertain. It is unclear if the EPA-calculated Effective Concentration 10% (EC10) values are supported; additional details on the modeling and the variability and fit of each EC10 model need to be provided. • **Is it consistent with the protection of freshwater aquatic life from acute, chronic, and bioaccumulative effects?** The criteria derived are aimed at protecting aquatic life (e.g., fish, invertebrates) from the direct acute and chronic toxicity of PFOS in water. Generally, the values applied are protective and are generally similar to protective values derived by other regulatory organizations and independent (i.e., academic, private sector) scientists. Although, as based on my comments, I believe there is room for improvement. The criteria derived for tissues attempt to provide criteria that take into account bioaccumulation so that measurements in tissue can be interpreted with respect to the potential for potential effects; however, the uncertainty with the tissue criteria is high. The water and tissue criteria are not intended protective of bioaccumulative effects that may affect higher trophic levels, such as wildlife that may consume aquatic life.
Reviewer 2	Yes, the technical approach used by the EPA to derive the criterion is logical and defensible. The approach is also clearly laid out in the document. Dividing the 5th centile of the acute GSD by 2 is sufficiently conservative to ensure the protection of 95% of species, based on the data currently available.

2.2. The Technical Approach Used to Derive the Draft Criterion for PFOS	
Reviewer	**Comments**
	Yes, I think the science supports the EPA's conclusions. However, there appears to be several studies that were not considered by the EPA. I have listed these studies below. Yes, I think the approach taken by the EPA is sufficiently conservative to be protective of freshwater aquatic life from acute, chronic, and bioaccumulative effects based on the data that was available at the time. It was a good idea to evaluate the influence on non-North American species on the derivation of the criteria.
Reviewer 3	• **Is the technical approach used to derive the criterion elements logical?** This is logical and follows the established GLRI guidance; however, both Canada and Australia utilize a species sensitivity distributions to determine the 95th and 99th percentile of species protection. Is there a defensible reason why EPA did not employ this approach or at the very least present these distributions and analysis that would support the currently drafted criteria? Additionally, thresholds from those SSDs (and others published) are lower than the draft guidance here, this should be addressed: Australia – 0.13 µg/L Canada – 6.8 µg/L Salice et al. 2018 – 1.12 µg/L Conder et al. 2020 – 5.85 µg/L *Giesy et al. 2010 – 5.1 µg/L (using CCC based on GLRI guidance) • **Does the science support the conclusions?** The GMCV for Zebrafish is 0.0165 mg/L, thus there are studies that result in chronic toxicity at concentrations lower than this mean; however, this is very close to the CCC of 0.014mg/L. This seems borderline protective when considering potential exposures to this species (and those more sensitive). • **Is it consistent with the protection of freshwater aquatic life from acute, chronic, and bioaccumulative effects?** I have confusion over Tables 3-9 and 4-6 calculations. How is it that the inclusion of *Lampsilis* with a higher GMAV results in a lower overall CMC (3.3 mg PFOS/L) compared to the CMC in table 3-9 (3.6 mg PFOS/L)? Actually, looking more closely at this, the ln(GMAV)^2 are inconsistent among the tables for *Xenopus*, this is likely are result of using table 3-6 as a template for 4-6. It is great to see the inclusion of the Burkhard et al. 2021 as this synthesis has been peer-reviewed and published and is an exceptional overview of PFOS bioaccumulation; unfortunately, there are not more current literature used within the draft document.
Reviewer 4	This EPA report provides a critical review of toxicity data identified in EPA's literature search for PFOS, including the anionic form (CAS No. 45298-90-6), the acid form (CAS No. 1763-23-1),

2.2. The Technical Approach Used to Derive the Draft Criterion for PFOS	
Reviewer	**Comments**
	potassium salt (CAS No. 2795-39-3), an ammonium salt (CAS No. 56773-42-3), sodium salt (CAS No. 4021-47-0), and a lithium salt (CAS No. 29457-72-5). It quantifies the toxicity of PFOS to aquatic life, and provides criteria intended to protect aquatic life from the acute and chronic toxic effects of PFOS. The detailed assessment is as follows: • These criteria have been derived using robust methods and the best available toxicity data on aquatic life. • The approach used to derive the draft criterion for PFOS is very logical and consistent with the protection offered by acute and chronic aquatic life criteria derived using empirical data, as prescribed in the 1985 *Guidelines.* • Exclusion and inclusion criteria are appropriately discussed in the context of the toxicological data reported in the literature and provide additional evidence on the selection of toxicity data criteria development. • With limited toxicity datasets to North American resident species, non-North American resident species were included for criteria development. For example, inclusion of non-resident species such as Planaria, *Dugesia japonica* and Japanese swamp shrimp, *Neocaridina denticulata* for calculating acute water quality criteria and zebra fish, *Danio rerio* for chronic criteria. The EPA team did not find any influence of excluding non-North American resident species in criteria derivations and decided to retain the full acute and chronic toxicity dataset. This was very rational decision and non-Northern American species served as surrogate species for the broad range of the thousands of untested species present in the freshwater environment in the U.S. • The acute measures of effect on aquatic organisms selected included the lethal concentration (LC50), effect concentration (EC50), or inhibitory concentration (IC50) estimated to produce a specific effect in 50 percent of the test organisms as per the *Guidelines*. • The endpoint for chronic exposures incorporated the effect concentration estimated to produce a chronic effect on survival, growth, or reproduction in 10 percent of the test organisms (EC_{10}). This approach has been also consistent with the harmonized guidelines from OECD and the generally preferred effect level for countries such as Canada, Australia, and New Zealand. • Reported (No Observed Effect Concentrations) (NOECs) and (Lowest Observed Effect Concentrations) (LOECs) were only used for the derivation of a chronic criterion when a robust EC_{10} could not be calculated for the genus. • Furthermore, EPA independently calculated these toxicity values if sufficient raw data were available for EPA to conduct statistical analyses. EPA's independently-calculated toxicity values were used preferentially, where available. • I agree with the authors' decision on not developing plant criteria based on their lesser sensitivity to PFOS than in comparison to aquatic vertebrates and invertebrates. The

2.2. The Technical Approach Used to Derive the Draft Criterion for PFOS	
Reviewer	**Comments**
	EPA team evaluated the toxicity data to plants as an additional line of evidence and confirmed that the proposed PFOS freshwater acute and chronic criteria are expected to be protective of freshwater plants. • EPA developed protective tissue-based criteria through a bioaccumulation factor approach. This was based on the application of evaluation criteria for screening bioaccumulation factors (BAFs). • Based on comprehensive toxicity data assessment, the EPA team has developed the following criteria using the procedures described in the 1985 *Guidelines*. The freshwater acute water column-based criterion magnitude is 3.6 mg/L and the chronic water column-based criterion magnitude is 0.014 mg/L. The chronic freshwater criterion also contains tissue-based criteria expressed as 43.0 mg/kg wet weight (ww) for fish whole-body, 25.3 mg/L ww for fish muscle tissue, and 12.3 mg/kg ww for invertebrate whole-body tissue. • Acute and chronic MDRs for PFOS estuarine/marine criteria derivation were not met and an estuarine/marine FAV could not be calculated to derive an estuarine/marine acute criterion. Further benchmark was developed using predictive approach and discussed in the follow-up question 3.
Reviewer 5	• **Is the technical approach used to derive the criterion elements logical?** Overall, I think the technical approach is relatively sound although there are instances where it was difficult to keep track of all the decisions with regard to data and whether these were consistent and logical. Admittedly, I think this is a tough chemical and a tough dataset and EPA did an excellent job with the background material and highlighted and used key studies (but more have been published since and will, I'm sure be included). Unfortunately, the technical approach to derive criterion elements is not universally logical. Moreover, as mentioned, using only the 4 most sensitive toxicity endpoints followed by a regression (what type? Was this specified?) seems less robust than using a full species sensitivity distribution with a more "natural" distribution of sensitivity (s-shaped, for example). This last statement may not be true so a reasonable compromise might be to include a full SSD as part of the characterization piece related to "considering other toxicity values impact on the FCV, etc.". What really confused me was that when EPA did what amounts to a sensitivity analysis of the FCV by replacing toxicity values, the FCV DECREASED when higher toxicity values were used. While I suspect this happened because switching to higher toxicity values steepened the slope (or something), it does not make intuitive sense to me and should be further explained. Alternatively, an explanation and justification, even brief, would be helpful in supporting the 4 most sensitive toxicity value approach. I am aware that the 1985 guidelines may include this but I suspect most users of the AWQC may not be familiar with the details of the guidelines. With regard to the tissue-based criteria, EPA mentions using "only PFOS studies in which organisms were exposed in the diet" (or similar; p. 88) but then go on to say the BAF approach was used. I would edit this section to start with mentioning that a BAF approach was used

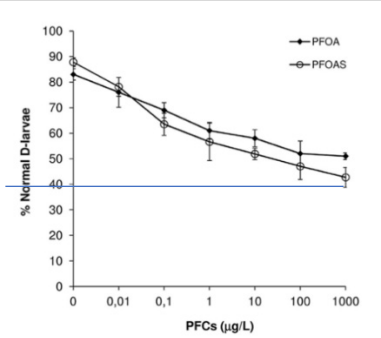
2.2. The Technical Approach Used to Derive the Draft Criterion for PFOS	
Reviewer	**Comments**
	because there were not enough tissue data from laboratory studies. I mention this because it was confusing – there was a lot of explanation of using only dietary exposures and then one sentence (basically) stating…EPA explored a BAF approach. • **Does the science support the conclusions?** Well, offhand, I think the final chronic value for freshwater organisms should likely be lower. Importantly, several studies have been published in 2021 that should likely be included as toxicity values and they may result in lower toxicity estimates. The fact that EPA's criteria are higher than all other published criteria is worrisome. We are all using the same data and many in the field are quite capable scientists. These two papers come immediately to mind but I am sure there are others. Sensitivity and Accumulation of Perfluorooctanesulfonate and Perfluorohexanesulfonic Acid in Fathead Minnows (*Pimephales promelas*) Exposed over Critical Life Stages of Reproduction and Development J.G. Suski, C.J. Salice, M.K. Chanov, J. Ayers, J. Rewerts, J. Field Environmental Toxicology and Chemistry, 2021, pp. 811-819. Toxicological Response of *Chironomus dilutus* in Single-Chemical and Binary Mixture Exposure Experiments with 6 Perfluoralkyl Substances Christopher J. McCarthy, Shaun A. Roark, Demitria Wright, Kelly O'Neal, Brett Muckey, Mike Stanaway, Justin N. Rewerts, Jennifer A. Field, Todd A. Anderson, Christopher J. Salice Environmental Toxicology and Chemistry, 2021, pp. 2319-2333. In my view, it is essential that EPA incorporate newly published toxicity data for PFOS (and PFOA). Furthermore, in several cases, EPA's decisions to use what look like higher estimates of toxicity seem somewhat arbitrary and not internally consistent. I also noted above and mention here again the sensitivity of the criteria development approach to changes in one of the 4 most sensitive taxa/toxicity values. • **Is it consistent with the protection of freshwater aquatic life from acute, chronic, and bioaccumulative effects?** I believe the criteria are "NEARLY" protective of freshwater aquatic life for acute, chronic and bioaccumulative effects of PFOS. I say "nearly" because it seems to me that the FCV, in particular, could and maybe should likely be lower. Also, below I comment on the appropriateness and utility of the frequency and duration elements of the criteria. Briefly, in my opinion the frequency and criteria elements of the criteria certainly help the criteria concentrations to be protective; it is unlikely that a 4-day exposure to the FCV would result in adverse effects to any taxa for which there are data; however, these data are not commonly reported (hourly or 4-day running average concentrations have never been reported to my knowledge).

2.3 Please comment on the approach used to derive the draft acute estuarine/marine benchmark for PFOS. Given the limited estuarine/marine test data available, a new approach method was used to support the derivation of an acute estuarine/marine benchmark to provide states and tribes with a protective value. Please provide detailed comments.

- **Is the technical approach used to derive the benchmark logical?**
- **Does the science support the conclusions?**
- **Is it consistent with the protection offered by acute estuarine/marine aquatic life criteria derived using empirical data, as prescribed in the 1985 *Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*?**

2.3. The Technical Approach used to Derive the Draft Acute Estuarine/Marine Benchmark for PFOS	
Reviewer	**Comments**
Reviewer 1	• **Is the technical approach used to derive the benchmark logical?** The derivation of the acute marine benchmarks (FAV and Criterion Maximum Concentration (CMC)) using the New Approach Method (NAM) is highly uncertain, and I would recommend this analysis not be included as in this document. I do not feel that the analysis and subsequent criteria have high confidence for use in a regulatory application. I understand that similar analyses with other chemicals have about a 90% probability of the predicted effect value being within a factor of 5 of the actual value (Raimondo et al., 2010 – cited in document). Given the calculated CMC (0.43 mg/L), this implies the CMC has about a 90% probability of being within 0.086 to 2.2 mg/L. If the NAM approach stays in the document, this uncertainty and range of values should be acknowledged in the discussion. I would rather see tentative or provisional acute criterion developed from the limited empirical marine acute data highlighted in Appendix B and other recently published marine acute data. This suggests a reasonable interim FAV of approximately 1 mg/L, which is similar to that calculated using the NAM approach. I place higher confidence in the limited empirical data and would suggest EPA emphasize it in addition to or in place of the values calculated by the NAM. I am hopeful that as new toxicity information on marine species are developed, these values can be supplanted with a proper and robust criteria calculation. If such a future analysis is possible, it should be noted. • **Does the science support the conclusions?** See above comment. • **Is it consistent with the protection offered by acute estuarine/marine aquatic life criteria derived using empirical data, as prescribed in the 1985 Guidelines for *Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*?**

2.3. The Technical Approach used to Derive the Draft Acute Estuarine/Marine Benchmark for PFOS	
Reviewer	**Comments**
	The approach seems to be consistent with the approach in the 1985 guidelines. As noted above, the uncertainty with regards to the predictive capability of the interspecies correlations should be acknowledged quantitatively.
Reviewer 2	The technical approach using Web-ICE to determine an acute benchmark for estuarine/marine species is logical. The science has shown that Web-ICE can effectively be used to derive effect measures for additional species using species for which data is available. I think the approach taken by EPA has included sufficient conservatism to address the relatively large amount of uncertainty around the acute toxicity of PFOS to estuarine and marine species. The proposed acute benchmark for estuarine and marine species is an order of magnitude lower than the acute benchmark for freshwater species, which I think underscores the conservatism used by EPA in determining an acute benchmark for estuarine and marine species. That said, the benchmark should be used cautiously due to the relatively large amount of uncertainty and effort should be made to generate acute and chronic toxicity data for PFOS on estuarine and marine species.
Reviewer 3	• **Is the technical approach used to derive the benchmark logical?** After potential inclusion of the data mentioned below this approach may be appropriate. In the current form with the limited data it may be misleading. Can this guidance be updated? I am aware of other researchers investigating PFAS on marine species (Ed Wirth, NOAA) and maybe others that will be coming out soon. • **Does the science support the conclusions?** I believe the data are incorrect for Fabbri et al. 2014. In table B.1 the reported effect concentration is recorded as >1 mg/L. However, looking at the paper, I read, "*The PFCs PFOA and PFOS induced a dose-dependent effect, with significant decreases in normal larval development from 0.1 µg/L (17% and 27%, respectively; P 0.01). Maximal effects were observed at 100 µg/L (about 40% and 50%, respectively; P 0.001) with no further decreases at higher concentrations*". There is a monotonic concentration-response curve. The associated figure also supports an effect at 0.1µg PFOS/L, see below. Furthermore, if the EC50 of the test organisms is a needed endpoint (as noted in the PFOA justification, for which is lacking support in the current form) looking at the figure below % of normal D-larvae for PFOS (although incorrectly referred to in the legend as PFOAS) could be inferred at 0.1 mg/L. Furthermore, has EPA considered calculating the MATC from this study?

2.3. The Technical Approach used to Derive the Draft Acute Estuarine/Marine Benchmark for PFOS	
Reviewer	**Comments**
	 **Fig. 4.** Effects of PFOA and PFOS (0.01–0.1–1–10–100–1000 µg/L) on *M. galloprovincialis* normal larval development in 96-multiwell plates. Data are reported as in Fig. 3a. I did not see data included or the study evaluated for: Robertson JC (1986) Potential for environmental impact of AFA-6 surfactant. Beak Consultants Ltd. Missassauga, Ontario, Canada. EPA Docket AR226-1030a043. There are data for saltwater spp in the ITRC from this citation. • **Is it consistent with the protection offered by acute estuarine/marine aquatic life criteria derived using empirical data, as prescribed in the 1985 *Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*?** No, this is a new approach; however, it follows the spirit of the 1985 guidelines.
Reviewer 4	• EPA applied the ICE model predictions to supplement the available test dataset to help fill missing MDRs and allow the derivation of acute estuarine/marine benchmark recommendations for aquatic life using procedures consistent with those in the 1985 *Guidelines*. A total of 3104 datapoints from 398 models were evaluated. • ICE model has been recommended to predict the sensitivity of an untested taxon (predicted taxa are represented by the y-axis) from the known, measured sensitivity of a surrogate species (represented by the x-axis). The ICE model approach used is very reasonable to predict toxicity of untested taxa. • As documented in Section L.1, ICE-predicted models have been used by multiple independent, international groups and further confirms that values developed from ICE-generated SSDs will provide a level of protection that is consistent with using measured laboratory data. • In addition, prediction accuracy and robustness of the model is evaluated using robust parameters (e.g., mean square error, R2), that fall within a defined range of acceptability, and with close prediction confidence intervals that facilitate evaluating the fit of the underlying data. This confirms the robustness of the model.

2.3. The Technical Approach used to Derive the Draft Acute Estuarine/Marine Benchmark for PFOS	
Reviewer	**Comments**
	• ICE models predicted with acceptable accuracy for PFOS when invertebrates were used to predict to invertebrate species and vertebrates were used to predict to vertebrate species in these comparisons. • The draft acute benchmark for estuarine/marine aquatic life developed using this approach is 0.43 mg/L PFOS, it is lower than the recommended acute freshwater criterion(3.6 mg/L), suggesting that estuarine/marine species may be more acutely sensitive to PFOS. This is in line with Hayman et al., (2021), confirming marine species have a higher sensitivity to PFOS than compared to the freshwater organisms. • In this report*, Mytilus galloprovincialis* was not used in the FAV calculation because the value was not definitive, and true sensitivity of this species is unknown. There are two more studies published reporting the toxicity values for marine/estuarine species, including *Mytilus galloprovincialis*. ○ Stuart L. Simpson, Yawen Liu, David A. Spadaro, Xinhong Wang; Rai S. Kookana and Graeme E. Batley Chronic effects and thresholds for estuarine and marine benthic organism exposure to perfluorooctane sulfonic acid (PFOS)-contaminated sediments: Influence of organic carbon and exposure routes https://doi.org/10.1016/j.scitotenv.2021.146008 ○ Nicholas T Hayman , Gunther Rosen , Marienne A Colvin , Jason Conder , Jennifer A Arblaster Aquatic toxicity evaluations of PFOS and PFOA for five standard marine endpoints. https://doi.org/10.1016/j.chemosphere.2021.129699 **It is recommended to assess the quality of the toxicity data on marine/estuarine species and recalculate estuarine criteria based on this recently available information.**
Reviewer 5	• **Is the technical approach used to derive the benchmark logical?** Yes, given the lack of PFOS toxicity data for acute estuarine/marine species, I think applying WEB-ICE is a REASONABLE APPROACH…perhaps the only approach that is defensible. Clearly, more (or some) data would be a wonderful contribution. WEB-ICE, as mentioned, has been reviewed and published quite a bit so I think, as an approach, it has merit and support of the scientific community. EPA also did a good job presenting the approach and being clear about the criteria being a draft. Overall, when data have been lacking, EPA has used state-of-the-art approaches to developing criteria (my concerns are mostly when sufficient data are available). • **Does the science support the conclusions?** Yes, the science supports the conclusions. Interestingly, the acute criteria for estuarine/marine species (o.43 mgPFOS/L) is almost an order of magnitude lower than the acute criteria for freshwater organisms (3.6 mg/L). Whether estuarine/marine species are truly more sensitive remains to be seen but, to me, it is more reasonable, given the lack of data, that the criteria draft is lower.

2.3. The Technical Approach used to Derive the Draft Acute Estuarine/Marine Benchmark for PFOS	
Reviewer	**Comments**
	• **Is it consistent with the protection offered by acute estuarine/marine aquatic life criteria derived using empirical data, as prescribed in the 1985 *Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*?** Yes, and EPA justified this in the explanation of WEB-ICE that occurs in Appendix L and, overall, the approach and resulting criteria are consistent with the protection of estuarine and marine species.

2.4 Please comment on the use of measured and unmeasured toxicity tests to derive the respective criterion. In particular please comment on the supporting justification for using unmeasured toxicity tests in Appendix O.

2.4. The Use of Measured and Unmeasured Toxicity Tests to Derive Respective Criterion	
Reviewer	**Comments**
Reviewer 1	The consideration of toxicity data from experiments in which PFOS measurements were not made seems appropriate. The Appendix O analysis is supportive of the general observation that actual concentrations in the toxicity test waters approximated nominal values for freshwater. I agree that actual concentrations in the toxicity test waters for the marine test may be lower than nominal values, thus, effect data originating from marine studies that only report nominal concentrations may be biased high in some cases. Given the tentative/temporary nature of the marine criteria developed in this study, this bias is manageable until additional empirical data from experiment with measured concentrations in water can be provided.
Reviewer 2	I am concerned with the approach of using the agreement of measured and nominal concentrations from studies that measured the concentration of PFOS in their tests to determine whether to use toxicity data from studies that did not measure the concentration PFOS in their tests. My concern stems from this approach having to assume that studies that did not measure the concentration of PFOS in their experiments performed the dosing of PFOS with the same care and skill as those studies that did measure the concentration of PFOS in their experiments and measured concentrations within 20% of nominal. My concern is compound by 58% and 65% of the freshwater and saltwater tests, respectively, only reporting nominal test concentrations. The EPA's approach uses the agreement of measured and nominal concentration in a minority of studies to determine whether to include the majority of studies on their assessment. I am assuming that there wouldn't be sufficient data to determine a criterion without using data from studies that did not measure the concentrations of PFOS in their experiment?

2.4. The Use of Measured and Unmeasured Toxicity Tests to Derive Respective Criterion	
Reviewer	**Comments**
	I think the approach that the EPA has used to determine the level of agreement between the nominal and measured concentration of PFOS in the studies that measured the concentration is logical and valid. It is encouraging that the agreement on average is high. Again, my largest concern is assuming this agreement in a minority of studies is present in all studies.
Reviewer 3	This seems acceptable for the time being. Having worked in the laboratory with PFOS, I can make a first-hand testament that mixing PFOS into exposures solutions does not guarantee a homogenous mixture despite working at solutions well below the solubility limit. There are nuances associated with achieving homogeneity of the exposure solution, we have developed a PFAS mixing protocol to reduce chemical clumping and this increases uniformity of the solutions. Furthermore, there is approximately 30% variability of PFOS quantitatively (see…Rewerts et al. 2020); so, the best measurement still has significant variability.
Reviewer 4	PFOS is a highly stable compound, resistant to hydrolysis, photolysis, volatilization, and biodegradation (as described in Section 1.1.1 of the Report) and, therefore, expected to vary only minimally in the course of a toxicity test. To determine if nominal and measured PFOS concentrations were typically in close agreement, pairs of nominal and corresponding measured PFOS concentrations were compared to one another through (1) linear correlation analysis and (2) an assessment of measured concentrations as a percent of its paired nominal concentration. The authors reported, 22 freshwater studies with PFOS measured concentrations, yielding 373 pairs of measured and nominal concentrations. In addition, there were 7 estuarine/marine studies with measured concentrations, yielding 142 pairs of measured and nominal concentrations. The data were grouped by classifications including water type (salt/fresh) and experimental conditions (acute/chronic; solvent/no solvent; fed/unfed, etc.). Data displayed a high degree of linear correlation and measured, and nominal concentrations were in close agreement The analysis conducted by EPA Team showed strong correlation (correlation = 0.9998) of the 326 pairs of nominal and measured concentrations from freshwater studies. In addition, the experimental conditions did not influence the correlation between nominal and measured concentrations. The detailed analyses of the data in Appendix O and the relevant Tables and Figures provide very comprehensive analyses – this is very useful information and will assist ecotoxioclogist in designing future experiments. ***This confirms inclusion of unmeasured PFOS toxicity tests for quantitative use in criteria derivation.*** Personal experience on analyzing PFOS in ecotoxicological studies using freshwater species have also exhibited strong correlation between nominal and measured concentrations. The authors reported the strong correlation (0.8993) of the 142 pairs of nominal and measured concentrations, the ratio of measured to nominal concentrations from the saltwater dataset showed bias with a geometric mean value of 0.6178. Additionally, the median percent difference between measured and nominal concentration was 30.82%. Furthermore, the saltwater comparison of nominal and measured concentrations indicated that these

2.4. The Use of Measured and Unmeasured Toxicity Tests to Derive Respective Criterion	
Reviewer	**Comments**
	experimental conditions (acute/chronic and unfed/fed) could influence the observed differences between measured and nominal concentrations. These results suggest that measured and nominal concentrations from saltwater tests were not in close agreement, but this analysis was based on limited set of data. The measured concentrations in the recently published paper on marine/estuarine toxicity of PFOS should also be included in this assessment: Nicholas T Hayman , Gunther Rosen , Marienne A Colvin , Jason Conder , Jennifer A Arblaster; Aquatic toxicity evaluations of PFOS and PFOA for five standard marine endpoints. https://doi.org/10.1016/j.chemosphere.2021.129699 The second paper is on benthic organisms and PFOS is measured in overlying water, porewater and sediment. This may provide further guidance on difference between PFOS measured and nominal concentrations. Stuart L. Simpson, Yawen Liu, David A. Spadaro, Xinhong Wang; Rai S. Kookana and Graeme E. Batley; Chronic effects and thresholds for estuarine and marine benthic organism exposure to perfluorooctane sulfonic acid (PFOS)-contaminated sediments: Influence of organic carbon and exposure routes https://doi.org/10.1016/j.scitotenv.2021.146008 **Additional information for Appendix O based on a recently published paper:** *According to **Rewerts et al., 2021** additional handling steps, which are not typically reported for ecotoxicological studies but may contribute to variability, include solution homogenization, subsampling procedures, and the container materials selected for storage*. https://doi.org/10.1002/etc.4667
Reviewer 5	I think the comparison of measured and nominal concentrations was an interesting read and a useful contribution. That said, many toxicologists focused on PFAS have commented that analytical confirmation is necessary for a high quality study – this was echoed (loudly) at the SETAC Workshop on Risk of PFAS that occurred in summer, 2019. As well, in my own experience there have been challenges in sometimes matching nominal and measured concentrations for aquatic exposures. The paper by Rewerts et al. 2020 highlights some of the challenges and provides recommendations for accurate solutions of PFAS. As a general rule, we have erred on the side of reporting measured concentrations. Two important thoughts. First, several very prominent analytical chemists that have made a career of measuring PFAS have indicated to me that the analytical method is only about 30% accurate – meaning that if the analytical measure was +/- 30% of nominal, they would be considered "the same". EPA used 20% as a threshold (for deciding nominal and measured were the same) and I'm not sure why this is. As far as I can tell, 30% is a more reasonable threshold. Second, in the review and derivation of toxicity values for the MacDonald et al. 2014 paper, EPA elected not to use the 20-day emergence rate endpoint, in part, because the nominal and measured did not agree. This makes no sense to me. As long as the solutions were confirmed

2.4. The Use of Measured and Unmeasured Toxicity Tests to Derive Respective Criterion	
Reviewer	**Comments**
	analytically and reported, that should be good enough and, in fact, preferred over nominal alone. Paper worth including in the section on nominal vs. measured PFOS concentrations: Key Considerations for Accurate Exposures in Ecotoxicological Assessments of Perfluorinated Carboxylates and Sulfonates. Justin N. Rewerts, Emerson C. Christie, Alix E. Robel, Todd A. Anderson, Christopher McCarthy, Christopher J. Salice, Jennifer A. Field Environmental Toxicology and Chemistry, 2020

2.5 Please comment on the toxicity data used to derive the draft criteria.

- **Were the data selected and/or excluded from the derivation of the criteria derivation appropriately utilized?**
- **Are there relevant data that you are aware of that should be added to the analyses (note that EPA is working on updating the toxicity data to reflect the data in ECOTOX between Sept. 2019 through the latest update)? If so, please provide references for consideration.**

In particular, please comment on:

2.5.a. The toxicity values used to derive the PFOS criteria, with a particular emphasis on:

2.5.a.i. the use of the qualitatively acceptable acute midge (Chironomus plumosus) data from Yang et al. (2014) to suggest aquatic insects are relatively tolerant to acute PFOS exposures. Specifically, Yang et al. (2014) conducted a 96-hour renewal, measured PFOS acute test with the midge, Chironomus plumosus. This study was not acceptable for quantitative use due to the potential problematic source of the organisms. The reported LC50 was 182 mg/L for PFOS indicating that insects may not be one of the more sensitive taxonomic groups. Therefore, this test was excluded from the acute criterion calculation, but used to waive the missing insect MDR.

2.5.a.ii. the use of the quantitatively acceptable chronic toxicity value for mussel (Lampsilis siliquoidea) from Hazelton et al. (2012). Specifically, Hazelton et al. (2012) conducted a 36-day renewal, measured PFOS chronic test with fatmucket, Lampsilis siliquoidea. The estimated EC10 was 0.05713 mg/L, which was extrapolated from the author-reported data and the exposure response slope from another PFOS toxicity study focused on another mussel species (Ellipto complamata) as explained in Section 3.1.1.3.3. Therefore, this test was used in the chronic criterion calculation.

2.5.a.iii. the use of the quantitatively acceptable chronic toxicity value for damselfly (Enallagma cyathigerum) from Bots et al. (2010). Bots et al. (2010) conducted a 320-day renewal, unmeasured PFOS chronic test with blue damselfly nymphs, Enallagma cyathigerum. The MATC was 0.03162 mg/L, which was calculated from the author-reported value for nymph survival as explained in Section 3.1.1.3.2. Therefore, this test was used in the chronic criterion calculation.

2.5.a.iv. the use of the quantitatively acceptable chronic toxicity value for midge (Chironomus dilutus) from MacDonald et al. (2004). MacDonald et al. (2004) conducted a 20-day renewal, measured PFOS chronic test with midge lava, Chironomus dilutus. The EC10 was 0.05963 mg/L, which was an EPA-calculated value for 10-day growth as explained in Section 3.1.1.3.4. Therefore, this test was used in the chronic criterion calculation.

2.5.b. EPA's approach for fitting concentration-response (C-R) data (described in Appendix K) as well as the specific acute LC50 values (Appendix A.2) and chronic EC10 values (Appendix C.2) that were estimated (for sensitive genera when C-R data were available) and used to derive criteria.

2.5. The Toxicity Data to Derive the Draft Criterion	
Reviewer	**Comments**
Reviewer 1	• **Were the data selected and/or excluded from the derivation of the criteria derivation appropriately utilized?** In most cases, yes. Please see detailed comments on particular studies and interpretations in response to other charge questions. • **Are there relevant data that you are aware of that should be added to the analyses (note that EPA is working on updating the toxicity data to reflect the data in ECOTOX between Sept. 2019 through the latest update)? If so, please provide references for consideration.** Hayman, N.T., Rosen, G., Colvin, M.A., Conder, J., Arblaster, J.A. 2021. Aquatic toxicity evaluations of PFOS and PFOA for five standard marine endpoints. Chemosphere 273:129699. **2.5.a.** **2.5.a.i.** I disagree with excluding this data point from the acute criteria calculations. I assume this data has been removed under the assumption that these animals may have been pre-exposed to PFOS and may have been more tolerant of PFOS exposures, which would result in biased-high median lethal concentration (LC50) values. If so, this should be explicitly stated. Assuming these *Chironomus* can develop tolerance to PFOS, it seems that they would have to be exposed to rather high mg/L ranges of PFOS in water given the reported 96-hour LC50 of 182 mg/L. Based on published literature, I am unaware of natural ecosystems in China (where the animals may have been originally harvested) with concentrations of PFOS that approach

2.5. The Toxicity Data to Derive the Draft Criterion	
Reviewer	**Comments**
	this order of magnitude range (in which they could build up a tolerance). The animals were obtained from a local market, so it is also possible that they were cultured for several generations, presumably using uncontaminated water (which would further reduce the chance that multiple generations were exposed at these levels). Overall, I think it is more reasonable to assume that the animals used in the experiment have not built up an acute lethal tolerance to PFOS, and the that LC50 result is unbiased. It does seem clearly show that insects may be less sensitive to acute lethality effects of PFOS. As such, I think it should be included as a quantitative endpoint. Additionally, it seems inconsistent to exclude this Yang et al (2014) study, when chronic data from an unpublished study by Funkhouser (2015) were included for quantitative consideration. As noted on page C-25, the animals in the Funkhouser (2015) study were "purchased from a private collector" and then kept for "several" generations prior to testing. The source of the animals is just as uncertain as the Yang et al (2014) animals, and it is unclear (if PFOS tolerance at lethal levels is possible) how many generations would be needed to shed adaptive tolerance and how it would compare to "several." Simply put, if data from experiments like Funkhouser (2015) are quantitatively included, those from Yang et al. (2014) should also be quantitatively included (with some notes on the uncertainty of the animal sources). **2.5.a.ii.** There were only three exposure levels in this experiment, including the control. One PFOS dose (4.5 µg/L) indicated an absence of detectable effects on metamorphosis, the other (69.5 µg/L) indicated an approximate 35% reduction relative to controls. This is not a definitive test; there is little dose response information to fully confirm the effects/absence of effects and predict an effective concentration (EC) value with a dose response model. Application of another study's dose response curve to generate EC10 values for this study does not address this fundamental shortcoming, and simply carries too much uncertainty. Although there are only two PFOS doses, which is highly uncertain, use of a Maximum Acceptable Toxicant Concentration (MATC) value may be a less uncertain path to including this study in quantitative calculations. This would result in a more conservative chronic value for this study (0.018 mg/L instead of 0.057 mg/L). Given the high uncertainty of using this result (due to only 2 PFOS doses), I believe this value should be caveated in some way and re-evaluated for use or excluded in future criteria derivation. For example, on page C-22, the Spachmo and Arukwe (2012) value (which also featured a limited PFOS dose design), the document notes that the limited doses "may limit its future use in the criteria derivation pending independent verification of the toxicity values by EPA." **2.5.a.iii.** I agree with the interpretation of the Bots et al. (2010) study and selection of the MATC. **2.5.a.iv.** First, on page 104, the document mentioned "an EC10 of 0.0586 mg/L for growth following 10-days of exposure", but on page 115, the document noted "10-day growth with an EC10 of 0.05963 mg/L". In Appendix C (page C-19), the document states "the independently-calculated 10-day EC10 for growth was 0.0586 mg/L." There's some inconsistencies in the value being obtained from EPA's EC10 modeling using this reference; please correct or clarify.

2.5. The Toxicity Data to Derive the Draft Criterion	
Reviewer	**Comments**
	One justification for using the 10-day EC10 growth result rather than other results is a lack of being able to calculate EC10s. I do think the emergence results should not be discounted, however. EPA notes "as for the emergence endpoint, there was a lack of a concentration-response relationship and there were very similar levels of observed effects (which ranged between 42.6 and 50.1%) despite the more than nine-fold increase in the mid-range treatment concentrations (0.0023, 0.0144, 0.0217 mg/L, respectively)." The magnitude of the effect (relative to controls), and the fact that there were statistically-detectable differences from controls in some of these doses (0.0144, 0.0217 mg/L) seems to indicate an ecologically meaningful adverse effect is occurring due to PFOS. This range of concentrations just might be a portion of dose response curve that is relatively flat. There is a very clear adverse effect at 0.0949 mg/L. I think it would be reasonable to select the MATC for emergence (0.0071 mg/L reported on page C-19) and treat it a second study point since it was a completely different experiment from the 10-day experiment used to provide the EC10 of 0.05963 (or 0.0586) mg/L value. A Species Mean Chronic Value using the 0.0071 and 0.0586 mg/L results would be 0.020 mg/L. This 0.020 mg/L value would seem to be protective while including the growth and emergence data from these two experiments. **2.5.b.** More details need to be provided on the dose response modeling using R. Appendix K is helpful for providing the reader with details on the general approach, but where EC10s are modeled by EPA, the model being used (out of the 22 available in the R software package) needs to be specified. Providing some indication of variability (such as a 95% confidence interval) for the model-generated EC10s is standard practice for dose response modeling, and this information should be provided somewhere in the document. Showing the R package output of the goodness of fit statistics (or equivalent) for the modeling in an Appendix would be helpful; since this was used to select the model used in each instance of an EC10 calculation, it must be available, so I would recommend including it for full transparency and to aid future efforts in understanding the aquatic toxicology of this chemical. Additionally, it would be helpful to show the selected model fits for all calculated EC10s (as shown for the most sensitive EC10s estimated). These steps would be helpful to ensure and demonstrate quality of the model fits and reproducibility of the modeling work. Additionally, somewhere in the document (Appendix K), the 22 dose response model equations should be provided to the reader. Alternately, a reference could be made to a document that clearly provides this information (ideally a peer-reviewed or EPA document) containing all 22 models.
Reviewer 2	• **Were the data selected and/or excluded from the derivation of the criteria derivation appropriately utilized?** I think the data used in the derivation of the criteria were appropriate. As mentioned above, I am a little concerned about the use of toxicity data from studies that did not measure the concentration of PFOS in their experiments, especially considering the proportion of studies that did not measure the concentrations. The confirmation of exposure concentrations is an important principle of sound ecotoxicology.

2.5. The Toxicity Data to Derive the Draft Criterion	
Reviewer	**Comments**
	• **Are there relevant data that you are aware of that should be added to the analyses (note that EPA is working on updating the toxicity data to reflect the data in ECOTOX between Sept. 2019 through the latest update)? If so, please provide references for consideration.** I have listed a number of papers below that were published in 2020 and 2021 that the EPA may want to consider in their assessment. Hayman, N.T., Rosen, G., Colvin, M.A., Conder, J., Arblaster, J.A., 2021. Aquatic toxicity evaluations of PFOS and PFOA for five standard marine endpoints. Chemosphere 273, 129699.. doi:10.1016/j.chemosphere.2021.129699 Logeshwaran, P., Sivaram, A.K., Surapaneni, A., Kannan, K., Naidu, R., Megharaj, M., 2021. Exposure to perfluorooctanesulfonate (PFOS) but not perflurorooctanoic acid (PFOA) at ppb concentration induces chronic toxicity in Daphnia carinata. Science of The Total Environment 769, 144577.. doi:10.1016/j.scitotenv.2020.144577 Simpson, S.L., Liu, Y., Spadaro, D.A., Wang, X., Kookana, R.S., Batley, G.E., 2021. Chronic effects and thresholds for estuarine and marine benthic organism exposure to perfluorooctane sulfonic acid (PFOS)-contaminated sediments: Influence of organic carbon and exposure routes. Science of The Total Environment 776, 146008.. doi:10.1016/j.scitotenv.2021.146008 Li, R., Tang, T., Qiao, W., Huang, J., 2020. Toxic effect of perfluorooctane sulfonate on plants in vertical-flow constructed wetlands. Journal of Environmental Sciences 92, 176–186.. doi:10.1016/j.jes.2020.02.018 Aquilina-Beck, A.A., Reiner, J.L., Chung, K.W., Delise, M.J., Key, P.B., Delorenzo, M.E., 2020. Uptake and Biological Effects of Perfluorooctane Sulfonate Exposure in the Adult Eastern Oyster Crassostrea virginica. Archives of Environmental Contamination and Toxicology 79, 333–342.. doi:10.1007/s00244-020-00765-4 Tornabene, B.J., Chislock, M.F., Gannon, M.E., Sepúlveda, M.S., Hoverman, J.T., 2021. Relative acute toxicity of three per- and polyfluoroalkyl substances on nine species of larval amphibians. Integrated Environmental Assessment and Management 17, 684–690.. doi:10.1002/ieam.4391 Suski, J.G., Salice, C.J., Chanov, M.K., Ayers, J., Rewerts, J., Field, J., 2021. Sensitivity and Accumulation of Perfluorooctanesulfonate and Perfluorohexanesulfonic Acid in Fathead Minnows (Pimephales promelas) Exposed over Critical Life Stages of Reproduction and Development. Environmental Toxicology and Chemistry 40, 811–819.. doi:10.1002/etc.4936 Mccarthy, C.J., Roark, S.A., Wright, D., O'Neal, K., Muckey, B., Stanaway, M., Rewerts, J.N., Field, J.A., Anderson, T.A., Salice, C.J., 2021. Toxicological Response of Chironomus dilutus in Single-Chemical and Binary Mixture Exposure Experiments with 6 Perfluoralkyl

2.5. The Toxicity Data to Derive the Draft Criterion	
Reviewer	**Comments**
	Substances. Environmental Toxicology and Chemistry 40, 2319–2333.. doi:10.1002/etc.5066 **2.5.a.** **2.5.a.i.** I think the EPA's decision that the data from Yang et al. (2014) was not acceptable for quantitative use was appropriate. The source of the larvae is problematic. However, I don't agree with the conclusion that insects may not be one of the most sensitive taxa. Chironomus tentans is a relatively sensitive taxa to chronic exposure to PFOS (MacDonald et al. 2004). In Table C.1, the EC10 for C. tentans is reported as 0.05963 mg/L. Chironomus tentans was also the fourth most sensitive species used in calculating the chronic freshwater criterion (Table 3-6). Also, another insect, Enallagma cyathigerum, another insect species, was the second most sensitive species used in calculating the chronic freshwater criterion (Table 3-6). **2.5.a.ii.** EPA used an EC10:EC35.4 from Drottar et al. (2000) for Elliptio complanata and applied this ratio to derive an EC10 from the data reported in Hazelton et al. (2012) for Lampsilis siliquoidea. The problem is that EPA have not clearly outlined in section 3.1.1.3.3 what endpoint that Drottar et al. (2000) was measuring in Elliptio complanata (also note that the genus and species are not spelled correctly in section 3.1.1.3.3). Is the endpoint measured in E. complanata the same as the endpoint measure in L. siliquoidea? I tried to look up the endpoint measure in Drottar et al. (2000) but I could not find the study and there was no reference provided in the reference section for Drottar et al. (2000). This missing information makes it difficult to comment on the validity of the approach that EPA has taken to derive an EC10 for L. siliquoidea. **2.5.a.iii.** I think the EPA's justification for the use of the survival data from Bots et al. (2010) is valid. While control mortality reached 40% in the control, the plateau in control mortality after 60 days, the total duration of the test being 200 days, and 82.57% survival in the control from day 60 to 200, justifies the inclusion of the MATC derived from Bots et al. (2010) for Enallagma cyathigerum in the derivation of a chronic criterion. **2.5.a.iv.**First, the wrong species is referenced in relation to the MacDonald et al. (2004) study. MacDonald et al. (2004) reported the toxicity of PFOS to Chironomus tentans. The EPA's derivation of a 10-d EC10 for Chironomus tentans using the data from MacDonald et al. (2004) is not clear. In Appendix C, section C.2.4, the EPA writes, "EPA could not fit a curve to independently verify the 10-day survival (due to a lack of a specific sample size for this endpoint as the number of replicates was not stated in the paper; however, the number of replicates was between 2 and 4 and EPA sought to obtain clarification and treatment level data from the study authors)" It is not clear how EPA got the information necessary, e.g., number of replicates, to fit a curve. It is also not clear what EPA means by "…and treatment level data from the study authors."? Did EPA acquire the raw data for growth from the 10-day toxicity test with C. tentans? If that is the case, they have not made that clear. If that is the case, it would also strengthen their independently derived EC10 for growth in C. tentans. I think the EPA needs to more clearly explain where they got the data necessary to derive the EC10 for C. tentans used in the chronic criterion.

2.5. The Toxicity Data to Derive the Draft Criterion	
Reviewer	**Comments**
	2.5.b. I think the approach that the EPA used to determine effect measure from concentration-response data was appropriate. The use of the drc package in R to fit 22 different models to the empirical data and then using several criteria (e.g., AIC, residual standard errors, confidence intervals) to evaluate the fit of the different models is robust. It would have been useful if the EPA reported the 22 different models in Appendix K. I think the LC50 and EC10 values determined by the EPA using the approach mentioned in the previous paragraph was appropriate. It is valid for these effect measures to be determined when the concentration-response data has been provided by the authors of the study. The EPA has also made is clear in Appendix A.2 and C.2 how they determined these effect measures using the concentration-response data provide in the studies. This generates a high level of transparency in the derivation of the criterion.
Reviewer 3	See collective responses below • **Were the data selected and/or excluded from the derivation of the criteria derivation appropriately utilized?** Data selection and waiving of the MDR for insect family in the FAV seem reasonable. • **Are there relevant data that you are aware of that should be added to the analyses (note that EPA is working on updating the toxicity data to reflect the data in ECOTOX between Sept. 2019 through the latest update)? If so, please provide references for consideration.** The data selection for the derivation of the draft criteria are limited to published and/or available studies from 2018 and prior. This significantly reduces the studies used in the derivation as a number of publications have become available in recent years. For example: A newly published study is available for fathead minnows exposed to PFOS for chronic duration and over the course of reproduction and development. Although, this study was static-renewal, PFOS concentrations are measured ; importantly, this study resulted in a NOEC of 88µg/L based on reduced biomass seen in the second generation (Suski et al. 2020). Importantly, follow-on work (in prep) indicates that this may be a maternal transfer impact as PFOS exposures to juvenile fish alone do not share results. Also, from the authors noted above is an ongoing full life-cycle fathead PFOS and PFAS mixture exposure. This study is being conducted under flow through conditions and is expected to reach termination in December 2021. McCarthy et al. 2021 published data on chironomids (EC20 = 1.7µg/L), these are also not included here.

2.5. The Toxicity Data to Derive the Draft Criterion	
Reviewer	**Comments**
	Bryan Brooks (Baylor) and Matt Simcik (UMN) also have acute data on the fathead minnow with measured concentrations, these are not published just yet. David Moore (Army Corps) is near completion of a full life-cycle fish study In particular, SERDP has been funding this research for years and those data are published, recently published or near final. EPA should reach out to SERDP PIs for data inquiries and potential inclusion in these draft criteria. **2.5.a.** **2.5.a.i.** This seems appropriate, the flower market is most definitely an odd place to purchase research organisms. **2.5.a.ii.** From Section 3.1.1.3.3: "*The in marsupia exposure was followed by a 24-hour free glochidia exposure consisting of a factorial design, such that free glochidia from the control group of the marsupia exposure were divided between a control and the two PFOS treatments and the PFOS treatments were split into control and the same PFOS treatment group as the marsupia exposure."* - Comment: This is an exceptionally long and confusing sentence please revise to help the reader understand this complex study and overall approach that EPA took. The approach seems ok given the limited data availability at this time. **2.5.a.iii.** Given the duration of the study the researchers likely hovered around the nominal concentrations of PFOS. Inclusion seems appropriate. **2.5.a.iv.** I am uncomfortable with this conclusion presented here, it may be more appropriate to use MacDonald et al. data from the 20-day endpoint considering recent publication from McCarthy et al. 2020 as noted above. **2.5.b.** This seems like a reasonable and defensible approach if it is applied consistently across genera.
Reviewer 4	The data selected to derive PFOS criteria are appropriate. Studies that did not fully meet the data quality objectives outlined in the 1985 *Guidelines* were not considered for inclusion in the criteria derivation, including some studies with other PFAS exposures, but were considered qualitatively as supporting information. A brief summary of each study describing the experimental conditions and summary tables providing all the relevant information such as strengths and limitations of each study, end points selected for deriving criteria are well documented by the EPA team and provides further confidence in data selection process. The key acceptable exclusion/inclusion criteria used to derive draft criteria are listed below: • Only single chemical toxicity tests with PFOS were considered for possible inclusion in criteria derivation, studies that tested chemical mixtures, including mixtures with PFAS compounds were excluded from criteria derivation.

2.5. The Toxicity Data to Derive the Draft Criterion	
Reviewer	**Comments**
	• Both controlled laboratory experiments and field observations/studies were included. • PFOS toxicity tests were not excluded from quantitative use in criteria derivation on the basis of unmeasured test concentrations alone. • Due to lower sensitivity, insect MDR was excluded from the criterion calculation, but were used to waive the missing insect MDR. • Further supporting information on acceptable and unused studies for acute and chronic endpoints and for freshwater and marine studies are documented and summarized as appendices in this report. Additional toxicity data published over the last six months is listed below: **Marine/estuarine** Nicholas T Hayman , Gunther Rosen , Marienne A Colvin , Jason Conder , Jennifer A Arblaster Aquatic toxicity evaluations of PFOS and PFOA for five standard marine endpoints. https://doi.org/10.1016/j.chemosphere.2021.129699 Stuart L. Simpson, Yawen Liu, David A. Spadaro, Xinhong Wang; Rai S. Kookana and Graeme E. Batley Chronic effects and thresholds for estuarine and marine benthic organism exposure to perfluorooctane sulfonic acid (PFOS)-contaminated sediments: Influence of organic carbon and exposure routes https://doi.org/10.1016/j.scitotenv.2021.146008 **Fresh water** Christopher J. McCarthy, Shaun A. Roark, Demitria Wright, Kelly O'Neal, Brett Muckey, Mike Stanaway, Justin N. Rewerts, Jennifer A. Field, Todd A. Anderson, Christopher J. Salice, Toxicological Response of *Chironomus dilutus* in Single-Chemical and Binary Mixture Exposure Experiments with 6 Perfluoralkyl Substances, Environmental Toxicology and Chemistry, 10.1002/etc.5066, **40**, 8, (2319-2333), (2021). https://doi.org/10.1002/etc.5066 **2.5.a.** **2.5.a.i.** Waiving an unfulfilled MDR when available data suggest it is not among the four most sensitive genera is consistent with previous EPA criteria documents, including U.S. EPA (2016). At this stage, I do not fully agree with the statement that midge larvae are tolerant to acute exposures. The OECD protocol recommends 48h acute test for midge larvae and the 48h exposure period is acceptable duration for assessing acute toxicity. The study by Olson (2017) has limitations but this study can't be fully ruled out. The chronic toxicity data exhibits sensitivity of insects to PFOS and this statement is also supported by the authors. In addition, Stefani et al. (2014), Macdonald et al. (2004), and Marziali et al. (2019) conducted chronic toxicity tests with *Chironomus* spp. and reported apical endpoints. *Results of these studies, taken together, also suggest that insects are among sensitive taxa to chronic PFOS exposures (with adverse effects reports at low ug/L)*

2.5. The Toxicity Data to Derive the Draft Criterion	
Reviewer	**Comments**
	I support the recommendation '*Additional insect toxicity data for PFOS would be very useful for further examining the relative sensitivity of insects to PFOS exposures''.* Unpublished work from our lab shows acute toxicity to midge larva, *Chironomus tepperi* at 1 mg/L PFOS (48 h EC50 value). **2.5.a.ii.** The authors have provided detailed assessment of this study and explained the approach used for the calculation of chronic toxicity value (section C.2.3-Third Sensitive Freshwater Genus for Chronic Toxicity: *Lampsilis siliquoidea* (mussel). Hazelton et al. 2012 used robust study design in spite of including only two concentration of PFOS in this study. The PFOS exposure concentration was measured, and metamorphosis success was used as an endpoint for inclusion in the criteria development. While viability of free glochidia at 24 hours post removal from females was a less sensitive endpoint and did not meet the acceptability criteria. The reduction in metamorphosis success at the 0.0695 mg/L was estimated to be 35.4% but EC10 could not be calculated based on only two PFOS concentrations tested in this study. The EPA team has calculated EC10 (0.05713 mg/L) using the exposure response slope from PFOS toxicity study on another mussel species (*Ellipto complamata*). The explanation and logic provided is reasonable to include the calculated EC10 value to derive the freshwater chronic criterion and to better understand the effects of PFOS on aquatic insects. **2.5.a.iii.** As a weight of evidence approach, EPA ran additional analyses with some of the other toxicity values for *E. cyathigerum* to understand the influence of this study on the overall chronic criterion. The 150-day MATC was more comparable to the other aquatic insect data and more representative of life cycle effects than the 10-day MATC or NOEC at 60 and 320 days of exposure (Table 4.3 of the report). EPA has concluded that the 150-day MATC should be used quantitatively to derive the chronic freshwater criterion toxicity. In addition, the control survival of test organisms was determined to be acceptable at this time point in the test. I am in agreement with this decision. **2.5.a.iv.** The observed effects of PFOS on *C. dilutus* reported in the paper by the study authors include survival and growth as weight (measured as mg of ash-free dry mass per individual) for both the 10-day and 20-day exposure durations and emergence and reproduction over the 20-day exposure duration. The author reported 10-day growth and survival EC10s for the study were 0.0492 and 0.1079 mg/L, respectively. The study authors also reported NOECs of 0.0491 mg/L, LOECs of 0.0962 mg/L, and MATCs of 0.0687 mg/L for both endpoints. The author reported 20-day EC_{10}s for growth, survival, and total emergence were 0.0882, 0.0864, and 0.0893 mg/L, respectively. And the study authors also reported NOECs of 0.0217 mg/L for growth and survival and < 0.0023 mg/L for emergence, LOECs of 0.0949 mg/L for growth and survival and 0.0217 mg/L for emergence, and MATCs of 0.0454 mg/L for growth and survival and 0.0071 mg/L for emergence. Independent statistical analyses were conducted by EPA Team for both the 10-day and 20-day exposure durations using data that were estimated The 20-day EC_{10}s for survival and emergence were not considered to be reliable endpoints given the disparities in the calculated EC_{10}s and the level of data that was presented in the paper, which made independent verification of the toxicity values less accurate. The dosing of the 20-day exposure was more of

2.5. The Toxicity Data to Derive the Draft Criterion	
Reviewer	**Comments**
	a concern than the 10-day exposure, which had measured concentrations that were much more in line with the expected nominal concentrations. The independently-calculated 10-day EC_{10} for growth was 0.0586 mg/L was used quantitatively to derive the chronic aquatic life criterion. The EPA team has reviewed publications by Stefani et al. (2014) and Marziali et al. (2019) as additional supporting information. These authors conducted chronic toxicity tests with *Chironomus* spp. and reported chronic apical endpoints (at low ug/l) but at only at one concentration. **Use of the chronic toxicity data for PFOS in a recent publication should also be considered to assess the reliability of 20-day endpoints (adverse effects reported at 2-3 µg/L) .** Christopher J. McCarthy, Shaun A. Roark, Demitria Wright, Kelly O'Neal, Brett Muckey, Mike Stanaway, Justin N. Rewerts, Jennifer A. Field, Todd A. Anderson, Christopher J. Salice, Toxicological Response of *Chironomus dilutus* in Single-Chemical and Binary Mixture Exposure Experiments with 6 Perfluoralkyl Substances, Environmental Toxicology and Chemistry, 10.1002/etc.5066, **40**, 8, (2319-2333), (2021). https://doi.org/10.1002/etc.5066 **2.5.b.** This is an excellent approach utilized by the EPA Team. EPA's approach for fitting concentration-response (C-R) data resulted in consistent approach across various ecotoxicological studies. The R drc package was used to fit 22 different models to each individual C-R dataset. A single model was then selected from the 22 models to serve as the representative C-R model. The selected model represented the most statistically-robust model available. In certain cases, this approach even improved and helped to select most sensitive toxicological endpoint. In depth analyses and associated dose-response graphs in Appendix A.2 and Appendix C.2 provides further in-depth information on the EPA's approach for fitting concentration-response (C-R) data. As noted in Section 8 some of the values are missing.
Reviewer 5	• **Were the data selected and/or excluded from the derivation of the criteria derivation appropriately utilized?** As mentioned, I feel that there are some inconsistencies with how some data were included or excluded. In the previous comment, for example, some data were excluded from the MacDonald et al. 2014 paper because there was some disagreement between nominal and measured. With regard to PFAS, I would say measured is almost always better than nominal and the fact that these sometimes don't agree should not be too big of a deal as long as they are not wildly different. EPA put substantial effort into sometimes justifying nominal – in all cases, excluding studies that had analytical confirmation is less defensible than including studies that only report nominal, in my opinion. This last statement is, of course, provided the analytical methods are robust.

2.5. The Toxicity Data to Derive the Draft Criterion	
Reviewer	**Comments**
	• **Are there relevant data that you are aware of that should be added to the analyses (note that EPA is working on updating the toxicity data to reflect the data in ECOTOX between Sept. 2019 through the latest update)? If so, please provide references for consideration.** Sensitivity and Accumulation of Perfluorooctanesulfonate and Perfluorohexanesulfonic Acid in Fathead Minnows (*Pimephales promelas*) Exposed over Critical Life Stages of Reproduction and Development. J.G. Suski, C.J. Salice, M.K. Chanov, J. Ayers, J. Rewerts, J. Field Environmental Toxicology and Chemistry, 2021, pp. 811-819. Toxicological Response of *Chironomus dilutus* in Single-Chemical and Binary Mixture Exposure Experiments with 6 Perfluoralkyl Substances. Christopher J. McCarthy, Shaun A. Roark, Demitria Wright, Kelly O'Neal, Brett Muckey, Mike Stanaway, Justin N. Rewerts, Jennifer A. Field, Todd A. Anderson, Christopher J. Salice. Environmental Toxicology and Chemistry, 2021, pp. 2319-2333. **2.5.a.** **2.5.a.i.** Given that insects are among the most sensitive organisms for the chronic exposures to PFOS, it seems the Yang et al. 2014 paper is not very consistent with the prevailing data. Additionally, McCarthy et al. 2021 reports toxicity to chironomids similar to that of MacDonald et al. Additionally, while the Olson 2017 data for Aedes species was not acceptable (for valid reasons), nonetheless the study shows very high sensitivity of another insect species to acute exposures to PFOS. That said, given the EPA's stance and justification that "nominal generally equal measured PFOS concentrations", I'm inclined to put more confidence in Olson's study. Same for the 20-day data in the MacDonald et al. paper on chironomids. In that case, there was a "relatively large difference between measured and nominal concentrations (p. 278)" and so the data were not used. This seems odd to me – as long as there are measured data, that's what I would suggest using. Regardless, the MacDonald et al. paper points to the sensitivity of insects so, collectively, I'd be disinclined to say that the Yang et al. paper shows insects are not sensitive and the data requirement can be waived. I wonder if it's possible to somehow estimate an acute toxicity value for aquatic insects based on chronic toxicity data? Basically, a reverse of the Acute/Chronic ratio approach. **2.5.a.ii.** Unfortunately, I cannot find the Drottar et al. (2000) paper which is the basis of estimating the EC10 from the EC35 generated in the Hazelton et al. (2012) study. And..I think the citation in the document is incorrect and this should be Drottar and Kreugar (2000g)… or check to make sure the citations in text and references match. Moroever, the Drottar paper appears to be an acute test which is VERY different than the Fatmucket study. This approach seems like a "stretch" and, again, somewhat inconsistent with the approaches and decision matrix EPA has used to utilize or discard other data and endpoints. **2.5.a.iii.** Well, clearly it would have been better to be able to estimate an EC10 but this appears appropriate. I note that for this study the exposure concentrations were an order of magnitude apart; in other cases, EPA has used "too big of a difference between exposure concentrations" to discard a study or two. Somewhere, it would be good to know at what point

2.5. The Toxicity Data to Derive the Draft Criterion	
Reviewer	**Comments**
	there is too great a difference among exposure concentrations (10x, 20x, ?) for the study to be deemed acceptable for use quantitatively. **2.5.a.iv.** I do not agree with the toxicity value used by EPA as obtained from the MacDonald et al. study. The lowest toxicity value is the MATC for 20-day emergence of 0.0071 mg PFOS/L. It is not clear why EPA did not use this value? Emergence is clearly extremely ecologically relevant, and the value generated seems as defensible as most of the other endpoints EPA has chosen to include? Additionally, and as mentioned above, see the paper by McCarthy et al. (2021) that was just published in Environmental Toxicology and Chemistry. Those data appear robust and should meet acceptability criteria. **2.5.b.** In general, the approach for fitting C-R data that EPA used is basically state-of-the-art. The drc package is very powerful and provides a way to test many different curves to then select the best fit model. Although EPA described some of this in the several sections related to "fitting x data (K 1.2)", I think more details would be warranted. The description for the criteria to select best fit models is rather vague. Perhaps a table of specific fit criteria would be helpful? Perhaps this is not doable because every dataset is different. When I teach modules on Akaike Information Criteria (AIC) I emphasize that the metric "penalizes" fit for more parameters within a model. So, using AIC can yield the simplest, best model that fits the data. This is because models with more parameters tend to yield a better fit purely based on statistical properties and not the actual phenomena being studied. I am not aware that AIC is a measure of the model fit to "true outcomes" which are only theoretical constructs, I think. If we knew the "true outcomes" we would not really need the model. Anyway, I would encourage the authors to review the AIC section and make edits if necessary and certainly cite the source of the explanation. For section K.2.2. are there actually any criteria (i.e., numbers) that are used to determine when a model fit is appropriate? As a simple example, maybe one would consider an r^2 of 0.8 or better to be a "good model" for linear regression? Some statements to this effect and any details regarding actual criteria used to select "good models" would be helpful. So, overall the curve fitting approach is appropriate but more, specific details would be helpful.

2.6 Please comment on the translation of the chronic water column criterion elements for aquatic life to derive the tissue-based criterion elements, considering the bioaccumulation of PFOA and PFOS. In particular, please comment on:

2.6.a. Uncertainty surrounding the bioaccumulation factors (BAFs) used to translate of the chronic water column criterion elements into tissue-based criterion elements.

2.6.b. EPA's determination of appropriate BAFs and the tissue types that the tissue criterion elements were based.

2.6. The Translation of the Chronic Water Column Criterion Elements for Aquatic Life to Derive the Tissue-Based Criterion Elements Considering Bioaccumulation	
Reviewer	**Comments**
Reviewer 1	The derivation of the tissue criteria in this manner is highly uncertain. To my knowledge this is the first time EPA has applied ambient water quality criteria protective of aquatic life direct toxicity with uptake factors (bioaccumulation factors (BAFs), bioconcentration factors (BCFs)) in this manner to calculate tissue criteria. References are made to the selenium tissue criteria, but those are used in the reverse (i.e., criteria based on measured concentrations in tissue used to calculate water criteria). The use of criteria for water with a assumed uptake factor carries a large amount of uncertainty, and in general, the use of measured concentrations in tissue linked to adverse effects is a more straightforward approach since it does not involve uptake model predictions. This needs to be noted in the text. Also, are the predicted tissue criteria meant to be a temporary stop-gap until tissue effect data become available? This should be discussed and clarified. **2.6.a.** The use of BAFs derived from field studies is inherently uncertain due to the wide variety of techniques used in the compiled studies, their analytical data quality, the differences in species and ecosystems, experimental designs, spatial uncertainties for mobile animals like fish, etc. That being said, the use of a BAF value (or BCF) in criteria derivation is consistent with other criteria developed by EPA. As noted above, the use of the tissue criteria needs to be considered carefully, and I think empirical tissue data from toxicity experiments should form the basis of a next iteration of a tissue criteria. **2.6.b.** The development of BAFs for invertebrates, fish (whole body), and fish (muscle) seems reasonable for the application in estimating a draft or interim tissue criteria until empirical tissue data can be used to calculate tissue criteria directly.
Reviewer 2	**2.6.a.** I think the EPA has sufficiently addressed the uncertainty around the use of BAFs and the chronic water column criterion in the derivation of tissue-based criterion. They have indicated that tissue-based criterion should only be observed once in 10 years. The use of the geometric mean of the reported BAFs incorporates the range of BAFs that may be present for different invertebrate and fish species. The use of the chronic water column criterion also builds in added conservatism to the tissue-based criterion. Prosser et al. (2016) reported BAFs for PFOA in three freshwater species (two invertebrates and one fish) (See Tables S29-31 in Supplementary Information), but it was not considered in this assessment. It is not clear why it was not considered. Prosser, R.S., Mahon, K., Sibley, P.K., Poirier, D., Watson-Leung, T., 2016. Bioaccumulation of perfluorinated carboxylates and sulfonates and polychlorinated biphenyls in laboratory-cultured Hexagenia spp., Lumbriculus variegatus and Pimephales promelas from field-collected sediments. Science of The Total Environment 543, 715–726. doi:10.1016/j.scitotenv.2015.11.062 **2.6.b.** The evaluation criteria for BAFs outline in Table 2-4 are appropriate and the decision to only use high and medium quality BAFs is justified based on the criteria that would make a BAF low quality. It was a good idea to use fish BAFs based on the concentration in muscle and

2.6. The Translation of the Chronic Water Column Criterion Elements for Aquatic Life to Derive the Tissue-Based Criterion Elements Considering Bioaccumulation	
Reviewer	**Comments**
	whole body (Table 3-12). Muscle tissue is usually exclusively sampled in large fish, especially as part of fish consumption guidelines. The whole body is more appropriate for small fish and invertebrate species, e.g., minnows, benthic macroinvertebrates.
Reviewer 3	Please clarify, the following sentence: "BAFs used in the derivation of the PFOS tissue criteria consisted of > 2 water and organism samples each and were collected within one year and 2 km distance." It is unclear if the >2 samples refer to the tissue & water samples being mismatched temporally or if there where >2 sets of water and tissue samples that were collected in different years. If the latter then this approach seems appropriate; if the former, EPA should discuss differences in water chemistry between years to alleviate any concerns with matching tissue concentration data to water samples that may have significant environmental temporal variability. A table summarizing the animal tissues used in deriving the BAFs would be helpful to assess the range of fish species and their dietary preferences.
Reviewer 4	**2.6.a.** The freshwater chronic PFOS toxicity data with measured tissue concentrations was limited, with no quantitatively acceptable tissue-based tests. Therefore, there were insufficient data to derive tissue-based criteria using a GSD approach from empirical tissue data from toxicity studies. Tissue criteria derived from the chronic water column concentration (CCC) with the use of bioaccumulation factors were developed by EPA. The chronic freshwater criterion also contains tissue-based criteria expressed as 43.0 mg/kg wet weight (ww) for fish whole-body, 25.3 mg/kg ww for fish muscle tissue, and 12.3 mg/kg ww for invertebrate whole-body tissue. EPA developed protective tissue-based criteria through a bioaccumulation factor approach. The authors reviewed PFOS BAF literature based on four criteria 1) number of water samples, 2) number of organism samples, 3) water and organism temporal coordination in sample collection, and 4) water and organism spatial coordination in sample collection and developed a ranking system. BAFs used in the derivation of the PFOS tissue-based criteria consisted of > 2 water and organism samples each and were collected within one year and 2 km distance. This scheme assured selection of only BAFs of high and medium quality to derive the tissue criteria. **2.6.b.** BAFs are different for muscle/fillet and whole-body tissues. Humans consume muscle/fillets from fish and soft tissues from bivalves, therefore the water quality criteria recommended by EPA used BAFs based on these tissues. In addition, muscle and whole-body are the most commonly sampled tissue types in monitoring programs. These criteria were developed based on the values reported for 50-60 samples (Table 3-12). Within the body, PFOS tends to bioaccumulate within protein-rich tissues, such as the blood serum proteins and liver. EPA Team calculated additional tissue values for liver, blood, and reproductive tissues by transforming the freshwater chronic water column criterion into

2.6. The Translation of the Chronic Water Column Criterion Elements for Aquatic Life to Derive the Tissue-Based Criterion Elements Considering Bioaccumulation	
Reviewer	**Comments**
	representative tissue concentrations using tissue-specific bioaccumulation factors (BAFs). Author's decision uses agreement on the use of female reproductive tissues due to its relevance for potential maternal transfer to offspring. These additional tissue-based values were calculated for comparative purposes and were not proposed as recommended criteria.
Reviewer 5	**2.6.a.** Using the BAF for PFOS to determine the tissue-based criterion elements is, I think, an interesting and useful approach given the lack of tissue-based metrics associated with toxicity data. The variability in observed bioaccumulation of PFOS is an active area of research but the work by Burkhard 2021 (also an author on the AWQC) provides an excellent synthesis and compendium of available BAFs for PFOS. That said, I noticed that the criteria for co-located tissue and water samples for PFOS was that they were collected within a year of each other and within 2 km distance (p. 134) – this likely contributes to significant variability in the BAFs. Although there are few published datasets, there are some datasets that show considerable temporal and spatial variability in PFAS water concentrations over the course of a few weeks and over a spatial distance of less than 0.5 km. I wonder if the variability in BAFs would decrease if the criteria were narrowed to co-collected samples measured at the same time? Might be worth the exercise. Given the variability in PFOS BAFs, why not use something like the 25% percentile BAF instead of the geometric mean? When developing a protective criteria and there is a very noisy data set, it might be beneficial to err on the side of caution until better data (many co-located samples in space and time) were available. So, in summary, I think the approach of using BAFs to estimate tissue-based criteria is reasonable but given the variability in BAFs, I would encourage using a lower BAF instead of the geometric mean or reconsidering the data that went into the BAFs used for criteria development. **2.6.b.** Invertebrate whole body, fish whole body, and fish muscle are appropriate tissues for the tissue-based criterion. These are the most commonly collected tissue types and are relevant to monitoring efforts and are even useful for considerations of fish advisories. That said, the only other tissue worth considering would be for liver in fish since this tissue accumulates considerably more PFOS than muscle – these are included in the appendices so this is appropriate. Overall, tissues used for the tissue-based criteria are appropriate.

2.7 Please comment on the frequency and duration of the criterion elements, in particular the tissue-based criterion elements.

2.7. The Frequency and Duration of the Criterion Elements	
Reviewer	**Comments**
Reviewer 1	The 4-day duration seems to be supported by the more sensitive chronic endpoints used to derive the CCC.

2.7. The Frequency and Duration of the Criterion Elements	
Reviewer	**Comments**
	For the tissue-based criterion (page 135), there is no clear support for assuming a 10-year exceedance frequency. Given the uncertainty with the BAF-predicted tissue criteria, and how little is known regarding the recalcitrance of PFOS in aquatic ecosystems and recovery time if PFOS inputs in water were halted, the assignment of a 10-year exceedance frequency at this stage seems completely arbitrary. We simply do not yet know the time frame over which aquatic ecosystems recover from PFOS. It is not technically supported to cite recovery times for selenium to support a 10-year recovery time for PFOS, these are completely different toxicants that have their own unique fate and behavior. USEPA (1985) guidance suggests assuming a 3-year frequency as a default, and the discussion on page 135-136 is not scientifically convincing enough to modify it to 10 years. Additionally, it should be noted that the exceedance frequency for another organic chemical, Tributyltin (TBT) was set at 3 years by EPA in derivation of that criteria. TBT exhibits uptake factors similar to PFOS (i.e., BCF of approximately 2,000 L/kg, wet weight for goldfish, as noted in the EPA TBT criteria document, which is similar to the PFOS BAFs of 1,800-3,100 L/kg, wet weight being used to calculate the fish tissue criteria). TBT is also persistent in aquatic ecosystems, as noted by EPA. Given TBT is at least an organic chemical, it is a closer analog than selenium, which is an element. As such, the exceedance frequency for the PFOS tissue criterion should be set at the default of 3 years unless EPA can provide convincing technical information specific to recovery times for PFOS. Additionally, on page 136, the paragraph that begins with "Metals and other chemical pollutants such as PFOS..." is not convincing as any quantitative support for EPA's 10-year exceedance frequency for the chronic tissue-based criteria. The text as written may give the reader the conclusion that PFOS recovery may be "on the order of decades", as EPA notes for selenium. There is no support for the conjecture that PFOS recovery may be "relatively slow" or require decades, as noted in my above comment.
Reviewer 2	As per Table 0-1, I think the chosen durations and frequencies for the acute and chronic criteria are appropriate. They will ensure protection of aquatic life. The duration of the tissue-based criterion is appropriate as the concentration will be measured when biota is collected. The 10-year frequency is appropriate considering that for biota to reach the tissue-based criteria, they would likely to have been exposed to concentrations at or above the chronic criteria for an extended period of time.
Reviewer 3	This is a not an easy statement to comment on, as it may be unlikely that the aquatic receptors will exceed or reach these tissue concentrations prior to exceedances from the CCC. What I am not clear on is, if tissue concentrations exceed these proposed thresholds yet, PFOS water concentrations do not exceed the CCC, what would be the proposed guidance?
Reviewer 4	PFOS concentrations in tissues are generally expected to change only gradually over time in response to environmental fluctuations. The chronic tissue-based criteria averaging periods, or duration components, were therefore specified as instantaneous, because tissue data provide

2.7. The Frequency and Duration of the Criterion Elements	
Reviewer	**Comments**
	point, or instantaneous, measurements that reflect integrative accumulation of PFOS over time and space in population(s) at a given site. It was appropriate for EPA to inform the recommended ten-year exceedance frequencies for the chronic tissue-based criteria given the large variation in possible biological and physical variable influencing ecological recovery.
Reviewer 5	In my opinion, the frequency and duration of criterion elements is among the most uncertain and potentially contentious elements of any type of protective criteria. The frequency and duration for tissue-based criteria is that the tissue-based criteria cannot be exceeded more than once in a 10 year period. This means that if the PFOS criterion for whole body in fish of 43 mg/kg bw is exceeded more than once in a 10 year period then the criteria is exceeded. This also means the fish was likely exposed to the 0.014 mg/l concentration for longer than an instantaneous exposure and likely longer than 4 days. So, to me, does this not mean that if 43 mg/kg bw was measured in a fish tissue, then the fish was likely exposed to 0.014 mg PFOS/L for longer than 4 days, doesn't it? And, this also means that if fish whole body concentrations were 42.5 mg/kg bw for 10 years, the criterion would not be exceeded. I would suspect that long term PFOS exposures that consistently lead to 42.5 mg/kg bw in fish would likely translate to adverse ecological impacts in some biota present in the same system. When I think of it this way, these criteria do not seem appropriately protective. In my view, the water column continuous exposure criteria should be adjusted downward which would then translate to a lower tissue-based criteria which might be more reasonable. Although, as mentioned, another protective approach would be to use something like the 25th percentile BAF or something other than the mean. That said, at least in many cases fish tissue monitoring occurs on a yearly basis so there is some potential for the criteria to be reasonably assessed against environmental data. It is still possible that fish tissue concentrations could be exceeded every year and this be missed by monitoring efforts. Nonetheless, because tissue concentrations are an intergrative measure and because many monitoring programs probably do measure fish every year, this is a better match than the water column criteria. When we consider the acute and chronic water column criteria, the frequency and duration elements are protective, in my opinion. The problem is that nobody knows if the criteria for acute toxicity are exceeded for more than 1 hour or whether the chronic criteria was exceeded for more than 4 days – this extent of temporal resolution (hourly concentrations or 4-day running averages) just does not exist. So while I agree that conceptually, the frequency and duration elements definitely would add to the protection of aquatic life…I just don't see how these can be implemented or regulated. Perhaps EPA is aware that my concern is not warranted because in relevant circumstances, appropriately timed environmental data are obtained.

2.8 Please provide any additional technical comments that you believe should be considered.

2.8. Additional Technical Comments to Consider	
Reviewer	**Comments**
Reviewer 1	I have the additional detailed comments: a) Please note that the comments provided in this file reflect a focus on of key portions of the "Draft of the Aquatic Life Water Quality Criterion…" document as directed by the above charge questions provided to me. Given time and resource constraints and the scope of my review, it was not feasible to provide a detailed review of the entire document and all of the supporting references and their associated results and conclusions. As such, I reserve my right to supplement or amend my comments in future, pending additional review or new information. Thank you for the opportunity to assist EPA in its work on this very important matter, and I was honored to be selected as a reviewer. b) In general, the document needs some quality copy editing effort. I found many typographical errors, issues with formatting, reference/citation issues, and in some cases, poorly-worded text. I have noted a few of these instances below. c) Page xv: "25.3 mg/L ww" is not correct units for a concentration in tissue. d) Page 6: "The carbon chain can be fully fluorinated…". Please specify that this applies to PFAS in general, not to PFOS. e) Page 6: The reference to "Table 2-1"; should that be Table 1-2? f) Page 9: Please note in Figure 2-1 that this is the linear isomer of PFOS. It would be helpful to note that the PFOS data in this study are likely from experiments with water spiked with the linear PFOS isomer. It is hypothesized that toxicity and bioaccumulation may differ between branched and linear forms of PFCAs and PFASs. Linear PFOS is thought to be more accumulative (as noted on Page 45) and potentially more toxic to aquatic life when the dose is expressed as an external water concentration. At some sites, a portion of the concentrations of PFOS in water (which are reported as the sum of branched and linear PFOS) is branched PFOS, so criteria derived from linear PFOS could be overly protective. Please include this uncertainty in the discussion in the document. g) Page 18: To my knowledge, FTSAs degrade to PFCAs, not PFSAs like PFOS. See Zhang et al. (2016): Zhang, S., Lu, X., Wang, N., & Buck, R. C. (2016). Biotransformation potential of 6:2 fluorotelomer sulfonate (6:2 FTSA) in aerobic and anaerobic sediment. Chemosphere, 154, 224–230. doi:10.1016/j.chemosphere.2016.03.062 h) Page 62: Regarding "The importance of the sediment pathway for PFOS bioaccumulation…" Larson et al. (2018) conducted some insightful food web modeling on benthic and pelagic sources of PFOS. See: Larson, E.S., Conder, J.M., Arblaster, J.A. 2018. Modeling avian exposures to perfluoroalkyl substances in aquatic habitats

2.8. Additional Technical Comments to Consider	
Reviewer	**Comments**
	impacted by historical aqueous film forming foam releases. https://doi.org/10.1016/j.chemosphere.2018.03.004 Chemosphere 201:335-341. i) Page 66: Starting here on this page and in the rest of this section, most of the units need to be specified for dry weight or wet weight for concentrations of PFOS in tissue. There were other instances of this error in the document as well. For units of every concentration of PFOS in tissue, please be sure to specify dry weight or wet weight. j) Page 73-75: There are a few scientific names on these pages that are not italicized. Also may occur in other portions of the document. k) Page 78: USEPA (1998) is not cited in the references section; I fear there may be other similar omissions. l) Page 78: Where the 1985 guidelines are mentioned, please cite to USEPA (1985). m) Page 78: Replace the "Stephan et al. (1985)" citation with USEPA (1985). Also, I believe the surname of senior author of USEPA (1985) is Stephen, not Stephan. n) Page 80: At the start of Section 2.10.2, it would be good to discuss linear and branched PFOS. o) Page 86: The use of EC10 values instead of effective concentration 20% (EC20) values for chronic values is inconsistent with EPA's general practice for developing aquatic life values. The selection of EC10s for the selenium criteria (EPA, 2016) was associated with the derivation of tissue guidelines. In the EPA (2016) document, EPA noted "EC20s have historically been used in the derivation of EPA criteria applicable to the water medium". As noted in the EPA (2016) selenium guidance EC10s were selected over EC20s "given the nature of exposure and effects for this bioaccumulative chemical." Additionally EPA (2016) selected EC10 for selenium because "it was found that the dose-response curves for selenium across a broad range of fish genera are very steep, such that a small change in selenium tissue concentration yielded a large increase in observed adverse effect." p) First, all the derivation of aquatic life criteria for "bioaccumulative chemicals" have not followed the process used for selenium, and there is no quantitative discussion in the current document that compares the bioaccumulation values for selenium to those of PFOS in a manner than justifies the use of EC10s. For example, EPA in its 2016 aquatic life criteria for cadmium noted that cadmium "can bioaccumulate in aquatic organisms", but EC20s (not EC10s) were used as chronic values in the derivation of aquatic life criteria in that document. Fundamentally, there is a logical disconnect between adding additional conservativism (i.e., using EC10s instead of EC20s) simply because a chemical has a higher bioaccumulative potential than another chemical or exceeds a BCF or BAF criteria used to determine a chemical has "bioaccumulative" status by typical chemical registration guidelines. The use of chronic exposure toxicology data generally assumes that concentrations in the organisms have reached steady state and, and thus, any bioaccumulation that has

2.8. Additional Technical Comments to Consider	
Reviewer	**Comments**
	occurred is accounted for and manifests in toxic action. Coincidentally, the general assumption is that toxic responses have plateaued as well and that effective doses (measured via external concentrations in water or concentrations in the organism) will not change significantly with additional exposure time. The bioaccumulative nature of the toxicant at that point is a moot point with regards to toxic effects in an aquatic organism, so there seems no need to add additional conservatism in the estimation of a threshold for potential ecologically-significant effects on aquatic life. Adding additional conservatism to the aquatic life criteria to protect other trophic levels (i.e. wildlife that consume aquatic life) or human consumers of aquatic life, which does involve bioaccumulation of chemicals in aquatic organisms, is not justified. Criteria to protect wildlife and humans exposed via exposure pathways involving bioaccumulation of chemicals in aquatic life are handled via separate approaches, and are completely disconnected from the acute and chronic toxicity data developed to evaluate the risks to aquatic invertebrates and lower trophic level vertebrates like fish and amphibians. q) Second, EPA has not provided any analysis of the dose response curves that demonstrates the need for EC10s versus EC20s (as was mentioned for selenium). Additionally, justification of the use of EC10s by simply referencing the regulatory policies of other countries seems to be insufficient as the basis for a US policy, and is unsatisfying from a scientific perspective. r) More discussion is needed to support the poorly-supported move from EC20s to EC10s, or alternately, EC20s need to be used in throughout the document, as consistent with past EPA practice in aquatic life criteria derivation. EC10s are more conservative than EC20s, but there is often greater variability and uncertainty associated with EC10 values given the typical 50% effect ranges that are generally targeted in the experimental designs of typical toxicological studies. Additionally, as noted in EPA's 2016 aquatic life criteria document for cadmium, EC10s are "rarely statistically significantly different from the control treatment." A 20% effect has often been discussed as a point of departure of ecologically-significant population- and community-level effects (e.g., Suter, 2000: Suter, G.W., Efroymson, R.A., Sample, B.E., & Jones, D.S. (2000). Ecological Risk Assessment for Contaminated Sites. CRC Press. April). s) Overall, the adoption of a more conservative 10% effect level (i.e., EC10) for chronic values used in criteria calculation carries large environmental management and policy implications. As noted above, clarification and careful justification is needed. EPA needs to clearly articulate (ideally with ample scientific support) why the additional conservatism is needed. This important potential policy matter deserves an open and earnest discourse among the scientific, stakeholder, and regulated communities. t) Page 88: It appears that only studies in which organisms exposed via diet were included for evaluation of tissue criteria. Is this correct? It is questionable to exclude effect concentrations in tissue from experiments in which exposure of PFAS was only via water. EPA (2016) took the "dietary exposure only" approach with selenium

2.8. Additional Technical Comments to Consider	
Reviewer	**Comments**
	because the primary exposure route for selenium has been shown to be via the diet in natural ecosystems. In contrast, for many aquatic animals (especially lower trophic level fish and invertebrates), a significant portion of the exposure to PFOS is via non-dietary pathways. Part of this is due to the fact that controlled studies (e.g., Martin et al., 2003 studies cited in the document) have found that water-to-organism BCFs for aquatic life such as fish are generally larger than diet-to-organism biomagnification factors (BMFs). Additionally, there is no reason to expect dietary or non-dietary exposure pathways would affect toxic responses given the relatively rapid internal kinetics of PFAS in aquatic life (i.e., half-life of hours or days), especially for small invertebrates and fish that are in relative equilibrium with their surrounding exposure water. u) Page 112: There's only one "Bots et al. 2010" in the references section. Multiple instances of "Bots et al 2010b" are cited in this document. I believe there is only one Bots et al. 2010 paper. Please clarify. v) Page 125: The Aedes data point is missing from Figure 3-5. If the qualitative Chironomus data point is included please include Aedes. w) Page 132: "expected to protect P. primulas from chronic time-variable PFOA exposures"… should that be "PFOS" instead? x) Page 135: A reference for "Appendix Q" is made. Please provide Appendix Q. y) The percentage effect for LOECs (relative to controls) needs to be clearly noted in the Appendices, for example, in Table C.1 and in the detailed summary text for the reviews of each paper. This should be provided when LOECs or MATCs are used as chronic values. z) Page 173: "Reduction in superoxide dismutase" and "Changes in protein expression" are atypical endpoints not well tied to ecologically significant effects. These should be removed from the table and subsequent discussion, or presented separately as qualitative analyses only. aa) Page 173: It is not appropriate to refer to the Gosner stage endpoint as "Length at metamorphosis" in the table. Refer to it as "Gosner stage" if it is to be included. bb) Page 176: Gosner stages are not a typical endpoint, and the use of the growth data would be much more supportable. See comment below regarding Page C-42. cc) Page C-2: When the MATC is used in tables in the Appendices, it would be helpful to provide the percent effect level (relative to control) for the LOEC associated with the MATC. Also, in cases in which the LOEC is provided as the chronic value, please provide the percent effect level. dd) Page C-3: How was the Species Mean Chronic Value (SMCV) for leopard frog calculated? There is only one chronic value that is bolded in the data, and it does not

2.8. Additional Technical Comments to Consider	
Reviewer	**Comments**
	equal the SMCV. Please add text to clearly discuss which values are included (and how the ">" values are used in subsequent calculations like geometric means). ee) Page C-7: For Wang et al., since this value is the lowest used in the criterion derivation, please share a table of the raw data graphed in the Figure on page C-7. ff) Page C-16: Appears to be a missing figure. gg) Page C-29: Typo "XX.XX mg/L". There are other typos like this in the document (search for "XX"). hh) Page C-42: The amount of detail for the review of the Hoover et al. (2017) experiment is insufficient. The selection of the Gosner stage as an endpoint requires additional detail. The relationship between Gosner stage and more typical endpoints clearly linked to ecological health (growth, reproduction, and survival is unclear). The effects on Gosner stage in this study are subtle; all dosed animals indicated they had reached tadpole stage (Gosner stages 25-41) at the 40-day endpoint noted. The maximum difference in Gosner stages noted in the study was approximately 2 (control Gosner stage result of ~30, 100 and 1000 µg/L Gosner stage results of ~28). A 7% difference in Gosner stages (especially when both 28 and 30 values fall within a tadpole Gosner stage development range) is difficult to translate to adverse ecological impact. As shown in the Gosner stage chart for anurans (Virginia Herpetological Society, http://www.virginiaherpetologicalsociety.com/amphibians/amphibian-development/amphibian-development.htm) the difference between stage 28 and stage 30 is the shape of the tail. It is unclear if this statistically detectable difference in the tail shape that distinguishes Gosner stage 28 and 30 would result in an ecologically significant decrease in the overall time period required to reach sexual maturity or ultimately translate to a developmental malformation that would result in an ecologically meaningful population-level effect (decrease in survival, decrease in reproductive output, etc.). The uncertainty with this atypical endpoint is high, and given the slight difference (~7%) between NOEC and LOEC exposures, I would recommend this datum be removed from the quantitative analysis. Notably, Figure S2 of this paper presents results for a measurement of growth via the Snout Vent Length (SVL) measurement endpoint. This endpoint provide more a continuous measurement of growth and is more typical of endpoints used in criteria derivation. ii) Page C-44: "In the later phases of the tests, (Bots et al. 2010a)" is repeated. jj) Page D-3: Regarding the Han et al (2015) study, I disagree with the selection of the less conservative growth endpoint. The reproductive effect does look to be valid and a reasonable endpoint to consider. EPA's reasoning to exclude it is not compelling and is unclear. The exposure duration was at least 10 days, which is likely sufficient for many marine invertebrates with relatively short life cycles (i.e., mysids). Perhaps EPA could reach out to the study authors to clarify the uncertainty around the exposure time (10 days or 20 days?). At any rate, I think the MATC should rely on the reproductive

<table>
<tr><th colspan="2">2.8. Additional Technical Comments to Consider</th></tr>
<tr><th>Reviewer</th><th>Comments</th></tr>
<tr><td></td><td>endpoint, and given the good dose-response for the reproductive data, a robust EC10 or EC20 value could likely be calculated.

kk) Page G-3: Seems like the Olson (2017) snail experiment provides some useful chronic (21-day exposure?) data for a relevant sublethal growth endpoints. Please explain why this data was excluded from the chronic evaluation. Simply listing "Duration" in this table does not provide enough detail.

ll) Page H-1: Please explain the acceptable duration acceptable for the urchin test and other tests. Simply listing "Duration too short" without noting the acceptable duration that would be considered is not helpful. Perhaps a summary table for acceptable durations for particular endpoints could be provided in this document.

mm) Page H-2: First use of "atypical duration" in the table. This entry is inconsistent with other entries (e.g., "duration too short") and does not clearly describe why the experiment is not considered. Please explain this table entry.

nn) Page P-12: The Hoover et al. (2017) paper is included twice. There may be more errors like this in the document, it needs to be reviewed closely by a technical editor.

oo) Appendix L: The references cited in this section seem to be missing.</td></tr>
<tr><td>Reviewer 2</td><td>I think the EPA's criteria for PFOS are very defensible based on the science and data available. I think they did a great job clearly laying out how they derived the criteria and providing all of the data that was used in the derivation.</td></tr>
<tr><td>Reviewer 3</td><td>All technical comments have been previously mentioned</td></tr>
<tr><td>Reviewer 4</td><td>Additional suggestions are listed below:

1. The species listed in the table is Mytilus galloprovincialis not M. edulis

Table 0-1. The Three Most Sensitive Acute Estuarine/Marine Genera.

Ranked Below from Most to Least Sensitive.
<table>
<tr><th>Rank</th><th>Genus</th><th>Species</th><th>GMAV (mg/L PFOS)</th><th>Comments</th></tr>
<tr><td></td><td>Mytilus[1]</td><td>Mediterranean mussel,
M. edulis
Mytilus galloprovincialis</td><td>> 1</td><td>Not a resident species in North America, but other species in this genus are resident, commercially, or ecologically important species</td></tr>
</table></td></tr>
</table>

2.8. Additional Technical Comments to Consider	
Reviewer	**Comments**
	2. Page 115- second paragraph (values highlighted in red and underlined are not consistent) The author reported 10-day growth and survival EC_{10}s for the study were 0.0492 and 0.1079 mg/L, respectively. The study authors also reported NOECs of 0.0491 mg/L, LOECs of 0.0962 mg/L, and MATCs of 0.0687 mg/L for both endpoints. And the author reported 20-day EC_{10}s for growth, survival, and total emergence were 0.0882, 0.0864, and 0.0893 mg/L, respectively. And the study authors also reported NOECs of 0.0217 mg/L for growth and survival and < 0.0023 mg/L for emergence, LOECs of 0.0949 mg/L for growth and survival and 0.0217 mg/L for emergence, and MATCs of 0.0454 mg/L for growth and survival and 0.0071 mg/L for emergence. Also, it should be noted, the paper reported contrasting NOECs for 20-day survival. The text in the paper stated that the NOEC was 0.0271 mg/L and Table 2 of the paper stated 0.0949 mg/L. EPA assumed the NOEC in Table 2 of the paper was not correct and that 0.0217 mg/L was the correct NOEC based on the data presented in Figure 3A of the paper. This assumption was applied to the summary of the study results presented in this PFOS draft criteria. 3. Page 138-middle of the paragraph The chronic freshwater criteria also contain tissue-based criteria expressed as 43.0 mg/kg wet weight (ww) for fish whole-body, 25.3 mg**/** --- ww for fish muscle tissue and 12.3 mg/kg ww for invertebrate whole-body tissue. 4. Page A-21 last paragraph The noted toxicity values provided in each study summary above (ADD NUMBERS), comprising of both author-reported and independently-calculated LC_{50} values, were used to calculate the GMAV value (as the geometric mean of the three LC_{50} values previously mentioned) of 22.48 mg/L, which was used to derive the freshwater aquatic life criterion. 5. Page A-24- Fourth line from bottom-The study author reported LC50 was 22.2 ± 4.6 mg/L for PFOS. The independently-calculated toxicity value was x.xx mg/L. The study author reported value was used quantitatively to derive the draft acute water column criterion. 6. Page A-25- Fourth line from bottom The study author reported 96-hour LC_{50} was 50.51 mg/L PFOS. The independently-calculated toxicity value was x.xx mg/L. The study author reported value was used quantitatively to derive the draft acute water column criterion. 7. Page A-27- Fourth line from bottom. The independently-calculated toxicity value was x.xx mg/L. The study author reported value was used quantitatively to derive the draft acute water column criterion. 8. Page A-29- First paragraph For comparison, the 7-day LC50 was 39.71 mg/L. The independently-calculated toxicity value was x.xx mg/L. The 96-hour study author reported value was used quantitatively to derive the draft acute water column criterion. 9. Page A-30- First paragraph The independently-calculated toxicity value was x.xx mg/L. The study author reported value was used quantitatively to derive the draft acute water column criterion. 10. Page A-36- 5th line from bottom in complete data x.xx mg/L.

2.8. Additional Technical Comments to Consider	
Reviewer	**Comments**
	11. Also at A-37 in complete data x.xx mg/L.

Table 0-3. Summary of Assessment Endpoints and Measures of Effect Used in the Criteria

Assessment Endpoints for the	
Aquatic Life: Survival, growth, and reproduction of freshwater and estuarine/marine aquatic life (i.e., fish, amphibians, aquatic invertebrates)	For effects from acute exposure: 1. LC_{50} concentrations in water, diet, and/or tissue (e.g.,) 2. NOEC and LOEC concentrations in water, diet, and/or tissue (e.g.,) For effects from chronic exposure: 1. EC_{10} concentrations in water, diet, and/or tissue (e.g.,) 2. NOEC and LOEC concentrations in water, diet, and/or tissue (e.g.,); *Only used when an EC_{10} could not be calculated for a genus.* *Note: only chronic exposures were considered for derivation of the tissue-based criteria since PFOS is a bioaccumulative chemical. These chronic tissue-based criteria are expected to be protective of acute effects, because acute effects were observed at much greater concentrations than chronic effects.*

Please review if the highlighted muscle, blood and egg would be relevant to this section in terms of LC50, EC10, LOEC and NOEC endpoints .

12. 1.1.2 and page 3- Previously Published Chronic Water Criteria for Direct Aqueous Exposure

The information on Australian guidelines to be updated based on NEMP2 published in 2020. I will attach it as a PDF. https://www.environment.gov.au/system/files/resources/2fadf1bc-b0b6-44cb-a192-78c522d5ec3f/files/pfas-nemp-2.pdf

> *"Previously published freshwater chronic values were available for two states (Minnesota and Michigan) and three countries or geographic regions (Australia/New Zealand, Canada, and Europe). These publicly available values for other jurisdictions were 0.019 mg/L and 0.14 mg/L for Minnesota (STS/MPCA 2007) and Michigan (EGLE 2010), respectively, and were ; EPAV 2016), 0.00680 mg/L in Canada (ECCC 2018), and 0.000023 mg/L in Europe (RIVM 2010). Previously published estuarine/marine chronic values were available for two geographic regions (Australia/New Zealand and Europe). These publicly available values were 0.0000046 mg/L in Europe (RIVM 2010) and*

2.8. Additional Technical Comments to Consider	
Reviewer	**Comments**
	The CRC marine guidelines are not valid as they are not based on the framework Freshwater values are to be used on an interim basis

13. Page 4- Table 1.1 to be updated accordingly

Exposure scenario	PFOS	Exposure scenario	Comments and source
Freshwater	0.00023 µg/L	99% species protection - high conservation value systems	Australian and New Zealand Guidelines for Fresh and Marine Water Quality - technical draft default guideline values for PFOS and PFOA. Note 1: The 99% species protection level for PFOS is close to the level of detection. Agencies may wish to apply a 'detect' threshold in such circumstances rather than a quantified measurement. Note 2: The draft guidelines do not account for effects which result from the biomagnification of toxicants in airbreathing animals or in animals which prey on aquatic organisms. Note 3: The WQGs advise[a] that the 99% level of protection be used for slightly to moderately disturbed systems. This approach is generally adopted for chemicals that bioaccumulate and biomagnify in wildlife. Regulators may specify or environmental legislation may prescribe the level of species protection required, rather than allowing for case by-case assessments.
	0.13 µg/L	95% species protection - slightly to moderately disturbed systems	
	2 µg/L	90% species protection - highly disturbed systems	
	31 µg/L	80% species protection - highly disturbed systems	

2.8. Additional Technical Comments to Consider

Reviewer	Comments

Exposure scenario	PFOS	Exposure scenario	Comments and source
Interim marine	0.00023 µg/L	99% species protection - high conservation value systems	As above. Freshwater values are to be used on an interim basis until final marine guideline values can be set using the nationally-agreed process under the Australian and New Zealand Guidelines for Fresh and Marine Water Quality. The WQG advise that in the case of estuaries, the most stringent of freshwater and marine criteria apply, taking account of any available salinity correction. Marine guideline values developed by CRC CARE are under consideration through the nationally-agreed water quality guideline development process.
	0.13 µg/L	95% species protection - slightly to moderately disturbed systems	
	2 µg/L	90% species protection - highly disturbed systems	
	31 µg/L	80% species protection - highly disturbed systems	

[a]https://www.waterquality.gov.au/anz-guidelines/guideline-values/default/water-quality-toxicants/local-conditions#bioaccumulation

14. Table 1 2. Two Primary Categories of PFAS
Please refer to OECD 2021 to be consistent with PFAS terminology/nomenclature

OECD (2021), Reconciling Terminology of the Universe of Per- and Polyfluoroalkyl Substances: Recommendations and Practical Guidance, OECD Series on Risk Management, No. 61, OECD Publishing, Paris

15. Table 1.3 Page 8
Please review Figure 9 OECD 2021 (also attached as PDF)

16. Conceptual Model of PFOS in the Aquatic Environment and Effects
Figure 2.9 page 77- Growth as an endpoint missing in the endpoints – first pentagon

Reviewer 5	Specific comments to the various elements of the PFOS AWQC are above. Here, I want to suggest that EPA revisit the 1985 Guidelines and publish either an updated version or an

2.8. Additional Technical Comments to Consider	
Reviewer	**Comments**
	amendment. Basing critically important criteria on documents published in 1985 and then using this to justify decisions seems like it would not pass muster in the scientific community. I've had papers rejected because they did not include enough recent citations, for example. Moreover, I've mentioned my concerns with the 4- most sensitive taxa + linear regression for criteria derivation. No paper I've read on generating the 5th percentile most sensitive species has used this approach. Granted, I may have missed them but my sense is that it is more common to use a full SSD. It would be helpful, for example, if the revised Guidelines explored this further or other means of criteria development (including new approach methods) and published, used, and cited and updated guidelines document. I'd like to think we still generally lead the world (more or less) in environmental protection so having an updated document would be welcomed.

APPENDIX A

CHARGE TO REVIEWERS

Technical Charge to External Peer Reviewers

Contract No. EP-C-17-017
Task Order 68HERH21F0090 (ERG Task 49)
June 2021

External Peer Review of EPA's Draft

Aquatic Life Water Quality Criterion for Perfluorooctane Sulfonate (PFOS)

BACKGROUND

PFOA and PFOS toxicity studies have been conducted on a limited number of aquatic organisms, including species of fish and invertebrates, indicating that exposure to elevated concentrations of PFOA and PFOS can cause effects on survival, growth, and reproduction. In these draft criteria documents EPA is proposing two separate water quality criteria to ensure the protection of aquatic life species from the exposure to PFOA and PFOS individually.

Background on the PFOA and PFOS Aquatic Life Criteria Development Process:

Toxicity studies used to derive the PFOA and PFOS criteria were carefully evaluated and thoroughly reviewed to ensure studies were of sufficient data quality to use in criteria derivation. Scientists from the Office of Water (OW) and Office of Research and Development (ORD) conducted an extensive review of the PFOA and PFOS toxicity studies. Additionally, EPA obtained replicate-level (or treatment-level, when replicates were unavailable) concentration-response (C-R) data from publications, supplemental materials, or via contacting authors so that EPA could independently fit C-R models to estimate acute LC50 and chronic EC10 values that were used to derive the criteria to ensure endpoints used were statistically sound. Individual C-R models and resultant point estimates were also reviewed and discussed between OW and ORD to ensure the most statistically robust models informed the derivation of the PFOA and PFOS criteria. In addition to contacting study authors for C-R data (when not reported in the open literature), EPA also consulted primary authors for methods clarifications in many instances during the data quality review phase to ensure that the studies used to derive criteria were of high quality.

Overall, due to the paucity of measured freshwater toxicity data, EPA included a number of tests with unmeasured treatments to derive criteria to ensure the dataset was representative of a range of taxa and there were sufficient data to develop criteria. EPA also conducted meta-analyses to evaluate the relationship between nominal and measured test concentrations using tests with measured treatment concentrations. These meta-analyses (described in detail as Appendix L of the PFOA criteria document and Appendix O of the PFOS criteria document) suggested measured concentrations were similar to nominal concentrations and that the use of unmeasured tests, in light of data limitations, was appropriate. Additionally, estuarine/marine toxicity data limitations did not allow for the direct derivation of acute or chronic estuarine/marine criteria for PFOA or PFOS. Therefore, to develop of recommendations that states and tribes could use in adopting protective values for estuarine/marine waters, EPA developed acute PFOA and PFOS protective benchmarks using a New Approach Methodology (detailed in Appendix K of the PFOA criteria document and Appendix L of the PFOS criteria document).

Addressing data limitations to derive robust criteria/benchmarks, extensively reviewing studies, and calculating point estimates meant that the derivation of the PFOA and PFOS aquatic life criteria were developed via comprehensive, rigorous process that included collaborations across EPA scientists in OW and

ORD. Beyond detailed discussions between OW and ORD, the PFOA and PFOS drafts also underwent two rounds of review with the EPA Scoping Workgroup (consisting of additional scientists from both OW and ORD) and one round of review with a group of internal EPA Reviewers that included representatives from the Office of Water, Office of Research and Development, other EPA Program Offices, and EPA Regions. In this peer review EPA is seeking to obtain a focused, objective evaluation of the two separate draft criteria documents, one for PFOA and the other for PFOS. Generally, the charge questions below are the same for EPA's PFOA and PFOS draft aquatic life water quality criteria. However, there are some unique questions specific to the individual drafts and therefore, there are two separate sets of charge questions.

CHARGE QUESTIONS

PFOS

Charge Questions for the draft Aquatic Life Water Quality Criterion for Perfluorooctane Sulfonate (PFOS) recommendations document

1) Please comment on the overall clarity of the document as it relates to the derivation of each criterion.

2) Please comment on the approach used to derive the draft criterion for PFOS. Please provide detailed comments.
    - Is the technical approach used to derive the criterion elements logical?
    - Does the science support the conclusions?
    - Is it consistent with the protection of freshwater aquatic life from acute, chronic, and bioaccumulative effects?

3) Please comment on the approach used to derive the draft acute estuarine/marine benchmark for PFOS. Given the limited estuarine/marine test data available, a new approach method was used to support the derivation of an acute estuarine/marine benchmark to provide states and tribes with a protective value. Please provide detailed comments.
    - Is the technical approach used to derive the benchmark logical?
    - Does the science support the conclusions?
    - Is it consistent with the protection offered by acute estuarine/marine aquatic life criteria derived using empirical data, as prescribed in the 1985 *Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*?

4) Please comment on the use of measured and unmeasured toxicity tests to derive the respective criterion. In particular please comment on the supporting justification for using unmeasured toxicity tests in Appendix O.

5) Please comment on the toxicity data used to derive the draft criteria.
    - Were the data selected and/or excluded from the derivation of the criteria derivation appropriately utilized?
    - Are there relevant data that you are aware of that should be added to the analyses (note that EPA is working on updating the toxicity data to reflect the data in ECOTOX between Sept. 2019 through the latest update)? If so, please provide references for consideration.

In particular, please comment on:

5a. The toxicity values used to derive the PFOS criteria, with a particular emphasis on:

i. the use of the qualitatively acceptable acute midge (*Chironomus plumosus*) data from Yang et al. (2014) to suggest aquatic insects are relatively tolerant to acute PFOS exposures. Specifically, Yang et al. (2014) conducted a 96-hour renewal, measured PFOS acute test with the midge, *Chironomus plumosus*. This study was not acceptable for quantitative use due to the potential problematic source of the organisms. The reported LC_{50} was 182 mg/L for PFOS indicating that insects may not be one of the more sensitive taxonomic groups. Therefore, this test was excluded from the acute criterion calculation, but used to waive the missing insect MDR.

ii. the use of the quantitatively acceptable chronic toxicity value for mussel (*Lampsilis siliquoidea*) from Hazelton et al. (2012). Specifically, Hazelton et al. (2012) conducted a 36-day renewal, measured PFOS chronic test with fatmucket, *Lampsilis siliquoidea.* The estimated EC_{10} was 0.05713 mg/L, which was extrapolated from the author-reported data and the exposure response slope from another PFOS toxicity study focused on another mussel species (*Ellipto complamata*) as explained in Section 3.1.1.3.3. Therefore, this test was used in the chronic criterion calculation.

iii. the use of the quantitatively acceptable chronic toxicity value for damselfly (*Enallagma cyathigerum*) from Bots et al. (2010). Bots et al. (2010) conducted a 320-day renewal, unmeasured PFOS chronic test with blue damselfly nymphs, *Enallagma cyathigerum.* The MATC was 0.03162 mg/L, which was calculated from the author-reported value for nymph survival as explained in Section 3.1.1.3.2. Therefore, this test was used in the chronic criterion calculation.

iv. the use of the quantitatively acceptable chronic toxicity value for midge (*Chironomus dilutus)* from MacDonald et al. (2004). MacDonald et al. (2004) conducted a 20-day renewal, measured PFOS chronic test with midge lava, *Chironomus dilutus.* The EC_{10} was 0.05963 mg/L, which was an EPA-calculated value for 10-day growth as explained in Section 3.1.1.3.4. Therefore, this test was used in the chronic criterion calculation.

5b. EPA's approach for fitting concentration-response (C-R) data (described in Appendix K) as well as the specific acute LC_{50} values (Appendix A.2) and chronic EC_{10} values (Appendix C.2) that were estimated (for sensitive genera when C-R data were available) and used to derive criteria.

6) Please comment on the translation of the chronic water column criterion elements for aquatic life to derive the tissue-based criterion elements, considering the bioaccumulation of PFOA and PFOS. In particular, please comment on:

6a. Uncertainty surrounding the bioaccumulation factors (BAFs) used to translate of the chronic water column criterion elements into tissue-based criterion elements.

6b. EPA's determination of appropriate BAFs and the tissue types that the tissue criterion elements were based.

7) Please comment on the frequency and duration of the criterion elements, in particular the tissue-based criterion elements.

8) Please provide any additional technical comments that you believe should be considered.

APPENDIX B

INDIVIDUAL REVIEWER COMMENTS

COMMENTS SUBMITTED BY

REVIEWER 1

External Peer Review of EPA's Draft Aquatic Life Water Quality Criterion for Perfluorooctane Sulfonate (PFOS)

1. Please comment on the overall clarity of the document as it relates to the derivation of each criterion.

Overall, the document is clear and the reader can follow the logic of criteria derivation, and track the values used back to the cited research articles or values calculated by EPA.

2. Please comment on the approach used to derive the draft criterion for PFOS. Please provide detailed comments.

- **Is the technical approach used to derive the criterion elements logical?**

  Yes, the technical approach used to derive the criteria elements is generally logical. I disagree with some of the elements of the analyses, as noted in my detailed comments (see below, responses to charge question 8).

- **Does the science support the conclusions?**

  In general, the science is supportive of the general conclusions. As noted in my below detailed responses to other charge questions, I believe the science is not supportive of the work in a few key instances including:

  1. I believe the Criterion Continuous Concentration (CCC) should be potentially re-calculated considering my comments provided in response to charge question 5a.
  2. The science does not support the assumption of a 10-year recovery time for PFOS in aquatic systems.
  3. The generation of tissue criteria is weakly supported, and the uncertainty associated with these criteria should be emphasized.
  4. The NAM-generated marine Final Acute Value (FAV) and FAV/2 values (Appendix L) are highly uncertain.
  5. It is unclear if the EPA-calculated Effective Concentration 10% (EC10) values are supported; additional details on the modeling and the variability and fit of each EC10 model need to be provided.

- **Is it consistent with the protection of freshwater aquatic life from acute, chronic, and bioaccumulative effects?**

  The criteria derived are aimed at protecting aquatic life (e.g., fish, invertebrates) from the direct acute and chronic toxicity of PFOS in water. Generally, the values applied are protective and are generally similar to protective values derived by other regulatory organizations and independent (i.e., academic, private sector) scientists. Although, as based on my comments, I believe there is room for improvement. The criteria derived for tissues attempt to provide criteria that take into account bioaccumulation so that measurements in tissue can be interpreted with respect to the potential for potential effects; however, the uncertainty with the tissue criteria is high. The water and tissue criteria are not intended protective of bioaccumulative effects that may affect higher trophic levels, such as wildlife that may consume aquatic life.

3. Please comment on the approach used to derive the draft acute estuarine/marine benchmark for PFOS. Given the limited estuarine/marine test data available, a new approach method was used to

support the derivation of an acute estuarine/marine benchmark to provide states and tribes with a protective value. Please provide detailed comments.

- **Is the technical approach used to derive the benchmark logical?**

  The derivation of the acute marine benchmarks (FAV and Criterion Maximum Concentration (CMC)) using the New Approach Method (NAM) is highly uncertain, and I would recommend this analysis not be included as in this document. I do not feel that the analysis and subsequent criteria have high confidence for use in a regulatory application. I understand that similar analyses with other chemicals have about a 90% probability of the predicted effect value being within a factor of 5 of the actual value (Raimondo et al., 2010 – cited in document). Given the calculated CMC (0.43 mg/L), this implies the CMC has about a 90% probability of being within 0.086 to 2.2 mg/L. If the NAM approach stays in the document, this uncertainty and range of values should be acknowledged in the discussion.

  I would rather see tentative or provisional acute criterion developed from the limited empirical marine acute data highlighted in Appendix B and other recently published marine acute data. This suggests a reasonable interim FAV of approximately 1 mg/L, which is similar to that calculated using the NAM approach. I place higher confidence in the limited empirical data and would suggest EPA emphasize it in addition to or in place of the values calculated by the NAM.

  I am hopeful that as new toxicity information on marine species are developed, these values can be supplanted with a proper and robust criteria calculation. If such a future analysis is possible, it should be noted.

- **Does the science support the conclusions?**

  See above comment.

- **Is it consistent with the protection offered by acute estuarine/marine aquatic life criteria derived using empirical data, as prescribed in the 1985 Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses?**

  The approach seems to be consistent with the approach in the 1985 guidelines. As noted above, the uncertainty with regards to the predictive capability of the interspecies correlations should be acknowledged quantitatively.

4. Please comment on the use of measured and unmeasured toxicity tests to derive the respective criterion. In particular please comment on the supporting justification for using unmeasured toxicity tests in Appendix O.

The consideration of toxicity data from experiments in which PFOS measurements were not made seems appropriate. The Appendix O analysis is supportive of the general observation that actual concentrations in the toxicity test waters approximated nominal values for freshwater. I agree that actual concentrations in the toxicity test waters for the marine test may be lower than nominal values, thus, effect data originating from marine studies that only report nominal concentrations may be biased high in some cases. Given the tentative/temporary nature of the marine criteria developed in this study, this bias is manageable until additional empirical data from experiment with measured concentrations in water can be provided.

5. Please comment on the toxicity data used to derive the draft criteria.

- **Were the data selected and/or excluded from the derivation of the criteria derivation appropriately utilized?**

In most cases, yes. Please see detailed comments on particular studies and interpretations in response to other charge questions.

- **Are there relevant data that you are aware of that should be added to the analyses (note that EPA is working on updating the toxicity data to reflect the data in ECOTOX between Sept. 2019 through the latest update)? If so, please provide references for consideration.**

  Hayman, N.T., Rosen, G., Colvin, M.A., Conder, J., Arblaster, J.A. 2021. Aquatic toxicity evaluations of PFOS and PFOA for five standard marine endpoints. Chemosphere 273:129699.

In particular, please comment on:

5a. The toxicity values used to derive the PFOS criteria, with a particular emphasis on:

i. the use of the qualitatively acceptable acute midge (*Chironomus plumosus*) data from Yang et al. (2014) to suggest aquatic insects are relatively tolerant to acute PFOS exposures. Specifically, Yang et al. (2014) conducted a 96-hour renewal, measured PFOS acute test with the midge, *Chironomus plumosus*. This study was not acceptable for quantitative use due to the potential problematic source of the organisms. The reported LC50 was 182 mg/L for PFOS indicating that insects may not be one of the more sensitive taxonomic groups. Therefore, this test was excluded from the acute criterion calculation, but used to waive the missing insect MDR.

I disagree with excluding this data point from the acute criteria calculations. I assume this data has been removed under the assumption that these animals may have been pre-exposed to PFOS and may have been more tolerant of PFOS exposures, which would result in biased-high median lethal concentration (LC50) values. If so, this should be explicitly stated. Assuming these *Chironomus* can develop tolerance to PFOS, it seems that they would have to be exposed to rather high mg/L ranges of PFOS in water given the reported 96-hour LC50 of 182 mg/L. Based on published literature, I am unaware of natural ecosystems in China (where the animals may have been originally harvested) with concentrations of PFOS that approach this order of magnitude range (in which they could build up a tolerance). The animals were obtained from a local market, so it is also possible that they were cultured for several generations, presumably using uncontaminated water (which would further reduce the chance that multiple generations were exposed at these levels). Overall, I think it is more reasonable to assume that the animals used in the experiment have not built up an acute lethal tolerance to PFOS, and the that LC50 result is unbiased. It does seem clearly show that insects may be less sensitive to acute lethality effects of PFOS. As such, I think it should be included as a quantitative endpoint.

Additionally, it seems inconsistent to exclude this Yang et al (2014) study, when chronic data from an unpublished study by Funkhouser (2015) were included for quantitative consideration. As noted on page C-25, the animals in the Funkhouser (2015) study were "purchased from a private collector" and then kept for "several" generations prior to testing. The source of the animals is just as uncertain as the Yang et al (2014) animals, and it is unclear (if PFOS tolerance at lethal levels is possible) how many generations would be needed to shed adaptive tolerance and how it would compare to "several." Simply put, if data from experiments like Funkhouser (2015) are quantitatively included, those from Yang et al. (2014) should also be quantitatively included (with some notes on the uncertainty of the animal sources).

ii. **the use of the quantitatively acceptable chronic toxicity value for mussel (*Lampsilis siliquoidea*) from Hazelton et al. (2012). Specifically, Hazelton et al. (2012) conducted a 36-day renewal, measured PFOS chronic test with fatmucket, *Lampsilis siliquoidea*. The estimated EC10 was**

0.05713 mg/L, which was extrapolated from the author-reported data and the exposure response slope from another PFOS toxicity study focused on another mussel species (*Ellipto complamata*) as explained in Section 3.1.1.3.3. Therefore, this test was used in the chronic criterion calculation.

There were only three exposure levels in this experiment, including the control. One PFOS dose (4.5 µg/L) indicated an absence of detectable effects on metamorphosis, the other (69.5 µg/L) indicated an approximate 35% reduction relative to controls. This is not a definitive test; there is little dose response information to fully confirm the effects/absence of effects and predict an effective concentration (EC) value with a dose response model. Application of another study's dose response curve to generate EC10 values for this study does not address this fundamental shortcoming, and simply carries too much uncertainty. Although there are only two PFOS doses, which is highly uncertain, use of a Maximum Acceptable Toxicant Concentration (MATC) value may be a less uncertain path to including this study in quantitative calculations. This would result in a more conservative chronic value for this study (0.018 mg/L instead of 0.057 mg/L). Given the high uncertainty of using this result (due to only 2 PFOS doses), I believe this value should be caveated in some way and re-evaluated for use or excluded in future criteria derivation. For example, on page C-22, the Spachmo and Arukwe (2012) value (which also featured a limited PFOS dose design), the document notes that the limited doses "may limit its future use in the criteria derivation pending independent verification of the toxicity values by EPA."

iii. **the use of the quantitatively acceptable chronic toxicity value for damselfly (*Enallagma cyathigerum*) from Bots et al. (2010). Bots et al. (2010) conducted a 320-day renewal, unmeasured PFOS chronic test with blue damselfly nymphs, *Enallagma cyathigerum*. The MATC was 0.03162 mg/L, which was calculated from the author-reported value for nymph survival as explained in Section 3.1.1.3.2. Therefore, this test was used in the chronic criterion calculation.**

I agree with the interpretation of the Bots et al. (2010) study and selection of the MATC.

iv. **the use of the quantitatively acceptable chronic toxicity value for midge (*Chironomus dilutus*) from MacDonald et al. (2004). MacDonald et al. (2004) conducted a 20-day renewal, measured PFOS chronic test with midge lava, *Chironomus dilutus*. The EC10 was 0.05963 mg/L, which was an EPA-calculated value for 10-day growth as explained in Section 3.1.1.3.4. Therefore, this test was used in the chronic criterion calculation.**

First, on page 104, the document mentioned "an EC10 of 0.0586 mg/L for growth following 10-days of exposure", but on page 115, the document noted "10-day growth with an EC10 of 0.05963 mg/L". In Appendix C (page C-19), the document states "the independently-calculated 10-day EC10 for growth was 0.0586 mg/L." There's some inconsistencies in the value being obtained from EPA's EC10 modeling using this reference; please correct or clarify.

One justification for using the 10-day EC10 growth result rather than other results is a lack of being able to calculate EC10s. I do think the emergence results should not be discounted, however. EPA notes "as for the emergence endpoint, there was a lack of a concentration-response relationship and there were very similar levels of observed effects (which ranged between 42.6 and 50.1%) despite the more than nine-fold increase in the mid-range treatment concentrations (0.0023, 0.0144, 0.0217 mg/L, respectively)." The magnitude of the effect (relative to controls), and the fact that there were statistically-detectable differences from controls in some of these doses (0.0144, 0.0217 mg/L) seems to indicate an ecologically meaningful adverse effect is occurring due to PFOS. This range of concentrations just might be a portion of dose response curve that is relatively flat. There is a very clear adverse effect at 0.0949 mg/L. I think it would be reasonable to select the

MATC for emergence (0.0071 mg/L reported on page C-19) and treat it a second study point since it was a completely different experiment from the 10-day experiment used to provide the EC10 of 0.05963 (or 0.0586) mg/L value. A Species Mean Chronic Value using the 0.0071 and 0.0586 mg/L results would be 0.020 mg/L. This 0.020 mg/L value would seem to be protective while including the growth and emergence data from these two experiments.

5b. EPA's approach for fitting concentration-response (C-R) data (described in Appendix K) as well as the specific acute LC50 values (Appendix A.2) and chronic EC10 values (Appendix C.2) that were estimated (for sensitive genera when C-R data were available) and used to derive criteria.

More details need to be provided on the dose response modeling using R. Appendix K is helpful for providing the reader with details on the general approach, but where EC10s are modeled by EPA, the model being used (out of the 22 available in the R software package) needs to be specified. Providing some indication of variability (such as a 95% confidence interval) for the model-generated EC10s is standard practice for dose response modeling, and this information should be provided somewhere in the document. Showing the R package output of the goodness of fit statistics (or equivalent) for the modeling in an Appendix would be helpful; since this was used to select the model used in each instance of an EC10 calculation, it must be available, so I would recommend including it for full transparency and to aid future efforts in understanding the aquatic toxicology of this chemical. Additionally, it would be helpful to show the selected model fits for all calculated EC10s (as shown for the most sensitive EC10s estimated). These steps would be helpful to ensure and demonstrate quality of the model fits and reproducibility of the modeling work.

Additionally, somewhere in the document (Appendix K), the 22 dose response model equations should be provided to the reader. Alternately, a reference could be made to a document that clearly provides this information (ideally a peer-reviewed or EPA document) containing all 22 models.

6. Please comment on the translation of the chronic water column criterion elements for aquatic life to derive the tissue-based criterion elements, considering the bioaccumulation of PFOA and PFOS.

The derivation of the tissue criteria in this manner is highly uncertain. To my knowledge this is the first time EPA has applied ambient water quality criteria protective of aquatic life direct toxicity with uptake factors (bioaccumulation factors (BAFs), bioconcentration factors (BCFs)) in this manner to calculate tissue criteria. References are made to the selenium tissue criteria, but those are used in the reverse (i.e., criteria based on measured concentrations in tissue used to calculate water criteria). The use of criteria for water with a assumed uptake factor carries a large amount of uncertainty, and in general, the use of measured concentrations in tissue linked to adverse effects is a more straightforward approach since it does not involve uptake model predictions. This needs to be noted in the text. Also, are the predicted tissue criteria meant to be a temporary stop-gap until tissue effect data become available? This should be discussed and clarified.

In particular, please comment on:

6a. Uncertainty surrounding the bioaccumulation factors (BAFs) used to translate of the chronic water column criterion elements into tissue-based criterion elements.

The use of BAFs derived from field studies is inherently uncertain due to the wide variety of techniques used in the compiled studies, their analytical data quality, the differences in species and ecosystems, experimental designs, spatial uncertainties for mobile animals like fish, etc. That being said, the use of a BAF value (or BCF) in criteria derivation is consistent with other criteria developed by EPA. As noted above, the use of the tissue criteria needs to be considered carefully, and I think empirical tissue data from toxicity experiments should form the basis of a next iteration of a tissue criteria.

6b. EPA's determination of appropriate BAFs and the tissue types that the tissue criterion elements were based.

The development of BAFs for invertebrates, fish (whole body), and fish (muscle) seems reasonable for the application in estimating a draft or interim tissue criteria until empirical tissue data can be used to calculate tissue criteria directly.

7. Please comment on the frequency and duration of the criterion elements, in particular the tissue-based criterion elements.

The 4-day duration seems to be supported by the more sensitive chronic endpoints used to derive the CCC.

For the tissue-based criterion (page 135), there is no clear support for assuming a 10-year exceedance frequency. Given the uncertainty with the BAF-predicted tissue criteria, and how little is known regarding the recalcitrance of PFOS in aquatic ecosystems and recovery time if PFOS inputs in water were halted, the assignment of a 10-year exceedance frequency at this stage seems completely arbitrary. We simply do not yet know the time frame over which aquatic ecosystems recover from PFOS. It is not technically supported to cite recovery times for selenium to support a 10-year recovery time for PFOS, these are completely different toxicants that have their own unique fate and behavior. USEPA (1985) guidance suggests assuming a 3-year frequency as a default, and the discussion on page 135-136 is not scientifically convincing enough to modify it to 10 years.

Additionally, it should be noted that the exceedance frequency for another organic chemical, Tributyltin (TBT) was set at 3 years by EPA in derivation of that criteria. TBT exhibits uptake factors similar to PFOS (i.e., BCF of approximately 2,000 L/kg, wet weight for goldfish, as noted in the EPA TBT criteria document, which is similar to the PFOS BAFs of 1,800-3,100 L/kg, wet weight being used to calculate the fish tissue criteria). TBT is also persistent in aquatic ecosystems, as noted by EPA. Given TBT is at least an organic chemical, it is a closer analog than selenium, which is an element. As such, the exceedance frequency for the PFOS tissue criterion should be set at the default of 3 years unless EPA can provide convincing technical information specific to recovery times for PFOS.

Additionally, on page 136, the paragraph that begins with "Metals and other chemical pollutants such as PFOS..." is not convincing as any quantitative support for EPA's 10-year exceedance frequency for the chronic tissue-based criteria. The text as written may give the reader the conclusion that PFOS recovery may be "on the order of decades", as EPA notes for selenium. There is no support for the conjecture that PFOS recovery may be "relatively slow" or require decades, as noted in my above comment.

8. Please provide any additional technical comments that you believe should be considered.

I have the additional detailed comments:

a) Please note that the comments provided in this file reflect a focus on of key portions of the "Draft of the Aquatic Life Water Quality Criterion…" document as directed by the above charge questions provided to me. Given time and resource constraints and the scope of my review, it was not feasible to provide a detailed review of the entire document and all of the supporting references and their associated results and conclusions. As such, I reserve my right to supplement or amend my comments in future, pending additional review or new information. Thank you for the opportunity to assist EPA in its work on this very important matter, and I was honored to be selected as a reviewer.

b) In general, the document needs some quality copy editing effort. I found many typographical errors, issues with formatting, reference/citation issues, and in some cases, poorly-worded text. I have noted a few of these instances below.

c) Page xv: "25.3 mg/L ww" is not correct units for a concentration in tissue.

d) Page 6: "The carbon chain can be fully fluorinated…". Please specify that this applies to PFAS in general, not to PFOS.

e) Page 6: The reference to "Table 2-1"; should that be Table 1-2?

f) Page 9: Please note in Figure 2-1 that this is the linear isomer of PFOS. It would be helpful to note that the PFOS data in this study are likely from experiments with water spiked with the linear PFOS isomer. It is hypothesized that toxicity and bioaccumulation may differ between branched and linear forms of PFCAs and PFASs. Linear PFOS is thought to be more accumulative (as noted on Page 45) and potentially more toxic to aquatic life when the dose is expressed as an external water concentration. At some sites, a portion of the concentrations of PFOS in water (which are reported as the sum of branched and linear PFOS) is branched PFOS, so criteria derived from linear PFOS could be overly protective. Please include this uncertainty in the discussion in the document.

g) Page 18: To my knowledge, FTSAs degrade to PFCAs, not PFSAs like PFOS. See Zhang et al. (2016): Zhang, S., Lu, X., Wang, N., & Buck, R. C. (2016). Biotransformation potential of 6:2 fluorotelomer sulfonate (6:2 FTSA) in aerobic and anaerobic sediment. Chemosphere, 154, 224–230. doi:10.1016/j.chemosphere.2016.03.062

h) Page 62: Regarding "The importance of the sediment pathway for PFOS bioaccumulation…" Larson et al. (2018) conducted some insightful food web modeling on benthic and pelagic sources of PFOS. See: Larson, E.S., Conder, J.M., Arblaster, J.A. 2018. Modeling avian exposures to perfluoroalkyl substances in aquatic habitats impacted by historical aqueous film forming foam releases. https://doi.org/10.1016/j.chemosphere.2018.03.004 Chemosphere 201:335-341.

i) Page 66: Starting here on this page and in the rest of this section, most of the units need to be specified for dry weight or wet weight for concentrations of PFOS in tissue. There were other instances of this error in the document as well. For units of every concentration of PFOS in tissue, please be sure to specify dry weight or wet weight.

j) Page 73-75: There are a few scientific names on these pages that are not italicized. Also may occur in other portions of the document.

k) Page 78: USEPA (1998) is not cited in the references section; I fear there may be other similar omissions.

l) Page 78: Where the 1985 guidelines are mentioned, please cite to USEPA (1985).

m) Page 78: Replace the "Stephan et al. (1985)" citation with USEPA (1985). Also, I believe the surname of senior author of USEPA (1985) is Stephen, not Stephan.

n) Page 80: At the start of Section 2.10.2, it would be good to discuss linear and branched PFOS.

o) Page 86: The use of EC10 values instead of effective concentration 20% (EC20) values for chronic values is inconsistent with EPA's general practice for developing aquatic life values. The selection of EC10s for the selenium criteria (EPA, 2016) was associated with the derivation of tissue guidelines. In the EPA (2016) document, EPA noted "EC20s have historically been used in the derivation of EPA

criteria applicable to the water medium". As noted in the EPA (2016) selenium guidance EC10s were selected over EC20s "given the nature of exposure and effects for this bioaccumulative chemical." Additionally EPA (2016) selected EC10 for selenium because "it was found that the dose-response curves for selenium across a broad range of fish genera are very steep, such that a small change in selenium tissue concentration yielded a large increase in observed adverse effect."

p) First, all the derivation of aquatic life criteria for "bioaccumulative chemicals" have not followed the process used for selenium, and there is no quantitative discussion in the current document that compares the bioaccumulation values for selenium to those of PFOS in a manner than justifies the use of EC10s. For example, EPA in its 2016 aquatic life criteria for cadmium noted that cadmium "can bioaccumulate in aquatic organisms", but EC20s (not EC10s) were used as chronic values in the derivation of aquatic life criteria in that document. Fundamentally, there is a logical disconnect between adding additional conservativism (i.e., using EC10s instead of EC20s) simply because a chemical has a higher bioaccumulative potential than another chemical or exceeds a BCF or BAF criteria used to determine a chemical has "bioaccumulative" status by typical chemical registration guidelines. The use of chronic exposure toxicology data generally assumes that concentrations in the organisms have reached steady state and, and thus, any bioaccumulation that has occurred is accounted for and manifests in toxic action. Coincidentally, the general assumption is that toxic responses have plateaued as well and that effective doses (measured via external concentrations in water or concentrations in the organism) will not change significantly with additional exposure time. The bioaccumulative nature of the toxicant at that point is a moot point with regards to toxic effects in an aquatic organism, so there seems no need to add additional conservatism in the estimation of a threshold for potential ecologically-significant effects on aquatic life. Adding additional conservatism to the aquatic life criteria to protect other trophic levels (i.e. wildlife that consume aquatic life) or human consumers of aquatic life, which does involve bioaccumulation of chemicals in aquatic organisms, is not justified. Criteria to protect wildlife and humans exposed via exposure pathways involving bioaccumulation of chemicals in aquatic life are handled via separate approaches, and are completely disconnected from the acute and chronic toxicity data developed to evaluate the risks to aquatic invertebrates and lower trophic level vertebrates like fish and amphibians.

q) Second, EPA has not provided any analysis of the dose response curves that demonstrates the need for EC10s versus EC20s (as was mentioned for selenium). Additionally, justification of the use of EC10s by simply referencing the regulatory policies of other countries seems to be insufficient as the basis for a US policy, and is unsatisfying from a scientific perspective.

r) More discussion is needed to support the poorly-supported move from EC20s to EC10s, or alternately, EC20s need to be used in throughout the document, as consistent with past EPA practice in aquatic life criteria derivation. EC10s are more conservative than EC20s, but there is often greater variability and uncertainty associated with EC10 values given the typical 50% effect ranges that are generally targeted in the experimental designs of typical toxicological studies. Additionally, as noted in EPA's 2016 aquatic life criteria document for cadmium, EC10s are "rarely statistically significantly different from the control treatment." A 20% effect has often been discussed as a point of departure of ecologically-significant population- and community-level effects (e.g., Suter, 2000: Suter, G.W., Efroymson, R.A., Sample, B.E., & Jones, D.S. (2000). Ecological Risk Assessment for Contaminated Sites. CRC Press. April).

s) Overall, the adoption of a more conservative 10% effect level (i.e., EC10) for chronic values used in criteria calculation carries large environmental management and policy implications. As noted above, clarification and careful justification is needed. EPA needs to clearly articulate (ideally with

ample scientific support) why the additional conservatism is needed. This important potential policy matter deserves an open and earnest discourse among the scientific, stakeholder, and regulated communities.

t) Page 88: It appears that only studies in which organisms exposed via diet were included for evaluation of tissue criteria. Is this correct? It is questionable to exclude effect concentrations in tissue from experiments in which exposure of PFAS was only via water. EPA (2016) took the "dietary exposure only" approach with selenium because the primary exposure route for selenium has been shown to be via the diet in natural ecosystems. In contrast, for many aquatic animals (especially lower trophic level fish and invertebrates), a significant portion of the exposure to PFOS is via non-dietary pathways. Part of this is due to the fact that controlled studies (e.g., Martin et al., 2003 studies cited in the document) have found that water-to-organism BCFs for aquatic life such as fish are generally larger than diet-to-organism biomagnification factors (BMFs). Additionally, there is no reason to expect dietary or non-dietary exposure pathways would affect toxic responses given the relatively rapid internal kinetics of PFAS in aquatic life (i.e., half-life of hours or days), especially for small invertebrates and fish that are in relative equilibrium with their surrounding exposure water.

u) Page 112: There's only one "Bots et al. 2010" in the references section. Multiple instances of "Bots et al 2010b" are cited in this document. I believe there is only one Bots et al. 2010 paper. Please clarify.

v) Page 125: The Aedes data point is missing from Figure 3-5. If the qualitative Chironomus data point is included please include Aedes.

w) Page 132: "expected to protect P. primulas from chronic time-variable PFOA exposures"… should that be "PFOS" instead?

x) Page 135: A reference for "Appendix Q" is made. Please provide Appendix Q.

y) The percentage effect for LOECs (relative to controls) needs to be clearly noted in the Appendices, for example, in Table C.1 and in the detailed summary text for the reviews of each paper. This should be provided when LOECs or MATCs are used as chronic values.

z) Page 173: "Reduction in superoxide dismutase" and "Changes in protein expression" are atypical endpoints not well tied to ecologically significant effects. These should be removed from the table and subsequent discussion, or presented separately as qualitative analyses only.

aa) Page 173: It is not appropriate to refer to the Gosner stage endpoint as "Length at metamorphosis" in the table. Refer to it as "Gosner stage" if it is to be included.

bb) Page 176: Gosner stages are not a typical endpoint, and the use of the growth data would be much more supportable. See comment below regarding Page C-42.

cc) Page C-2: When the MATC is used in tables in the Appendices, it would be helpful to provide the percent effect level (relative to control) for the LOEC associated with the MATC. Also, in cases in which the LOEC is provided as the chronic value, please provide the percent effect level.

dd) Page C-3: How was the Species Mean Chronic Value (SMCV) for leopard frog calculated? There is only one chronic value that is bolded in the data, and it does not equal the SMCV. Please add text to clearly discuss which values are included (and how the ">" values are used in subsequent calculations like geometric means).

ee) Page C-7: For Wang et al., since this value is the lowest used in the criterion derivation, please share a table of the raw data graphed in the Figure on page C-7.

ff) Page C-16: Appears to be a missing figure.

gg) Page C-29: Typo "XX.XX mg/L". There are other typos like this in the document (search for "XX").

hh) Page C-42: The amount of detail for the review of the Hoover et al. (2017) experiment is insufficient. The selection of the Gosner stage as an endpoint requires additional detail. The relationship between Gosner stage and more typical endpoints clearly linked to ecological health (growth, reproduction, and survival is unclear). The effects on Gosner stage in this study are subtle; all dosed animals indicated they had reached tadpole stage (Gosner stages 25-41) at the 40-day endpoint noted. The maximum difference in Gosner stages noted in the study was approximately 2 (control Gosner stage result of ~30, 100 and 1000 µg/L Gosner stage results of ~28). A 7% difference in Gosner stages (especially when both 28 and 30 values fall within a tadpole Gosner stage development range) is difficult to translate to adverse ecological impact. As shown in the Gosner stage chart for anurans (Virginia Herpetological Society, http://www.virginiaherpetologicalsociety.com/amphibians/amphibian-development/amphibian-development.htm) the difference between stage 28 and stage 30 is the shape of the tail. It is unclear if this statistically detectable difference in the tail shape that distinguishes Gosner stage 28 and 30 would result in an ecologically significant decrease in the overall time period required to reach sexual maturity or ultimately translate to a developmental malformation that would result in an ecologically meaningful population-level effect (decrease in survival, decrease in reproductive output, etc.). The uncertainty with this atypical endpoint is high, and given the slight difference (~7%) between NOEC and LOEC exposures, I would recommend this datum be removed from the quantitative analysis. Notably, Figure S2 of this paper presents results for a measurement of growth via the Snout Vent Length (SVL) measurement endpoint. This endpoint provide more a continuous measurement of growth and is more typical of endpoints used in criteria derivation.

ii) Page C-44: "In the later phases of the tests, (Bots et al. 2010a)" is repeated.

jj) Page D-3: Regarding the Han et al (2015) study, I disagree with the selection of the less conservative growth endpoint. The reproductive effect does look to be valid and a reasonable endpoint to consider. EPA's reasoning to exclude it is not compelling and is unclear. The exposure duration was at least 10 days, which is likely sufficient for many marine invertebrates with relatively short life cycles (i.e., mysids). Perhaps EPA could reach out to the study authors to clarify the uncertainty around the exposure time (10 days or 20 days?). At any rate, I think the MATC should rely on the reproductive endpoint, and given the good dose-response for the reproductive data, a robust EC10 or EC20 value could likely be calculated.

kk) Page G-3: Seems like the Olson (2017) snail experiment provides some useful chronic (21-day exposure?) data for a relevant sublethal growth endpoints. Please explain why this data was excluded from the chronic evaluation. Simply listing "Duration" in this table does not provide enough detail.

ll) Page H-1: Please explain the acceptable duration acceptable for the urchin test and other tests. Simply listing "Duration too short" without noting the acceptable duration that would be considered is not helpful. Perhaps a summary table for acceptable durations for particular endpoints could be provided in this document.

mm) Page H-2: First use of "atypical duration" in the table. This entry is inconsistent with other entries (e.g., "duration too short") and does not clearly describe why the experiment is not considered. Please explain this table entry.

nn) Page P-12: The Hoover et al. (2017) paper is included twice. There may be more errors like this in the document, it needs to be reviewed closely by a technical editor.

oo) Appendix L: The references cited in this section seem to be missing.

COMMENTS SUBMITTED BY

REVIEWER 2

External Peer Review of EPA's Draft Aquatic Life Water Quality Criterion for Perfluorooctane Sulfonate (PFOS)

1) Please comment on the overall clarity of the document as it relates to the derivation of each criterion.

I thought that the document was well written and laid out. I thought that the document clearly laid out the approach that the EPA used to derive each criterion. I thought it clearly outlined the approach that the EPA chose in deciding which data to use in their derivation and how these data would be used in derivation.

The appendices are very useful in providing added detail and the data that were used in the derivation of the criteria. The appendices allow for a high level of transparency around how the criteria were generated.

In Table 3-1, the acronym "GMAV" was used as a heading in the table, but I could not locate where this acronym was defined earlier in the document.

The captions of figures and tables are not sufficiently detailed. Figures and tables should be able to stand on their own. Also, the use of acronyms in the caption and headings of tables and figures decreases clarity, e.g., Figs 3-1, Tables 3-1, 3-2, 3-3, 3-4, 3-5. The use of acronyms in the figure or table is valid to save space, as long as they are defined in the caption of the figure or table.

2) Please comment on the approach used to derive the draft criterion for PFOS. Please provide detailed comments.

- **Is the technical approach used to derive the criterion elements logical?**
- **Does the science support the conclusions?**
- **Is it consistent with the protection of freshwater aquatic life from acute, chronic, and bioaccumulative effects?**

Yes, the technical approach used by the EPA to derive the criterion is logical and defensible. The approach is also clearly laid out in the document. Dividing the 5th centile of the acute GSD by 2 is sufficiently conservative to ensure the protection of 95% of species, based on the data currently available.

Yes, I think the science supports the EPA's conclusions. However, there appears to be several studies that were not considered by the EPA. I have listed these studies below.

Yes, I think the approach taken by the EPA is sufficiently conservative to be protective of freshwater aquatic life from acute, chronic, and bioaccumulative effects based on the data that was available at the time. It was a good idea to evaluate the influence on non-North American species on the derivation of the criteria.

3) Please comment on the approach used to derive the draft acute estuarine/marine benchmark for PFOS. Given the limited estuarine/marine test data available, a new approach method was used to support the derivation of an acute estuarine/marine benchmark to provide states and tribes with a protective value. Please provide detailed comments.

- **Is the technical approach used to derive the benchmark logical?**
- **Does the science support the conclusions?**
- **Is it consistent with the protection offered by acute estuarine/marine aquatic life criteria derived using empirical data, as prescribed in the 1985 Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses?**

The technical approach using Web-ICE to determine an acute benchmark for estuarine/marine species is logical. The science has shown that Web-ICE can effectively be used to derive effect measures for additional

species using species for which data is available. I think the approach taken by EPA has included sufficient conservatism to address the relatively large amount of uncertainty around the acute toxicity of PFOS to estuarine and marine species. The proposed acute benchmark for estuarine and marine species is an order of magnitude lower than the acute benchmark for freshwater species, which I think underscores the conservatism used by EPA in determining an acute benchmark for estuarine and marine species. That said, the benchmark should be used cautiously due to the relatively large amount of uncertainty and effort should be made to generate acute and chronic toxicity data for PFOS on estuarine and marine species.

4) Please comment on the use of measured and unmeasured toxicity tests to derive the respective criterion. In particular please comment on the supporting justification for using unmeasured toxicity tests in Appendix O.

I am concerned with the approach of using the agreement of measured and nominal concentrations from studies that measured the concentration of PFOS in their tests to determine whether to use toxicity data from studies that did not measure the concentration PFOS in their tests. My concern stems from this approach having to assume that studies that did not measure the concentration of PFOS in their experiments performed the dosing of PFOS with the same care and skill as those studies that did measure the concentration of PFOS in their experiments and measured concentrations within 20% of nominal. My concern is compound by 58% and 65% of the freshwater and saltwater tests, respectively, only reporting nominal test concentrations. The EPA's approach uses the agreement of measured and nominal concentration in a minority of studies to determine whether to include the majority of studies on their assessment.

I am assuming that there wouldn't be sufficient data to determine a criterion without using data from studies that did not measure the concentrations of PFOS in their experiment?

I think the approach that the EPA has used to determine the level of agreement between the nominal and measured concentration of PFOS in the studies that measured the concentration is logical and valid. It is encouraging that the agreement on average is high. Again, my largest concern is assuming this agreement in a minority of studies is present in all studies.

5) Please comment on the toxicity data used to derive the draft criteria.

- **Were the data selected and/or excluded from the derivation of the criteria derivation appropriately utilized?**

I think the data used in the derivation of the criteria were appropriate. As mentioned above, I am a little concerned about the use of toxicity data from studies that did not measure the concentration of PFOS in their experiments, especially considering the proportion of studies that did not measure the concentrations. The confirmation of exposure concentrations is an important principle of sound ecotoxicology.

- **Are there relevant data that you are aware of that should be added to the analyses (note that EPA is working on updating the toxicity data to reflect the data in ECOTOX between Sept. 2019 through the latest update)? If so, please provide references for consideration.**

I have listed a number of papers below that were published in 2020 and 2021 that the EPA may want to consider in their assessment.

Hayman, N.T., Rosen, G., Colvin, M.A., Conder, J., Arblaster, J.A., 2021. Aquatic toxicity evaluations of PFOS and PFOA for five standard marine endpoints. Chemosphere 273, 129699. doi:10.1016/j.chemosphere.2021.129699

Logeshwaran, P., Sivaram, A.K., Surapaneni, A., Kannan, K., Naidu, R., Megharaj, M., 2021. Exposure to perfluorooctanesulfonate (PFOS) but not perflurorooctanoic acid (PFOA) at ppb concentration induces chronic toxicity in Daphnia carinata. Science of The Total Environment 769, 144577.. doi:10.1016/j.scitotenv.2020.144577

Simpson, S.L., Liu, Y., Spadaro, D.A., Wang, X., Kookana, R.S., Batley, G.E., 2021. Chronic effects and thresholds for estuarine and marine benthic organism exposure to perfluorooctane sulfonic acid (PFOS)-contaminated sediments: Influence of organic carbon and exposure routes. Science of The Total Environment 776, 146008.. doi:10.1016/j.scitotenv.2021.146008

Li, R., Tang, T., Qiao, W., Huang, J., 2020. Toxic effect of perfluorooctane sulfonate on plants in vertical-flow constructed wetlands. Journal of Environmental Sciences 92, 176–186.. doi:10.1016/j.jes.2020.02.018

Aquilina-Beck, A.A., Reiner, J.L., Chung, K.W., Delise, M.J., Key, P.B., Delorenzo, M.E., 2020. Uptake and Biological Effects of Perfluorooctane Sulfonate Exposure in the Adult Eastern Oyster Crassostrea virginica. Archives of Environmental Contamination and Toxicology 79, 333–342. doi:10.1007/s00244-020-00765-4

Tornabene, B.J., Chislock, M.F., Gannon, M.E., Sepúlveda, M.S., Hoverman, J.T., 2021. Relative acute toxicity of three per- and polyfluoroalkyl substances on nine species of larval amphibians. Integrated Environmental Assessment and Management 17, 684–690. doi:10.1002/ieam.4391

Suski, J.G., Salice, C.J., Chanov, M.K., Ayers, J., Rewerts, J., Field, J., 2021. Sensitivity and Accumulation of Perfluorooctanesulfonate and Perfluorohexanesulfonic Acid in Fathead Minnows (Pimephales promelas) Exposed over Critical Life Stages of Reproduction and Development. Environmental Toxicology and Chemistry 40, 811–819. doi:10.1002/etc.4936

Mccarthy, C.J., Roark, S.A., Wright, D., O'Neal, K., Muckey, B., Stanaway, M., Rewerts, J.N., Field, J.A., Anderson, T.A., Salice, C.J., 2021. Toxicological Response of Chironomus dilutus in Single-Chemical and Binary Mixture Exposure Experiments with 6 Perfluoralkyl Substances. Environmental Toxicology and Chemistry 40, 2319–2333. doi:10.1002/etc.5066

In particular, please comment on:

5a. The toxicity values used to derive the PFOS criteria, with a particular emphasis on:

i. the use of the qualitatively acceptable acute midge (Chironomus plumosus) data from Yang et al. (2014) to suggest aquatic insects are relatively tolerant to acute PFOS exposures. Specifically, Yang et al. (2014) conducted a 96-hour renewal, measured PFOS acute test with the midge, Chironomus plumosus. This study was not acceptable for quantitative use due to the potential problematic source of the organisms. The reported LC_{50} was 182 mg/L for PFOS indicating that insects may not be one of the more sensitive taxonomic groups. Therefore, this test was excluded from the acute criterion calculation, but used to waive the missing insect MDR.

I think the EPA's decision that the data from Yang et al. (2014) was not acceptable for quantitative use was appropriate. The source of the larvae is problematic. However, I don't agree with the conclusion that insects may not be one of the most sensitive taxa. Chironomus tentans is a relatively sensitive taxa to chronic exposure to PFOS (MacDonald et al. 2004). In Table C.1, the EC10 for C. tentans is reported as 0.05963 mg/L. Chironomus tentans was also the fourth most sensitive species used in calculating the chronic freshwater criterion (Table 3-6). Also, another insect, Enallagma cyathigerum, another insect

species, was the second most sensitive species used in calculating the chronic freshwater criterion (Table 3-6).

ii. **the use of the quantitatively acceptable chronic toxicity value for mussel (Lampsilis siliquoidea) from Hazelton et al. (2012). Specifically, Hazelton et al. (2012) conducted a 36-day renewal, measured PFOS chronic test with fatmucket, Lampsilis siliquoidea. The estimated EC_{10} was 0.05713 mg/L, which was extrapolated from the author-reported data and the exposure response slope from another PFOS toxicity study focused on another mussel species (Ellipto complamata) as explained in Section 3.1.1.3.3. Therefore, this test was used in the chronic criterion calculation.**

EPA used an EC10:EC35.4 from Drottar et al. (2000) for Elliptio complanata and applied this ratio to derive an EC10 from the data reported in Hazelton et al. (2012) for Lampsilis siliquoidea. The problem is that EPA have not clearly outlined in section 3.1.1.3.3 what endpoint that Drottar et al. (2000) was measuring in Elliptio complanata (also note that the genus and species are not spelled correctly in section 3.1.1.3.3). Is the endpoint measured in E. complanata the same as the endpoint measure in L. siliquoidea? I tried to look up the endpoint measure in Drottar et al. (2000) but I could not find the study and there was no reference provided in the reference section for Drottar et al. (2000). This missing information makes it difficult to comment on the validity of the approach that EPA has taken to derive an EC10 for L. siliquoidea.

iii. **the use of the quantitatively acceptable chronic toxicity value for damselfly (Enallagma cyathigerum) from Bots et al. (2010). Bots et al. (2010) conducted a 320-day renewal, unmeasured PFOS chronic test with blue damselfly nymphs, Enallagma cyathigerum. The MATC was 0.03162 mg/L, which was calculated from the author-reported value for nymph survival as explained in Section 3.1.1.3.2. Therefore, this test was used in the chronic criterion calculation.**

I think the EPA's justification for the use of the survival data from Bots et al. (2010) is valid. While control mortality reached 40% in the control, the plateau in control mortality after 60 days, the total duration of the test being 200 days, and 82.57% survival in the control from day 60 to 200, justifies the inclusion of the MATC derived from Bots et al. (2010) for Enallagma cyathigerum in the derivation of a chronic criterion.

iv. **the use of the quantitatively acceptable chronic toxicity value for midge (Chironomus dilutus) from MacDonald et al. (2004). MacDonald et al. (2004) conducted a 20-day renewal, measured PFOS chronic test with midge lava, Chironomus dilutus. The EC10 was 0.05963 mg/L, which was an EPA-calculated value for 10-day growth as explained in Section 3.1.1.3.4. Therefore, this test was used in the chronic criterion calculation.**

First, the wrong species is referenced in relation to the MacDonald et al. (2004) study. MacDonald et al. (2004) reported the toxicity of PFOS to Chironomus tentans. The EPA's derivation of a 10-d EC10 for Chironomus tentans using the data from MacDonald et al. (2004) is not clear. In Appendix C, section C.2.4, the EPA writes, "EPA could not fit a curve to independently verify the 10-day survival (due to a lack of a specific sample size for this endpoint as the number of replicates was not stated in the paper; however, the number of replicates was between 2 and 4 and EPA sought to obtain clarification and treatment level data from the study authors)" It is not clear how EPA got the information necessary, e.g., number of replicates, to fit a curve. It is also not clear what EPA means by "…and treatment level data from the study authors."? Did EPA acquire the raw data for growth from the 10-day toxicity test with C. tentans? If that is the case, they have not made that clear. If that is the case, it would also strengthen their independently derived EC10 for growth in C. tentans. I think the EPA needs to more clearly explain where they got the data necessary to derive the EC10 for C. tentans used in the chronic criterion.

5b. EPA's approach for fitting concentration-response (C-R) data (described in Appendix K) as well as the specific acute LC50 values (Appendix A.2) and chronic EC10 values (Appendix C.2) that were estimated (for sensitive genera when C-R data were available) and used to derive criteria.

I think the approach that the EPA used to determine effect measure from concentration-response data was appropriate. The use of the drc package in R to fit 22 different models to the empirical data and then using several criteria (e.g., AIC, residual standard errors, confidence intervals) to evaluate the fit of the different models is robust. It would have been useful if the EPA reported the 22 different models in Appendix K.

I think the LC50 and EC10 values determined by the EPA using the approach mentioned in the previous paragraph was appropriate. It is valid for these effect measures to be determined when the concentration-response data has been provided by the authors of the study. The EPA has also made is clear in Appendix A.2 and C.2 how they determined these effect measures using the concentration-response data provide in the studies. This generates a high level of transparency in the derivation of the criterion.

6) Please comment on the translation of the chronic water column criterion elements for aquatic life to derive the tissue-based criterion elements, considering the bioaccumulation of PFOA and PFOS. In particular, please comment on:

6a. Uncertainty surrounding the bioaccumulation factors (BAFs) used to translate of the chronic water column criterion elements into tissue-based criterion elements.

I think the EPA has sufficiently addressed the uncertainty around the use of BAFs and the chronic water column criterion in the derivation of tissue-based criterion. They have indicated that tissue-based criterion should only be observed once in 10 years. The use of the geometric mean of the reported BAFs incorporates the range of BAFs that may be present for different invertebrate and fish species. The use of the chronic water column criterion also builds in added conservatism to the tissue-based criterion.

Prosser et al. (2016) reported BAFs for PFOA in three freshwater species (two invertebrates and one fish) (See Tables S29-31 in Supplementary Information), but it was not considered in this assessment. It is not clear why it was not considered.

Prosser, R.S., Mahon, K., Sibley, P.K., Poirier, D., Watson-Leung, T., 2016. Bioaccumulation of perfluorinated carboxylates and sulfonates and polychlorinated biphenyls in laboratory-cultured Hexagenia spp., Lumbriculus variegatus and Pimephales promelas from field-collected sediments. Science of The Total Environment 543, 715–726. doi:10.1016/j.scitotenv.2015.11.062

6b. EPA's determination of appropriate BAFs and the tissue types that the tissue criterion elements were based.

The evaluation criteria for BAFs outline in Table 2-4 are appropriate and the decision to only use high and medium quality BAFs is justified based on the criteria that would make a BAF low quality. It was a good idea to use fish BAFs based on the concentration in muscle and whole body (Table 3-12). Muscle tissue is usually exclusively sampled in large fish, especially as part of fish consumption guidelines. The whole body is more appropriate for small fish and invertebrate species, e.g., minnows, benthic macroinvertebrates.

7) Please comment on the frequency and duration of the criterion elements, in particular the tissue-based criterion elements.

As per Table 0-1, I think the chosen durations and frequencies for the acute and chronic criteria are appropriate. They will ensure protection of aquatic life. The duration of the tissue-based criterion is appropriate as the concentration will be measured when biota is collected. The 10-year frequency is

appropriate considering that for biota to reach the tissue-based criteria, they would likely to have been exposed to concentrations at or above the chronic criteria for an extended period of time.

8) Please provide any additional technical comments that you believe should be considered.

I think the EPA's criteria for PFOS are very defensible based on the science and data available. I think they did a great job clearly laying out how they derived the criteria and providing all of the data that was used in the derivation.

COMMENTS SUBMITTED BY

REVIEWER 3

External Peer Review of EPA's Draft
Aquatic Life Water Quality Criterion for Perfluorooctane Sulfonate (PFOS)

1) Please comment on the overall clarity of the document as it relates to the derivation of each criterion.

- Main Question: Perhaps this was missed in the draft document, but, is there guidance when one criteria is exceeded and the other is not? For example, if the tissue based criteria are exceeded yet, the CCC is not; perhaps this is an unlikely scenario as those receptors have been accumulating PFOS for a duration likely under higher water concentrations (> 0.014mg/L). Although if sediment concentrations remain elevated (but not water column concentrations) this may also be a likely route of PFOS exposure to fish with sediment dwelling prey.

- There are additional domestic criteria missing from the previously published criteria section; please review those for Texas, Florida and California.

- Are there two Sharpe et al.'s, I believe this is only one publication but flipping between Sharpe et al. 2010, Sharpe et al. 2010a and Sharpe et al. 2010b throughout the document. If the goal is to distinguish between supplemental vs the manuscript proper I suggest just clarifying in the text instead of the reader looking for two pubs by Sharpe et al. 2010.

- Page 238 – error: The study authors reported a 96-hour LC_{50} of 58.47 mg/L PFOS, based on the results of the range finding test. The independently-calculated toxicity value was x.xx mg/L.

- Page 296 – error: The independently-calculated toxicity value was x.xx mg/L.

- Table 3-6 is not referenced/described in the text. Additionally, the title reads "Six" most sensitive and lists "Seven".

- Overall comment: ranking of sensitive genera flips back and forth between most and least sensitive among tables, consistency would help the reader.

2) Please comment on the approach used to derive the draft criterion for PFOS. Please provide detailed comments.

- **Is the technical approach used to derive the criterion elements logical?**
  - This is logical and follows the established GLRI guidance; however, both Canada and Australia utilize a species sensitivity distributions to determine the 95th and 99th percentile of species protection. Is there a defensible reason why EPA did not employ this approach or at the very least present these distributions and analysis that would support the currently drafted criteria?
  - Additionally, thresholds from those SSDs (and others published) are lower than the draft guidance here, this should be addressed:
    - Australia – 0.13 µg/L
    - Canada – 6.8 µg/L
    - Salice et al. 2018 – 1.12 µg/L
    - Conder et al. 2020 – 5.85 µg/L

- *Giesy et al. 2010 – 5.1 µg/L (using CCC based on GLRI guidance)

- **Does the science support the conclusions?**
  - The GMCV for Zebrafish is 0.0165 mg/L, thus there are studies that result in chronic toxicity at concentrations lower than this mean; however, this is very close to the CCC of 0.014mg/L. This seems borderline protective when considering potential exposures to this species (and those more sensitive).

- **Is it consistent with the protection of freshwater aquatic life from acute, chronic, and bioaccumulative effects?**
  - I have confusion over Tables 3-9 and 4-6 calculations. How is it that the inclusion of *Lampsilis* with a higher GMAV results in a lower overall CMC (3.3 mg PFOS/L) compared to the CMC in table 3-9 (3.6 mg PFOS/L)? Actually, looking more closely at this, the ln(GMAV)^2 are inconsistent among the tables for *Xenopus*, this is likely are result of using table 3-6 as a template for 4-6.

  - It is great to see the inclusion of the Burkhard et al. 2021 as this synthesis has been peer-reviewed and published and is an exceptional overview of PFOS bioaccumulation; unfortunately, there are not more current literature used within the draft document.

3) Please comment on the approach used to derive the draft acute estuarine/marine benchmark for PFOS. Given the limited estuarine/marine test data available, a new approach method was used to support the derivation of an acute estuarine/marine benchmark to provide states and tribes with a protective value. Please provide detailed comments.

- **Is the technical approach used to derive the benchmark logical?**
  - After potential inclusion of the data mentioned below this approach may be appropriate. In the current form with the limited data it may be misleading. Can this guidance be updated? I am aware of other researchers investigating PFAS on marine species (Ed Wirth, NOAA) and maybe others that will be coming out soon.

- **Does the science support the conclusions?**
  - I believe the data are incorrect for Fabbri et al. 2014. In table B.1 the reported effect concentration is recorded as >1 mg/L. However, looking at the paper, I read, "*The PFCs PFOA and PFOS induced a dose-dependent effect, with significant decreases in normal larval development from 0.1 µg/L (17% and 27%, respectively; P 0.01). Maximal effects were observed at 100 µg/L (about 40% and 50%, respectively; P 0.001) with no further decreases at higher concentrations*". There is a monotonic concentration-response curve. The associated figure also supports an effect at 0.1µg PFOS/L, see below. Furthermore, if the EC50 of the test organisms is a needed endpoint (as noted in the PFOA justification, for which is lacking support in the current form) looking at the figure below % of normal D-larvae for PFOS (although incorrectly referred to in the legend as PFOAS) could be inferred at 0.1 mg/L. Furthermore, has EPA considered calculating the MATC from this study?

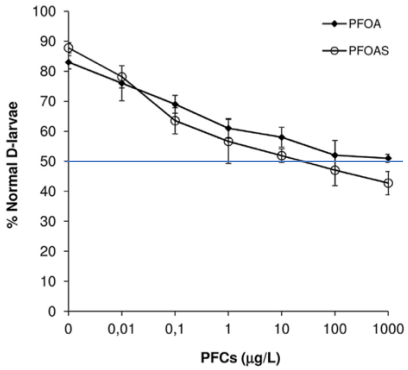


Fig. 4. Effects of PFOA and PFOS (0.01–0.1–1–10–100–1000 µg/L) on *M. galloprovincialis* normal larval development in 96-multiwell plates. Data are reported as in Fig. 3a.

- I did not see data included or the study evaluated for: Robertson JC (1986) Potential for environmental impact of AFA-6 surfactant. Beak Consultants Ltd. Missassauga, Ontario, Canada. EPA Docket AR226-1030a043.
  - There are data for saltwater spp in the ITRC from this citation.

- **Is it consistent with the protection offered by acute estuarine/marine aquatic life criteria derived using empirical data, as prescribed in the 1985 *Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*?**
  - No, this is a new approach; however, it follows the spirit of the 1985 guidelines.

4) **Please comment on the use of measured and unmeasured toxicity tests to derive the respective criterion. In particular please comment on the supporting justification for using unmeasured toxicity tests in Appendix O.**
   - This seems acceptable for the time being. Having worked in the laboratory with PFOS, I can make a first-hand testament that mixing PFOS into exposures solutions does not guarantee a homogenous mixture despite working at solutions well below the solubility limit. There are nuances associated with achieving homogeneity of the exposure solution, we have developed a PFAS mixing protocol to reduce chemical clumping and this increases uniformity of the solutions. Furthermore, there is approximately 30% variability of PFOS quantitatively (see…Rewerts et al. 2020); so, the best measurement still has significant variability.

5) **Please comment on the toxicity data used to derive the draft criteria.**
   - See collective responses below
   - **Were the data selected and/or excluded from the derivation of the criteria derivation appropriately utilized?**
     - Data selection and waiving of the MDR for insect family in the FAV seem reasonable.
   - **Are there relevant data that you are aware of that should be added to the analyses (note that EPA is working on updating the toxicity data to reflect the data in ECOTOX between Sept. 2019 through the latest update)? If so, please provide references for consideration.**

- The data selection for the derivation of the draft criteria are limited to published and/or available studies from 2018 and prior. This significantly reduces the studies used in the derivation as a number of publications have become available in recent years.

For example:

- A newly published study is available for fathead minnows exposed to PFOS for chronic duration and over the course of reproduction and development. Although, this study was static-renewal, PFOS concentrations are measured ; importantly, this study resulted in a NOEC of 88µg/L based on reduced biomass seen in the second generation (Suski et al. 2020). Importantly, follow-on work (in prep) indicates that this may be a maternal transfer impact as PFOS exposures to juvenile fish alone do not share results.

- Also, from the authors noted above is an ongoing full life-cycle fathead PFOS and PFAS mixture exposure. This study is being conducted under flow through conditions and is expected to reach termination in December 2021.

- McCarthy et al. 2021 published data on chironomids (EC20 = 1.7µg/L), these are also not included here.

- Bryan Brooks (Baylor) and Matt Simcik (UMN) also have acute data on the fathead minnow with measured concentrations, these are not published just yet.

- David Moore (Army Corps) is near completion of a full life-cycle fish study

- In particular, SERDP has been funding this research for years and those data are published, recently published or near final. EPA should reach out to SERDP PIs for data inquiries and potential inclusion in these draft criteria.

In particular, please comment on:

5a. The toxicity values used to derive the PFOS criteria, with a particular emphasis on:

i. the use of the qualitatively acceptable acute midge (*Chironomus plumosus*) data from Yang et al. (2014) to suggest aquatic insects are relatively tolerant to acute PFOS exposures. Specifically, Yang et al. (2014) conducted a 96-hour renewal, measured PFOS acute test with the midge, *Chironomus plumosus*. This study was not acceptable for quantitative use due to the potential problematic source of the organisms. The reported LC_{50} was 182 mg/L for PFOS indicating that insects may not be one of the more sensitive taxonomic groups. Therefore, this test was excluded from the acute criterion calculation, but used to waive the missing insect MDR.

- This seems appropriate, the flower market is most definitely an odd place to purchase research organisms.

ii. the use of the quantitatively acceptable chronic toxicity value for mussel (*Lampsilis siliquoidea*) from Hazelton et al. (2012). Specifically, Hazelton et al. (2012) conducted a 36-day renewal, measured PFOS chronic test with fatmucket, *Lampsilis siliquoidea.* The estimated EC_{10} was 0.05713 mg/L, which was extrapolated from the author-reported data and the exposure response slope from another PFOS toxicity study focused on another mussel species (*Ellipto complamata*) as explained in Section 3.1.1.3.3. Therefore, this test was used in the chronic criterion calculation.

- From Section 3.1.1.3.3: "*The in marsupia exposure was followed by a 24-hour free glochidia exposure consisting of a factorial design, such that free glochidia from the control group of the marsupia exposure were divided between a control and the two PFOS treatments and the PFOS treatments were split into control and the same PFOS treatment group as the marsupia exposure."* - Comment: This is an exceptionally long and confusing sentence please revise to help the reader understand this complex study and overall approach that EPA took.
- The approach seems ok given the limited data availability at this time.

iii. the use of the quantitatively acceptable chronic toxicity value for damselfly (*Enallagma cyathigerum*) from Bots et al. (2010). Bots et al. (2010) conducted a 320-day renewal, unmeasured PFOS chronic test with blue damselfly nymphs, *Enallagma cyathigerum.* The MATC was 0.03162 mg/L, which was calculated from the author-reported value for nymph survival as explained in Section 3.1.1.3.2. Therefore, this test was used in the chronic criterion calculation.

- Given the duration of the study the researchers likely hovered around the nominal concentrations of PFOS. Inclusion seems appropriate.

iv. the use of the quantitatively acceptable chronic toxicity value for midge (*Chironomus dilutus)* from MacDonald et al. (2004). MacDonald et al. (2004) conducted a 20-day renewal, measured PFOS chronic test with midge lava, *Chironomus dilutus.* The EC_{10} was 0.05963 mg/L, which was an EPA-calculated value for 10-day growth as explained in Section 3.1.1.3.4. Therefore, this test was used in the chronic criterion calculation.

- I am uncomfortable with this conclusion presented here, it may be more appropriate to use MacDonald et al. data from the 20-day endpoint considering recent publication from McCarthy et al. 2020 as noted above.

5b. EPA's approach for fitting concentration-response (C-R) data (described in Appendix K) as well as the specific acute LC_{50} values (Appendix A.2) and chronic EC_{10} values (Appendix C.2) that were estimated (for sensitive genera when C-R data were available) and used to derive criteria.

- This seems like a reasonable and defensible approach if it is applied consistently across genera.

6) Please comment on the translation of the chronic water column criterion elements for aquatic life to derive the tissue-based criterion elements, considering the bioaccumulation of PFOA and PFOS. In particular, please comment on:

6a. Uncertainty surrounding the bioaccumulation factors (BAFs) used to translate of the chronic water column criterion elements into tissue-based criterion elements.

6b. EPA's determination of appropriate BAFs and the tissue types that the tissue criterion elements were based.

- Please clarify, the following sentence: "BAFs used in the derivation of the PFOS tissue criteria consisted of > 2 water and organism samples each and were collected within one year and 2 km distance." It is unclear if the >2 samples refer to the tissue & water samples being mismatched temporally or if there where >2 sets of water and tissue samples that were collected in different years.
  - If the latter then this approach seems appropriate; if the former, EPA should discuss differences in water chemistry between years to alleviate any concerns with matching tissue concentration data to water samples that may have significant environmental temporal variability.

- A table summarizing the animal tissues used in deriving the BAFs would be helpful to assess the range of fish species and their dietary preferences.

7) Please comment on the frequency and duration of the criterion elements, in particular the tissue-based criterion elements.

- This is a not an easy statement to comment on, as it may be unlikely that the aquatic receptors will exceed or reach these tissue concentrations prior to exceedances from the CCC.
- What I am not clear on is, if tissue concentrations exceed these proposed thresholds yet, PFOS water concentrations do not exceed the CCC, what would be the proposed guidance?

8) Please provide any additional technical comments that you believe should be considered.

- All technical comments have been previously mentioned

COMMENTS SUBMITTED BY

REVIEWER 4

External Peer Review of EPA's Draft Aquatic Life Water Quality Criterion for Perfluorooctane Sulfonate (PFOS)

1) Please comment on the overall clarity of the document as it relates to the derivation of each criterion.

EPA has drafted the PFOS aquatic life criteria to be consistent with methods described in EPA's "*Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*" (U.S. EPA 1985). I congratulate the EPA Team for a very thorough, comprehensive analysis of the toxicological data to derive each criterion.

- The report is technically sound and is very clearly written.
- The criteria have been derived using strong science-based evidence.
- Sub-sections on overview of PFAS, PFAS nomenclature, problem formulation, exposure pathways, transformation and degradation of PFOS precursors in the aquatic environment sources, concentration reported in environment and existing criteria (both national and international) help to set the scene before toxicological data is presented and assessed for developing various criterion.
- The freshwater acute water column-based criterion, the chronic water column-based chronic criterion, the chronic fish whole-body tissue criterion, the chronic fish muscle tissue criterion and the chronic invertebrate whole-body tissue criterion have been developed and documented in this report are based on comprehensive assessment of the toxicological data and consistent with the *Guidelines*.
- Acute and chronic MDRs for PFOS estuarine/marine criteria derivation were not met due to fewer empirical PFOS toxicity data. To address this gap, the EPA Team developed an acute aquatic life benchmark for estuarine/marine environments based on Interspecies Correlation Estimation (ICE) model. Such predictive models should be used when there is limited toxicity data.
- EPA Team has provided extensive background information on toxicity data assessment and collated this information in various Appendices as additional line of evidence.
- Tables and Figures are very well laid out throughout the document and provide additional information of the toxicity data used in developing Water Quality Criteria for PFOS.

2) Please comment on the approach used to derive the draft criterion for PFOS. Please provide detailed comments.

- **Is the technical approach used to derive the criterion elements logical?**
- **Does the science support the conclusions?**
- **Is it consistent with the protection of freshwater aquatic life from acute, chronic, and bioaccumulative effects?**

This EPA report provides a critical review of toxicity data identified in EPA's literature search for PFOS, including the anionic form (CAS No. 45298-90-6), the acid form (CAS No. 1763-23-1), potassium salt (CAS No. 2795-39-3), an ammonium salt (CAS No. 56773-42-3), sodium salt (CAS No. 4021-47-0), and a lithium salt (CAS No. 29457-72-5). It quantifies the toxicity of PFOS to aquatic life, and provides criteria intended to protect aquatic life from the acute and chronic toxic effects of PFOS. The detailed assessment is as follows:

- These criteria have been derived using robust methods and the best available toxicity data on aquatic life.
- The approach used to derive the draft criterion for PFOS is very logical and consistent with the protection offered by acute and chronic aquatic life criteria derived using empirical data, as prescribed in the 1985 *Guidelines.*

- Exclusion and inclusion criteria are appropriately discussed in the context of the toxicological data reported in the literature and provide additional evidence on the selection of toxicity data criteria development.
- With limited toxicity datasets to North American resident species, non-North American resident species were included for criteria development. For example, inclusion of non-resident species such as Planaria, *Dugesia japonica* and Japanese swamp shrimp, *Neocaridina denticulata* for calculating acute water quality criteria and zebra fish, *Danio rerio* for chronic criteria. The EPA team did not find any influence of excluding non-North American resident species in criteria derivations and decided to retain the full acute and chronic toxicity dataset. This was very rational decision and non-Northern American species served as surrogate species for the broad range of the thousands of untested species present in the freshwater environment in the U.S.
- The acute measures of effect on aquatic organisms selected included the lethal concentration (LC50), effect concentration (EC50), or inhibitory concentration (IC50) estimated to produce a specific effect in 50 percent of the test organisms as per the *Guidelines*.
- The endpoint for chronic exposures incorporated the effect concentration estimated to produce a chronic effect on survival, growth, or reproduction in 10 percent of the test organisms (EC_{10}). This approach has been also consistent with the harmonized guidelines from OECD and the generally preferred effect level for countries such as Canada, Australia, and New Zealand.
- Reported (No Observed Effect Concentrations) (NOECs) and (Lowest Observed Effect Concentrations) (LOECs) were only used for the derivation of a chronic criterion when a robust EC_{10} could not be calculated for the genus.
- Furthermore, EPA independently calculated these toxicity values if sufficient raw data were available for EPA to conduct statistical analyses. EPA's independently-calculated toxicity values were used preferentially, where available.
- I agree with the authors' decision on not developing plant criteria based on their lesser sensitivity to PFOS than in comparison to aquatic vertebrates and invertebrates. The EPA team evaluated the toxicity data to plants as an additional line of evidence and confirmed that the proposed PFOS freshwater acute and chronic criteria are expected to be protective of freshwater plants.
- EPA developed protective tissue-based criteria through a bioaccumulation factor approach. This was based on the application of evaluation criteria for screening bioaccumulation factors (BAFs).
- Based on comprehensive toxicity data assessment, the EPA team has developed the following criteria using the procedures described in the 1985 *Guidelines*. The freshwater acute water column-based criterion magnitude is 3.6 mg/L and the chronic water column-based criterion magnitude is 0.014 mg/L. The chronic freshwater criterion also contains tissue-based criteria expressed as 43.0 mg/kg wet weight (ww) for fish whole-body, 25.3 mg/L ww for fish muscle tissue, and 12.3 mg/kg ww for invertebrate whole-body tissue.
- Acute and chronic MDRs for PFOS estuarine/marine criteria derivation were not met and an estuarine/marine FAV could not be calculated to derive an estuarine/marine acute criterion. Further benchmark was developed using predictive approach and discussed in the follow-up question 3.

3) Please comment on the approach used to derive the draft acute estuarine/marine benchmark for PFOS. Given the limited estuarine/marine test data available, a new approach method was used to support the derivation of an acute estuarine/marine benchmark to provide states and tribes with a protective value. Please provide detailed comments.

- **Is the technical approach used to derive the benchmark logical?**
- **Does the science support the conclusions?**

- **Is it consistent with the protection offered by acute estuarine/marine aquatic life criteria derived using empirical data, as prescribed in the 1985** *Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses***?**
  - EPA applied the ICE model predictions to supplement the available test dataset to help fill missing MDRs and allow the derivation of acute estuarine/marine benchmark recommendations for aquatic life using procedures consistent with those in the 1985 *Guidelines***.** A total of 3104 datapoints from 398 models were evaluated.
  - ICE model has been recommended to predict the sensitivity of an untested taxon (predicted taxa are represented by the y-axis) from the known, measured sensitivity of a surrogate species (represented by the x-axis). The ICE model approach used is very reasonable to predict toxicity of untested taxa.
  - As documented in Section L.1, ICE-predicted models have been used by multiple independent, international groups and further confirms that values developed from ICE-generated SSDs will provide a level of protection that is consistent with using measured laboratory data.
  - In addition, prediction accuracy and robustness of the model is evaluated using robust parameters (e.g., mean square error, R2), that fall within a defined range of acceptability, and with close prediction confidence intervals that facilitate evaluating the fit of the underlying data. This confirms the robustness of the model.
  - ICE models predicted with acceptable accuracy for PFOS when invertebrates were used to predict to invertebrate species and vertebrates were used to predict to vertebrate species in these comparisons.
  - The draft acute benchmark for estuarine/marine aquatic life developed using this approach is 0.43 mg/L PFOS, it is lower than the recommended acute freshwater criterion(3.6 mg/L), suggesting that estuarine/marine species may be more acutely sensitive to PFOS. This is in line with Hayman et al., (2021), confirming marine species have a higher sensitivity to PFOS than compared to the freshwater organisms.
  - In this report*, Mytilus galloprovincialis* was not used in the FAV calculation because the value was not definitive, and true sensitivity of this species is unknown. There are two more studies published reporting the toxicity values for marine/estuarine species, including *Mytilus galloprovincialis*.
    - Stuart L. Simpson, Yawen Liu, David A. Spadaro, Xinhong Wang; Rai S. Kookana and Graeme E. Batley Chronic effects and thresholds for estuarine and marine benthic organism exposure to perfluorooctane sulfonic acid (PFOS)-contaminated sediments: Influence of organic carbon and exposure routes https://doi.org/10.1016/j.scitotenv.2021.146008
    - Nicholas T Hayman , Gunther Rosen , Marienne A Colvin , Jason Conder , Jennifer A Arblaster Aquatic toxicity evaluations of PFOS and PFOA for five standard marine endpoints. https://doi.org/10.1016/j.chemosphere.2021.129699

It is recommended to assess the quality of the toxicity data on marine/estuarine species and recalculate estuarine criteria based on this recently available information.

4) Please comment on the use of measured and unmeasured toxicity tests to derive the respective criterion. In particular please comment on the supporting justification for using unmeasured toxicity tests in Appendix O.

PFOS is a highly stable compound, resistant to hydrolysis, photolysis, volatilization, and biodegradation (as described in Section 1.1.1 of the Report) and, therefore, expected to vary only minimally in the course of a toxicity test. To determine if nominal and measured PFOS concentrations were typically in close agreement, pairs of nominal and corresponding measured PFOS concentrations were compared to one another through (1) linear correlation analysis and (2) an assessment of measured concentrations as a percent of its paired nominal concentration. The authors reported, 22 freshwater studies with PFOS measured concentrations, yielding 373 pairs of measured and nominal concentrations. In addition, there were 7 estuarine/marine studies with measured concentrations, yielding 142 pairs of measured and nominal concentrations. The data were grouped by classifications including water type (salt/fresh) and experimental conditions (acute/chronic; solvent/no solvent; fed/unfed, etc.). Data displayed a high degree of linear correlation and measured, and nominal concentrations were in close agreement

The analysis conducted by EPA Team showed strong correlation (correlation = 0.9998) of the 326 pairs of nominal and measured concentrations from freshwater studies. In addition, the experimental conditions did not influence the correlation between nominal and measured concentrations. The detailed analyses of the data in Appendix O and the relevant Tables and Figures provide very comprehensive analyses – this is very useful information and will assist ecotoxioclogist in designing future experiments.

This confirms inclusion of unmeasured PFOS toxicity tests for quantitative use in criteria derivation.

Personal experience on analyzing PFOS in ecotoxicological studies using freshwater species have also exhibited strong correlation between nominal and measured concentrations.

The authors reported the strong correlation (0.8993) of the 142 pairs of nominal and measured concentrations, the ratio of measured to nominal concentrations from the saltwater dataset showed bias with a geometric mean value of 0.6178. Additionally, the median percent difference between measured and nominal concentration was 30.82%. Furthermore, the saltwater comparison of nominal and measured concentrations indicated that these experimental conditions (acute/chronic and unfed/fed) could influence the observed differences between measured and nominal concentrations. These results suggest that measured and nominal concentrations from saltwater tests were not in close agreement, but this analysis was based on limited set of data.

The measured concentrations in the recently published paper on marine/estuarine toxicity of PFOS should also be included in this assessment:

- Nicholas T Hayman , Gunther Rosen , Marienne A Colvin , Jason Conder , Jennifer A Arblaster Aquatic toxicity evaluations of PFOS and PFOA for five standard marine endpoints. https://doi.org/10.1016/j.chemosphere.2021.129699

The second paper is on benthic organisms and PFOS is measured in overlying water, porewater and sediment. This may provide further guidance on difference between PFOS measured and nominal concentrations.

- Stuart L. Simpson, Yawen Liu, David A. Spadaro, Xinhong Wang; Rai S. Kookana and Graeme E. Batley Chronic effects and thresholds for estuarine and marine benthic organism exposure to perfluorooctane sulfonic acid (PFOS)-contaminated sediments: Influence of organic carbon and exposure routes https://doi.org/10.1016/j.scitotenv.2021.146008

Additional information for Appendix O based on a recently published paper:

*According to **Rewerts et al., 2021** additional handling steps, which are not typically reported for ecotoxicological studies but may contribute to variability, include solution homogenization, subsampling procedures, and the container materials selected for storage*. https://doi.org/10.1002/etc.4667

5) Please comment on the toxicity data used to derive the draft criteria.

- **Were the data selected and/or excluded from the derivation of the criteria derivation appropriately utilized?**
- **Are there relevant data that you are aware of that should be added to the analyses (note that EPA is working on updating the toxicity data to reflect the data in ECOTOX between Sept. 2019 through the latest update)? If so, please provide references for consideration.**

The data selected to derive PFOS criteria are appropriate. Studies that did not fully meet the data quality objectives outlined in the 1985 *Guidelines* were not considered for inclusion in the criteria derivation, including some studies with other PFAS exposures, but were considered qualitatively as supporting information. A brief summary of each study describing the experimental conditions and summary tables providing all the relevant information such as strengths and limitations of each study, end points selected for deriving criteria are well documented by the EPA team and provides further confidence in data selection process.

The key acceptable exclusion/inclusion criteria used to derive draft criteria are listed below:

- Only single chemical toxicity tests with PFOS were considered for possible inclusion in criteria derivation, studies that tested chemical mixtures, including mixtures with PFAS compounds were excluded from criteria derivation.
- Both controlled laboratory experiments and field observations/studies were included.
- PFOS toxicity tests were not excluded from quantitative use in criteria derivation on the basis of unmeasured test concentrations alone.
- Due to lower sensitivity, insect MDR was excluded from the criterion calculation, but were used to waive the missing insect MDR.
- Further supporting information on acceptable and unused studies for acute and chronic endpoints and for freshwater and marine studies are documented and summarized as appendices in this report.

Additional toxicity data published over the last six months is listed below:

Marine/estuarine

- Nicholas T Hayman , Gunther Rosen , Marienne A Colvin , Jason Conder , Jennifer A Arblaster Aquatic toxicity evaluations of PFOS and PFOA for five standard marine endpoints. https://doi.org/10.1016/j.chemosphere.2021.129699
- Stuart L. Simpson, Yawen Liu, David A. Spadaro, Xinhong Wang; Rai S. Kookana and Graeme E. Batley Chronic effects and thresholds for estuarine and marine benthic organism exposure to perfluorooctane sulfonic acid (PFOS)-contaminated sediments: Influence of organic carbon and exposure routes https://doi.org/10.1016/j.scitotenv.2021.146008

Fresh water

- Christopher J. McCarthy, Shaun A. Roark, Demitria Wright, Kelly O'Neal, Brett Muckey, Mike Stanaway, Justin N. Rewerts, Jennifer A. Field, Todd A. Anderson, Christopher J. Salice, Toxicological Response of *Chironomus dilutus* in Single-Chemical and Binary Mixture Exposure Experiments with 6 Perfluoralkyl Substances, Environmental Toxicology and Chemistry, 10.1002/etc.5066, **40**, 8, (2319-2333), (2021). https://doi.org/10.1002/etc.5066

In particular, please comment on:

5a. The toxicity values used to derive the PFOS criteria, with a particular emphasis on:

i. the use of the qualitatively acceptable acute midge (*Chironomus plumosus*) data from Yang et al. (2014) to suggest aquatic insects are relatively tolerant to acute PFOS exposures. Specifically, Yang et al. (2014) conducted a 96-hour renewal, measured PFOS acute test with the midge, *Chironomus plumosus*. This study was not acceptable for quantitative use due to the potential problematic source of the organisms. The reported LC_{50} was 182 mg/L for PFOS indicating that insects may not be one of the more sensitive taxonomic groups. Therefore, this test was excluded from the acute criterion calculation, but used to waive the missing insect MDR.

Waiving an unfulfilled MDR when available data suggest it is not among the four most sensitive genera is consistent with previous EPA criteria documents, including U.S. EPA (2016). At this stage, I do not fully agree with the statement that midge larvae are tolerant to acute exposures. The OECD protocol recommends 48h acute test for midge larvae and the 48h exposure period is acceptable duration for assessing acute toxicity. The study by Olson (2017) has limitations but this study can't be fully ruled out. The chronic toxicity data exhibits sensitivity of insects to PFOS and this statement is also supported by the authors. In addition, Stefani et al. (2014), Macdonald et al. (2004), and Marziali et al. (2019) conducted chronic toxicity tests with *Chironomus* spp. and reported apical endpoints. *Results of these studies, taken together, also suggest that insects are among sensitive taxa to chronic PFOS exposures (with adverse effects reports at low ug/L)*

I support the recommendation '*Additional insect toxicity data for PFOS would be very useful for further examining the relative sensitivity of insects to PFOS exposures''.*

Unpublished work from our lab shows acute toxicity to midge larva, *Chironomus tepperi* at 1 mg/L PFOS (48 h EC50 value).

ii. the use of the quantitatively acceptable chronic toxicity value for mussel (*Lampsilis siliquoidea*) from Hazelton et al. (2012). Specifically, Hazelton et al. (2012) conducted a 36-day renewal, measured PFOS chronic test with fatmucket, *Lampsilis siliquoidea.* The estimated EC_{10} was 0.05713 mg/L, which was extrapolated from the author-reported data and the exposure response slope from another PFOS toxicity study focused on another mussel species (*Ellipto complamata*) as explained in Section 3.1.1.3.3. Therefore, this test was used in the chronic criterion calculation.

The authors have provided detailed assessment of this study and explained the approach used for the calculation of chronic toxicity value (section C.2.3-Third Sensitive Freshwater Genus for Chronic Toxicity: *Lampsilis siliquoidea* (mussel). Hazelton et al. 2012 used robust study design in spite of including only two concentration of PFOS in this study. The PFOS exposure concentration was measured, and metamorphosis success was used as an endpoint for inclusion in the criteria development. While viability of free glochidia at 24 hours post removal from females was a less sensitive endpoint and did not meet the acceptability criteria. The reduction in metamorphosis success at the 0.0695 mg/L was estimated to be 35.4% but EC10 could not be calculated based on only two PFOS concentrations tested in this study. The EPA team has calculated EC10 (0.05713 mg/L) using the exposure response slope from PFOS toxicity study on another mussel species (*Ellipto complamata*). The explanation and logic provided is reasonable to include the calculated EC10 value to derive the freshwater chronic criterion and to better understand the effects of PFOS on aquatic insects.

iii. the use of the quantitatively acceptable chronic toxicity value for damselfly (*Enallagma cyathigerum*) from Bots et al. (2010). Bots et al. (2010) conducted a 320-day renewal, unmeasured PFOS chronic test with blue damselfly nymphs, *Enallagma cyathigerum.* The MATC was 0.03162 mg/L, which was calculated from the author-reported value for nymph survival as explained in Section 3.1.1.3.2. Therefore, this test was used in the chronic criterion calculation.

As a weight of evidence approach, EPA ran additional analyses with some of the other toxicity values for *E. cyathigerum* to understand the influence of this study on the overall chronic criterion. The 150-day MATC was more comparable to the other aquatic insect data and more representative of life cycle effects than the 10-day MATC or NOEC at 60 and 320 days of exposure (Table 4.3 of the report). EPA has concluded that the 150-day MATC should be used quantitatively to derive the chronic freshwater criterion toxicity. In addition, the control survival of test organisms was determined to be acceptable at this time point in the test. I am in agreement with this decision.

iv. the use of the quantitatively acceptable chronic toxicity value for midge (Chironomus dilutus) from MacDonald et al. (2004). MacDonald et al. (2004) conducted a 20-day renewal, measured PFOS chronic test with midge lava, Chironomus dilutus. The EC10 was 0.05963 mg/L, which was an EPA-calculated value for 10-day growth as explained in Section 3.1.1.3.4. Therefore, this test was used in the chronic criterion calculation.

The observed effects of PFOS on *C. dilutus* reported in the paper by the study authors include survival and growth as weight (measured as mg of ash-free dry mass per individual) for both the 10-day and 20-day exposure durations and emergence and reproduction over the 20-day exposure duration. The author reported 10-day growth and survival EC10s for the study were 0.0492 and 0.1079 mg/L, respectively. The study authors also reported NOECs of 0.0491 mg/L, LOECs of 0.0962 mg/L, and MATCs of 0.0687 mg/L for both endpoints. The author reported 20-day EC_{10}s for growth, survival, and total emergence were 0.0882, 0.0864, and 0.0893 mg/L, respectively. And the study authors also reported NOECs of 0.0217 mg/L for growth and survival and < 0.0023 mg/L for emergence, LOECs of 0.0949 mg/L for growth and survival and 0.0217 mg/L for emergence, and MATCs of 0.0454 mg/L for growth and survival and 0.0071 mg/L for emergence.

Independent statistical analyses were conducted by EPA Team for both the 10-day and 20-day exposure durations using data that were estimated The 20-day EC_{10}s for survival and emergence were not considered to be reliable endpoints given the disparities in the calculated EC_{10}s and the level of data that was presented in the paper, which made independent verification of the toxicity values less accurate. The dosing of the 20-day exposure was more of a concern than the 10-day exposure, which had measured concentrations that were much more in line with the expected nominal concentrations. The independently-calculated 10-day EC_{10} for growth was 0.0586 mg/L was used quantitatively to derive the chronic aquatic life criterion.

The EPA team has reviewed publications by Stefani et al. (2014) and Marziali et al. (2019) as additional supporting information. These authors conducted chronic toxicity tests with *Chironomus* spp. and reported chronic apical endpoints (at low ug/l) but at only at one concentration.

Use of the chronic toxicity data for PFOS in a recent publication should also be considered to assess the reliability of 20-day endpoints (adverse effects reported at 2-3 µg/L).

Christopher J. McCarthy, Shaun A. Roark, Demitria Wright, Kelly O'Neal, Brett Muckey, Mike Stanaway,Justin N. Rewerts, Jennifer A. Field, Todd A. Anderson, Christopher J. Salice, Toxicological Response of *Chironomus dilutus* in Single-Chemical and Binary Mixture Exposure Experiments with 6 Perfluoralkyl Substances, Environmental Toxicology and Chemistry, 10.1002/etc.5066, **40**, 8, (2319-2333), (2021). https://doi.org/10.1002/etc.5066

5b. EPA's approach for fitting concentration-response (C-R) data (described in Appendix K) as well as the specific acute LC_{50} values (Appendix A.2) and chronic EC_{10} values (Appendix C.2) that were estimated (for sensitive genera when C-R data were available) and used to derive criteria.

This is an excellent approach utilized by the EPA Team. EPA's approach for fitting concentration-response (C-R) data resulted in consistent approach across various ecotoxicological studies. The R drc package was used to fit 22 different models to each individual C-R dataset. A single model was then selected from the 22 models to serve as the representative C-R model. The selected model represented the most statistically-robust model available. In certain cases, this approach even improved and helped to select most sensitive toxicological endpoint.

In depth analyses and associated dose-response graphs in Appendix A.2 and Appendix C.2 provides further in-depth information on the EPA's approach for fitting concentration-response (C-R) data. As noted in Section 8 some of the values are missing.

6) Please comment on the translation of the chronic water column criterion elements for aquatic life to derive the tissue-based criterion elements, considering the bioaccumulation of PFOA and PFOS. In particular, please comment on:

6a. Uncertainty surrounding the bioaccumulation factors (BAFs) used to translate of the chronic water column criterion elements into tissue-based criterion elements.

The freshwater chronic PFOS toxicity data with measured tissue concentrations was limited, with no quantitatively acceptable tissue-based tests. Therefore, there were insufficient data to derive tissue-based criteria using a GSD approach from empirical tissue data from toxicity studies.

Tissue criteria derived from the chronic water column concentration (CCC) with the use of bioaccumulation factors were developed by EPA. The chronic freshwater criterion also contains tissue-based criteria expressed as 43.0 mg/kg wet weight (ww) for fish whole-body, 25.3 mg/kg ww for fish muscle tissue, and 12.3 mg/kg ww for invertebrate whole-body tissue.

EPA developed protective tissue-based criteria through a bioaccumulation factor approach. The authors reviewed PFOS BAF literature based on four criteria 1) number of water samples, 2) number of organism samples, 3) water and organism temporal coordination in sample collection, and 4) water and organism spatial coordination in sample collection and developed a ranking system. BAFs used in the derivation of the PFOS tissue-based criteria consisted of > 2 water and organism samples each and were collected within one year and 2 km distance. This scheme assured selection of only BAFs of high and medium quality to derive the tissue criteria.

6b. EPA's determination of appropriate BAFs and the tissue types that the tissue criterion elements were based.

BAFs are different for muscle/fillet and whole-body tissues. Humans consume muscle/fillets from fish and soft tissues from bivalves, therefore the water quality criteria recommended by EPA used BAFs based on these tissues. In addition, muscle and whole-body are the most commonly sampled tissue types in monitoring programs. These criteria were developed based on the values reported for 50-60 samples (Table 3-12).

Within the body, PFOS tends to bioaccumulate within protein-rich tissues, such as the blood serum proteins and liver. EPA Team calculated additional tissue values for liver, blood, and reproductive tissues by transforming the freshwater chronic water column criterion into representative tissue concentrations using tissue-specific bioaccumulation factors (BAFs). Author's decision uses agreement on the use of female

reproductive tissues due to its relevance for potential maternal transfer to offspring. These additional tissue-based values were calculated for comparative purposes and were not proposed as recommended criteria.

7) Please comment on the frequency and duration of the criterion elements, in particular the tissue-based criterion elements.

PFOS concentrations in tissues are generally expected to change only gradually over time in response to environmental fluctuations. The chronic tissue-based criteria averaging periods, or duration components, were therefore specified as instantaneous, because tissue data provide point, or instantaneous, measurements that reflect integrative accumulation of PFOS over time and space in population(s) at a given site. It was appropriate for EPA to inform the recommended ten-year exceedance frequencies for the chronic tissue-based criteria given the large variation in possible biological and physical variable influencing ecological recovery.

8) Please provide any additional technical comments that you believe should be considered.

Additional suggestions are listed below:

1- The species listed in the table is *Mytilus galloprovincialis* not *M. edulis*

Table 0-1. The Three Most Sensitive Acute Estuarine/Marine Genera.

Ranked Below from Most to Least Sensitive.

Rank	Genus	Species	GMAV (mg/L PFOS)	Comments
	Mytilus[1]	Mediterranean mussel, *M. edulis* *Mytilus galloprovincialis*	> 1	Not a resident species in North America, but other species in this genus are resident, commercially, or ecologically important species

2. Page 115- second paragraph (values highlighted in red and underlined are not consistent)

The author reported 10-day growth and survival EC_{10}s for the study were 0.0492 and 0.1079 mg/L, respectively. The study authors also reported NOECs of 0.0491 mg/L, LOECs of 0.0962 mg/L, and MATCs of 0.0687 mg/L for both endpoints. And the author reported 20-day EC_{10}s for growth, survival, and total emergence were 0.0882, 0.0864, and 0.0893 mg/L, respectively. And the study authors also reported NOECs of 0.0217 mg/L for growth and survival and < 0.0023 mg/L for emergence, LOECs of 0.0949 mg/L for growth and survival and 0.0217 mg/L for emergence, and MATCs of 0.0454 mg/L for growth and survival and 0.0071 mg/L for emergence. Also, it should be noted, the paper reported contrasting NOECs for 20-day survival. The text in the paper stated that the NOEC was 0.0271 mg/L and Table 2 of the paper stated 0.0949 mg/L. EPA assumed the NOEC in Table 2 of the paper was not correct and that 0.0217 mg/L was the correct NOEC based on the data presented in Figure 3A of the paper. This assumption was applied to the summary of the study results presented in this PFOS draft criteria.

3. Page 138-middle of the paragraph The chronic freshwater criteria also contain tissue-based criteria expressed as 43.0 mg/kg wet weight (ww) for fish whole-body, 25.3 mg**/** ---ww for fish muscle tissue and 12.3 mg/kg ww for invertebrate whole-body tissue.

4. Page A-21 last paragraph The noted toxicity values provided in each study summary above (ADD NUMBERS), comprising of both author-reported and independently-calculated LC_{50} values, were used to

calculate the GMAV value (as the geometric mean of the three LC_{50} values previously mentioned) of 22.48 mg/L, which was used to derive the freshwater aquatic life criterion.

5. Page A-24- Fourth line from bottom-The study author reported LC50 was 22.2 ± 4.6 mg/L for PFOS. The independently-calculated toxicity value was x.xx mg/L. The study author reported value was used quantitatively to derive the draft acute water column criterion.
6. Page A-25- Fourth line from bottom The study author reported 96-hour LC_{50} was 50.51 mg/L PFOS. The independently-calculated toxicity value was x.xx mg/L. The study author reported value was used quantitatively to derive the draft acute water column criterion.
7. Page A-27- Fourth line from bottom. The independently-calculated toxicity value was x.xx mg/L. The study author reported value was used quantitatively to derive the draft acute water column criterion.
8. Page A-29- First paragraph For comparison, the 7-day LC50 was 39.71 mg/L. The independently-calculated toxicity value was x.xx mg/L. The 96-hour study author reported value was used quantitatively to derive the draft acute water column criterion.
9. Page A-30- First paragraph The independently-calculated toxicity value was x.xx mg/L. The study author reported value was used quantitatively to derive the draft acute water column criterion.
10. Page A-36- 5th line from bottom in complete data x.xx mg/L.
11. Also at A-37 in complete data x.xx mg/L.

Table 0-3. Summary of Assessment Endpoints and Measures of Effect Used in the Criteria Derivation for PFOS

Assessment Endpoints for the Aquatic Community	Measures of Effect
Aquatic Life: Survival, growth, and reproduction of freshwater and estuarine/marine aquatic life (i.e., fish, amphibians, aquatic invertebrates)	For effects from acute exposure: 1. LC_{50} concentrations in water, diet, and/or tissue (e.g., muscle, blood, egg) 2. NOEC and LOEC concentrations in water, diet, and/or tissue (e.g., muscle, blood, egg) For effects from chronic exposure: 1. EC_{10} concentrations in water, diet, and/or tissue (e.g., muscle, blood, egg) 2. NOEC and LOEC concentrations in water, diet, and/or tissue (e.g., muscle, blood, egg); *Only used when an EC_{10} could not be calculated for a genus.* *Note: only chronic exposures were considered for derivation of the tissue-based criteria since PFOS is a bioaccumulative chemical. These chronic tissue-based criteria are expected to be protective of acute effects, because acute effects were observed at much greater concentrations than chronic effects.*

Please review if the highlighted muscle, blood and egg would be relevant to this section in terms of LC50, EC10, LOEC and NOEC endpoints .

12. 1.1.2 and page 3- Previously Published Chronic Water Criteria for Direct Aqueous Exposure

The information on Australian guidelines to be updated based on NEMP2 published in 2020. I will attach it as a PDF. https://www.environment.gov.au/system/files/resources/2fadf1bc-b0b6-44cb-a192-78c522d5ec3f/files/pfas-nemp-2.pdf

> *"Previously published freshwater chronic values were available for two states (Minnesota and Michigan) and three countries or geographic regions (Australia/New Zealand, Canada, and Europe). These publicly available values for other jurisdictions were 0.019 mg/L and 0.14 mg/L for Minnesota (STS/MPCA 2007) and Michigan (EGLE 2010), respectively, and were 0.00013 mg/L in Australia/New Zealand (CRC CARE 2017; EPAV 2016), 0.00680 mg/L in Canada (ECCC 2018), and 0.000023 mg/L in Europe (RIVM 2010). Previously published estuarine/marine chronic values were available for two geographic regions (Australia/New Zealand and Europe). These publicly available values were 0.0000046 mg/L in Europe (RIVM 2010) and 0.0078 mg/L in Australia/New Zealand (CRC CARE 2017; EPAV 2016)"*

The CRC marine guidelines are not valid as they are not based on the framework Freshwater values are to be used on an interim basis

13. Page 4- Table 1.1 to be updated accordingly

Exposure scenario	PFOS	Exposure scenario	Comments and source
Freshwater	0.00023 µg/L	99% species protection - high conservation value systems	Australian and New Zealand Guidelines for Fresh and Marine Water Quality - technical draft default guideline values for PFOS and PFOA. Note 1: The 99% species protection level for PFOS is close to the level of detection. Agencies may wish to apply a 'detect' threshold in such circumstances rather than a quantified measurement. Note 2: The draft guidelines do not account for effects which result from the biomagnification of toxicants in airbreathing animals or in animals which prey on aquatic organisms. Note 3: The WQGs advise[a] that the 99% level of protection be used for slightly to moderately disturbed systems. This approach is generally adopted for chemicals that bioaccumulate and biomagnify in wildlife. Regulators may specify or environmental legislation may
	0.13 µg/L	95% species protection - slightly to moderately disturbed systems	
	2 µg/L	90% species protection - highly disturbed systems	
	31 µg/L	80% species protection - highly disturbed systems	

			prescribe the level of species protection required, rather than allowing for case by-case assessments.
Exposure scenario	**PFOS**	**Exposure scenario**	**Comments and source**
Interim marine	0.00023 µg/L	99% species protection - high conservation value systems	As above. Freshwater values are to be used on an interim basis until final marine guideline values can be set using the nationally-agreed process under the Australian and New Zealand Guidelines for Fresh and Marine Water Quality. The WQG advise that in the case of estuaries, the most stringent of freshwater and marine criteria apply, taking account of any available salinity correction. Marine guideline values developed by CRC CARE are under consideration through the nationally-agreed water quality guideline development process.
	0.13 µg/L	95% species protection - slightly to moderately disturbed systems	
	2 µg/L	90% species protection - highly disturbed systems	
	31 µg/L	80% species protection - highly disturbed systems	

[a]https://www.waterquality.gov.au/anz-guidelines/guideline-values/default/water-quality-toxicants/local-conditions#bioaccumulation

14. Table 1 2. Two Primary Categories of PFAS

Please refer to OECD 2021 to be consistent with PFAS terminology/nomenclature

OECD (2021), Reconciling Terminology of the Universe of Per- and Polyfluoroalkyl Substances: Recommendations and Practical Guidance, OECD Series on Risk Management, No. 61, OECD Publishing, Paris

15. Table 1.3 Page 8

Please review Figure 9 OECD 2021 (also attached as PDF)

16. Conceptual Model of PFOS in the Aquatic Environment and Effects

Figure 2.9 page 77- Growth as an endpoint missing in the endpoints – first pentagon

COMMENTS SUBMITTED BY

REVIEWER 5

External Peer Review of EPA's Draft Aquatic Life Water Quality Criterion for Perfluorooctane Sulfonate (PFOS)

1) Please comment on the overall clarity of the document as it relates to the derivation of each criterion.

RESPONSE: Overall, the document is clearly written and generally free of grammatical errors. I applaud the scientists and EPA for compiling an impressive amount of work and communicating it in a reasonably clean and coherent way. That said, there are a few instances of redundancy – literally, the same sentences repeated. I have made note of these in the actual report and will include that along with this document, if requested. Although these are easy enough to see with careful review. The document is VERY LONG and very detailed so any efforts to shorten or make more concise would be welcomed.

As for the clarity of technical elements of the document, I feel that many of the decisions to use or not use data or endpoints could be more consistent and/or communicated better. For example, in some cases the geometric mean of endpoints for a certain taxa is used for the chronic value (pimephales) but for other taxa, this is not the case (e.g. daphnia; zebrafish). In other cases, the decision to not use certain data seems as if it could be communicated more clearly. So while the language of the document is pretty clear, the actual technical aspects are less so.

One major concern I have is with the overall approach of using the 4 most sensitive toxicity values to then derive the final acute and final chronic values. Using this approach it would seem the AWQC are then very sensitive to changes in any 1 of the 4 toxicity values. For example, when EPA explored the impact of different toxicity values on the chronic freshwater water column criteria, using higher toxicity values (e.g., pimephales in place of fatmucket Table 4-3) resulted in a lower chronic criteria. This is nonintuitive and suggests a possible flaw in the approach. It's possible this is not the case and it makes sense both mathematically (steeper slope) and perhaps even from a protection standpoint. Either way, EPA should explain why this happens and what it means for the overall approach. I suspect the EPA is somewhat constrained by the 1985 guidelines in developing the AWQC but I also see that New Approach Methods (WEB-ICE) were used to derive criteria with limited data. I wonder if using a species sensitivity distribution approach in which all the chronic or acute (freshwater)data are used would result in more defensible criteria that are less impacted by changes in any one toxicity value? At the very least, I think including a full SSD would be useful for comparison as part of the characterization piece. In the PFOA document I mention revisiting and publishing and updated 1985 guidelines…this is warranted when EPA has the bandwidth to do so. Having said all this, I am aware that EPA likely has justification for their approach of using the 4 most sensitive tox values but it would perhaps be good to mention this again as I suspect a lot of people that may read the AWQC will not also read the 1985 guidelines.

Lastly, I had a hard time keeping track of all the decisions to use or not use data for each of the tox values that supported the criteria. I think a more detailed table with all the tox values considered (data shown in Figure 3-3) and including whether the data used were author-reported, re-calculated by EPA, along with the lowest reported/calculated value that wasn't used and why. This may be asking a lot and this information is throughout the document but not in a single, easy to locate and read location.

2) Please comment on the approach used to derive the draft criterion for PFOS. Please provide detailed comments.

- **Is the technical approach used to derive the criterion elements logical?**

RESPONSE: Overall, I think the technical approach is relatively sound although there are instances where it was difficult to keep track of all the decisions with regard to data and whether these were consistent and logical. Admittedly, I think this is a tough chemical and a tough dataset and EPA did an excellent job with the background material and highlighted and used key studies (but more have been published since and will, I'm sure be included). Unfortunately, the technical approach to derive criterion elements is not universally logical. Moreover, as mentioned, using only the 4 most sensitive toxicity endpoints followed by a regression (what type? Was this specified?) seems less robust than using a full species sensitivity distribution with a more "natural" distribution of sensitivity (s-shaped, for example). This last statement may not be true so a reasonable compromise might be to include a full SSD as part of the characterization piece related to "considering other toxicity values impact on the FCV, etc.". What really confused me was that when EPA did what amounts to a sensitivity analysis of the FCV by replacing toxicity values, the FCV DECREASED when higher toxicity values were used. While I suspect this happened because switching to higher toxicity values steepened the slope (or something), it does not make intuitive sense to me and should be further explained. Alternatively, an explanation and justification, even brief, would be helpful in supporting the 4 most sensitive toxicity value approach. I am aware that the 1985 guidelines may include this but I suspect most users of the AWQC may not be familiar with the details of the guidelines.

With regard to the tissue-based criteria, EPA mentions using "only PFOS studies in which organisms were exposed in the diet" (or similar; p. 88) but then go on to say the BAF approach was used. I would edit this section to start with mentioning that a BAF approach was used because there were not enough tissue data from laboratory studies. I mention this because it was confusing – there was a lot of explanation of using only dietary exposures and then one sentence (basically) stating…EPA explored a BAF approach.

- **Does the science support the conclusions?**

RESPONSE: Well, offhand, I think the final chronic value for freshwater organisms should likely be lower. Importantly, several studies have been published in 2021 that should likely be included as toxicity values and they may result in lower toxicity estimates. The fact that EPA's criteria are higher than all other published criteria is worrisome. We are all using the same data and many in the field are quite capable scientists.

These two papers come immediately to mind but I am sure there are others.

> Sensitivity and Accumulation of Perfluorooctanesulfonate and Perfluorohexanesulfonic Acid in Fathead Minnows (*Pimephales promelas*) Exposed over Critical Life Stages of Reproduction and Development J.G. Suski, C.J. Salice, M.K. Chanov, J. Ayers, J. Rewerts, J. Field Environmental Toxicology and Chemistry, 2021, pp. 811-819.
>
> Toxicological Response of *Chironomus dilutus* in Single-Chemical and Binary Mixture Exposure Experiments with 6 Perfluoralkyl Substances Christopher J. McCarthy, Shaun A. Roark, Demitria Wright, Kelly O'Neal, Brett Muckey, Mike Stanaway, Justin N. Rewerts, Jennifer A. Field, Todd A. Anderson, Christopher J. Salice Environmental Toxicology and Chemistry, 2021, pp. 2319-2333.

In my view, it is essential that EPA incorporate newly published toxicity data for PFOS (and PFOA).

Furthermore, in several cases, EPA's decisions to use what look like higher estimates of toxicity seem somewhat arbitrary and not internally consistent. I also noted above and mention here again the sensitivity of the criteria development approach to changes in one of the 4 most sensitive taxa/toxicity values.

- **Is it consistent with the protection of freshwater aquatic life from acute, chronic, and bioaccumulative effects?**

RESPONSE: I believe the criteria are "NEARLY" protective of freshwater aquatic life for acute, chronic and bioaccumulative effects of PFOS. I say "nearly" because it seems to me that the FCV, in particular, could and maybe should likely be lower. Also, below I comment on the appropriateness and utility of the frequency and duration elements of the criteria. Briefly, in my opinion the frequency and criteria elements of the criteria certainly help the criteria concentrations to be protective; it is unlikely that a 4-day exposure to the FCV would result in adverse effects to any taxa for which there are data; however, these data are not commonly reported (hourly or 4-day running average concentrations have never been reported to my knowledge).

3) **Please comment on the approach used to derive the draft acute estuarine/marine benchmark for PFOS. Given the limited estuarine/marine test data available, a new approach method was used to support the derivation of an acute estuarine/marine benchmark to provide states and tribes with a protective value. Please provide detailed comments.**

- **Is the technical approach used to derive the benchmark logical?**

RESPONSE: Yes, given the lack of PFOS toxicity data for acute estuarine/marine species, I think applying WEB-ICE is a REASONABLE APPROACH…perhaps the only approach that is defensible. Clearly, more (or some) data would be a wonderful contribution. WEB-ICE, as mentioned, has been reviewed and published quite a bit so I think, as an approach, it has merit and support of the scientific community. EPA also did a good job presenting the approach and being clear about the criteria being a draft. Overall, when data have been lacking, EPA has used state-of-the-art approaches to developing criteria (my concerns are mostly when sufficient data are available).

- **Does the science support the conclusions?**

RESPONSE: Yes, the science supports the conclusions. Interestingly, the acute criteria for estuarine/marine species (o.43 mgPFOS/L) is almost an order of magnitude lower than the acute criteria for freshwater organisms (3.6 mg/L). Whether estuarine/marine species are truly more sensitive remains to be seen but, to me, it is more reasonable, given the lack of data, that the criteria draft is lower.

- **Is it consistent with the protection offered by acute estuarine/marine aquatic life criteria derived using empirical data, as prescribed in the 1985** *Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses***?**

RESPONSE: Yes, and EPA justified this in the explanation of WEB-ICE that occurs in Appendix L and, overall, the approach and resulting criteria are consistent with the protection of estuarine and marine species.

4) **Please comment on the use of measured and unmeasured toxicity tests to derive the respective criterion. In particular please comment on the supporting justification for using unmeasured toxicity tests in Appendix O.**

RESPONSE: I think the comparison of measured and nominal concentrations was an interesting read and a useful contribution. That said, many toxicologists focused on PFAS have commented that analytical confirmation is necessary for a high quality study – this was echoed (loudly) at the SETAC Workshop on Risk of PFAS that occurred in summer, 2019. As well, in my own experience there have been challenges in sometimes matching nominal and measured concentrations for aquatic exposures. The paper by Rewerts et al. 2020 highlights some of the challenges and provides recommendations for accurate solutions of PFAS. As a general rule, we have erred on the side of reporting measured concentrations.

Two important thoughts. First, several very prominent analytical chemists that have made a career of measuring PFAS have indicated to me that the analytical method is only about 30% accurate – meaning that if the analytical measure was +/- 30% of nominal, they would be considered "the same". EPA used 20% as a threshold (for deciding nominal and measured were the same) and I'm not sure why this is. As far as I can tell, 30% is a more reasonable threshold.

Second, in the review and derivation of toxicity values for the MacDonald et al. 2014 paper, EPA elected not to use the 20-day emergence rate endpoint, in part, because the nominal and measured did not agree. This makes no sense to me. As long as the solutions were confirmed analytically and reported, that should be good enough and, in fact, preferred over nominal alone.

Paper worth including in the section on nominal vs. measured PFOS concentrations:

> Key Considerations for Accurate Exposures in Ecotoxicological Assessments of Perfluorinated Carboxylates and Sulfonates. Justin N. Rewerts, Emerson C. Christie, Alix E. Robel, Todd A. Anderson, Christopher McCarthy, Christopher J. Salice, Jennifer A. Field Environmental Toxicology and Chemistry, 2020

5) Please comment on the toxicity data used to derive the draft criteria.

- **Were the data selected and/or excluded from the derivation of the criteria derivation appropriately utilized?**

RESPONSE: As mentioned, I feel that there are some inconsistencies with how some data were included or excluded. In the previous comment, for example, some data were excluded from the MacDonald et al. 2014 paper because there was some disagreement between nominal and measured. With regard to PFAS, I would say measured is almost always better than nominal and the fact that these sometimes don't agree should not be too big of a deal as long as they are not wildly different. EPA put substantial effort into sometimes justifying nominal – in all cases, excluding studies that had analytical confirmation is less defensible than including studies that only report nominal, in my opinion. This last statement is, of course, provided the analytical methods are robust.

- **Are there relevant data that you are aware of that should be added to the analyses (note that EPA is working on updating the toxicity data to reflect the data in ECOTOX between Sept. 2019 through the latest update)? If so, please provide references for consideration.**

  Sensitivity and Accumulation of Perfluorooctanesulfonate and Perfluorohexanesulfonic Acid in Fathead Minnows (*Pimephales promelas*) Exposed over Critical Life Stages of Reproduction and Development. J.G. Suski, C.J. Salice, M.K. Chanov, J. Ayers, J. Rewerts, J. Field Environmental Toxicology and Chemistry, 2021, pp. 811-819.

  Toxicological Response of *Chironomus dilutus* in Single-Chemical and Binary Mixture Exposure Experiments with 6 Perfluoralkyl Substances. Christopher J. McCarthy, Shaun A. Roark, Demitria Wright, Kelly O'Neal, Brett Muckey, Mike Stanaway, Justin N. Rewerts, Jennifer A. Field, Todd A. Anderson, Christopher J. Salice. Environmental Toxicology and Chemistry, 2021, pp. 2319-2333.

In particular, please comment on:

5a. The toxicity values used to derive the PFOS criteria, with a particular emphasis on:

i. the use of the qualitatively acceptable acute midge (*Chironomus plumosus*) data from Yang et al. (2014) to suggest aquatic insects are relatively tolerant to acute PFOS exposures. Specifically, Yang et al. (2014) conducted a 96-hour renewal, measured PFOS acute test with the midge,

***Chironomus plumosus*. This study was not acceptable for quantitative use due to the potential problematic source of the organisms. The reported LC_{50} was 182 mg/L for PFOS indicating that insects may not be one of the more sensitive taxonomic groups. Therefore, this test was excluded from the acute criterion calculation, but used to waive the missing insect MDR.**

RESPONSE: Given that insects are among the most sensitive organisms for the chronic exposures to PFOS, it seems the Yang et al. 2014 paper is not very consistent with the prevailing data. Additionally, McCarthy et al. 2021 reports toxicity to chironomids similar to that of MacDonald et al. Additionally, while the Olson 2017 data for Aedes species was not acceptable (for valid reasons), nonetheless the study shows very high sensitivity of another insect species to acute exposures to PFOS. That said, given the EPA's stance and justification that "nominal generally equal measured PFOS concentrations", I'm inclined to put more confidence in Olson's study. Same for the 20-day data in the MacDonald et al. paper on chironomids. In that case, there was a "relatively large difference between measured and nominal concentrations (p. 278)" and so the data were not used. This seems odd to me – as long as there are measured data, that's what I would suggest using. Regardless, the MacDonald et al. paper points to the sensitivity of insects so, collectively, I'd be disinclined to say that the Yang et al. paper shows insects are not sensitive and the data requirement can be waived. I wonder if it's possible to somehow estimate an acute toxicity value for aquatic insects based on chronic toxicity data? Basically, a reverse of the Acute/Chronic ratio approach.

ii. the use of the quantitatively acceptable chronic toxicity value for mussel (*Lampsilis siliquoidea*) from Hazelton et al. (2012). Specifically, Hazelton et al. (2012) conducted a 36-day renewal, measured PFOS chronic test with fatmucket, *Lampsilis siliquoidea.* The estimated EC_{10} was 0.05713 mg/L, which was extrapolated from the author-reported data and the exposure response slope from another PFOS toxicity study focused on another mussel species (*Ellipto complamata*) as explained in Section 3.1.1.3.3. Therefore, this test was used in the chronic criterion calculation.

RESPONSE: Unfortunately, I cannot find the Drottar et al. (2000) paper which is the basis of estimating the EC10 from the EC35 generated in the Hazelton et al. (2012) study. And..I think the citation in the document is incorrect and this should be Drottar and Kreugar (2000g)… or check to make sure the citations in text and references match. Moroever, the Drottar paper appears to be an acute test which is VERY different than the Fatmucket study. This approach seems like a "stretch" and, again, somewhat inconsistent with the approaches and decision matrix EPA has used to utilize or discard other data and endpoints.

iii. the use of the quantitatively acceptable chronic toxicity value for damselfly (*Enallagma cyathigerum*) from Bots et al. (2010). Bots et al. (2010) conducted a 320-day renewal, unmeasured PFOS chronic test with blue damselfly nymphs, *Enallagma cyathigerum.* The MATC was 0.03162 mg/L, which was calculated from the author-reported value for nymph survival as explained in Section 3.1.1.3.2. Therefore, this test was used in the chronic criterion calculation.

RESPONSE: Well, clearly it would have been better to be able to estimate an EC10 but this appears appropriate. I note that for this study the exposure concentrations were an order of magnitude apart; in other cases, EPA has used "too big of a difference between exposure concentrations" to discard a study or two. Somewhere, it would be good to know at what point there is too great a difference among exposure concentrations (10x, 20x, ?) for the study to be deemed acceptable for use quantitatively.

iv. the use of the quantitatively acceptable chronic toxicity value for midge (*Chironomus dilutus)* from MacDonald et al. (2004). MacDonald et al. (2004) conducted a 20-day renewal, measured PFOS chronic test with midge lava, *Chironomus dilutus.* The EC_{10} was 0.05963 mg/L, which was

an EPA-calculated value for 10-day growth as explained in Section 3.1.1.3.4. Therefore, this test was used in the chronic criterion calculation.

RESPONSE: I do not agree with the toxicity value used by EPA as obtained from the MacDonald et al. study. The lowest toxicity value is the MATC for 20-day emergence of 0.0071 mg PFOS/L. It is not clear why EPA did not use this value? Emergence is clearly extremely ecologically relevant, and the value generated seems as defensible as most of the other endpoints EPA has chosen to include?

Additionally, and as mentioned above, see the paper by McCarthy et al. (2021) that was just published in Environmental Toxicology and Chemistry. Those data appear robust and should meet acceptability criteria.

5b. EPA's approach for fitting concentration-response (C-R) data (described in Appendix K) as well as the specific acute LC_{50} values (Appendix A.2) and chronic EC_{10} values (Appendix C.2) that were estimated (for sensitive genera when C-R data were available) and used to derive criteria.

RESPONSE: In general, the approach for fitting C-R data that EPA used is basically state-of-the-art. The drc package is very powerful and provides a way to test many different curves to then select the best fit model. Although EPA described some of this in the several sections related to "fitting x data (K 1.2)", I think more details would be warranted. The description for the criteria to select best fit models is rather vague. Perhaps a table of specific fit criteria would be helpful? Perhaps this is not doable because every dataset is different.

When I teach modules on Akaike Information Criteria (AIC) I emphasize that the metric "penalizes" fit for more parameters within a model. So, using AIC can yield the simplest, best model that fits the data. This is because models with more parameters tend to yield a better fit purely based on statistical properties and not the actual phenomena being studied. I am not aware that AIC is a measure of the model fit to "true outcomes" which are only theoretical constructs, I think. If we knew the "true outcomes" we would not really need the model. Anyway, I would encourage the authors to review the AIC section and make edits if necessary and certainly **cite the source of the explanation.**

For section K.2.2. are there actually any criteria (i.e., numbers) that are used to determine when a model fit is appropriate? As a simple example, maybe one would consider an r^2 of 0.8 or better to be a "good model" for linear regression? Some statements to this effect and any details regarding actual criteria used to select "good models" would be helpful. So, overall the curve fitting approach is appropriate but more, specific details would be helpful.

6) Please comment on the translation of the chronic water column criterion elements for aquatic life to derive the tissue-based criterion elements, considering the bioaccumulation of PFOA and PFOS. In particular, please comment on:

6a. Uncertainty surrounding the bioaccumulation factors (BAFs) used to translate of the chronic water column criterion elements into tissue-based criterion elements.

RESPONSE: Using the BAF for PFOS to determine the tissue-based criterion elements is, I think, an interesting and useful approach given the lack of tissue-based metrics associated with toxicity data. The variability in observed bioaccumulation of PFOS is an active area of research but the work by Burkhard 2021 (also an author on the AWQC) provides an excellent synthesis and compendium of available BAFs for PFOS. That said, I noticed that the criteria for co-located tissue and water samples for PFOS was that they were collected within a year of each other and within 2 km distance (p. 134) – this likely contributes to significant variability in the BAFs. Although there are few published datasets, there are some datasets that show considerable temporal and spatial variability in PFAS water concentrations over the course of a few weeks and over a spatial distance of less than 0.5 km. I wonder if the variability in BAFs would decrease if the

criteria were narrowed to co-collected samples measured at the same time? Might be worth the exercise. Given the variability in PFOS BAFs, why not use something like the 25% percentile BAF instead of the geometric mean? When developing a protective criteria and there is a very noisy data set, it might be beneficial to err on the side of caution until better data (many co-located samples in space and time) were available. So, in summary, I think the approach of using BAFs to estimate tissue-based criteria is reasonable but given the variability in BAFs, I would encourage using a lower BAF instead of the geometric mean or reconsidering the data that went into the BAFs used for criteria development.

6b. EPA's determination of appropriate BAFs and the tissue types that the tissue criterion elements were based.

RESPONSE: Invertebrate whole body, fish whole body, and fish muscle are appropriate tissues for the tissue-based criterion. These are the most commonly collected tissue types and are relevant to monitoring efforts and are even useful for considerations of fish advisories. That said, the only other tissue worth considering would be for liver in fish since this tissue accumulates considerably more PFOS than muscle – these are included in the appendices so this is appropriate. Overall, tissues used for the tissue-based criteria are appropriate.

7) Please comment on the frequency and duration of the criterion elements, in particular the tissue-based criterion elements.

RESPONSE: In my opinion, the frequency and duration of criterion elements is among the most uncertain and potentially contentious elements of any type of protective criteria. The frequency and duration for tissue-based criteria is that the tissue-based criteria cannot be exceeded more than once in a 10 year period. This means that if the PFOS criterion for whole body in fish of 43 mg/kg bw is exceeded more than once in a 10 year period then the criteria is exceeded. This also means the fish was likely exposed to the 0.014 mg/l concentration for longer than an instantaneous exposure and likely longer than 4 days. So, to me, does this not mean that if 43 mg/kg bw was measured in a fish tissue, then the fish was likely exposed to 0.014 mg PFOS/L for longer than 4 days, doesn't it? And, this also means that if fish whole body concentrations were 42.5 mg/kg bw for 10 years, the criterion would not be exceeded. I would suspect that long term PFOS exposures that consistently lead to 42.5 mg/kg bw in fish would likely translate to adverse ecological impacts in some biota present in the same system. When I think of it this way, these criteria do not seem appropriately protective. In my view, the water column continuous exposure criteria should be adjusted downward which would then translate to a lower tissue-based criteria which might be more reasonable. Although, as mentioned, another protective approach would be to use something like the 25th percentile BAF or something other than the mean. That said, at least in many cases fish tissue monitoring occurs on a yearly basis so there is some potential for the criteria to be reasonably assessed against environmental data. It is still possible that fish tissue concentrations could be exceeded every year and this be missed by monitoring efforts. Nonetheless, because tissue concentrations are an intergrative measure and because many monitoring programs probably do measure fish every year, this is a better match than the water column criteria.

When we consider the acute and chronic water column criteria, the frequency and duration elements are protective, in my opinion. The problem is that nobody knows if the criteria for acute toxicity are exceeded for more than 1 hour or whether the chronic criteria was exceeded for more than 4 days – this extent of temporal resolution (hourly concentrations or 4-day running averages) just does not exist. So while I agree that conceptually, the frequency and duration elements definitely would add to the protection of aquatic life…I just don't see how these can be implemented or regulated. Perhaps EPA is aware that my concern is not warranted because in relevant circumstances, appropriately timed environmental data are obtained.

8) Please provide any additional technical comments that you believe should be considered.

RESPONSE: Specific comments to the various elements of the PFOS AWQC are above. Here, I want to suggest that EPA revisit the 1985 Guidelines and publish either an updated version or an amendment. Basing critically important criteria on documents published in 1985 and then using this to justify decisions seems like it would not pass muster in the scientific community. I've had papers rejected because they did not include enough recent citations, for example. Moreover, I've mentioned my concerns with the 4- most sensitive taxa + linear regression for criteria derivation. No paper I've read on generating the 5^{th} percentile most sensitive species has used this approach. Granted, I may have missed them but my sense is that it is more common to use a full SSD. It would be helpful, for example, if the revised Guidelines explored this further or other means of criteria development (including new approach methods) and published, used, and cited and updated guidelines document. I'd like to think we still generally lead the world (more or less) in environmental protection so having an updated document would be welcomed.