Final Comment Report
External Peer Review of Acute Protective Freshwater Benchmarks for Selected PFAS Using a New Approach Methods

Contract No. 68HERH23A0021
Task Order 68HERH23F0371

Prepared for:

U.S. Environmental Protection Agency
Office of Water, Office of Science and Technology
1200 Pennsylvania Avenue, NW, MC 4304T
Washington, DC 20460
Attn: Mike Elias

Prepared by:

Versar, Inc.
1025 Vermont Ave NW Suite 500
Washington, DC 20005

Peer Reviewers:
David Buchwalter, Ph.D.
Anupama Kumar, Ph.D.
Håkon Austad Langberg, Ph.D.
Ryan Prosser, Ph.D.
William Stubblefield, Ph.D.

January 5, 2024

TABLE OF CONTENTS

I. INTRODUCTION

Versar Global Solutions, an independent contractor for the Environmental Protection Agency (EPA), coordinated an external letter peer review of the *Derivation of Acute Protective Freshwater Benchmarks for Selected PFAS through a New Approach Method* report. The peer review was conducted for EPA's Office of Water, Office of Science and Technology.

Background
The U.S. Environmental Protection Agency (EPA) Office of Water is charged with protecting human health and the environment from chemicals in water, under the purview of the Clean Water Act (CWA). In accordance with this mission, EPA is working to develop acute protective freshwater benchmark values for eight PFAS: Perfluorobutanoic acid (PFBA), Perfluorobutanesulfonic acid (PFBS), Perfluorononanoic acid (PFNA), Perfluorodecanoic acid (PFDA), Perfluorohexanoic acid (PFHxA), Perfluorohexanesulfonic acid (PFHxS), Hexadecafluoro-2-decenoic acid (8:2 FTUCA), and Pentadecafluorodecanoic acid (7:3 FTCA). Empirical data are limited for these eight PFAS chemicals to support the development of aquatic life criteria. EPA accordingly developed draft acute protective benchmarks for these eight PFAS using available freshwater species empirical test data in conjunction with the application of a New Approach Method (NAM), specifically EPA's Office of Research and Development's (ORD) peer-reviewed web-based Interspecies Correlation Estimate tool (Web-ICE; Version 4.0; https://www.epa.gov/webice/)(Raimondo et al. 2010). EPA additionally investigated the approach described in Giddings et al. (2019) to determine whether it could be applied to the above PFAS to derive protective values for carboxylic acid PFAS and sulfonic acid PFAS.

Versar conducted an independent search for scientific experts with expertise in one or more of the following disciplines: a) application of NAMs to the derivation of protective aquatic life benchmark values; b) toxicity of PFAS to aquatic life; c) aquatic ecotoxicology; and d) the acceptability of methods, statistical analyses and data interpretation applied to the determination of data and methods acceptability.

As a result of this search, Versar identified and contacted 25 experts. Of these experts, Versar received eight positive responses expressing interest and availability to participate. The remaining 17 experts were not available during the peer review timeframe or did not respond to the invitation. For each interested and available peer reviewer, Versar evaluated their qualifications and conducted conflict of interest (COI) screening to ensure that the experts had no COI.

Versar selected the following five scientific experts to serve as peer reviewers:

Peer Reviewers:

David Buchwalter, Ph.D.
North Carolina State University

Anupama Kumar, Ph.D.
CSIRO Environment, Australia

Håkon Austad Langberg, Ph.D.
Norwegian Geotechnical Institute, Norway

Ryan Prosser, Ph.D.
University of Guelph, Canada

William Stubblefield, Ph.D.
Oregon State University

II. CHARGE TO PEER REVIEWERS

1) Please comment on the overall clarity of the documents and construction as it relates to assessing the effects and derivation of acute benchmarks for the eight selected PFAS.

2) Please comment on **each of the technical approaches** used to derive the draft benchmark values for the eight selected PFAS presented in *EPA's* ***Derivation of Acute Protective Freshwater Benchmarks for Selected PFAS through a New Approach Method*** (ICE-based approach using extrapolation [Sections 3-5], ICE-based approach using scaled data [Appendix F], and PFAS benchmark calculations using a data binning approach [Appendix G]).

    a. Are the technical approaches used to derive the benchmark values logical?
    b. Does the science support the conclusions?
    c. Are the approaches and resulting values consistent with the protection of aquatic life?
    d. For the ICE-based models in particular, please compare and contrast the strengths and weaknesses of the extrapolation approach [Sections 3-5] with those of the approach using scaled data [Appendix F].

3) Please comment on the empirical (direct test) and ICE-generated toxicity data used to derive the benchmark values presented in the draft document.

    a. Were the data adequately used and sufficiently comprehensive to represent risks to sensitive aquatic life?
    b. Were the data selected and/or excluded from the benchmark values derivation appropriately utilized?
    c. Are there relevant data that you are aware of that should be included? If so, please provide for derivation of benchmark values.

4) Are the derived benchmark values appropriately protective of sensitive aquatic life?

III. PEER REVIEWER COMMENTS TABLE

I. General Impressions	
REVIEWER	**REVIEWER COMMENT**
1	This report describes the process and results for acquiring draft acute recommended freshwater aquatic life benchmarks for eight PFAS: PFBA, PFHxA, PFNA, PFDA, PFBS, PFHxS, 7:3 FTCA, and 8:2 FTUCA.
1	Acute benchmark values were derived using the procedure described in EPA's "Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses" where sensitivity distributions based on toxicity databases are used to derive acute freshwater criteria. Due to the lack of acceptable empirical data for constructing the sensitivity distributions, Interspecies Correlation Estimation (ICE) models were applied to complete the toxicity databases.
1	The applied methods are transparent and the methods have undergone previous peer review. The approaches and data used are reported in a clear manner, making it possible to examine the whole process of acquiring the benchmarks.
1	Some clarification would, however, strengthen the report and make it less likely that readers will misunderstand what is the appropriate use of these benchmarks: 1) It should be more clearly expressed in the summary that these benchmarks are for acute exposure, i.e., to protect aquatic life in freshwater from acute toxic effects of PFAS. For example, these benchmarks are not sufficient to prevent pollution of the environment, adverse effects of chronic exposure to aquatic animals, human exposure, etc. This is important information as the problems with PFAS pollution are often due to the combination of their persistency, mobility, potential for bioaccumulation, and long-term toxicological effects.
1	2) The assumptions for the use of the ICE models should be summarized in the main report.
1	3) The resulting benchmark values should be discussed in more detail. Especially trends for acute toxicity depending on PFAS group and chain length.
1	4) There is a lack of clear conclusions regarding the different methods explored for deriving the benchmarks.
2	Establishing environmental standards for the protection human health and the environment is critical. Unfortunately, our laws and regulatory approaches regarding the release of chemicals into the environment have allowed for the generation of thousands of chemicals used in commerce without the requirement that they be tested for safety. This leads us to our current situation where the pace of new chemistries being introduced to the world is far outpacing our ability to evaluate their toxicity to all forms of life including aquatic life. This document reflects this current state-of-affairs. PFAS as a chemical class are ubiquitous with new chemistries being introduced a rapid pace. Few environmental standards for their concentrations in surface waters have been established and toxicity data are limited.

I. General Impressions	
REVIEWER	**REVIEWER COMMENT**
	Methods to extrapolate toxicity data are unfortunately necessary in light of this data-limited situation. Here, the authors use a "New Approach Method" of Interspecies Correlation Estimation developed by Raimondo and others as a way of generating predictions of toxicity to aquatic organisms such that acute benchmarks for selected PFAS could be established based on the 1985 *Guidelines for the Protection of Aquatic Life*. As these compounds are generally not acutely toxic, and the benchmarks proposed in the document appear to be orders of magnitude higher than expected environmental concentrations, it is unclear what the rationale is for proposing these benchmarks in the absence of chronic benchmarks. By releasing these benchmarks as the only protective values available for the compounds in question, it is possible that the discovery of environmental concentrations that are well below these benchmarks (but could be chronically toxic) might not receive an appropriate response by states and tribes. I generally like the ICE approach to fill data gaps such that environmental standards can be set. It just does not feel like these acute standards are particularly relevant in light of the expected environmental concentrations.
2	In general, the document requires the reader to consult some of the original literature on ICE models and there is not enough explanation of these approaches contained within the document itself. I would encourage the authors to add more technical information about how ICE models work, what their limitations are, and perhaps build in some uncertainty factors given the language in the Forward section of the document.
3	The document is generally well-written and easy to follow, especially in the discussion of the data used and derivation of the benchmark. Greater detail and perhaps some examples would help in the discussion of the difference between "extrapolated" and "scaled" ICE-based data. It is not clear exactly what the difference is between these, or at least the implications of the two approaches other than to increase the number of species represented. Reviewing both the Raimondo et al (2010) and Willming et al (2016) papers did not address this topic. Raimondo et al (202?) is "in review" and perhaps will address the issues; however, the manuscript was not included in the review materials. The following discussion applies to the document and approach and was not specifically addressed in the charge questions.
3	To sum up the issue, the problem is that there is insufficient high-quality empirical data available to derive AWQC for the "selected" PFAS compounds. This issue is not new, data limitations in deriving Ambient Water Quality Criteria (AWQC) have been an issue since shortly after the implementation of the Clean Water Act (Kimerle et. al., 1985[1]) and became a greater concern with the reduction of AWQC data development at the EPA-ORD research laboratories. EPA previously proposed an approach to address this issue in the Type II standards methodology developed as part of

[1] *Kimerle, R. A., Werner, A. F., and Adams, W. J., "Aquatic Hazard Evaluation Principles Applied to the Development of Water Quality Criteria," Aquatic Toxicology and Hazard Assessment: Seventh Symposium, ASTMSTP854, R. D. Cardwell, R. Purdy, and R. C. Bahner, Eds., American Society for Testing and Materials, Philadelphia, 1985, pp. 538-547*

<table>
<tr><th colspan="2">I. General Impressions</th></tr>
<tr><th>REVIEWER</th><th>REVIEWER COMMENT</th></tr>
<tr><td></td><td>the Great Lakes Initiative (GLI) published in 1995 (USEPA 40 CFR 9, 122, 123, 131, and 132, Final Water Quality Guidance for the Great Lakes System; Final Rule, March 23 1995). EPA presented a method to develop Secondary Maximum Concentrations (SMC) and the Secondary Continuous Concentrations (SCC) based on data sets that were insufficient to satisfy the eight minimum data requirements (MDR) to derive a national AWQC. Briefly, a secondary acute value (SAV) is calculated by dividing the lowest GMAV in the database by a Secondary Acute Factor (SAF) that is designated in Table A-1 in the document (ranging from 4.3 to 21.9) based on the number of satisfied MDRs available for the compound. This approach is somewhat crude and certainly lacks a great deal of technical basis; nonetheless, it probably should be discussed in the current document. In addition, application of the GLI technique to the PFAS compounds in this document has been previously conducted and presented in Gripo et al (2021)[2]. Resulting values between the Gripo et al (2021) report and application of the method to EPA's data result in slightly differing values, likely due to the acceptance and availability of different empirical data. In Table 1, a comparison of the benchmarks reported in this document (using the ICE approach) is made with the values calculated using the GLI approach (based on the empirical data reported in this document).
Table 1. Comparison of benchmark values using the GLI and ICE-based derivation methods.
<table>
<tr><th>Chemical</th><th>Lowest empirical value</th><th>GLI factor</th><th>Tier II GLI calculated value (mg/L)</th><th>EPA Benchmark (Extrapolation) FAV/2 (mg/L)</th></tr>
<tr><td>Perfluorobutanoic acid (PFBA)</td><td>110</td><td>8</td><td>13.75</td><td>83</td></tr>
<tr><td>Perfluorobutanesulfonic acid (PFBS)</td><td>1938</td><td>13</td><td>149.1</td><td>183</td></tr>
<tr><td>Perfluorononanoic acid (PFNA)</td><td>27.84</td><td>13</td><td>2.14</td><td>10.3</td></tr>
<tr><td>Perfluorodecanoic acid (PFDA)</td><td>32</td><td>8</td><td>4</td><td>7.9</td></tr>
<tr><td>Perfluorohexanoic acid (PFHxA)</td><td>140</td><td>8</td><td>17.5</td><td>75</td></tr>
<tr><td>Perfluorohexanesulfonic acid (PFHxS)</td><td>22.5</td><td>13</td><td>1.7</td><td>9.1</td></tr>
<tr><td>Hexadecafluoro-2-decenoic acid (8:2 FTUCA)</td><td>3.2</td><td>13</td><td>0.24</td><td>0.58</td></tr>
<tr><td>Pentadecafluorodecanoic acid (7:3 FTCA)</td><td>0.959</td><td>13</td><td>0.074</td><td>0.18</td></tr>
</table></td></tr>
<tr><td>3</td><td>Other comments not specifically addressed in the charge questions are provided below:</td></tr>
</table>

[2] *M. Grippo, J. Hayse, I. Hlohowskyj, and K. Picel. 2021. Derivation of PFAS Ecological Screening Values. Environmental Science Division. Argonne National Laboratory. September 2021.*

<table>
<tr><th colspan="2">I. General Impressions</th></tr>
<tr><th>REVIEWER</th><th>REVIEWER COMMENT</th></tr>
<tr><td></td><td>

- In at least three cases (i.e., PFBS, PFNA, and PFHxS) the derived criteria, did not comply with the MDR minimum of n=8, it appears that EPA disregarded the MDR minimum for these materials and calculated the benchmark with fewer MDRs, i.e., 7. This is not keeping with the 1985 guidance that states: "*Similarly. if all required data are not available, a numerical criterion should not be derived except in special cases."* This should be acknowledged in the text and some statement regarding the minimum number of MDRs to calculate a benchmark addressed.

</td></tr>
<tr><td>3</td><td>

- Table 2 summarizes the MDRs available for each of the 8 PFAS compounds. In all cases 62.5 to 75% of the MDR data used in deriving the SSD-based benchmark was estimated using the ICE model method. Thus, most of the data used for MDRs and to develop SSDs, are estimated values. It is interesting that in the best case, only 1 empirical data point was available among the 4 most sensitive species for 50% of the materials. The other 4 materials had no empirical data represented among the most sensitive species. Although there may be good correlations between species making estimation for one species based on data from another possible, questions remain regarding the extent of the role that estimated values should play in the derivation of water quality criteria, standards, or benchmarks.

Table 2. Numbers of MDRs used in derivation of benchmarks

Chemical	# of empirical MDRs	# of estimated MDRs	Total MDRs met	# quantitively accepted empirical data in 4 lowest species
Perfluorobutanoic acid (PFBA)	3	5	8	1
Perfluorobutanesulfonic acid (PFBS)	2	5	7	0
Perfluorononanoic acid (PFNA)	2	5	7	1
Perfluorodecanoic acid (PFDA)	3	5	8	0
Perfluorohexanoic acid (PFHxA)	3	5	8	1
Perfluorohexanesulfonic acid (PFHxS)	2	3	5	1
Hexadecafluoro-2-decenoic acid (8:2 FTUCA)	2	6	8	0
Pentadecafluorodecanoic acid (7:3 FTCA)	2	6	8	0

</td></tr>
</table>

I. General Impressions	
REVIEWER	**REVIEWER COMMENT**
3	• If EPA is going to revise the AWQC minimum MDR data requirement (i.e., 8) for the purpose of "Benchmark" derivation, then EPA should develop guidance regarding the minimum quantity and quality of **empirical** data required before a benchmark can be derived; this should potentially include: o a minimum number of empirical data required to be contained among the 4 lowest species (should benchmarks should be derived based solely on "estimated" values?). Is there a minimum percent of empirical to estimated data that should be met to establish a benchmark? o a requirement for empirical data with a representation among a base set of organisms that have historically been shown to be sensitive to a range of toxicants, e.g., *Ceriodaphnia dubia*, fathead minnow, would be useful. Kimerle et al (1985)[3] suggested that a minimum base data set composed of an algae, daphnid, and fish could consistently predict the most sensitive species based on available data at the time. Use of the ICE database could help to identify species consistently shown to be among those predicted to be most sensitive among chemical groups with a common mode-of-action. The table below, composed from the ICE-modeled data for PFAS compounds, suggests that a base data set composed of a freshwater mussel, cladoceran, and fish, would cover the majority of predicted most-sensitive species for PFAS compounds. o a minimum number of empirically derived MDR data points should be established, and a maximum number of ICE-estimated values that are allowed to be considered in an SSD should be established.

[3] *Kimerle, R. A., Werner, A. F., and Adams, W. J., "Aquatic Hazard Evaluation Principles Applied to the Development of Water Quality Criteria," Aquatic Toxicology and Hazard Assessment: Seventh Symposium, ASTMSTP854, R. D. Cardwell, R. Purdy, and R. C. Bahner, Eds., American Society for Testing and Materials, Philadelphia, 1985, pp. 538-547.*

<table>
<tr><th colspan="2">I. General Impressions</th></tr>
<tr><th>REVIEWER</th><th>REVIEWER COMMENT</th></tr>
<tr><td></td><td>
Table 3. Comparison of the species sensitivity ranking from the ICE model for each of the PFAS compounds
<table>
<tr><th rowspan="2">Chemical</th><th colspan="4">Species Sensitivity Rank from ICE model</th></tr>
<tr><th>1</th><th>2</th><th>3</th><th>4</th></tr>
<tr><td>Perfluorobutanoic acid (PFBA)</td><td>Brachionus calyciflorus (rotifer)</td><td>Oncorhynchus mykiss (rainbow trout)</td><td>Gammarus fasciatus (amphipod)</td><td>Amblema plicata (mussel)</td></tr>
<tr><td>Perfluorobutanesulfonic acid (PFBS)</td><td>Amblema plicata (mussel)</td><td>Gammarus fasciatus (amphipod)</td><td>Hyalella azteca (amphipod)</td><td>Ceriodaphnia dubia (cladoceran)</td></tr>
<tr><td>Perfluorononanoic acid (PFNA)</td><td>Amblema plicata (mussel)</td><td>Chydorus sphaericus (cladoceran)</td><td>Megalonias nervosa (mussel)</td><td>Oncorhynchus mykiss (rainbow trout)</td></tr>
<tr><td>Perfluorodecanoic acid (PFDA)</td><td>Caecidotea brevicauda (isopod)</td><td>Micropterus salmoides (bass)</td><td>Perca flavescens (yellow perch)</td><td>Salvelinus fontinalis (brook trout)</td></tr>
<tr><td>Perfluorohexanoic acid (PFHxA)</td><td>Brachionus calyciflorus (rotifer)</td><td>Amblema plicata (mussel)</td><td>Gammarus fasciatus (amphipod)</td><td>Chydorus sphaericus (cladoceran)</td></tr>
<tr><td>Perfluorohexanesulfonic acid (PFHxS)</td><td>Danio rerio (zebrafish)</td><td>Jordanella floridae (flagfish)</td><td>Daphnia magna (cladoceran)</td><td>Limnodrilus hoffmeisteri (oligochaete)</td></tr>
<tr><td>Hexadecafluoro-2-decenoic acid (8:2 FTUCA)</td><td>Amblema plicata (mussel)</td><td>Palaemonetes kadiakensis (grass shrimp)</td><td>Chydorus sphaericus (cladoceran)</td><td>Megalonias nervosa (mussel)</td></tr>
<tr><td>Pentadecafluorodecanoic acid (7:3 FTCA)</td><td>Amblema plicata (mussel)</td><td>Macrobrachium nipponense (river shrimp)</td><td>Chydorus sphaericus (cladoceran)</td><td>Megalonias nervosa (mussel)</td></tr>
</table>
</td></tr>
<tr><td>3</td><td>

If EPA chooses to develop criteria/benchmarks for materials that have limited empirical data, then a two-tiered approach should be adopted much like that previously adopted by EPA for the GLI program. It is unclear from the document what the long-term intent and regulatory status will be for “benchmarks.” Those materials that do not have sufficient empirical data to permit derivation of a “Tier 1” criteria, could be addressed by a “Tier II” benchmark, as suggested. The Tier II benchmark could be derived using the proposed ICE-based methodology or a method like that used for the GLI Secondary Acute Values. That said, will benchmarks serve the same purpose as the current AWQC? They do not have the same scientific basis as “Tier I” AWQC, but if adopted as “standards” by states and tribes, they will have the same regulatory/legal status. This is briefly addressed in the document’s forward; however, greater clarity regarding the “scientific confidence” and “regulatory validity” could be provided. Questions regarding use of the approach should be considered, for

</td></tr>
</table>

I. General Impressions	
REVIEWER	**REVIEWER COMMENT**
	example, if a state developed a proposed standard for a chemical based on limited empirical data and relying predominately on ICE-estimated data (e.g., 7:3 FTCA), would EPA approve it?
4	The draft *Derivation of Acute Protective Freshwater Benchmarks for Selected PFAS Compounds through a New Approach Method (NAM)* was well organized and well written. The accuracy of each element of the derivation process was satisfactory. The structure and writing of the draft document clearly communicated the rationale and the process of deriving the benchmarks. I think that the proposed benchmarks are reasonable and protective of aquatic life based on the acceptable empirical acute toxicity data available for the eight PFAS. The major source of uncertainty for the derived benchmarks is the lack of acceptable empirical acute toxicity data on freshwater primary producers and freshwater invertebrates for the eight PFAS. I think the "New Approach Method" is reasonable when there is a lack of acceptable empirical data, and a benchmark needs to be derived. However, the NAM should not replace the derivation of benchmarks with empirical data.
5	This document provides draft Acute Protective Freshwater Benchmarks for the following eight PFAS Compounds using New Approach Method (NAM): 1. Perfluorobutanoic acid (PFBA) 2. Perfluorobutanesulfonic acid (PFBS) 3. Perfluorononanoic acid (PFNA) 4. Perfluorodecanoic acid (PFDA) 5. Perfluorohexanoic acid (PFHxA) 6. Perfluorohexanesulfonic acid (PFHxS) 7. Hexadecafluoro-2-decenoic acid (8:2 FTUCA), and 8. Pentadecafluorodecanoic acid (7:3 FTCA) The detailed methodology used for the derivation of benchmarks has been thoroughly explained. The process used and results of a systematic review of available empirical toxicity data for aquatic organisms identified via EPA's literature search for the eight PFAS has been adequately addressed.
5	The aquatic life benchmarks for the eight PFAS compounds have been developed using the empirical and Web-ICE data for these chemicals and were calculated by applying statistical methods. This method aligns with the EPA's objective to decrease reliance on animal testing by employing NAMS in toxicity assessment. In addition, the EPA applied 'binning' approach to calculate protective benchmark values for six PFAS, utilizing combined carboxylic acid (PFBA, PFNA, PFDA, PFHxA) and sulfonic acid (PFBS, PFHxS) groupings to facilitate value derivation.

I. General Impressions	
REVIEWER	**REVIEWER COMMENT**
5	The use of estimated data suggests a proactive approach in addressing gaps in empirical data. It also demonstrates agencies' commitment to methodological rigor and adaptability in the face of data challenges when deriving protective values for PFAS compounds.

II. Response to Charge Questions	

Charge Question 1: ***Please comment on the overall clarity of the documents and construction as it relates to assessing the effects and derivation of acute benchmarks for the eight selected PFAS.***	
REVIEWER	**REVIEWER COMMENT**
1	The overall methodology and data used is clearly presented. The conclusions for the explored approaches, i.e., the use of extrapolation versus the "scaled" approach and the use of the binning approach should be made clearer. These conclusions should be included in the summary.
1	The report would be strengthened by including an introduction to PFAS and its uses (and emissions). The brief introduction in the summary is very good, however a version with some more details in the main report would be good to include. Further, I would encourage the authors to include a brief discussion on the trends for the benchmarks for the different PFAS (including PFOA and PFOS) compared to the scientific literature (i.e., trends for toxicity depending on group and chain length).
2	It would be useful to better explain what a Benchmark is related to a Water Quality Criterion. The forward states that benchmarks are "less certain than Water Quality Criteria", but the reader should also be informed about enforcement differences between benchmarks and WQC.
3	In general, the document is clear and well-organized. The sections of the document follow a "template" making the subsections parallel to each other making it easy to compare. As previously suggested, some additional detail regarding the extrapolation and scaled estimation technique would be helpful. This may be contained in the Raimondo et al (in review) document but it was not provided for this review. Also, a brief discussion of the "binning" technique (Giddings et al 2019) was made in Section 5.10 and in Appendix G and a comparison of the benchmarks derived using the binning vs ICE-based techniques is provided, but discussion or assessment of the utility/advantages/disadvantages of the technique is not provided.
4	I thought the overall clarity of the writing and construction of the document were good. I found the structure logical and easy to follow in the progression of the derivation process. I would not recommend any changes to the overall writing or structure of the document. There were a few spelling and grammatical errors but all very minor.

Charge Question 1: ***Please comment on the overall clarity of the documents and construction as it relates to assessing the effects and derivation of acute benchmarks for the eight selected PFAS.***	
REVIEWER	**REVIEWER COMMENT**
5	Great job on the overall structure and organization of the document! The logical flow and seamless transitions between sections significantly enhance the readability and understanding of the content. Information has been laid out in appendices with detailed information on the approaches and examples for deriving benchmarks. The list of Tables and Figures provides information on all the empirical data and acceptable ICE models used in deriving aquatic life acute benchmarks of all eight compounds. The lowest quantitatively acceptable empirical toxicity studies used to derive aquatic life benchmarks for eight PFAS compounds were detailed in the appendix. Ranked GMAVs and FAVs have been provided for all eight PFAS compounds. Data incorporated in SSDs have been listed and all figures are self-explanatory.
5	It is apparent that considerable thought and effort were invested in crafting a document with a well-considered and smooth progression.

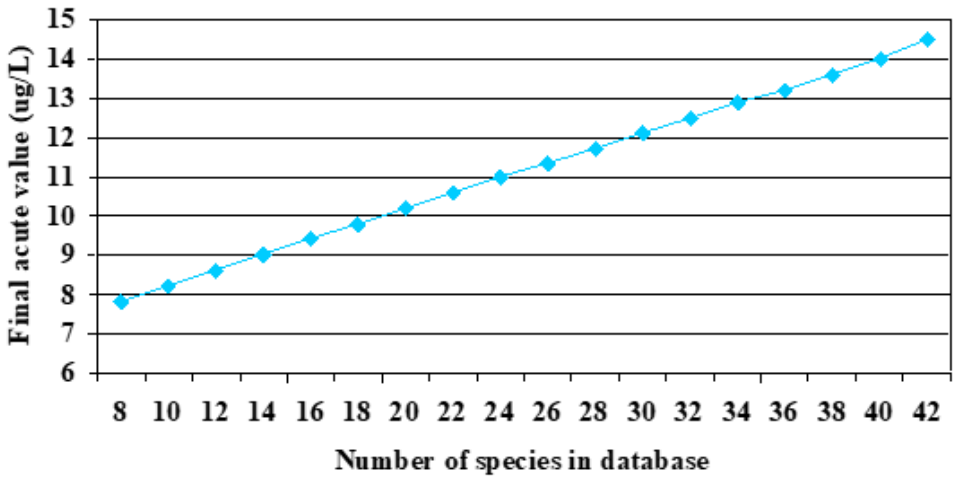
Charge Question 2: ***Please comment on each of the technical approaches used to derive the draft benchmark values for the eight selected PFAS presented in EPA's Derivation of Acute Protective Freshwater Benchmarks for Selected PFAS through a New Approach Method (ICE-based approach using extrapolation [Sections 3-5], ICE-based approach using scaled data [Appendix F], and PFAS benchmark calculations using a data binning approach [Appendix G]).***

2.a. Are the technical approaches used to derive the benchmark values logical?	
REVIEWER	**REVIEWER COMMENT**
1	All these 3 methods are logical. However, I am missing a discussion comparing these methods. It is mentioned that the "scaled" approach is undergoing evaluation by Raimondo et al. (in review), however no information from this work is reported. The lack of a discussion and clear conclusions for the comparisons of these methods makes it difficult for the reader.
2	I think the reader does not have enough information to evaluate ICE-based approaches without consulting the original literature. The introduction to Web-ICE (p. 17) is remarkably brief in explaining the technical approach and this section could be expanded significantly. It is helpful that the reader is given references to read that point to the successful applications of the approach, but this document should be self-contained with respect to describing the technical approach in detail.

2.a. Are the technical approaches used to derive the benchmark values logical?	
REVIEWER	**REVIEWER COMMENT**
2	The same criticism can be applied to the scaled data and data binning approaches. These technical approaches should be explained in more detail particularly in terms of the mechanisms by which these different approaches could yield different toxicity estimates.
2	In general, I favor action on the creation of environmental standards, even when data are limiting. The approach of making toxicity predictions is logical and the process follows the 1985 Guidelines. However, these guidelines are in need of modernization, and it is unclear to me how aggregated, individual species toxicity tests (that ignore dietary exposure pathways and species interactions) provide compelling evidence for protecting aquatic communities in nature.
3	The approaches used by the three methods are logical and creative methods to address the issue of data limitations. Although the calculated data are provided in the document and in Appendix F, it is difficult for this reviewer to fully understand the technical differences in the ICE-based approach using extrapolation or scaled data. Perhaps hands on evaluation of the models or review of the Raimondo et al (in review) manuscript would help. At the least, an example showing calculations both ways would help the reader.
3	PFAS benchmark calculations using the data binning approach presented in Giddings et al (2019) are logical and provide a method to expand the quantity of empirical data provided that the assumption of a common mode of action (MOA) is valid. Raimondo et al (2010) notes that MOA-specific models are more robust and improve the fit of the ICE model approach.
3	Approaching the problem of missing data using the ICE model and binning techniques are more elegant than the previous GLI technique and are more generally scientifically supportable.
4	I think the technical approaches taken to derive the benchmark values were logical. In the absence of acceptable empirical data, the use of the ICE models to generate a data-rich SSD is a logical approach. Obviously, the benchmarks could be re-evaluated if and when acceptable empirical data is available to regulators.
5	The methodologies employed to determine the benchmark values are rational and sound from a technical perspective. The available empirical data for the eight PFAS under consideration fulfill only 2-3 Minimum Data Requirements (MDRs). Consequently, the EPA opted to employ the peer-reviewed Interspecies Correlation Estimation (ICE) models developed by Raimondo et al. in 2010. The primary objective of this application was to provide acute toxicity data to fulfill MDRs in instances where direct toxicity data were not at hand. The ICE models underwent rigorous evaluation based on acceptance parameters, including mean square error (MSE), R^2, and slope, as delineated in Box 1. Only models meeting these predefined acceptance criteria were utilized in the derivation of species-specific toxicity

2.a. Are the technical approaches used to derive the benchmark values logical?	
REVIEWER	**REVIEWER COMMENT**
	data. This data, when integrated with empirical toxicity data, served to strengthen the process of establishing benchmark values.
5	The EPA's investigation into the "binning" approach for establishing protective values for grouped carboxylic acid PFASs and grouped sulfonic acid PFAS is grounded in the precedent established by Giddings et al. in 2019. A similar methodology proved successful for pyrethroids in that study. This strategic approach involved consolidating chemicals with shared modes of action, offering advantages in scenarios where limitations in available data present challenges to value determination. The calculated values were based on the amalgamation of carboxylic acid compounds (PFBA, PFNA, PFDA, PFHxA) and sulfonic acid compounds (PFBS, PFHxS), thereby substantiating the derivation of these values. Calculated benchmark values for carboxylic acids and sulfonic acids consistently demonstrated higher values when utilizing the SSD generator in comparison to the Guidelines-based approach. However, these benchmarks displayed variability in magnitude when contrasted with the ICE-based benchmark values. This might be influenced by the constrained empirical datasets for certain PFAS and the restricted number of data points available for the species employed in normalization.
5	By stating that the derived benchmarks are considered less certain than ambient water quality criteria, the authors acknowledge a level of uncertainty. This acknowledgment is crucial in providing a realistic assessment of the reliability of the benchmarks.

2.b. Does the science support the conclusions?	
REVIEWER	**REVIEWER COMMENT**
1	The applied methods have undergone peer review in previous publications. Previous testing of ICE model performance indicate validity for the assumptions of this approach: 1) that the relationship of inherent sensitivity between two species is conserved across chemicals, mechanisms of action, and ranges of toxicity; and 2) that the nature of a contaminant that was tested on the surrogate reflects the nature of the contaminant in the predicted species.
1	A deeper scientific understanding would require in-depth knowledge of the toxicological mechanisms. Modes of action (MOA) specific models have previously been reported to be more robust. However, such mechanistic understanding, including data for MOA specific toxic effects, is not available. Therefore, the applied methodology can be considered as the best available option.
2	The science tells us that these compounds are not acutely toxic and the benchmark values could give states and tribes a false sense of safety if they encounter high concentrations that are below the benchmark. The language in the

2.b. Does the science support the conclusions?	
REVIEWER	**REVIEWER COMMENT**
	document is explicit about the application as a 1 hour maximum every 3 years, but in the absence of more environmentally relevant standards, I'm not sure what these benchmarks do for environmental protection.
3	I am not sure what is being asked since there are no specific conclusions called out in the document. The approach of estimating species sensitivities to toxins is logical and scientifically supportable, provided that sufficient underpinning data are available to support the models. Derivation of 1985 AWQC compliant values (FAV/FCVs) are dependent on 5 datapoints; the four lowest Genus Mean Values (GMAV/GMCV) and the total number of species represented in the database. The calculation is more sensitive to the relationship of the 4 lowest values than to the total number of species represented. As an example, Figure 1 provides an example of increasing the size of the database for a compound (related to the number of available ICE models); increasing from 8 GMAVs to 42 results in a slightly less than a doubling in the calculated FAV (8 to 14.5). So, the choice of extrapolated vs scaled ICE models may result in a slight increase in the calculated benchmark due to acceptance of more models. Far more important is the validity and relationship of the 4 lowest GMAVs. It is critical that these values be valid and as accurate as possible. Relying on estimated values can introduce a large degree of uncertainty in the resulting benchmark value.  *Figure 1. Final Acute and Chronic values are sensitive to the number of species in the database*
4	Overall, yes, the conclusions are supported by the available science.

2.b. Does the science support the conclusions?	
REVIEWER	**REVIEWER COMMENT**
4	An important element of the science supporting the conclusions is the validation of the ICE models to predict the acute toxicity of PFAS. In the benchmark document on page 17, an unpublished work by Raimonda et al. is cited to support the validation of the ICE models to predict acute toxicity of PFAS. The documents states, "*ICE models have been developed from a broad range of chemicals (e.g., metals and other inorganics, pesticides, solvents, and reactive chemicals) and across a wide range of toxicity values and have been validated as accurate predictors of PFAS acute toxicity when model criteria parameters are followed (Raimondo et al., in review).*" As the unpublished manuscript by Raimondo et al. is not available as part of this review, I am left to assume that the statement made in the benchmark document about the validation of the ICE models to predict the acute toxicity of PFAS to be accurate.
4	Another question is whether the statement "*…have been validated as accurate predictors of PFAS acute toxicity…"* is solely based on data with PFOS and PFOA?
5	This is a valuable contribution to the current scientific knowledge on the toxicity of PFAS compounds to aquatic life, even while acknowledging the inherent uncertainties associated with using estimated data in the derivation process.

2.c. Are the approaches and resulting values consistent with the protection of aquatic life?	
REVIEWER	**REVIEWER COMMENT**
1	In my opinion, the approaches and resulting values are protective for acute toxic effects on aquatic life as defined in the report (i.e., to be protective of 95% of freshwater genera potentially exposed to these specific PFAS under short-term conditions of one-hour duration, if the one-hour average magnitude is not exceeded more than once in three years).
1	However, it is important to note that the most problematic properties of PFAS are not the acute toxicological effects but rather the combination of their persistency, mobility, potential for bioaccumulation (for some PFAS), and long-term toxicological effects. Emissions of PFAS resulting in recipient concentrations comparable to the threshold values in this report may (depending on the volume of water in the recipient) be substantial sources of PFAS pollution to the environment. As many PFAS (PFCA and PFSA) are both persistent and relatively mobile, such emissions may result in problematic pollution of drinking water and wildlife with adverse long-term consequences for both human health and the environment.
2	See comments above. Adhering to an outdated (1985) understanding of toxicology and species sensitivity differences remains an unfortunate state-of-affairs at EPA.
3	The table below summarizes the benchmarks calculated using the various methods described in the reviewed document and the values calculated using the GLI method. Values vary substantially, in some cases as much as an

2.c. Are the approaches and resulting values consistent with the protection of aquatic life?	
REVIEWER	**REVIEWER COMMENT**
	order of magnitude. The Tier II GLI value frequently provided the lowest calculated value; however, the values were not inconsistent with the other methods. Given the minimal amount of empirical data available for these materials, it is difficult to identify if the values are *"consistent with the protection of aquatic life."* Perhaps conducting an analysis with a data rich compound (e.g., copper or a pesticide), using only a limited portion of the available data followed by a comparison to the full AWQC database would give some insight into the comparability of the benchmark and the standard AWQC approach.

Table 5. Summary of calculated benchmark values

Chemical	EPA Benchmark (Extrapolation) (mg/L)	EPA Benchmark (Scaled) (mg/L)	Binning approach (Guidelines-based) (mg/L)	Binning approach (SSD-based) (mg/L)	Tier II GLI calculated value (mg/L)
Perfluorobutanoic acid (PFBA)	83	174	194	467	13.75
Perfluorobutanesulfonic acid (PFBS)	183	237	24	102	149.1
Perfluorononanoic acid (PFNA)	10.3	12	3.4	8.3	2.14
Perfluorodecanoic acid (PFDA)	7.9	10	4.9	12	4
Perfluorohexanoic acid (PFHxA)	75	95	43	103	17.5
Perfluorohexanesulfonic acid (PFHxS)	9.1	9.4	0.18	0.76	1.7
Hexadecafluoro-2-decenoic acid (8:2 FTUCA)	0.58	0.65			0.24
Pentadecafluorodecanoic acid (7:3 FTCA)	0.18	0.23			0.074

4	Yes, I think the approaches and resulting values are consistent with the protection of aquatic life based on the acceptable empirical data that was available to the assessors.
5	Yes, the approaches and resulting values align with the protection of aquatic life. The aquatic life benchmarks for the eight PFAS compounds were established using empirical and Web-ICE data, employing statistical methods for

2.c. Are the approaches and resulting values consistent with the protection of aquatic life?	
REVIEWER	**REVIEWER COMMENT**
	calculation. This approach aligns with the EPA's goal to reduce reliance on animal testing by incorporating NAMS in toxicity assessment.
5	Detailed response as in 2a.
5	*Limitation* ICE models have not been developed for chronic toxicity data and therefore only acute criteria were developed.

2.d. For the ICE-based models in particular, please compare and contrast the strengths and weaknesses of the extrapolation approach [Sections 3-5] with those of the approach using scaled data [Appendix F].	
REVIEWER	**REVIEWER COMMENT**
1	The use of extrapolation beyond the model range may result in large confidence intervals and hence, many ICE models will not meet the acceptability parameters. However, in my opinion, this method is the most intuitive. The "scaled" approach produced more models that met the acceptability parameters. Nevertheless, based on the results, both these methods provided similar benchmark values.
1	It would strengthen the report to include some of the conclusions from Raimondo et al. (in review) regarding the use of the scaling method.
2	We are not given enough technical information to make this comparison. The scaled approach information we are given seems more like instructions for how to run the model when values fall out of environmental realism than a detailed description of how it differs technically from the normal model.
3	As stated above, the technical approach between these techniques is not clear. The implication of the use of the scaled approach rather that the extrapolation method results in the acceptance of more GMAVs, thus resulting in an increased benchmark.
4	Based on the benchmarks derived using the two different approaches, the extrapolation approach generated lower benchmarks across the eight PFAS compared to using scaled data. For the purpose of the protection of aquatic life, the extrapolation approach would be more protective than the scaled approach. I don't know if this would be the case for other groups of chemicals, but it appears that for PFAS, the extrapolation approach is a more protective approach. In the absence of acceptable empirical data, the more protective approach should be selected. This is critical to avoiding a type II error (i.e., false negative), which is an important consideration in risk assessment.

<table>
<tr><td colspan="2">2.d. For the ICE-based models in particular, please compare and contrast the strengths and weaknesses of the extrapolation approach [Sections 3-5] with those of the approach using scaled data [Appendix F].</td></tr>
<tr><td>REVIEWER</td><td>REVIEWER COMMENT</td></tr>
<tr><td>4</td><td>Benchmarks derived using extrapolation approach:
<table>
<tr><th>Chemical[1]</th><th>PFBA (mg/L)</th><th>PFBS (mg/L)</th><th>PFNA (mg/L)</th><th>PFDA (mg/L)</th><th>PFHxA (mg/L)</th><th>PFHxS (mg/L)</th><th>8:2 FTUCA (mg/L)</th><th>7:3 FTCA (mg/L)</th></tr>
<tr><td>Magnitude</td><td>83</td><td>183</td><td>10.3</td><td>7.9</td><td>75</td><td>9.1</td><td>0.58</td><td>0.18</td></tr>
<tr><td>Duration</td><td colspan="8">One hour average</td></tr>
<tr><td>Frequency</td><td colspan="8">Not to be exceeded more than once in three years on average</td></tr>
</table></td></tr>
<tr><td>4</td><td>Benchmarks derived using scaled approach:
<table>
<tr><th>Chemical[1]</th><th>PFBA (mg/L)</th><th>PFBS (mg/L)</th><th>PFNA (mg/L)</th><th>PFDA (mg/L)</th><th>PFHxA (mg/L)</th><th>PFHxS (mg/L)</th><th>8:2 FTUCA (mg/L)</th><th>7:3 FTCA (mg/L)</th></tr>
<tr><td>Magnitude</td><td>174</td><td>237</td><td>12</td><td>10</td><td>95</td><td>9.4</td><td>0.65</td><td>0.23</td></tr>
<tr><td>Duration</td><td colspan="8">One hour average</td></tr>
<tr><td>Frequency</td><td colspan="8">Not to be exceeded more than once in three years on average</td></tr>
</table></td></tr>
<tr><td>4</td><td>Based on the reported validation that has been conducted by Raimondo et al., both approaches seem reasonable. It would be nice to be able to see the validation that has been conducted by Raimondo et al., but it appears that this manuscript is currently in review.</td></tr>
<tr><td>5</td><td>The acknowledgment of potential challenges, such as large confidence intervals and potential limitations in accepting ICE models beyond the model range, demonstrates transparency and a thorough understanding of the modeling process.</td></tr>
<tr><td>5</td><td>The decision to select the "scaled" approach as an alternative approach for deriving benchmark values, showcases a proactive and meticulous approach. The alternative scaled approach modifies toxicity values, as needed, to align them with the ICE model range, avoiding the extrapolation of regressions beyond the established model range. There is close agreement between the benchmark values calculated using either approach (as listed in Table 5-26).</td></tr>
</table>

2.d. For the ICE-based models in particular, please compare and contrast the strengths and weaknesses of the extrapolation approach [Sections 3-5] with those of the approach using scaled data [Appendix F].	
REVIEWER	**REVIEWER COMMENT**
5	This consideration of alternative methods highlights a commitment to rigorous evaluation and continuous improvement, reflecting a commendable scientific rigor in the approach to deriving benchmark values.

Charge Question 3: ***Please comment on the empirical (direct test) and ICE-generated toxicity data used to derive the benchmark values presented in the draft document.***

3.a. ***Were the data adequately used and sufficiently comprehensive to represent risks to sensitive aquatic life?***	
REVIEWER	**REVIEWER COMMENT**
1	Yes. The data covers the eight MDRs, and hence a wide range of taxa with different characteristics in aquatic ecosystems. This, combined with the statistical approach which is focusing on the lowest GMAVs ensures that acute toxic effects on sensitive aquatic life are taken into account. For PFBS and PFHxS, only seven and six, respectively, of the eight MDRs were fulfilled and hence, these benchmarks are associated with greater uncertainty. This information should be included in the summary (could be footnotes in Table Ex1-1).
2	Probably not. There are too few empirical data to be secure in understanding which species in the real world might be sensitive.
3	As stated, the problem is a lack of empirical data and a reliance on data estimation techniques. The table below provides the data for the 4 most sensitive species used to derive the benchmark values for the PFAS materials. Actual empirical data are highlighted (4 of 32 data points, 12.5%); as you can see, most of the data for the range of species are estimated values. Without additional confirmatory experimental data, it is difficult to say anything about how comprehensive the data represent sensitive aquatic organisms.

3.a. ***Were the data adequately used and sufficiently comprehensive to represent risks to sensitive aquatic life?***

REVIEWER	REVIEWER COMMENT
	Table 4. Species sensitivity ranking for PFAS compounds. Empirical data are in bold, all others are ICE-estimated values.

Chemical	Species Sensitivity Rank from ICE model[1]			
	1	2	3	4
Perfluorobutanoic acid (PFBA)	***Brachionus calyciflorus*** **(rotifer)**	*Oncorhynchus mykiss* (rainbow trout)	*Gammarus fasciatus* (amphipod)	*Amblema plicata* (mussel)
Perfluorobutanesulfonic acid (PFBS)	*Amblema plicata* (mussel)	*Gammarus fasciatus* (amphipod)	*Hyalella azteca* (amphipod)	*Ceriodaphnia dubia* (cladoceran)
Perfluorononanoic acid (PFNA)	*Amblema plicata* (mussel)	***Chydorus sphaericus*** **(cladoceran)**	*Megalonias nervosa* (mussel)	*Oncorhynchus mykiss* (rainbow trout)
Perfluorodecanoic acid (PFDA)	*Caecidotea brevicauda* (isopod)	*Micropterus salmoides* (bass)	*Perca flavescens* (yellow perch)	*Salvelinus fontinalis* (brook trout)
Perfluorohexanoic acid (PFHxA)	***Brachionus calyciflorus*** **(rotifer)**	*Amblema plicata* (mussel)	*Gammarus fasciatus* (amphipod)	*Chydorus sphaericus* (cladoceran)
Perfluorohexanesulfonic acid (PFHxS)	***Danio rerio*** **(zebrafish)**	*Jordanella floridae* (flagfish)	*Daphnia magna* (cladoceran)	*Limnodrilus hoffmeisteri* (oligochaete)
Hexadecafluoro-2-decenoic acid (8:2 FTUCA)	*Amblema plicata* (mussel)	*Palaemonetes kadiakensis* (grass shrimp)	*Chydorus sphaericus* (cladoceran)	*Megalonias nervosa* (mussel)
Pentadecafluorodecanoic acid (7:3 FTCA)	*Amblema plicata* (mussel)	*Macrobrachium nipponense* (river shrimp)	*Chydorus sphaericus* (cladoceran)	*Megalonias nervosa* (mussel)

[1] It should also be noted that the 4 data points listed in bold are all based on nominal concentrations, so the accuracy of the EC50 values may be questioned.

4	I think the available empirical was adequately used.
4	However, there was an overall lack of empirical acute toxicity data, particularly for freshwater primary producers and freshwater invertebrates. For example, there was no acute toxicity data on the eight PFAS for freshwater primary producers. Freshwater invertebrates were also under-represented in the empirical data set. For example, there was only three empirical data points for PFHxS and all three were for freshwater vertebrates. When empirical toxicity data on the eight PFAS was available for a freshwater invertebrate species, it was usually *Daphnia magna*. Consequently, I don't think that the data is sufficiently comprehensive to represent risk to sensitive aquatic life.
5	The EPA employed both empirical test data and ICE values, derived for missing Minimum Data Requirements (MDRs), to determine acute freshwater benchmark recommendations for aquatic life. The utilization of ICE-predicted values by various independent, international groups to establish protective values for aquatic life confirms that values

3.a. ***Were the data adequately used and sufficiently comprehensive to represent risks to sensitive aquatic life?***	
REVIEWER	**REVIEWER COMMENT**
	derived from ICE-generated Species Sensitivity Distributions (SSDs) offer a consistent level of protection comparable to using directly measured laboratory data.

3.b. ***Were the data selected and/or excluded from the benchmark values derivation appropriately utilized?***	
REVIEWER	**REVIEWER COMMENT**
1	The data that was deemed of sufficient quality to be used (i.e., quantitatively acceptable freshwater acute toxicity studies in appendix A) were appropriately utilized as these data were used to produce the sensitivity distributions. However, the qualitative freshwater acute toxicity studies in appendix C should be included in a discussion about the derived benchmarks.
2	I think the authors did the best they could with the available data on hand. It is a shame that more resources are not being deployed to generate more empirical data.
3	Much of the data accepted would not meet current standards for data acceptability or criteria derivation. The authors have tried to maintain a degree of fidelity to the Stephan et al (1985) AWQC methodology; however, several concerns exist with the data considered: • The benchmark document states: "Toxicity studies accessed from the ECOTOX database were further evaluated by Office of Water. Studies were evaluated for data quality as described by EPA OW's data quality standard operating procedure (SOP), and consistent with OW's data quality review approach U.S. EPA (1985), and EPA's Office of Chemical Safety and Pollution Prevention (OPP)'s Ecological Effects Test Guidelines (U.S. EPA 2016c)." These documents were not included in the reference list and were not provided.
3	• In general, most of the accepted empirical studies are reported based on nominal exposure concentrations rather than analytically measured concentrations; this is not consistent with state-of-the-science standards of acceptability for empirical toxicology data. In fact, one of the studies that measured exposure concentrations (Ding et al 2012) ultimately reported test endpoint data (EC50) based on nominal concentrations rather than measured values. The reported analytical data indicates that test concentrations differed from nominals by 10-20%, so the value reported based on nominals is likely to be 10-20% off.
3	• Some of the toxicity data used in the derivation of the aquatic benchmarks comes from studies that used non-native species (i.e., zebrafish, *Danio rerio*), which adds uncertainty associated with the representativeness of such species to native North American aquatic fauna. Stephan et al (1985) states: II. G. "Questionable data, data on formulated mixtures and emulsifiable concentrates, and data obtained with non-resident species in

3.b. ***Were the data selected and/or excluded from the benchmark values derivation appropriately utilized?***	
REVIEWER	**REVIEWER COMMENT**
	North America or previously exposed organisms may be used to provide auxiliary information **but should not be used in the derivation of criteria**." Appendix I Resident North American Species of Aquatic Animals Used in Toxicity and Bioconcentration Tests defines zebrafish (*Danio rerio*) as "Non-resident" species and therefore should not be included for criteria derivation. Use of Non-resident species is briefly discussed in the report and a reference to US EPA 2018b is cited; however, this reference is not included in the reference list. The zebrafish is in the family *Cyprinidae*, which all North American native minnows (including the fathead minnow), shiners, and dace belong. Although not native to North America, EPA seems to have decided that in the absence of suitable data on native cyprinids, the zebrafish is an acceptable representative. However, given that zebrafish are frequently among the more sensitive species and at least some studies with PFAS compounds have suggested that fathead minnows may be more sensitive[4], it would be good to have some comparative additional data with NA species.
3	• Some of the test methods used are not consistent with the 1985 guidance. EPA 1985 states that "Acute EC50s that are based on effects that are not severe, such as reduction in shell deposition and reduction in growth, are not used in calculating the Final Acute Value." The zebrafish tests included in the benchmark document (Annunziato et al 2020) followed the OECD 236 method and reported results based on a growth rather than survival endpoint. These data would not be acceptable for derivation of an FAV based on the 1985 guideline.
3	• In addition to the above concerns, it was noted that at least two studies (Ding et. al. 2012, Annunziato et al 2020) that reported tests with PFBA, PFBS, PFNA, PFDA, and PFHxS conducted their studies using dimethylsulfoxide (DMSO) as a carrier solvent. In EPA's current test guidelines, it is recommended that if a carrier solvent must be used, *"Preferred solvents are dimethyl formamide, triethylene glycol, methanol, acetone, or ethanol. Solvent use should be avoided if possible."* DMSO is known to transport nonionized molecules thorough many biological membranes (Jacob and Herschler 1985[5]). Although the authors of the lab tests conducted a "solvent control" this does not control for possible synergistic interactions of DMSO acting as a membrane carrier, thus potentially increasing observed toxicity. Because these tests represent a large portion of the quantitatively acceptable freshwater toxicity tests (20%, 7 of 36), EPA should consider the potential for inclusion of these data resulting in lower than desired criteria, due to an overestimation of toxicity due to DMSO synergy.

[4] Suski et al. 2023. Ecotoxicity and Accumulation of Perfluorononanoic Acid in the Fathead Minnow (*Pimephales promelas*) and an Approach to Developing Protective Thresholds in the Aquatic Environment Through Species Sensitivity Distribution. Environ Toxicol Chem. https://doi.org/10.1002/etc.5692

[5] *Jacob, S. W., & Herschler, R. (1986). Pharmacology of DMSO. Cryobiology, 23(1), 14-27.*

3.b. ***Were the data selected and/or excluded from the benchmark values derivation appropriately utilized?***	
REVIEWER	**REVIEWER COMMENT**
3	EPA should reassess the quality and acceptability of the available data for regulatory purposes.
4	Yes, the process of including and excluding empirical data for derivation of the benchmark values was clearly explained, logical, and well established.
5	Quantitatively acceptable empirical acute toxicity data available for each of the eight PFAS was tabulated for each individual study. All toxicity values, including LC values, EC values, NOECs, LOECs, and species- and genus-mean values, were presented with four significant figures. This practice avoided round-off errors in subsequent calculations. Studies that were determined to be qualitatively acceptable as supporting information, but not acceptable for quantitative use were listed with deficiencies in each study. Furthermore, studies that were deemed unsuitable for either quantitative or qualitative were also cited. I endorse the choice to incorporate toxicity data for studies solely based on unmeasured test concentrations. This decision is rooted in findings for PFOA and PFOS (U.S. EPA 2022a, b), leading the EPA to determine that nominal test concentrations effectively represent real PFAS exposures in standard acute laboratory-based toxicity tests. In addition, Hoke et al.,2012 (https://doi.org/10.1016/j.chemosphere.2011.12.066) also reported mean measured test concentrations were similar (within 80–120% of nominal) to the targeted nominal test concentrations for fluorinated acids with the exception of the 5:3 acid.
5	The authors applied the criteria recommended by Willming et al., 2016 to enhance models' reliability and robustness (Box 1). Models adhering to these acceptance parameters were employed to generate species toxicity data, which were then combined with empirical toxicity data to strengthen the derivation of benchmark values. This approach demonstrated logical and consistent application of standard criteria across all eight PFAS compounds.

3.c. ***Are there relevant data that you are aware of that should be included? If so, please provide for derivation of benchmark values.***	
REVIEWER	**REVIEWER COMMENT**
1	I am not aware of data not already included in the study.
2	There is some mayfly data from the Soucek lab that does not seem to be acknowledged here. The general lack of insect data is a systemic problem – particularly when a single midge is used to represent the toxicity of an entire class or organisms that is likely close to 10,000 species in N. America.

3.c. ***Are there relevant data that you are aware of that should be included? If so, please provide for derivation of benchmark values.***	
REVIEWER	**REVIEWER COMMENT**
3	Several recent publications have critically reviewed the available data for PFAS compounds, e.g., Pandelides et al. 2023[6]. The references below are just an example of amphibian references included in one of the review articles, these include acute and chronic endpoints: Abercrombie, S. A., de Perre, C., Choi, Y. J., Tornabene, B. J., Sepúlveda, M. S., Lee, L. S., & Hoverman, J. T. (2019). Larval amphibians rapidly bioaccumulate poly- and perfluoroalkyl substances. *Ecotoxicology and Environmental Safety*, **178**, 137–145. https://doi.org/10.1016/j.ecoenv.2019.04.022; Ankley, G. T., Kuehl, D. W., Kahl, M. D., Jensen, K. M., Butterworth, B. C., & Nichols, J. W. (2004). Partial life-cycle toxicity and bioconcentration modeling of perfluorooctane sulfonate in the northern leopard frog (*Rana pipiens*). *Environmental Toxicology and Chemistry*, **23**, 2745. https://doi.org/10.1897/03-667 Brown, S. R., Flynn, R. W., & Hoverman, J. T. (2021). Perfluoroalkyl substances increase susceptibility of northern leopard frog tadpoles to trematode infection. *Environmental Toxicology and Chemistry*, **40**, 689–694. https://doi.org/10.1002/etc.4678 Flynn, R. W., Chislock, M. F., Gannon, M. E., Bauer, S. J., Tornabene, B. J., Hoverman, J. T., & Sepúlveda, M. S. (2019). Acute and chronic effects of perfluoroalkyl substance mixtures on larval American bullfrogs (*Rana catesbeiana*). *Chemosphere*, **236**, 124350. https://doi.org/10.1016/j.chemosphere.2019.124350 Flynn, R. W., Hoover, G., Iacchetta, M., Guffey, S., de Perre, C., Huerta, B., Li, W., Hoverman, J. T., Lee, L., & Sepúlveda, M. S. (2022). Comparative toxicity of aquatic per- and polyfluoroalkyl substance exposure in three species of amphibians. *Environmental Toxicology and Chemistry*, **41**, 1407–1415. https://doi.org/10.1002/etc.5319 Flynn, R. W., Iacchetta, M., Perre, C., Lee, L., Sepúlveda, M. S., & Hoverman, J. T. (2021). Chronic per-/polyfluoroalkyl substance exposure under environmentally relevant conditions delays

[6] *Pandelides Z, J Conder, Y Choi, E Allmon, T Hoskins, L Lee, J Hoverman, M Sepúlveda. 2023. A Critical Review of Amphibian Per- and Polyfluoroalkyl Substance Ecotoxicity Research Studies: Identification of Screening Levels in Water and Other Useful Resources for Site-Specific Ecological Risk Assessments. Environ Toxicol Chem. https://doi.org/10.1002/etc.5695*

3.c. ***Are there relevant data that you are aware of that should be included? If so, please provide for derivation of benchmark values.***	
REVIEWER	**REVIEWER COMMENT**
	development in northern leopard frog (*Rana pipiens*) larvae. *Environmental Toxicology and Chemistry*, **40**, 711–716. https://doi.org/10.1002/etc.4690 Foguth, R. M., Hoskins, T. D., Clark, G. C., Nelson, M., Flynn, R. W., de Perre, C., Hoverman, J. T., Lee, L. S., Sepúlveda, M. S., & Cannon, J. R. (2020). Single and mixture per- and polyfluoroalkyl substances accumulate in developing northern leopard frog brains and produce complex neurotransmission alterations. *Neurotoxicology and Teratology*, **81**, 106907. https://doi.org/10.1016/j.ntt.2020.106907 Fort, D. J., Mathis, M. B., Guiney, P. D., & Weeks, J. A. (2019). Evaluation of the developmental toxicity of perfluorooctane sulfonate in the Anuran, *Silurana tropicalis*. *Journal of Applied Toxicology*, **39**, 365–374. https://doi.org/10.1002/jat.3727 Hoover, G. M., Chislock, M. F., Tornabene, B. J., Guffey, S. C., Choi, Y. J., De Perre, C., Hoverman, J. T., Lee, L. S., & Sepúlveda, M. S. (2017). Uptake and depuration of four per/polyfluoroalkyl substances (PFAS) in northern leopard frog *Rana pipiens* tadpoles. *Environmental Science and Technology Letters*, **4**, 399–403. https://doi.org/10.1021/acs.estlett.7b00339 Hoskins, T. D., Allmon, E. B., Flynn, R. W., Lee, L. S., Choi, Y., Hoverman, J. T., & Sepúlveda, M. S.(2022). An environmentally relevant mixture of perfluorooctane sulfonic acid and perfluorohexane sulfonic acid does not conform to additivity in northern leopard frogs exposed through metamorphosis. *Environmental Toxicology and Chemistry*, **41**, 3007–3016. https://doi.org/10.1002/etc.5486 Lech, M. E., Choi, Y. J., Lee, L. S., Sepúlveda, M. S., & Hoverman, J. T. (2022). Effects of per- and polyfluoroalkyl substance mixtures on the susceptibility of larval American bullfrogs to parasites. *Environmental Science & Technology*, **56**, 15953–15959. https://doi.org/10.1021/acs.est.2c04574
4	I am not aware of relevant data that should be included in this process. An extensive and complete review of available data has been conducted in preparation for this process of deriving benchmarks.
5	Below, a recent chronic study by Kadlec et al., 20203 has been listed as an additional reference, some aspects may be relevant. **https://doi.org/10.1002/etc.5784** Sarah M. Kadlec, Will J. Backe, Russell J. Erickson, J. Russell Hockett, Sarah E. Howe, Ian D. Mundy, Edward Piasecki, Henry Sluka, Lauren K. Votava, David R. Mount (2023) Sublethal Toxicity

3.c. ***Are there relevant data that you are aware of that should be included? If so, please provide for derivation of benchmark values.***	
REVIEWER	**REVIEWER COMMENT**
	of 17 Per- and Polyfluoroalkyl Substances with Diverse Structures to *Ceriodaphnia dubia, Hyalella azteca*, and *Chironomus dilutus*

Charge Question 4: ***Are the derived benchmark values appropriately protective of sensitive aquatic life?***	
REVIEWER	**REVIEWER COMMENT**
1	In my opinion, the derived benchmark values are protective for acute toxic effects on aquatic life as defined in the report (i.e., to be protective of 95% of freshwater genera potentially exposed to these specific PFAS under short-term conditions of one-hour duration, if the one-hour average magnitude is not exceeded more than once in three years).
1	However, it is important to note that the most problematic properties of PFAS are not the acute toxicological effects but rather the combination of their persistency, mobility, potential for bioaccumulation (for some PFAS), and long-term toxicological effects. Emissions of PFAS resulting in recipient concentrations comparable to the threshold values in this report may (depending on the volume of water in the recipient) be substantial sources of PFAS pollution to the environment. As many PFAS (PFCA and PFSA) are both persistent and relatively mobile, such emissions may result in problematic pollution of drinking water and wildlife with adverse long-term consequences for both human health and the environment.
2	As applied to a one-hour maximum concentration not to be exceeded every 3 years, the values are likely protective to most species.
3	The benchmark methods attempt to maintain compliance with the EPA's 1985 method for derivation of AWQC and to the extent that the 1985 method was "*appropriately protective of sensitive aquatic life"* the benchmark approach should be as well. However, one major difference between the 1985 guidance and the new benchmark approach is that the requirements for high-quality empirical data for a minimal range of aquatic species have been reduced or eliminated. The benchmark approach seems to rely on existing data or extrapolation of limited data from similar compounds. ICE methods are extremely useful and important in estimating values for species where we cannot generate empirical data, e.g., T&E species (Willming et al 2016). However, the benchmark approach proposed seems to minimize the utility and need for empirical data. Modelling techniques that are based on robust empirical data are extremely useful for **supplementing** extant data for species-of-concern that cannot be easily or cost-effectively tested, or tested due to regulatory restrictions, but they should not suppliant the need for chemical-specific empirical data. Is there a minimum amount of empirical data that are needed to derive a benchmark? In theory, a single acute toxicity test may be sufficient, using ICE-models, to derive regulatory benchmarks. To gain confidence in the proposed

Charge Question 4: ***Are the derived benchmark values appropriately protective of sensitive aquatic life?***	
REVIEWER	**REVIEWER COMMENT**
	approach, EPA should conduct testing to further confirm the accuracy of the ICE estimates, especially for the most sensitive species in the benchmark data sets.
4	I do have concerns about the lack of acceptable empirical acute toxicity data for freshwater primary producers and freshwater invertebrates. I think the process of deriving benchmarks for the eight PFAS described by the USEPA is appropriate for the empirical data that is available. I think they have done their best with the data that is available to them.
5	The establishment of aquatic life benchmarks for the eight PFAS compounds involved the utilization of empirical and Web-ICE data, incorporating statistical methods for calculation. This strategy is in accordance with the EPA's objective of minimizing dependence on animal testing by integrating NAMS into toxicity assessments.
5	In stating that the benchmarks derived are regarded as less certain than ambient water quality criteria, the authors are acknowledging a degree of uncertainty. This recognition is essential for offering a logical evaluation of the benchmarks' reliability.

III. Specific Observations			
Reviewer	**Page**	**Paragraph**	**Comment or Question**
1	Front page	Title	As “PFAS” is an abbreviation for “per- and polyfluorinated substances”, the term “PFAS Compounds” would mean “per- and polyfluorinated substances Compounds”. Here and throughout the document, I suggest to just use “PFAS” instead of “PFAS Compounds”, “PFAS substances”, or “PFAS chemicals”
1	viii	Acronym list	Several of the acronyms used in the report are missing. Examples of missing acronyms are listed in the following, however, there are likely more missing acronyms. This should be reviewed and corrected before publication. Examples of missing acronyms: SMAV, all the eight PFAS in focus (PFBA, PFBS, PFHxA, PFHxS, etc.), EPA, DOD, MSE, SMAV
1	x	1	Please define what is meant by the term Water Quality Criteria in this context
1	x	1	Please define the difference between “draft ambient water quality benchmarks” and “Water Quality Criteria”
1	xi	1	It is somewhat confusing for the reader to understand what was done in the present study and what has been done previously. I suggest making this clearer by using terms such as “in the present study” or similar.

III. Specific Observations			
Reviewer	**Page**	**Paragraph**	**Comment or Question**
1	xi	1	Here, and throughout the document, it would be more logical to present the eight PFAS sorted according to group and number of perfluorinated carbon atoms in the structures. I.e., PFBA, PFHxA, PFNA, PFDA, PFBS, PFHxS, 7:3 FTCA, 8:2 FTUCA.
1	xi	1	It should be made clearer in the summary that the benchmarks reported here are for acute exposure. I.e.: "...protective of 95% of freshwater genera potentially exposed to the specific PFAS under short-term conditions of one-hour duration, if the one-hour average magnitude is not exceeded more than once in three years"
1	xi	2	The sentence that starts with "EPA's "*Guidelines for Deriving Numerical*" needs to be revised as it should likely be at least two sentences.
1	xii	Table Ex-1-1	Is the superscript "1" referring to a footnote?
1	xii	Table Ex-1-1	It would be more logical to present the eight PFAS sorted according to group and number of perfluorinated carbon atoms in the structures. I.e., PFBA, PFHxA, PFNA, PFDA, PFBS, PFHxS, 7:3 FTCA, 8:2 FTUCA.
1	xiii	1	Here, and throughout the document, I suggest using the terms perfluoroalkyl carboxylic acids (PFCA) and perfluoroalkyl sulfonic acids (PFSA) as this is the common terminology.
1	xiii	1	A conclusion for the binning approach should be included in the summary.
1	14 (or 1?)	Page numbers	The first page in the section "Background" should probably be 1 (it has page number 14 in the version I received)
1	14	1	Here, and throughout the document, it would be more logical to present the eight PFAS sorted according to group and number of perfluorinated carbon atoms in the structures. I.e., PFBA, PFHxA, PFNA, PFDA, PFBS, PFHxS, 7:3 FTCA, 8:2 FTUCA.
1	14	2	The acute water quality benchmark concentrations for PFOS and PFOA should be stated here
1	14	2	It would be good to include information on how the selection of these eight PFAS was done. Why these exact substances? Were other PFAS considered, but not included due to limited information available?
1	14	3	The reference U.S. EPA 1985 is not included in the reference list.
1	14	3	It would make it clearer for the reader to state "toxicity data" instead of just "data"

III. Specific Observations			
Reviewer	**Page**	**Paragraph**	**Comment or Question**
1	15	2	It would be good to include a reference to Table 4-2 to show which MDRs are fulfilled for which PFAS.
1	15	3	Would it be better to use the term “aquatic life benchmarks” instead of “aquatic life values” ?
1	17	2	Raimondo et al. 2023 is not in the References list. Is it the same as “Raimondo et al., in review”?
1	17	3	Wilming et al 2016 does not mention PFOS and PFOA
1	17	3	Wilming et al 2016 is not included in the References list
1	17	3	Wilming et al 2016 defines parameters for listed species. The use of these as general parameters is likely unproblematic, but it should be stated that these parameters are used in a slightly different context here compared to in Wilming et al
1	18	1	Bejarano and Wheeler, 2020 is not in the References list
1	18	Figure 2-1	The figure is not referred to in the text. A reference to the figure should be included where appropriate.
1	18	Figure 2-1	The last sentence in the figure text is not written in bold and appears to be incomplete. Should it be “...develop a log-linear model”?
1	21	Table 3-1	It would be good to include references in this table, alternatively to refer to appendix A in the table text.
1	21	Table 3-1	The heading “Toxicity” is a bit ambiguous. It would be more intuitive to write e.g., EC50/LC50
1	23	1	The sentence “The Office of Water completed a Data Evaluation Record (DER) for each species by chemical combination from the studies identified by ECOTOX for the eight PFAS compounds undergoing evaluation.” is not clear to me. What does “chemical combination” refer to in this context?
1	23	1	The sentence “Further, only single chemical toxicity tests with PFOA were considered for possible inclusion in benchmark derivation” is confusing. Is this statement correct?
1	24	2	The title for Appendix A used here (“Appendix A: Acceptable Freshwater Acute PFOA Toxicity Studies”) is not correct.

III. Specific Observations			
Reviewer	**Page**	**Paragraph**	**Comment or Question**
1	25	2	Were all toxicity values (e.g., LC, EC, NOEC, LOEC) treated as the same? Please explain
1	25	2	The last sentence is missing a word ("is"?) (i.e., "of the corresponding benchmarks is stated for each study at the end"?)
1	27	2	The EPA 1985 approach should be summarized here. At least the calculation procedure (as detailed in EPA 1985) should be stated. It should be clear to the reader why the four most sensitive values are focused on.
1	28	1	In the first sentence, I suggest reminding the reader that eight MDR groups are required fulfilled.
1	28	2	Please explain why the FAV was divided by two (what is the reasoning behind this approach?)
1	40	1	The sentence "A Web ICE model was not available to fulfill the MDR for a third family in the phylum chordata." Is somewhat confusing as it is according to the table already fulfilled using empirical data.
1	40	2	Here and for the other PFAS, explain why it is stated that "GMAVs for the four most sensitive genera were within a factor of 2.0 of each other" (i.e., it is according to the criteria in the guidelines)
1	81	1	A brief discussion on the results would be appropriate. For example: - The long chained PFAS are more acute toxic than shorter (PFBA, PFBS). This is in agreement with scientific literature. Previously published values for PFOS and PFOA indicate that these are the most toxic of the PFSA and PFCA, respectively. - According to the results published here, FTUCA and FTCA are the most toxic. Is this as expected?
1	82	1	The sentence "The resulting acute benchmarks, although consistently higher, were also small, with each of the benchmarks falling within a factor of < 2.1 of one another, indicating close agreement between values calculated using either approach." Should be revised. What were small, the difference between values calculated using the different approaches?
1	83	1	A conclusion for the use of the binning method should be included.

III. Specific Observations			
Reviewer	**Page**	**Paragraph**	**Comment or Question**
1	86	References	The reference list needs to be updated as several references used in the text are not included here.
1	Appendix A		The references are not included in any Reference list. This should be corrected.
1	Appendix B		Please explain why the four most sensitive values are summarized
1	C-2	Green alga	The text explaining the Deficiency for this study "Initially identified as Quantitative" is confusing.
1	C-3	Green alga	Why is the test duration written in red?
1	E-1		Only the Web-ICE version 3.3 is available online. I assume that the small differences I found between online model parameters and parameters in the Table is due to the updated values in the v4 model?
1	F-1		The claim "In these situations, a user can either enter the measured toxicity value (LC50/EC50) into the ICE model as µg/L and allow the regression to extrapolate beyond the range of the model or enter a "scaled" toxicity value (i.e. enter the measured LC50 value as mg/L)." should be supported by a literature reference.
1	F-73	1	Using the scaled approach, the eight MDRs were fulfilled for seven of the evaluated compounds (not six)?
1	G-7	Footnote b	It would be more intuitive for the reader to use "Species x SMAV", similar to what is done for PFAS
1	G-9		The table lacks references to the footnotes
1	G-12		Figures G-1 and G-2 are too small. Consider showing panel B below panel A
2	19	Box	Some classes are quite species rich and a single representative is likely not sufficient (see insects)
2	20	3	When is Raimondo et al (in review) going to be released? It seems like this should be available information for the reader
2	throughout		Why are no extra uncertainty factors used other than FAV/2 given the uncertainties associated with the process?

III. Specific Observations			
Reviewer	**Page**	**Paragraph**	**Comment or Question**
2	throughout		Within class extrapolation is a whole lot of biodiversity to lump together
3	x	2	Are benchmarks expected to carry the same weight as AWQC if adopted as state or tribal standards?
3	14	1	Should there be inclusion of EPA's GLI approach?
3	17	2	It is difficult to evaluate this statement since the Raimondo et al report is unpublished and not supplied.
3	23	1	The document references USEPA 2016c for information on how data were evaluated. The reference is not in the reference list and was not provided. It is critical to assess the acceptability of the empirical data accepted in Appendix A. USEPA 1985 was not included in the reference list.
3	23	2	The document references USEPA 2018b however, the reference is not in the reference list and was not provided.
4	xi	Second	Period missing at the end of "…*minimum data requirements (MDRs) to calculate aquatic life criteria*"
4	xii	First	Space needed between "…*(Guidelines)(U.S. EPA 1985*)."
4	xii	Table Ex-1-1	Not clear how the "*Duration*" and "*Frequency*" were determined for the recommended benchmarks. Is this standard for USEPA acute freshwater aquatic life benchmarks?
4	14	First	A closing bracket missing in "• *Perfluorohezanesulfonic acid (PFHxS (CAS# 355464, 108427538, 3871996, 82382125)"*
4	15	First	The first word in f), g), and h) is not capitalized as in a) to e). *a) "insect (e.g., mayfly, dragonfly, damselfly, stonefly, caddisfly, mosquito, midge, etc.)* *b) family in a phylum other than Arthropoda or Chordata (e.g., Rotifera, Annelida, Mollusca, etc.)* c) *family in any order of insect or any phylum not already represented*
4	20	Final sentence on the page	The final sentence of this section is "*Benchmark values for the eight PFAS using this alternative approach are summarized in* ***Section 5.10****."*

III. Specific Observations			
Reviewer	**Page**	**Paragraph**	**Comment or Question**
			I would specifically reference Table 5-26 in section 5.10 at the end of the sentence above. It would make it easier for the reader to find the benchmarks derived using the scaled approach. Or reference section F.9 and/or Table F-29.
4	23	First	The first sentence in the paragraph states, "*Empirical studies available for the eight PFAS were identified using the ECOTOXicology Knowledgebase*…" I assume that the "eight PFAS" refers to the compounds for which the benchmarks are being set. However, later in the paragraph, there is a sentence that states, "*Further, only single chemical toxicity tests with PFOA were considered for possible inclusion in benchmark derivation*." PFOA is not one of the eight PFAS for which a benchmark is being set, so it is not clear why acute toxicity data for PFOA is being used. It was stated earlier in section 3 that validation was conducted using measured and predicted values for PFOS and PFOA, but it is not clear at this point in section 4 how acute toxicity data for PFOA will be used in deriving benchmarks for the eight PFAS that are the focus of this document. You may want to make that clear to the reader.
4	24	First	The paragraph makes references to whether this process should consider studies that only report nominal concentrations of the PFAS in the toxicity study. The rationale given for choosing to consider studies that only report nominal concentrations is a case study that was conducted with measured and nominal concentrations for PFOS and PFOA. While the rationale is sound for considering studies that only include nominal concentrations of PFOS or PFOA, care should be taken to extrapolate to the entire class of chemicals, i.e., PFAS. The eight PFAS that are focus of this process have different physical and chemical properties than PFOS and PFOA, consequently, the probability of the nominal concentrations being with 20% of the measured concentrations for the eight PFAS may be different than PFOS and PFOA, which was 82 and 83%, respectively.
4	27	First	Comma needed between "*toxicity database*" and "*benchmark values"* in the first sentence on page 27.
4	27	Second	It would be useful to report at some point the number of acceptable empirical LC50 or EC50 used in each SSD that were based on nominal concentrations. The data is available in the document but the reader would have to take a great deal of time to compile these numbers.

III. Specific Observations			
Reviewer	**Page**	**Paragraph**	**Comment or Question**
4	27	Second	Why not use the lowest acute value for a species instead of the mean? Using the lowest acute value for a species would be a more conservative approach in terms of protection of sensitive species. The same question could be asked about the genus mean acute values.
4	Entire document	Entire document	Review the document to ensure that Greek letters are used consistently, e.g., µg/L vs. ug/L
4	Appendices A to C		I thought these were very valuable appendices. They clearly laid out the studies that were considered for inclusion in the derivation of the benchmarks and why studies were eventually not included.
5	70 and 80	Figure 5-7 and Figure 5-8	Is this a bimodal response- as this is model based SSD, it is challenging to confirm? Invertebrates and fish may have different mode of action for 8:2 FTUCA and 7:3 FTCA

Appendix A. Individual Reviewer Comments

Reviewer 1

Peer Review Comments on Acute Protective Freshwater Benchmarks for Selected PFAS Using a New Approach Methods

Reviewer 1

I. GENERAL IMPRESSIONS

This report describes the process and results for acquiring draft acute recommended freshwater aquatic life benchmarks for eight PFAS: PFBA, PFHxA, PFNA, PFDA, PFBS, PFHxS, 7:3 FTCA, and 8:2 FTUCA.

Acute benchmark values were derived using the procedure described in EPA's "Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses" where sensitivity distributions based on toxicity databases are used to derive acute freshwater criteria. Due to the lack of acceptable empirical data for constructing the sensitivity distributions, Interspecies Correlation Estimation (ICE) models were applied to complete the toxicity databases.

The applied methods are transparent and the methods have undergone previous peer review. The approaches and data used are reported in a clear manner, making it possible to examine the whole process of acquiring the benchmarks.

Some clarification would, however, strengthen the report and make it less likely that readers will misunderstand what is the appropriate use of these benchmarks:

1) It should be more clearly expressed in the summary that these benchmarks are for acute exposure, i.e., to protect aquatic life in freshwater from acute toxic effects of PFAS. For example, these benchmarks are not sufficient to prevent pollution of the environment, adverse effects of chronic exposure to aquatic animals, human exposure, etc. This is important information as the problems with PFAS pollution are often due to the combination of their persistency, mobility, potential for bioaccumulation, and long-term toxicological effects.
2) The assumptions for the use of the ICE models should be summarized in the main report.
3) The resulting benchmark values should be discussed in more detail. Especially trends for acute toxicity depending on PFAS group and chain length.
4) There is a lack of clear conclusions regarding the different methods explored for deriving the benchmarks.

Additional comments and suggestions are listed in the following.

II. RESPONSE TO CHARGE QUESTIONS

1. ***Please comment on the overall clarity of the documents and construction as it relates to assessing the effects and derivation of acute benchmarks for the eight selected PFAS.***

The overall methodology and data used is clearly presented. The conclusions for the explored approaches, i.e., the use of extrapolation versus the "scaled" approach and the use of the binning approach should be made clearer. These conclusions should be included in the summary.

The report would be strengthened by including an introduction to PFAS and its uses (and emissions). The brief introduction in the summary is very good, however a version with some more details in the main report would be good to include. Further, I would encourage the authors to include a brief discussion on the trends for the benchmarks for the different PFAS (including PFOA and PFOS) compared to the scientific literature (i.e., trends for toxicity depending on group and chain length).

2. ***Please comment on each of the technical approaches used to derive the draft benchmark values for the eight selected PFAS presented in EPA's Derivation of Acute Protective Freshwater Benchmarks for Selected PFAS through a New Approach Method (ICE-based approach using extrapolation [Sections 3-5], ICE-based approach using scaled data [Appendix F], and PFAS benchmark calculations using a data binning approach [Appendix G]).***

    a. ***Are the technical approaches used to derive the benchmark values logical?***

    All these 3 methods are logical. However, I am missing a discussion comparing these methods. It is mentioned that the "scaled" approach is undergoing evaluation by Raimondo et al. (in review), however no information from this work is reported. The lack of a discussion and clear conclusions for the comparisons of these methods makes it difficult for the reader.

    b. ***Does the science support the conclusions?***

    The applied methods have undergone peer review in previous publications. Previous testing of ICE model performance indicate validity for the assumptions of this approach: 1) that the relationship of inherent sensitivity between two species is conserved across chemicals, mechanisms of action, and ranges of toxicity; and 2) that the nature of a contaminant that was tested on the surrogate reflects the nature of the contaminant in the predicted species.

    A deeper scientific understanding would require in-depth knowledge of the toxicological mechanisms. Modes of action (MOA) specific models have previously been reported to be more robust. However, such mechanistic understanding, including data for MOA specific toxic effects, is not available. Therefore, the applied methodology can be considered as the best available option.

c. Are the approaches and resulting values consistent with the protection of aquatic life?

In my opinion, the approaches and resulting values are protective for acute toxic effects on aquatic life as defined in the report (i.e., to be protective of 95% of freshwater genera potentially exposed to these specific PFAS under short-term conditions of one-hour duration, if the one-hour average magnitude is not exceeded more than once in three years).

However, it is important to note that the most problematic properties of PFAS are not the acute toxicological effects but rather the combination of their persistency, mobility, potential for bioaccumulation (for some PFAS), and long-term toxicological effects. Emissions of PFAS resulting in recipient concentrations comparable to the threshold values in this report may (depending on the volume of water in the recipient) be substantial sources of PFAS pollution to the environment. As many PFAS (PFCA and PFSA) are both persistent and relatively mobile, such emissions may result in problematic pollution of drinking water and wildlife with adverse long-term consequences for both human health and the environment.

d. For the ICE-based models in particular, please compare and contrast the strengths and weaknesses of the extrapolation approach [Sections 3-5] with those of the approach using scaled data [Appendix F].

The use of extrapolation beyond the model range may result in large confidence intervals and hence, many ICE models will not meet the acceptability parameters. However, in my opinion, this method is the most intuitive. The "scaled" approach produced more models that met the acceptability parameters. Nevertheless, based on the results, both these methods provided similar benchmark values.

It would strengthen the report to include some of the conclusions from Raimondo et al. (in review) regarding the use of the scaling method.

3. Please comment on the empirical (direct test) and ICE-generated toxicity data used to derive the benchmark values presented in the draft document.

a. Were the data adequately used and sufficiently comprehensive to represent risks to sensitive aquatic life?

Yes. The data covers the eight MDRs, and hence a wide range of taxa with different characteristics in aquatic ecosystems. This, combined with the statistical approach which is focusing on the lowest GMAVs ensures that acute toxic effects on sensitive aquatic life are taken into account. For PFBS and PFHxS, only seven and six, respectively, of the eight MDRs were fulfilled and hence, these benchmarks are associated with greater uncertainty. This information should be included in the summary (could be footnotes in Table Ex1-1).

b. Were the data selected and/or excluded from the benchmark values derivation appropriately utilized?

The data that was deemed of sufficient quality to be used (i.e., quantitatively acceptable freshwater acute toxicity studies in appendix A) were appropriately utilized as these data were used to produce the sensitivity distributions. However, the qualitative freshwater acute toxicity studies in appendix C should be included in a discussion about the derived benchmarks.

c. Are there relevant data that you are aware of that should be included? If so, please provide for derivation of benchmark values.

I am not aware of data not already included in the study.

4. Are the derived benchmark values appropriately protective of sensitive aquatic life?

In my opinion, the derived benchmark values are protective for acute toxic effects on aquatic life as defined in the report (i.e., to be protective of 95% of freshwater genera potentially exposed to these specific PFAS under short-term conditions of one-hour duration, if the one-hour average magnitude is not exceeded more than once in three years).

However, it is important to note that the most problematic properties of PFAS are not the acute toxicological effects but rather the combination of their persistency, mobility, potential for bioaccumulation (for some PFAS), and long-term toxicological effects. Emissions of PFAS resulting in recipient concentrations comparable to the threshold values in this report may (depending on the volume of water in the recipient) be substantial sources of PFAS pollution to the environment. As many PFAS (PFCA and PFSA) are both persistent and relatively mobile, such emissions may result in problematic pollution of drinking water and wildlife with adverse long-term consequences for both human health and the environment.

III. SPECIFIC OBSERVATIONS

Page	Paragraph	Comments or Questions
Front page	Title	As “PFAS” is an abbreviation for “per- and polyfluorinated substances”, the term “PFAS Compounds” would mean “per- and polyfluorinated substances Compounds”. Here and throughout the document, I suggest to just use “PFAS” instead of “PFAS Compounds”, “PFAS substances”, or “PFAS chemicals”
viii	Acronym list	Several of the acronyms used in the report are missing. Examples of missing acronyms are listed in the following, however, there are likely more missing acronyms. This should be reviewed and corrected before publication. Examples of missing acronyms:

Page	Paragraph	Comments or Questions
		SMAV, all the eight PFAS in focus (PFBA, PFBS, PFHxA, PFHxS, etc.), EPA, DOD, MSE, SMAV
x	1	Please define what is meant by the term Water Quality Criteria in this context
x	1	Please define the difference between "draft ambient water quality benchmarks" and "Water Quality Criteria"
xi	1	It is somewhat confusing for the reader to understand what was done in the present study and what has been done previously. I suggest making this clearer by using terms such as "in the present study" or similar.
xi	1	Here, and throughout the document, it would be more logical to present the eight PFAS sorted according to group and number of perfluorinated carbon atoms in the structures. I.e., PFBA, PFHxA, PFNA, PFDA, PFBS, PFHxS, 7:3 FTCA, 8:2 FTUCA.
xi	1	It should be made clearer in the summary that the benchmarks reported here are for acute exposure. I.e.: "...protective of 95% of freshwater genera potentially exposed to the specific PFAS under short-term conditions of one-hour duration, if the one-hour average magnitude is not exceeded more than once in three years"
xi	2	The sentence that starts with "EPA's "*Guidelines for Deriving Numerical*" needs to be revised as it should likely be at least two sentences.
xii	Table Ex-1-1	Is the superscript "1" referring to a footnote?
xii	Table Ex-1-1	It would be more logical to present the eight PFAS sorted according to group and number of perfluorinated carbon atoms in the structures. I.e., PFBA, PFHxA, PFNA, PFDA, PFBS, PFHxS, 7:3 FTCA, 8:2 FTUCA.
xiii	1	Here, and throughout the document, I suggest using the terms perfluoroalkyl carboxylic acids (PFCA) and perfluoroalkyl sulfonic acids (PFSA) as this is the common terminology.
xiii	1	A conclusion for the binning approach should be included in the summary.
14 (or 1?)	Page numbers	The first page in the section "Background" should probably be 1 (it has page number 14 in the version I received)
14	1	Here, and throughout the document, it would be more logical to present the eight PFAS sorted according to group and number of perfluorinated carbon atoms in the structures. I.e., PFBA, PFHxA, PFNA, PFDA, PFBS, PFHxS, 7:3 FTCA, 8:2 FTUCA.
14	2	The acute water quality benchmark concentrations for PFOS and PFOA should be stated here
14	2	It would be good to include information on how the selection of these eight PFAS was done. Why these exact substances? Were other PFAS considered, but not included due to limited information available?

Page	Paragraph	Comments or Questions
14	3	The reference U.S. EPA 1985 is not included in the reference list.
14	3	It would make it clearer for the reader to state "toxicity data" instead of just "data"
15	2	It would be good to include a reference to Table 4-2 to show which MDRs are fulfilled for which PFAS.
15	3	Would it be better to use the term "aquatic life benchmarks" instead of "aquatic life values" ?
17	2	Raimondo et al. 2023 is not in the References list. Is it the same as "Raimondo et al., in review"?
17	3	Wilming et al 2016 does not mention PFOS and PFOA
17	3	Wilming et al 2016 is not included in the References list
17	3	Wilming et al 2016 defines parameters for listed species. The use of these as general parameters is likely unproblematic, but it should be stated that these parameters are used in a slightly different context here compared to in Wilming et al
18	1	Bejarano and Wheeler, 2020 is not in the References list
18	Figure 2-1	The figure is not referred to in the text. A reference to the figure should be included where appropriate.
18	Figure 2-1	The last sentence in the figure text is not written in bold and appears to be incomplete. Should it be "...develop a log-linear model"?
21	Table 3-1	It would be good to include references in this table, alternatively to refer to appendix A in the table text.
21	Table 3-1	The heading "Toxicity" is a bit ambiguous. It would be more intuitive to write e.g., EC50/LC50
23	1	The sentence "The Office of Water completed a Data Evaluation Record (DER) for each species by chemical combination from the studies identified by ECOTOX for the eight PFAS compounds undergoing evaluation." is not clear to me. What does "chemical combination" refer to in this context?
23	1	The sentence "Further, only single chemical toxicity tests with PFOA were considered for possible inclusion in benchmark derivation" is confusing. Is this statement correct?
24	2	The title for Appendix A used here ("Appendix A: Acceptable Freshwater Acute PFOA Toxicity Studies") is not correct.
25	2	Were all toxicity values (e.g., LC, EC, NOEC, LOEC) treated as the same? Please explain
25	2	The last sentence is missing a word ("is"?) (i.e., "of the corresponding benchmarks is stated for each study at the end"?)
27	2	The EPA 1985 approach should be summarized here. At least the calculation procedure (as detailed in EPA 1985) should be stated. It should be clear to the reader why the four most sensitive values are focused on.

Page	Paragraph	Comments or Questions
28	1	In the first sentence, I suggest reminding the reader that eight MDR groups are required fulfilled.
28	2	Please explain why the FAV was divided by two (what is the reasoning behind this approach?)
40	1	The sentence "A Web ICE model was not available to fulfill the MDR for a third family in the phylum chordata." Is somewhat confusing as it is according to the table already fulfilled using empirical data.
40	2	Here and for the other PFAS, explain why it is stated that "GMAVs for the four most sensitive genera were within a factor of 2.0 of each other" (i.e., it is according to the criteria in the guidelines)
81	1	A brief discussion on the results would be appropriate. For example: - The long chained PFAS are more acute toxic than shorter (PFBA, PFBS). This is in agreement with scientific literature. Previously published values for PFOS and PFOA indicate that these are the most toxic of the PFSA and PFCA, respectively. - According to the results published here, FTUCA and FTCA are the most toxic. Is this as expected?
82	1	The sentence "The resulting acute benchmarks, although consistently higher, were also small, with each of the benchmarks falling within a factor of < 2.1 of one another, indicating close agreement between values calculated using either approach." Should be revised. What were small, the difference between values calculated using the different approaches?
83	1	A conclusion for the use of the binning method should be included.
86	References	The reference list needs to be updated as several references used in the text are not included here.
Appendix A		The references are not included in any Reference list. This should be corrected.
Appendix B		Please explain why the four most sensitive values are summarized
C-2	Green alga	The text explaining the Deficiency for this study "Initially identified as Quantitative" is confusing.
C-3	Green alga	Why is the test duration written in red?
E-1		Only the Web-ICE version 3.3 is available online. I assume that the small differences I found between online model parameters and parameters in the Table is due to the updated values in the v4 model?
F-1		The claim "In these situations, a user can either enter the measured toxicity value (LC50/EC50) into the ICE model as µg/L and allow the regression to extrapolate beyond the range of the model or enter a "scaled" toxicity value (i.e. enter the measured

Page	Paragraph	Comments or Questions
		LC50 value as mg/L).” should be supported by a literature reference.
F-73	1	Using the scaled approach, the eight MDRs were fulfilled for seven of the evaluated compounds (not six)?
G-7	Footnote b	It would be more intuitive for the reader to use “Species x SMAV”, similar to what is done for PFAS
G-9		The table lacks references to the footnotes
G-12		Figures G-1 and G-2 are too small. Consider showing panel B below panel A

Reviewer 2

Peer Review Comments on Acute Protective Freshwater Benchmarks for Selected PFAS Using a New Approach Methods

Reviewer 2

I. GENERAL IMPRESSIONS

Establishing environmental standards for the protection human health and the environment is critical. Unfortunately, our laws and regulatory approaches regarding the release of chemicals into the environment have allowed for the generation of thousands of chemicals used in commerce without the requirement that they be tested for safety. This leads us to our current situation where the pace of new chemistries being introduced to the world is far outpacing our ability to evaluate their toxicity to all forms of life including aquatic life. This document reflects this current state-of-affairs. PFAS as a chemical class are ubiquitous with new chemistries being introduced a rapid pace. Few environmental standards for their concentrations in surface waters have been established and toxicity data are limited. Methods to extrapolate toxicity data are unfortunately necessary in light of this data-limited situation. Here, the authors use a "New Approach Method" of Interspecies Correlation Estimation developed by Raimondo and others as a way of generating predictions of toxicity to aquatic organisms such that acute benchmarks for selected PFAS could be established based on the 1985 *Guidelines for the Protection of Aquatic Life*. As these compounds are generally not acutely toxic, and the benchmarks proposed in the document appear to be orders of magnitude higher than expected environmental concentrations, it is unclear what the rationale is for proposing these benchmarks in the absence of chronic benchmarks. By releasing these benchmarks as the only protective values available for the compounds in question, it is possible that the discovery of environmental concentrations that are well below these benchmarks (but could be chronically toxic) might not receive an appropriate response by states and tribes. I generally like the ICE approach to fill data gaps such that environmental standards can be set. It just does not feel like these acute standards are particularly relevant in light of the expected environmental concentrations.

In general, the document requires the reader to consult some of the original literature on ICE models and there is not enough explanation of these approaches contained within the document itself. I would encourage the authors to add more technical information about how ICE models work, what their limitations are, and perhaps build in some uncertainty factors given the language in the Forward section of the document.

II. RESPONSE TO CHARGE QUESTIONS

1. ***Please comment on the overall clarity of the documents and construction as it relates to assessing the effects and derivation of acute benchmarks for the eight selected PFAS.***

   It would be useful to better explain what a Benchmark is related to a Water Quality Criterion. The forward states that benchmarks are "less certain than Water Quality Criteria", but the reader should also be informed about enforcement differences between benchmarks and WQC.

2. ***Please comment on each of the technical approaches used to derive the draft benchmark values for the eight selected PFAS presented in EPA's Derivation of Acute Protective Freshwater Benchmarks for Selected PFAS through a New Approach Method (ICE-based approach using extrapolation [Sections 3-5], ICE-based approach using scaled data [Appendix F], and PFAS benchmark calculations using a data binning approach [Appendix G]).***

    a. ***Are the technical approaches used to derive the benchmark values logical?***

    I think the reader does not have enough information to evaluate ICE-based approaches without consulting the original literature. The introduction to Web-ICE (p. 17) is remarkably brief in explaining the technical approach and this section could be expanded significantly. It is helpful that the reader is given references to read that point to the successful applications of the approach, but this document should be self-contained with respect to describing the technical approach in detail.

    The same criticism can be applied to the scaled data and data binning approaches. These technical approaches should be explained in more detail particularly in terms of the mechanisms by which these different approaches could yield different toxicity estimates.

    In general, I favor action on the creation of environmental standards, even when data are limiting. The approach of making toxicity predictions is logical and the process follows the 1985 Guidelines. However, these guidelines are in need of modernization, and it is unclear to me how aggregated, individual species toxicity tests (that ignore dietary exposure pathways and species interactions) provide compelling evidence for protecting aquatic communities in nature.

    b. ***Does the science support the conclusions?***

    The science tells us that these compounds are not acutely toxic and the benchmark values could give states and tribes a false sense of safety if they encounter high concentrations that are below the benchmark. The language in the document is explicit about the application as a 1 hour maximum every 3 years, but in the absence of more environmentally relevant standards, I'm not sure what these benchmarks do for environmental protection.

    c. ***Are the approaches and resulting values consistent with the protection of aquatic life?***

    See comments above. Adhering to an outdated (1985) understanding of toxicology and species sensitivity differences remains an unfortunate state-of-affairs at EPA.

*d. For the ICE-based models in particular, please compare and contrast the strengths and weaknesses of the extrapolation approach [Sections 3-5] with those of the approach using scaled data [**Appendix F**].*

We are not given enough technical information to make this comparison. The scaled approach information we are given seems more like instructions for how to run the model when values fall out of environmental realism than a detailed description of how it differs technically from the normal model.

3. Please comment on the empirical (direct test) and ICE-generated toxicity data used to derive the benchmark values presented in the draft document.

a. Were the data adequately used and sufficiently comprehensive to represent risks to sensitive aquatic life?

Probably not. There are too few empirical data to be secure in understanding which species in the real world might be sensitive.

b. Were the data selected and/or excluded from the benchmark values derivation appropriately utilized?

I think the authors did the best they could with the available data on hand. It is a shame that more resources are not being deployed to generate more empirical data.

c. Are there relevant data that you are aware of that should be included? If so, please provide for derivation of benchmark values.

There is some mayfly data from the Soucek lab that does not seem to be acknowledged here. The general lack of insect data is a systemic problem – particularly when a single midge is used to represent the toxicity of an entire class or organisms that is likely close to 10,000 species in N. America.

4. Are the derived benchmark values appropriately protective of sensitive aquatic life?

As applied to a one-hour maximum concentration not to be exceeded every 3 years, the values are likely protective to most species.

III. SPECIFIC OBSERVATIONS

Page	Paragraph	Comments or Questions
19	Box	Some classes are quite species rich and a single representative is likely not sufficient (see insects)
20	3	When is Raimondo et al (in review) going to be released? It seems like this should be available information for the reader
throughout		Why are no extra uncertainty factors used other than FAV/2 given the uncertainties associated with the process?

Page	Paragraph	Comments or Questions
throughout		Within class extrapolation is a whole lot of biodiversity to lump together

Reviewer 3

Peer Review Comments on Acute Protective Freshwater Benchmarks for Selected PFAS Using a New Approach Methods

Reviewer 3

I. GENERAL IMPRESSIONS

The document is generally well-written and easy to follow, especially in the discussion of the data used and derivation of the benchmark. Greater detail and perhaps some examples would help in the discussion of the difference between "extrapolated" and "scaled" ICE-based data. It is not clear exactly what the difference is between these, or at least the implications of the two approaches other than to increase the number of species represented. Reviewing both the Raimondo et al (2010) and Willming et al (2016) papers did not address this topic. Raimondo et al (202?) is "in review" and perhaps will address the issues; however, the manuscript was not included in the review materials. The following discussion applies to the document and approach and was not specifically addressed in the charge questions.

To sum up the issue, the problem is that there is insufficient high-quality empirical data available to derive AWQC for the "selected" PFAS compounds. This issue is not new, data limitations in deriving Ambient Water Quality Criteria (AWQC) have been an issue since shortly after the implementation of the Clean Water Act (Kimerle et. al., 1985[1]) and became a greater concern with the reduction of AWQC data development at the EPA-ORD research laboratories. EPA previously proposed an approach to address this issue in the Type II standards methodology developed as part of the Great Lakes Initiative (GLI) published in 1995 (USEPA 40 CFR 9, 122, 123, 131, and 132, Final Water Quality Guidance for the Great Lakes System; Final Rule, March 23 1995). EPA presented a method to develop Secondary Maximum Concentrations (SMC) and the Secondary Continuous Concentrations (SCC) based on data sets that were insufficient to satisfy the eight minimum data requirements (MDR) to derive a national AWQC. Briefly, a secondary acute value (SAV) is calculated by dividing the lowest GMAV in the database by a Secondary Acute Factor (SAF) that is designated in Table A-1 in the document (ranging from 4.3 to 21.9) based on the number of satisfied MDRs available for the compound. This approach is somewhat crude and certainly lacks a great deal of technical basis; nonetheless, it probably should be discussed in the current document. In addition, application of the GLI technique to the PFAS compounds in this document has been previously conducted and presented in Gripo et al (2021)[2]. Resulting values between the Gripo et al (2021) report and application of the method to EPA's data result in slightly differing values, likely due to the acceptance and availability of different empirical data. In Table 1, a comparison of the benchmarks reported in this document (using the ICE approach) is made with the values calculated using the GLI approach (based on the empirical data reported in this document).

[1] *Kimerle, R. A., Werner, A. F., and Adams, W. J., "Aquatic Hazard Evaluation Principles Applied to the Development of Water Quality Criteria," Aquatic Toxicology and Hazard Assessment: Seventh Symposium, ASTMSTP854, R. D. Cardwell, R. Purdy, and R. C. Bahner, Eds., American Society for Testing and Materials, Philadelphia, 1985, pp. 538-547*

[2] *M. Grippo, J. Hayse, I. Hlohowskyj, and K. Picel. 2021. Derivation of PFAS Ecological Screening Values. Environmental Science Division. Argonne National Laboratory. September 2021.*

Table 1. Comparison of benchmark values using the GLI and ICE-based derivation methods.

Chemical	Lowest empirical value	GLI factor	Tier II GLI calculated value (mg/L)	EPA Benchmark (Extrapolation) FAV/2 (mg/L)
Perfluorobutanoic acid (PFBA)	110	8	13.75	83
Perfluorobutanesulfonic acid (PFBS)	1938	13	149.1	183
Perfluorononanoic acid (PFNA)	27.84	13	2.14	10.3
Perfluorodecanoic acid (PFDA)	32	8	4	7.9
Perfluorohexanoic acid (PFHxA)	140	8	17.5	75
Perfluorohexanesulfonic acid (PFHxS)	22.5	13	1.7	9.1
Hexadecafluoro-2-decenoic acid (8:2 FTUCA)	3.2	13	0.24	0.58
Pentadecafluorodecanoic acid (7:3 FTCA)	0.959	13	0.074	0.18

Other comments not specifically addressed in the charge questions are provided below:

- In at least three cases (i.e., PFBS, PFNA, and PFHxS) the derived criteria, did not comply with the MDR minimum of n=8, it appears that EPA disregarded the MDR minimum for these materials and calculated the benchmark with fewer MDRs, i.e., 7. This is not keeping with the 1985 guidance that states: "*Similarly. if all required data are not available, a numerical criterion should not be derived except in special cases."* This should be acknowledged in the text and some statement regarding the minimum number of MDRs to calculate a benchmark addressed.
- Table 2 summarizes the MDRs available for each of the 8 PFAS compounds. In all cases 62.5 to 75% of the MDR data used in deriving the SSD-based benchmark was estimated using the ICE model method. Thus, most of the data used for MDRs and to develop SSDs, are estimated values. It is interesting that in the best case, only 1 empirical data point was available among the 4 most sensitive species for 50% of the materials. The other 4 materials had no empirical data represented among the most sensitive species. Although there may be good correlations between species making estimation for one species based on data from another possible, questions remain regarding the extent of the role that estimated values should play in the derivation of water quality criteria, standards, or benchmarks.

Table 2. Numbers of MDRs used in derivation of benchmarks

Chemical	# of empirical MDRs	# of estimated MDRs	Total MDRs met	# quantitively accepted empirical data in 4 lowest species
Perfluorobutanoic acid (PFBA)	3	5	8	1
Perfluorobutanesulfonic acid (PFBS)	2	5	7	0
Perfluorononanoic acid (PFNA)	2	5	7	1
Perfluorodecanoic acid (PFDA)	3	5	8	0
Perfluorohexanoic acid (PFHxA)	3	5	8	1
Perfluorohexanesulfonic acid (PFHxS)	2	3	5	1
Hexadecafluoro-2-decenoic acid (8:2 FTUCA)	2	6	8	0
Pentadecafluorodecanoic acid (7:3 FTCA)	2	6	8	0

- If EPA is going to revise the AWQC minimum MDR data requirement (i.e., 8) for the purpose of "Benchmark" derivation, then EPA should develop guidance regarding the minimum quantity and quality of **empirical** data required before a benchmark can be derived; this should potentially include:
    - a minimum number of empirical data required to be contained among the 4 lowest species (should benchmarks should be derived based solely on "estimated" values?). Is there a minimum percent of empirical to estimated data that should be met to establish a benchmark?
    - a requirement for empirical data with a representation among a base set of organisms that have historically been shown to be sensitive to a range of toxicants, e.g., *Ceriodaphnia dubia*, fathead minnow, would be useful. Kimerle et al (1985)[3] suggested that a minimum base data set composed of an algae, daphnid, and fish could consistently predict the most sensitive species based on available data at the time. Use of the ICE database could help to identify species consistently shown to be among those predicted to be most sensitive among chemical groups with a common mode-of-action. The table below, composed from the ICE-modeled data for PFAS compounds, suggests that a base data set composed of a freshwater mussel, cladoceran, and fish, would cover the majority of predicted most-sensitive species for PFAS compounds.

Table 3. Comparison of the species sensitivity ranking from the ICE model for each of the PFAS compounds

Chemical	Species Sensitivity Rank from ICE model			
	1	2	3	4
Perfluorobutanoic acid (PFBA)	*Brachionus calyciflorus* (rotifer)	*Oncorhynchus mykiss* (rainbow trout)	*Gammarus fasciatus* (amphipod)	*Amblema plicata* (mussel)
Perfluorobutanesulfonic acid (PFBS)	*Amblema plicata* (mussel)	*Gammarus fasciatus* (amphipod)	*Hyalella azteca* (amphipod)	*Ceriodaphnia dubia* (cladoceran)
Perfluorononanoic acid (PFNA)	*Amblema plicata* (mussel)	*Chydorus sphaericus* (cladoceran)	*Megalonias nervosa* (mussel)	*Oncorhynchus mykiss* (rainbow trout)
Perfluorodecanoic acid (PFDA)	*Caecidotea brevicauda* (isopod)	*Micropterus salmoides* (bass)	*Perca flavescens* (yellow perch)	*Salvelinus fontinalis* (brook trout)
Perfluorohexanoic acid (PFHxA)	*Brachionus calyciflorus* (rotifer)	*Amblema plicata* (mussel)	*Gammarus fasciatus* (amphipod)	*Chydorus sphaericus* (cladoceran)
Perfluorohexanesulfonic acid (PFHxS)	*Danio rerio* (zebrafish)	*Jordanella floridae* (flagfish)	*Daphnia magna* (cladoceran)	*Limnodrilus hoffmeisteri* (oligochaete)
Hexadecafluoro-2-decenoic acid (8:2 FTUCA)	*Amblema plicata* (mussel)	*Palaemonetes kadiakensis* (grass shrimp)	*Chydorus sphaericus* (cladoceran)	*Megalonias nervosa* (mussel)
Pentadecafluorodecanoic acid (7:3 FTCA)	*Amblema plicata* (mussel)	*Macrobrachium nipponense* (river shrimp)	*Chydorus sphaericus* (cladoceran)	*Megalonias nervosa* (mussel)

    - a minimum number of empirically derived MDR data points should be established, and a maximum number of ICE-estimated values that are allowed to be considered in an SSD should be established.
- If EPA chooses to develop criteria/benchmarks for materials that have limited empirical data, then a two-tiered approach should be adopted much like that previously adopted by EPA for the GLI program. It is unclear from the document what the long-term intent and

[3] *Kimerle, R. A., Werner, A. F., and Adams, W. J., "Aquatic Hazard Evaluation Principles Applied to the Development of Water Quality Criteria," Aquatic Toxicology and Hazard Assessment: Seventh Symposium, ASTMSTP854, R. D. Cardwell, R. Purdy, and R. C. Bahner, Eds., American Society for Testing and Materials, Philadelphia, 1985, pp. 538-547.*

regulatory status will be for "benchmarks." Those materials that do not have sufficient empirical data to permit derivation of a "Tier 1" criteria, could be addressed by a "Tier II" benchmark, as suggested. The Tier II benchmark could be derived using the proposed ICE-based methodology or a method like that used for the GLI Secondary Acute Values. That said, will benchmarks serve the same purpose as the current AWQC? They do not have the same scientific basis as "Tier I" AWQC, but if adopted as "standards" by states and tribes, they will have the same regulatory/legal status. This is briefly addressed in the document's forward; however, greater clarity regarding the "scientific confidence" and "regulatory validity" could be provided. Questions regarding use of the approach should be considered, for example, if a state developed a proposed standard for a chemical based on limited empirical data and relying predominately on ICE-estimated data (e.g., 7:3 FTCA), would EPA approve it?

II. RESPONSE TO CHARGE QUESTIONS

5. ***Please comment on the overall clarity of the documents and construction as it relates to assessing the effects and derivation of acute benchmarks for the eight selected PFAS.***

   In general, the document is clear and well-organized. The sections of the document follow a "template" making the subsections parallel to each other making it easy to compare. As previously suggested, some additional detail regarding the extrapolation and scaled estimation technique would be helpful. This may be contained in the Raimondo et al (in review) document but it was not provided for this review. Also, a brief discussion of the "binning" technique (Giddings et al 2019) was made in Section 5.10 and in Appendix G and a comparison of the benchmarks derived using the binning vs ICE-based techniques is provided, but discussion or assessment of the utility/advantages/disadvantages of the technique is not provided.

6. ***Please comment on each of the technical approaches used to derive the draft benchmark values for the eight selected PFAS presented in EPA's Derivation of Acute Protective Freshwater Benchmarks for Selected PFAS through a New Approach Method (ICE-based approach using extrapolation [Sections 3-5], ICE-based approach using scaled data [Appendix F], and PFAS benchmark calculations using a data binning approach [Appendix G]).***

   a. ***Are the technical approaches used to derive the benchmark values logical?***

      The approaches used by the three methods are logical and creative methods to address the issue of data limitations. Although the calculated data are provided in the document and in Appendix F, it is difficult for this reviewer to fully understand the technical differences in the ICE-based approach using extrapolation or scaled data. Perhaps hands on evaluation of the models or review of the Raimondo et al (in review) manuscript would help. At the least, an example showing calculations both ways would help the reader.

PFAS benchmark calculations using the data binning approach presented in Giddings et al (2019) are logical and provide a method to expand the quantity of empirical data provided that the assumption of a common mode of action (MOA) is valid. Raimondo et al (2010) notes that MOA-specific models are more robust and improve the fit of the ICE model approach.

Approaching the problem of missing data using the ICE model and binning techniques are more elegant than the previous GLI technique and are more generally scientifically supportable.

b. Does the science support the conclusions?

I am not sure what is being asked since there are no specific conclusions called out in the document. The approach of estimating species sensitivities to toxins is logical and scientifically supportable, provided that sufficient underpinning data are available to support the models. Derivation of 1985 AWQC compliant values (FAV/FCVs) are dependent on 5 datapoints; the four lowest Genus Mean Values (GMAV/GMCV) and the total number of species represented in the database. The calculation is more sensitive to the relationship of the 4 lowest values than to the total number of species represented. As an example, Figure 1 provides an example of increasing the size of the database for a compound (related to the number of available ICE models); increasing from 8 GMAVs to 42 results in a slightly less than a doubling in the calculated FAV (8 to 14.5). So, the choice of extrapolated vs scaled ICE models may result in a slight increase in the calculated benchmark due to acceptance of more models. Far more important is the validity and relationship of the 4 lowest GMAVs. It is critical that these values be valid and as accurate as possible. Relying on estimated values can introduce a large degree of uncertainty in the resulting benchmark value.

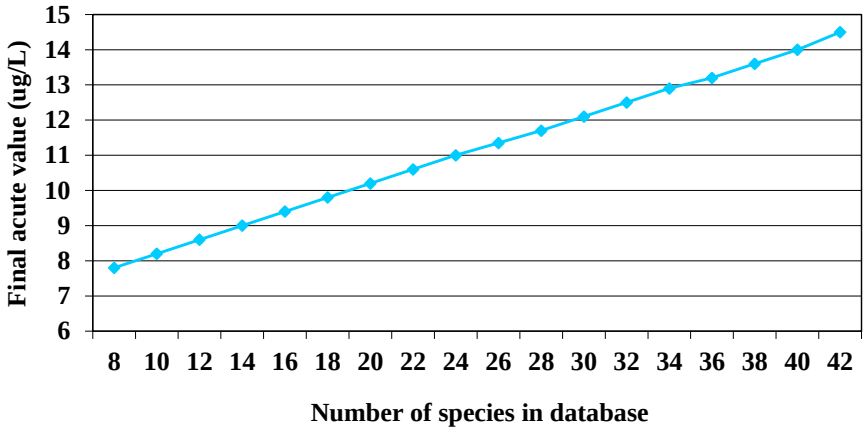


Figure 1. Final Acute and Chronic values are sensitive to the number of species in the database

c. Are the approaches and resulting values consistent with the protection of aquatic life?

The table below summarizes the benchmarks calculated using the various methods described in the reviewed document and the values calculated using the GLI method. Values vary substantially, in some cases as much as an order of magnitude. The Tier II GLI value frequently provided the lowest calculated value; however, the values were not inconsistent with the other methods. Given the minimal amount of empirical data available for these materials, it is difficult to identify if the values are *"consistent with the protection of aquatic life."* Perhaps conducting an analysis with a data rich compound (e.g., copper or a pesticide), using only a limited portion of the available data followed by a comparison to the full AWQC database would give some insight into the comparability of the benchmark and the standard AWQC approach.

Table 4. Summary of calculated benchmark values

Chemical	EPA Benchmark (Extrapolation) (mg/L)	EPA Benchmark (Scaled) (mg/L)	Binning approach (Guidelines-based) (mg/L)	Binning approach (SSD-based) (mg/L)	Tier II GLI calculated value (mg/L)
Perfluorobutanoic acid (PFBA)	83	174	194	467	13.75
Perfluorobutanesulfonic acid (PFBS)	183	237	24	102	149.1
Perfluorononanoic acid (PFNA)	10.3	12	3.4	8.3	2.14
Perfluorodecanoic acid (PFDA)	7.9	10	4.9	12	4
Perfluorohexanoic acid (PFHxA)	75	95	43	103	17.5
Perfluorohexanesulfonic acid (PFHxS)	9.1	9.4	0.18	0.76	1.7
Hexadecafluoro-2-decenoic acid (8:2 FTUCA)	0.58	0.65			0.24
Pentadecafluorodecanoic acid (7:3 FTCA)	0.18	0.23			0.074

d. For the ICE-based models in particular, please compare and contrast the strengths and weaknesses of the extrapolation approach [Sections 3-5] with those of the approach using scaled data [Appendix F].

As stated above, the technical approach between these techniques is not clear. The implication of the use of the scaled approach rather that the extrapolation method results in the acceptance of more GMAVs, thus resulting in an increased benchmark.

7. Please comment on the empirical (direct test) and ICE-generated toxicity data used to derive the benchmark values presented in the draft document.

a. Were the data adequately used and sufficiently comprehensive to represent risks to sensitive aquatic life?

As stated, the problem is a lack of empirical data and a reliance on data estimation techniques. The table below provides the data for the 4 most sensitive species used to derive the benchmark values for the PFAS materials. Actual empirical data are highlighted (4 of 32 data points, 12.5%); as you can see, most of the data for the range of species are estimated values. Without additional confirmatory experimental data, it is difficult to say anything about how comprehensive the data represent sensitive aquatic organisms.

Table 5. Species sensitivity ranking for PFAS compounds. Empirical data are in bold, all others are ICE-estimated values.

Chemical	Species Sensitivity Rank from ICE model[1]			
	1	2	3	4
Perfluorobutanoic acid (PFBA)	***Brachionus calyciflorus* (rotifer)**	*Oncorhynchus mykiss* (rainbow trout)	*Gammarus fasciatus* (amphipod)	*Amblema plicata* (mussel)
Perfluorobutanesulfonic acid (PFBS)	*Amblema plicata* (mussel)	*Gammarus fasciatus* (amphipod)	*Hyalella azteca* (amphipod)	*Ceriodaphnia dubia* (cladoceran)
Perfluorononanoic acid (PFNA)	*Amblema plicata* (mussel)	***Chydorus sphaericus* (cladoceran)**	*Megalonias nervosa* (mussel)	*Oncorhynchus mykiss* (rainbow trout)
Perfluorodecanoic acid (PFDA)	*Caecidotea brevicauda* (isopod)	*Micropterus salmoides* (bass)	*Perca flavescens* (yellow perch)	*Salvelinus fontinalis* (brook trout)
Perfluorohexanoic acid (PFHxA)	***Brachionus calyciflorus* (rotifer)**	*Amblema plicata* (mussel)	*Gammarus fasciatus* (amphipod)	*Chydorus sphaericus* (cladoceran)
Perfluorohexanesulfonic acid (PFHxS)	***Danio rerio* (zebrafish)**	*Jordanella floridae* (flagfish)	*Daphnia magna* (cladoceran)	*Limnodrilus hoffmeisteri* (oligochaete)
Hexadecafluoro-2-decenoic acid (8:2 FTUCA)	*Amblema plicata* (mussel)	*Palaemonetes kadiakensis* (grass shrimp)	*Chydorus sphaericus* (cladoceran)	*Megalonias nervosa* (mussel)
Pentadecafluorodecanoic acid (7:3 FTCA)	*Amblema plicata* (mussel)	*Macrobrachium nipponense* (river shrimp)	*Chydorus sphaericus* (cladoceran)	*Megalonias nervosa* (mussel)

[1] It should also be noted that the 4 data points listed in bold are all based on nominal concentrations, so the accuracy of the EC50 values may be questioned.

b. Were the data selected and/or excluded from the benchmark values derivation appropriately utilized?

Much of the data accepted would not meet current standards for data acceptability or criteria derivation. The authors have tried to maintain a degree of fidelity to the Stephan et al (1985) AWQC methodology; however, several concerns exist with the data considered:

- The benchmark document states: "Toxicity studies accessed from the ECOTOX database were further evaluated by Office of Water. Studies were evaluated for data quality as described by EPA OW's data quality standard operating procedure (SOP), and consistent with OW's data quality review approach U.S. EPA (1985), and EPA's Office of Chemical Safety and Pollution Prevention (OPP)'s Ecological Effects Test Guidelines (U.S. EPA 2016c)." These documents were not included in the reference list and were not provided.

- In general, most of the accepted empirical studies are reported based on nominal exposure concentrations rather than analytically measured concentrations; this is not consistent with state-of-the-science standards of acceptability for empirical toxicology data. In fact, one of the studies that measured exposure concentrations (Ding et al 2012) ultimately reported test endpoint data (EC50) based on nominal concentrations rather than measured values. The reported analytical data indicates that test concentrations differed from nominals by 10-20%, so the value reported based on nominals is likely to be 10-20% off.
- Some of the toxicity data used in the derivation of the aquatic benchmarks comes from studies that used non-native species (i.e., zebrafish, *Danio rerio*), which adds uncertainty associated with the representativeness of such species to native North American aquatic fauna. Stephan et al (1985) states: II. G. "Questionable data, data on formulated mixtures and emulsifiable concentrates, and data obtained with non-resident species in North America or previously exposed organisms may be used to provide auxiliary information **but should not be used in the derivation of criteria**." Appendix I Resident North American Species of Aquatic Animals Used in Toxicity and Bioconcentration Tests defines zebrafish (*Danio rerio*) as "Non-resident" species and therefore should not be included for criteria derivation. Use of Non-resident species is briefly discussed in the report and a reference to US EPA 2018b is cited; however, this reference is not included in the reference list. The zebrafish is in the family *Cyprinidae*, which all North American native minnows (including the fathead minnow), shiners, and dace belong. Although not native to North America, EPA seems to have decided that in the absence of suitable data on native cyprinids, the zebrafish is an acceptable representative. However, given that zebrafish are frequently among the more sensitive species and at least some studies with PFAS compounds have suggested that fathead minnows may be more sensitive[4], it would be good to have some comparative additional data with NA species.
- Some of the test methods used are not consistent with the 1985 guidance. EPA 1985 states that "Acute EC50s that are based on effects that are not severe, such as reduction in shell deposition and reduction in growth, are not used in calculating the Final Acute Value." The zebrafish tests included in the benchmark document (Annunziato et al 2020) followed the OECD 236 method and reported results based on a growth rather than survival endpoint. These data would not be acceptable for derivation of an FAV based on the 1985 guideline.
- In addition to the above concerns, it was noted that at least two studies (Ding et. al. 2012, Annunziato et al 2020) that reported tests with PFBA, PFBS, PFNA, PFDA, and PFHxS conducted their studies using

[4] Suski et al. 2023. Ecotoxicity and Accumulation of Perfluorononanoic Acid in the Fathead Minnow (*Pimephales promelas*) and an Approach to Developing Protective Thresholds in the Aquatic Environment Through Species Sensitivity Distribution. Environ Toxicol Chem. https://doi.org/10.1002/etc.5692

dimethylsulfoxide (DMSO) as a carrier solvent. In EPA's current test guidelines, it is recommended that if a carrier solvent must be used, *"Preferred solvents are dimethyl formamide, triethylene glycol, methanol, acetone, or ethanol. Solvent use should be avoided if possible."* DMSO is known to transport nonionized molecules thorough many biological membranes (Jacob and Herschler 1985[5]). Although the authors of the lab tests conducted a "solvent control" this does not control for possible synergistic interactions of DMSO acting as a membrane carrier, thus potentially increasing observed toxicity. Because these tests represent a large portion of the quantitatively acceptable freshwater toxicity tests (20%, 7 of 36), EPA should consider the potential for inclusion of these data resulting in lower than desired criteria, due to an overestimation of toxicity due to DMSO synergy.

EPA should reassess the quality and acceptability of the available data for regulatory purposes.

c. Are there relevant data that you are aware of that should be included? If so, please provide for derivation of benchmark values.

Several recent publications have critically reviewed the available data for PFAS compounds, e.g., Pandelides et al. 2023[6]. The references below are just an example of amphibian references included in one of the review articles, these include acute and chronic endpoints:

Abercrombie, S. A., de Perre, C., Choi, Y. J., Tornabene, B. J., Sepúlveda, M. S., Lee, L. S., & Hoverman, J. T. (2019). Larval amphibians rapidly bioaccumulate poly- and perfluoroalkyl substances. *Ecotoxicology and Environmental Safety*, **178**, 137–145. https://doi.org/10.1016/j.ecoenv.2019.04.022;

Ankley, G. T., Kuehl, D. W., Kahl, M. D., Jensen, K. M., Butterworth, B. C., & Nichols, J. W. (2004). Partial life-cycle toxicity and bioconcentration modeling of perfluorooctane sulfonate in the northern leopard frog (*Rana pipiens*). *Environmental Toxicology and Chemistry*, **23**, 2745. https://doi.org/10.1897/03-667

Brown, S. R., Flynn, R. W., & Hoverman, J. T. (2021). Perfluoroalkyl substances increase susceptibility of northern leopard frog tadpoles to trematode infection. *Environmental Toxicology and Chemistry*, **40**, 689–694. https://doi.org/10.1002/etc.4678

Flynn, R. W., Chislock, M. F., Gannon, M. E., Bauer, S. J., Tornabene, B. J., Hoverman, J. T., & Sepúlveda, M. S. (2019). Acute and chronic effects of

[5] *Jacob, S. W., & Herschler, R. (1986). Pharmacology of DMSO. Cryobiology, 23(1), 14-27.*

6 *Pandelides Z, J Conder, Y Choi, E Allmon, T Hoskins, L Lee, J Hoverman, M Sepúlveda. 2023. A Critical Review of Amphibian Per- and Polyfluoroalkyl Substance Ecotoxicity Research Studies: Identification of Screening Levels in Water and Other Useful Resources for Site-Specific Ecological Risk Assessments. Environ Toxicol Chem. https://doi.org/10.1002/etc.5695*

perfluoroalkyl substance mixtures on larval American bullfrogs (*Rana catesbeiana*). *Chemosphere*, **236**, 124350. https://doi.org/10.1016/j.chemosphere.2019.124350

Flynn, R. W., Hoover, G., Iacchetta, M., Guffey, S., de Perre, C., Huerta, B., Li, W., Hoverman, J. T., Lee, L., & Sepúlveda, M. S. (2022). Comparative toxicity of aquatic per- and polyfluoroalkyl substance exposure in three species of amphibians. *Environmental Toxicology and Chemistry*, **41**, 1407–1415. https://doi.org/10.1002/etc.5319

Flynn, R. W., Iacchetta, M., Perre, C., Lee, L., Sepúlveda, M. S., & Hoverman, J. T. (2021). Chronic per-/polyfluoroalkyl substance exposure under environmentally relevant conditions delays development in northern leopard frog (*Rana pipiens*) larvae. *Environmental Toxicology and Chemistry*, **40**, 711–716. https://doi.org/10.1002/etc.4690

Foguth, R. M., Hoskins, T. D., Clark, G. C., Nelson, M., Flynn, R. W., de Perre, C., Hoverman, J. T., Lee, L. S., Sepúlveda, M. S., & Cannon, J. R. (2020). Single and mixture per- and polyfluoroalkyl substances accumulate in developing northern leopard frog brains and produce complex neurotransmission alterations. *Neurotoxicology and Teratology*, **81**, 106907. https://doi.org/10.1016/j.ntt.2020.106907

Fort, D. J., Mathis, M. B., Guiney, P. D., & Weeks, J. A. (2019). Evaluation of the developmental toxicity of perfluorooctane sulfonate in the Anuran, *Silurana tropicalis*. *Journal of Applied Toxicology*, **39**, 365–374. https://doi.org/10.1002/jat.3727

Hoover, G. M., Chislock, M. F., Tornabene, B. J., Guffey, S. C., Choi, Y. J., De Perre, C., Hoverman, J. T., Lee, L. S., & Sepúlveda, M. S. (2017). Uptake and depuration of four per/polyfluoroalkyl substances (PFAS) in northern leopard frog *Rana pipiens* tadpoles. *Environmental Science and Technology Letters*, **4**, 399–403. https://doi.org/10.1021/acs.estlett.7b00339

Hoskins, T. D., Allmon, E. B., Flynn, R. W., Lee, L. S., Choi, Y., Hoverman, J. T., & Sepúlveda, M. S.(2022). An environmentally relevant mixture of perfluorooctane sulfonic acid and perfluorohexane sulfonic acid does not conform to additivity in northern leopard frogs exposed through metamorphosis. *Environmental Toxicology and Chemistry*, **41**, 3007–3016. https://doi.org/10.1002/etc.5486

Lech, M. E., Choi, Y. J., Lee, L. S., Sepúlveda, M. S., & Hoverman, J. T. (2022). Effects of per- and polyfluoroalkyl substance mixtures on the susceptibility of larval American bullfrogs to parasites. *Environmental Science & Technology*, **56**, 15953–15959. https://doi.org/10.1021/acs.est.2c04574

8. Are the derived benchmark values appropriately protective of sensitive aquatic life?

The benchmark methods attempt to maintain compliance with the EPA's 1985 method for derivation of AWQC and to the extent that the 1985 method was "*appropriately protective of sensitive aquatic life"* the benchmark approach should be as well. However, one major difference between the 1985 guidance and the new benchmark approach is that the requirements for high-quality empirical data for a minimal range of aquatic species have been reduced or eliminated. The benchmark approach seems to rely on existing data

or extrapolation of limited data from similar compounds. ICE methods are extremely useful and important in estimating values for species where we cannot generate empirical data, e.g., T&E species (Willming et al 2016). However, the benchmark approach proposed seems to minimize the utility and need for empirical data. Modelling techniques that are based on robust empirical data are extremely useful for **supplementing** extant data for species-of-concern that cannot be easily or cost-effectively tested, or tested due to regulatory restrictions, but they should not suppliant the need for chemical-specific empirical data. Is there a minimum amount of empirical data that are needed to derive a benchmark? In theory, a single acute toxicity test may be sufficient, using ICE-models, to derive regulatory benchmarks. To gain confidence in the proposed approach, EPA should conduct testing to further confirm the accuracy of the ICE estimates, especially for the most sensitive species in the benchmark data sets.

III. SPECIFIC OBSERVATIONS

Page	Paragraph	Comments or Questions
x	2	Are benchmarks expected to carry the same weight as AWQC if adopted as state or tribal standards?
14	1	Should there be inclusion of EPA's GLI approach?
17	2	It is difficult to evaluate this statement since the Raimondo et al report is unpublished and not supplied.
23	1	The document references USEPA 2016c for information on how data were evaluated. The reference is not in the reference list and was not provided. It is critical to assess the acceptability of the empirical data accepted in Appendix A. USEPA 1985 was not included in the reference list.
23	2	The document references USEPA 2018b however, the reference is not in the reference list and was not provided.

Reviewer 4

Peer Review Comments on Acute Protective Freshwater Benchmarks for Selected PFAS Using a New Approach Methods

Reviewer 4

I. GENERAL IMPRESSIONS

The draft *Derivation of Acute Protective Freshwater Benchmarks for Selected PFAS Compounds through a New Approach Method (NAM)* was well organized and well written. The accuracy of each element of the derivation process was satisfactory. The structure and writing of the draft document clearly communicated the rationale and the process of deriving the benchmarks. I think that the proposed benchmarks are reasonable and protective of aquatic life based on the acceptable empirical acute toxicity data available for the eight PFAS. The major source of uncertainty for the derived benchmarks is the lack of acceptable empirical acute toxicity data on freshwater primary producers and freshwater invertebrates for the eight PFAS. I think the "New Approach Method" is reasonable when there is a lack of acceptable empirical data, and a benchmark needs to be derived. However, the NAM should not replace the derivation of benchmarks with empirical data.

II. RESPONSE TO CHARGE QUESTIONS

1. ***Please comment on the overall clarity of the documents and construction as it relates to assessing the effects and derivation of acute benchmarks for the eight selected PFAS.***

   I thought the overall clarity of the writing and construction of the document were good. I found the structure logical and easy to follow in the progression of the derivation process. I would not recommend any changes to the overall writing or structure of the document. There were a few spelling and grammatical errors but all very minor.

2. ***Please comment on each of the technical approaches used to derive the draft benchmark values for the eight selected PFAS presented in EPA's Derivation of Acute Protective Freshwater Benchmarks for Selected PFAS through a New Approach Method (ICE-based approach using extrapolation [Sections 3-5], ICE-based approach using scaled data [Appendix F], and PFAS benchmark calculations using a data binning approach [Appendix G]).***

   a. ***Are the technical approaches used to derive the benchmark values logical?***

      I think the technical approaches taken to derive the benchmark values were logical. In the absence of acceptable empirical data, the use of the ICE models to generate a data-rich SSD is a logical approach. Obviously, the benchmarks could be re-evaluated if and when acceptable empirical data is available to regulators.

   b. ***Does the science support the conclusions?***

      Overall, yes, the conclusions are supported by the available science.

An important element of the science supporting the conclusions is the validation of the ICE models to predict the acute toxicity of PFAS. In the benchmark document on page 17, an unpublished work by Raimonda et al. is cited to support the validation of the ICE models to predict acute toxicity of PFAS. The documents states, "*ICE models have been developed from a broad range of chemicals (e.g., metals and other inorganics, pesticides, solvents, and reactive chemicals) and across a wide range of toxicity values and have been validated as accurate predictors of PFAS acute toxicity when model criteria parameters are followed (Raimondo et al., in review).*" As the unpublished manuscript by Raimondo et al. is not available as part of this review, I am left to assume that the statement made in the benchmark document about the validation of the ICE models to predict the acute toxicity of PFAS to be accurate.

Another question is whether the statement "…*have been validated as accurate predictors of PFAS acute toxicity…"* is solely based on data with PFOS and PFOA?

c. Are the approaches and resulting values consistent with the protection of aquatic life?

Yes, I think the approaches and resulting values are consistent with the protection of aquatic life based on the acceptable empirical data that was available to the assessors.

d. For the ICE-based models in particular, please compare and contrast the strengths and weaknesses of the extrapolation approach [Sections 3-5] with those of the approach using scaled data [Appendix F].

Based on the benchmarks derived using the two different approaches, the extrapolation approach generated lower benchmarks across the eight PFAS compared to using scaled data. For the purpose of the protection of aquatic life, the extrapolation approach would be more protective than the scaled approach. I don't know if this would be the case for other groups of chemicals, but it appears that for PFAS, the extrapolation approach is a more protective approach. In the absence of acceptable empirical data, the more protective approach should be selected. This is critical to avoiding a type II error (i.e., false negative), which is an important consideration in risk assessment.

Benchmarks derived using extrapolation approach:

Chemical[1]	**PFBA** (mg/L)	**PFBS** (mg/L)	**PFNA** (mg/L)	**PFDA** (mg/L)	**PFHxA** (mg/L)	**PFHxS** (mg/L)	**8:2 FTUCA** (mg/L)	**7:3 FTCA** (mg/L)
Magnitude	83	183	10.3	7.9	75	9.1	0.58	0.18
Duration	One hour average							
Frequency	Not to be exceeded more than once in three years on average							

Benchmarks derived using scaled approach:

Chemical[1]	PFBA (mg/L)	PFBS (mg/L)	PFNA (mg/L)	PFDA (mg/L)	PFHxA (mg/L)	PFHxS (mg/L)	8:2 FTUCA (mg/L)	7:3 FTCA (mg/L)
Magnitude	174	237	12	10	95	9.4	0.65	0.23
Duration	One hour average							
Frequency	Not to be exceeded more than once in three years on average							

Based on the reported validation that has been conducted by Raimondo et al., both approaches seem reasonable. It would be nice to be able to see the validation that has been conducted by Raimondo et al., but it appears that this manuscript is currently in review.

3. ***Please comment on the empirical (direct test) and ICE-generated toxicity data used to derive the benchmark values presented in the draft document.***

   a. ***Were the data adequately used and sufficiently comprehensive to represent risks to sensitive aquatic life?***

   I think the available empirical was adequately used.

   However, there was an overall lack of empirical acute toxicity data, particularly for freshwater primary producers and freshwater invertebrates. For example, there was no acute toxicity data on the eight PFAS for freshwater primary producers. Freshwater invertebrates were also under-represented in the empirical data set. For example, there was only three empirical data points for PFHxS and all three were for freshwater vertebrates. When empirical toxicity data on the eight PFAS was available for a freshwater invertebrate species, it was usually *Daphnia magna*. Consequently, I don't think that the data is sufficiently comprehensive to represent risk to sensitive aquatic life.

   b. ***Were the data selected and/or excluded from the benchmark values derivation appropriately utilized?***

   Yes, the process of including and excluding empirical data for derivation of the benchmark values was clearly explained, logical, and well established.

   c. ***Are there relevant data that you are aware of that should be included? If so, please provide for derivation of benchmark values.***

   I am not aware of relevant data that should be included in this process. An extensive and complete review of available data has been conducted in preparation for this process of deriving benchmarks.

4. Are the derived benchmark values appropriately protective of sensitive aquatic life?

I do have concerns about the lack of acceptable empirical acute toxicity data for freshwater primary producers and freshwater invertebrates. I think the process of deriving benchmarks for the eight PFAS described by the USEPA is appropriate for the empirical data that is available. I think they have done their best with the data that is available to them.

III. SPECIFIC OBSERVATIONS

Page	Paragraph	Comments or Questions
xi	Second	Period missing at the end of "…*minimum data requirements (MDRs) to calculate aquatic life criteria*"
xii	First	Space needed between "…*(Guidelines)(U.S. EPA 1985*)."
xii	Table Ex-1-1	Not clear how the "*Duration*" and "*Frequency*" were determined for the recommended benchmarks. Is this standard for USEPA acute freshwater aquatic life benchmarks?
14	First	A closing bracket missing in "• *Perfluorohezanesulfonic acid (PFHxS (CAS# 355464, 108427538, 3871996, 82382125)"*
15	First	The first word in f), g), and h) is not capitalized as in a) to e). *d) "insect (e.g., mayfly, dragonfly, damselfly, stonefly, caddisfly, mosquito, midge, etc.)* *e) family in a phylum other than Arthropoda or Chordata (e.g., Rotifera, Annelida, Mollusca, etc.)* *f) family in any order of insect or any phylum not already represented*
20	Final sentence on the page	The final sentence of this section is "*Benchmark values for the eight PFAS using this alternative approach are summarized in* ***Section 5.10****."* I would specifically reference Table 5-26 in section 5.10 at the end of the sentence above. It would make it easier for the reader to find the benchmarks derived using the scaled approach. Or reference section F.9 and/or Table F-29.
23	First	The first sentence in the paragraph states, "*Empirical studies available for the eight PFAS were identified using the ECOTOXicology Knowledgebase…*" I assume that the "eight PFAS" refers to the compounds for which the benchmarks are being set. However, later in the paragraph, there is a sentence that states, "*Further, only single chemical toxicity tests with PFOA were considered for possible inclusion in benchmark derivation*." PFOA is not one of the eight PFAS for which a benchmark is being set, so it is not clear why acute toxicity data for PFOA is being used. It was stated earlier in section 3 that validation was conducted using measured and predicted values for PFOS and

Page	Paragraph	Comments or Questions
		PFOA, but it is not clear at this point in section 4 how acute toxicity data for PFOA will be used in deriving benchmarks for the eight PFAS that are the focus of this document. You may want to make that clear to the reader.
24	First	The paragraph makes references to whether this process should consider studies that only report nominal concentrations of the PFAS in the toxicity study. The rationale given for choosing to consider studies that only report nominal concentrations is a case study that was conducted with measured and nominal concentrations for PFOS and PFOA. While the rationale is sound for considering studies that only include nominal concentrations of PFOS or PFOA, care should be taken to extrapolate to the entire class of chemicals, i.e., PFAS. The eight PFAS that are focus of this process have different physical and chemical properties than PFOS and PFOA, consequently, the probability of the nominal concentrations being with 20% of the measured concentrations for the eight PFAS may be different than PFOS and PFOA, which was 82 and 83%, respectively.
27	First	Comma needed between "*toxicity database*" and "*benchmark values"* in the first sentence on page 27.
27	Second	It would be useful to report at some point the number of acceptable empirical LC50 or EC50 used in each SSD that were based on nominal concentrations. The data is available in the document but the reader would have to take a great deal of time to compile these numbers.
27	Second	Why not use the lowest acute value for a species instead of the mean? Using the lowest acute value for a species would be a more conservative approach in terms of protection of sensitive species. The same question could be asked about the genus mean acute values.
Entire document	Entire document	Review the document to ensure that Greek letters are used consistently, e.g., µg/L vs. ug/L
Appendices A to C		I thought these were very valuable appendices. They clearly laid out the studies that were considered for inclusion in the derivation of the benchmarks and why studies were eventually not included.

Reviewer 5

Peer Review Comments on Acute Protective Freshwater Benchmarks for Selected PFAS Using a New Approach Methods

Reviewer 5

I. GENERAL IMPRESSIONS

This document provides draft Acute Protective Freshwater Benchmarks for the following eight PFAS Compounds using New Approach Method (NAM):

1. Perfluorobutanoic acid (PFBA)
2. Perfluorobutanesulfonic acid (PFBS)
3. Perfluorononanoic acid (PFNA)
4. Perfluorodecanoic acid (PFDA)
5. Perfluorohexanoic acid (PFHxA)
6. Perfluorohexanesulfonic acid (PFHxS)
7. Hexadecafluoro-2-decenoic acid (8:2 FTUCA), and
8. Pentadecafluorodecanoic acid (7:3 FTCA)

The detailed methodology used for the derivation of benchmarks has been thoroughly explained. The process used and results of a systematic review of available empirical toxicity data for aquatic organisms identified via EPA's literature search for the eight PFAS has been adequately addressed.

The aquatic life benchmarks for the eight PFAS compounds have been developed using the empirical and Web-ICE data for these chemicals and were calculated by applying statistical methods. This method aligns with the EPA's objective to decrease reliance on animal testing by employing NAMS in toxicity assessment. In addition, the EPA applied 'binning' approach to calculate protective benchmark values for six PFAS, utilizing combined carboxylic acid (PFBA, PFNA, PFDA, PFHxA) and sulfonic acid (PFBS, PFHxS) groupings to facilitate value derivation.

The use of estimated data suggests a proactive approach in addressing gaps in empirical data. It also demonstrates agencies' commitment to methodological rigor and adaptability in the face of data challenges when deriving protective values for PFAS compounds.

II. RESPONSE TO CHARGE QUESTIONS

1. ***Please comment on the overall clarity of the documents and construction as it relates to assessing the effects and derivation of acute benchmarks for the eight selected PFAS.***

   Great job on the overall structure and organization of the document! The logical flow and seamless transitions between sections significantly enhance the readability and understanding of the content. Information has been laid out in appendices with detailed information on the approaches and examples for deriving benchmarks. The list of Tables and Figures provides information on all the empirical data and acceptable ICE models used in deriving aquatic life acute benchmarks of all eight compounds. The lowest quantitatively acceptable empirical toxicity studies used to derive aquatic life benchmarks

for eight PFAS compounds were detailed in the appendix. Ranked GMAVs and FAVs have been provided for all eight PFAS compounds. Data incorporated in SSDs have been listed and all figures are self-explanatory.

It is apparent that considerable thought and effort were invested in crafting a document with a well-considered and smooth progression.

2. ***Please comment on each of the technical approaches used to derive the draft benchmark values for the eight selected PFAS presented in EPA's Derivation of Acute Protective Freshwater Benchmarks for Selected PFAS through a New Approach Method (ICE-based approach using extrapolation [Sections 3-5], ICE-based approach using scaled data [Appendix F], and PFAS benchmark calculations using a data binning approach [Appendix G]).***

    a. ***Are the technical approaches used to derive the benchmark values logical?***

    The methodologies employed to determine the benchmark values are rational and sound from a technical perspective. The available empirical data for the eight PFAS under consideration fulfill only 2-3 Minimum Data Requirements (MDRs). Consequently, the EPA opted to employ the peer-reviewed Interspecies Correlation Estimation (ICE) models developed by Raimondo et al. in 2010. The primary objective of this application was to provide acute toxicity data to fulfill MDRs in instances where direct toxicity data were not at hand. The ICE models underwent rigorous evaluation based on acceptance parameters, including mean square error (MSE), R^2, and slope, as delineated in Box 1. Only models meeting these predefined acceptance criteria were utilized in the derivation of species-specific toxicity data. This data, when integrated with empirical toxicity data, served to strengthen the process of establishing benchmark values.

    The EPA's investigation into the "binning" approach for establishing protective values for grouped carboxylic acid PFASs and grouped sulfonic acid PFAS is grounded in the precedent established by Giddings et al. in 2019. A similar methodology proved successful for pyrethroids in that study. This strategic approach involved consolidating chemicals with shared modes of action, offering advantages in scenarios where limitations in available data present challenges to value determination. The calculated values were based on the amalgamation of carboxylic acid compounds (PFBA, PFNA, PFDA, PFHxA) and sulfonic acid compounds (PFBS, PFHxS), thereby substantiating the derivation of these values. Calculated benchmark values for carboxylic acids and sulfonic acids consistently demonstrated higher values when utilizing the SSD generator in comparison to the Guidelines-based approach. However, these benchmarks displayed variability in magnitude when contrasted with the ICE-based benchmark values. This might be influenced by the constrained empirical datasets for certain PFAS and the restricted number of data points available for the species employed in normalization.

By stating that the derived benchmarks are considered less certain than ambient water quality criteria, the authors acknowledge a level of uncertainty. This acknowledgment is crucial in providing a realistic assessment of the reliability of the benchmarks.

b. Does the science support the conclusions?

This is a valuable contribution to the current scientific knowledge on the toxicity of PFAS compounds to aquatic life, even while acknowledging the inherent uncertainties associated with using estimated data in the derivation process.

c. Are the approaches and resulting values consistent with the protection of aquatic life?

Yes, the approaches and resulting values align with the protection of aquatic life. The aquatic life benchmarks for the eight PFAS compounds were established using empirical and Web-ICE data, employing statistical methods for calculation. This approach aligns with the EPA's goal to reduce reliance on animal testing by incorporating NAMS in toxicity assessment.

Detailed response as in 2a.

Limitation
ICE models have not been developed for chronic toxicity data and therefore only acute criteria were developed.

d. For the ICE-based models in particular, please compare and contrast the strengths and weaknesses of the extrapolation approach [Sections 3-5] with those of the approach using scaled data [Appendix F].

The acknowledgment of potential challenges, such as large confidence intervals and potential limitations in accepting ICE models beyond the model range, demonstrates transparency and a thorough understanding of the modeling process.

The decision to select the "scaled" approach as an alternative approach for deriving benchmark values, showcases a proactive and meticulous approach. The alternative scaled approach modifies toxicity values, as needed, to align them with the ICE model range, avoiding the extrapolation of regressions beyond the established model range. There is close agreement between the benchmark values calculated using either approach (as listed in Table 5-26).

This consideration of alternative methods highlights a commitment to rigorous evaluation and continuous improvement, reflecting a commendable scientific rigor in the approach to deriving benchmark values.

3. Please comment on the empirical (direct test) and ICE-generated toxicity data used to derive the benchmark values presented in the draft document.

a. Were the data adequately used and sufficiently comprehensive to represent risks to sensitive aquatic life?

The EPA employed both empirical test data and ICE values, derived for missing Minimum Data Requirements (MDRs), to determine acute freshwater benchmark recommendations for aquatic life. The utilization of ICE-predicted values by various independent, international groups to establish protective values for aquatic life confirms that values derived from ICE-generated Species Sensitivity Distributions (SSDs) offer a consistent level of protection comparable to using directly measured laboratory data.

b. Were the data selected and/or excluded from the benchmark values derivation appropriately utilized?

Quantitatively acceptable empirical acute toxicity data available for each of the eight PFAS was tabulated for each individual study. All toxicity values, including LC values, EC values, NOECs, LOECs, and species- and genus-mean values, were presented with four significant figures. This practice avoided round-off errors in subsequent calculations. Studies that were determined to be qualitatively acceptable as supporting information, but not acceptable for quantitative use were listed with deficiencies in each study. Furthermore, studies that were deemed unsuitable for either quantitative or qualitative were also cited. I endorse the choice to incorporate toxicity data for studies solely based on unmeasured test concentrations. This decision is rooted in findings for PFOA and PFOS (U.S. EPA 2022a, b), leading the EPA to determine that nominal test concentrations effectively represent real PFAS exposures in standard acute laboratory-based toxicity tests. In addition, Hoke et al.,2012 (https://doi.org/10.1016/j.chemosphere.2011.12.066) also reported mean measured test concentrations were similar (within 80–120% of nominal) to the targeted nominal test concentrations for fluorinated acids with the exception of the 5:3 acid.

The authors applied the criteria recommended by Willming et al., 2016 to enhance models' reliability and robustness (Box 1). Models adhering to these acceptance parameters were employed to generate species toxicity data, which were then combined with empirical toxicity data to strengthen the derivation of benchmark values. This approach demonstrated logical and consistent application of standard criteria across all eight PFAS compounds.

c. Are there relevant data that you are aware of that should be included? If so, please provide for derivation of benchmark values.

Below, a recent chronic study by Kadlec et al., 20203 has been listed as an additional reference, some aspects may be relevant.

https://doi.org/10.1002/etc.5784 Sarah M. Kadlec, Will J. Backe, Russell J. Erickson, J. Russell Hockett, Sarah E. Howe, Ian D. Mundy, Edward Piasecki, Henry Sluka, Lauren K. Votava, David R. Mount (2023) Sublethal Toxicity of 17 Per- and Polyfluoroalkyl Substances with Diverse Structures to *Ceriodaphnia dubia*, *Hyalella azteca*, and *Chironomus dilutus*

4. ***Are the derived benchmark values appropriately protective of sensitive aquatic life?***

The establishment of aquatic life benchmarks for the eight PFAS compounds involved the utilization of empirical and Web-ICE data, incorporating statistical methods for calculation. This strategy is in accordance with the EPA's objective of minimizing dependence on animal testing by integrating NAMS into toxicity assessments.

In stating that the benchmarks derived are regarded as less certain than ambient water quality criteria, the authors are acknowledging a degree of uncertainty. This recognition is essential for offering a logical evaluation of the benchmarks' reliability.

III. SPECIFIC OBSERVATIONS

Page	Paragraph	Comments or Questions
70 and 80	Figure 5-7 and Figure 5-8	Is this a bimodal response- as this is model based SSD, it is challenging to confirm? Invertebrates and fish may have different mode of action for 8:2 FTUCA and 7:3 FTCA