TASK ORDER 65 UNDER CONTRACT EP-C-12-029

EXTERNAL PEER REVIEW OF AQUATIC LIFE AMBIENT WATER QUALITY CRITERIA FOR ALUMINUM – 2017

PEER REVIEW SUMMARY REPORT

April 4, 2017

Submitted to:
U.S. Environmental Protection Agency
Office of Water, Office of Science and Technology
Health and Ecological Criteria Division
1200 Pennsylvania Avenue, NW
Washington, DC 20460
Attn: Diana Eignor
Eignor.Diana@epa.gov

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

CONTENTS

1.0 INTRODUCTION

This report documents the results of an independent peer review of *Aquatic Life Ambient Water Quality Criteria for Aluminum - 2017* for the U.S. Environmental Protection Agency (EPA), Office of Water (OW). Eastern Research Group, Inc. (ERG, a contractor to EPA) organized this review and developed this report. Sections 2.1 to 2.4 of this report present, for each charge question, the individual reviewer comments and a summary of those comments. New information (e.g., references) provided by reviewers is presented in Section 3. Appendices A and B provide, respectively, the charge to reviewers and the complete set of comments submitted by each reviewer.

1.1 Background

EPA establishes national recommended Ambient Water Quality Criteria (AWQC) under the Clean Water Act (CWA). Section 304(a)(1) aquatic life criteria serve as recommendations to states and tribes by defining ambient water concentrations that will protect against unacceptable adverse ecological effects to aquatic life from exposure to pollutants in water. Aquatic life criteria address the CWA goals of providing for protection and propagation of fish and shellfish. Once EPA publishes final §304(a) recommended water quality criteria, states and authorized tribes may adopt these criteria into their water quality standards to protect designated uses of water bodies. As required by the CWA, EPA periodically reviews and revises §304(a) AWQC to ensure they are consistent with the latest scientific information. In support of this mission, EPA is working to update water quality criteria to protect aquatic life from aluminum in freshwater environments.

The current aluminum criteria are from 1988 and applied to pH range 6.5 – 9.0 and across all hardness and dissolved organic carbon (DOC) ranges. The approach described in the review document is a multi-linear regression (MLR) model using pH, hardness, and DOC, with interaction terms between these parameters for freshwater aluminum criteria. The review document includes freshwater acute and chronic mussel tests and additional toxicity studies of exposure to different combinations of pH, hardness, and DOC ranges.

1.2 Peer Reviewers

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

- Gretchen K. Bielmyer-Fraser, Ph.D.; Associate Professor, Department of Biology, Valdosta State University.
- David Buchwalter, Ph.D.; Associate Professor, Department of Biological Sciences, North Carolina State University.
- Gregory Möller, Ph.D.; Professor of Environmental Chemistry and Toxicology, University of Idaho.
- Scott Smith, Ph.D.; Associate Professor, Department of Chemistry and Biochemistry, Wilfrid Laurier University, Canada.
- Randall Wentzel, Ph.D.; Private Consultant.

ERG provided reviewers with instructions, the draft aluminum criteria document, the aluminum criteria model calculator (an Excel spreadsheet), and the charge to reviewers (Appendix A of this report) prepared by EPA. Reviewers worked individually to develop written comments in response to the charge questions. After receiving reviewer comments, ERG summarized reviewers' responses to the charge questions, noting areas of agreement and disagreement, where relevant (see Section 2).

2.0 SUMMARY OF REVIEWER COMMENTS ORGANIZED BY CHARGE QUESTION

This section presents summaries of reviewer comments organized by charge question. Each summary is followed by a table presenting the individual reviewer comments by charge question. Individual comments are copied directly from the written comments submitted by each reviewer (presented in Appendix B).

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

All five reviewers had positive comments about the document's overall clarity and construction. Reviewer 2 commended the authors for improving the report's technical basis and application by incorporating very recent scientific approaches. Reviewer 3 commented positively on each section of the document, using phrases such as "sufficient detail," "well-developed and informative," "thorough and comprehensive," "fair and transparent," "clear and concise," and "reasonable and well supported." The reviewer advised that the document may need to reflect changes to the final versions of the "in-review" manuscripts cited in Section 2.4.1. The reviewer also noted that Section 2.7.1 and Figure 3 do not have citations. Reviewer 4 said the document was well prepared. The reviewer suggested adding an appendix with aluminum chemistry and inserting newer references into the initial sections that reused material from the 1985 document. Reviewer 5 commented that "the document very clearly explained the derivation of each criterion."

Reviewer	Comments	Response to Comments
Reviewer 1	Overall, the document is well constructed and reasonably clear in its presentation.	
Reviewer 2	I would first like to commend the authors for producing this draft report on the Water Quality Criteria (WQC) for aluminum. It is a complex topic and the authors incorporate very recent scientific approaches which improves the reports technical basis and its application to states and other users of the criteria. The Environmental Protection Agency and the Office of Water, in particular, have a successful history of improving the quality and utilization of our nation's waters and while others may forget turbid, oily, lakes and streams with detergent bubbles floating on them, I do not, and I thank you for continuing your important role in the protection of people and the environment in which they live. The purpose of the document, in the Executive Summary, is listed as improving scientific knowledge. I would suggest adding – and to increase the utilization by states, tribes, and other users. Developing the criteria using the multiple linear regression (MLR) approach will use water chemistry parameters that the vast majority of the states currently measure in their water monitoring programs. The relevance of these draft criteria will be very high because the states and tribes can easily apply them in their water quality programs.	

Reviewer	Comments	Response to Comments
	In the Executive Summary, I would suggest a paragraph on the history of utilization of the previous aluminum WQC (or others) by the states to show how useful the documents are at protecting our Nation's water. Overall the clarity of the writing is very good. The document organization follows the format of the ecological risk assessment framework and that process is familiar to the readers of the document. There is not enough data to conduct WQC for estuaries and marine environments, however discussion sections continue through the document. I would suggest stating that there is not enough data and then combine the further text and data on estuary/marine systems into an appendix. I think the reader needs to brought along with the progression of the use of the biotic ligand model (BLM) in the copper WQC to the current selection of the MLR approach. An expansion of the text on page 13 would benefit the report. A discussion on the development and application of the full BLM, which could include text on the collaborative effort to produce the BLM approach and its utilization of aquatic chemistry data. Implementation issues concerned the amount of water quality data required to run the full BLM has been a significant for the states and tribes. Then text on potential solutions to the implementation issues could include a discussion of analysis of the partial BLM approach and analysis using the MLR. This text would help the reader understand the amount of effort, in coordination with the scientific community, that went into the development of a more relevant approach.	
Reviewer 3	Overall the draft criteria document is well organized and clearly written. It contains a satisfying level of breadth and depth in analyzing the results of a comprehensive literature review and successfully performing a critical analysis of a large peer-reviewed data set concerning aluminum risk to aquatic ecosystems. The MLR model approach and the thoroughness of the input data selection/rejection are strengths of this criteria development document. **Specific Document Section Comments** **Executive Summary** The Executive Summary has sufficient detail to provide readers with the regulatory premise, background, critical concepts, derivation and modeling approaches and the criteria conclusions	

Reviewer	Comments	Response to Comments
	presented in the document. I would suggest "getting to the point" in the opening paragraph of the Executive Summary with the following edit: "This update establishes a freshwater criteria magnitude that is affected by total hardness, pH, and DOC and expands on the toxicity database to include those studies beyond the pH range of 6.5-9.0. The criteria results are presented in a series of look-up tables in Appendix K." **Introduction and Background** This is a well-developed and informative section. The regulatory premise and application of criteria are discussed in clear exposition. The active web link to National Recommended Water Quality Criteria - Aquatic Life Criteria Table will be helpful in establishing context to some readers, especially non-experts, who may access this document with limited background knowledge or as their first introduction to water quality criteria. **2 Problem Formulation** This section establishes cause and framework for the document in a concise manner. **2.1 Overview of Stressor Sources and Occurrence** This section is a well-cited review. The analysis is complete and representative. The supporting information and in context explanations (i.e., dissolved Al samples) are helpful without being pedantic. **2.2 Stressors of Concern** The treatment of 2009 total recoverable metal differences of aquatic Al concentrations and the potential for "conservative" risk assessment is balanced and offers a fair treatment of potential for bias from a change in sampling and analysis methodology. The potential for bias is managed in the selection of published controlled exposure studies cited in the draft document that uses the same dose construction, sampling, and analysis approaches. **2.3 Environmental Fate and Transport** Aluminum environmental chemistry is complex, and Figure 2 is very helpful in demonstrating this complexity to the reader. **2.4 Mode of Action and Toxicity** This section is a well-developed analysis and introduces the reader to the aquatic chemistry challenges of Al, as it relates to	

Reviewer	Comments	Response to Comments
	exposure studies. It establishes the context for study criteria to be included in model inputs. The section is comprehensive in exploring the practical challenges of performing controlled Al dosing studies. The discussion of pH effects and limits for study inclusion is a strength. **2.4.1 Water Quality Parameters Affecting Toxicity** This section introduces BLM and MLR modeling approaches with sufficient depth and breadth and is well cited. The MLR approach is well documented and clearly discussed in this section. The in-review manuscripts of Cardwell 2017, Gensemer 2017, Brix 2017 and DeForest 2017 are a critical foundation needed to support the draft document. Substantive changes to the final peer-reviewed papers will need to be reflected in changes to this draft criteria document. Because of the entwined fate of those papers and this draft document, acceptance for publication with minor changes is a hopeful outcome of that review process. My read of the DeForest 2017 manuscript suggests it is a strong work. **2.5 Measurement Endpoints** This section establishes a context for the approach. The outline of process and procedures is clearly developed and clearly written with a sufficient and satisfying level of explanation for a wide range of readers. **2.5.1 Overview of Toxicity Data Requirements** The use of bullets in this section is effective to outline an overview of data requirements. **2.6 Measures of Effect** The section does a solid job in explaining the data and study research search methodology that is used to determine test acceptability the development of the criteria. **2.6.1 Acute measures of effect** The section solidly introduces the reader to fundamental terms and resources found in the draft document and appendices. **2.6.2 Chronic measures of effect** The section solidly introduces the reader to fundamental terms and resources found in the draft document and appendices. Table 1 is a useful organization and summary of concepts.	

Reviewer	Comments	Response to Comments
	2.6.3 Aluminum Toxicity Data Fulfilling Minimum Data Requirements The section introduces the summary of acceptable toxicity data in Table 2 and the insufficient minimum dataset requirements for estuarine/marine criterion value determination with sufficient robustness. Communicating the need for additional acute and chronic toxicity testing on estuarine/marine taxa in this section is helpful in supporting research proposals to develop that data gap. **2.7 Conceptual Model** The approach of conceptual models is introduced with clarity and conciseness. **2.7.1 Conceptual Diagram** Figure 3 and the related discussion is an excellent resource for organizing the sources, pathways, receptors and controls of environmental Al. This section and figure do not have citations, and this is a curious omission as Al is a primary resource in common use. The authors may want to explore, review and cite similar conceptual diagrams and models that may exist in other quality resources. **2.8 Analysis Plan** This section is a thorough and comprehensive discussion of the background and approach for developing the data set used in modeling input and criteria development. The complexity of the topic is well treated in this exposition, and the authors present clear explanations with sufficient detail and definitions. The section is supportive of the criteria development goals of the draft document. **2.9 Identification of Data Gaps and Uncertainties for Aquatic Organisms** This section presents a fair and transparent treatment of data gaps and uncertainties in Al criterion development. Assumptions are clearly and openly discussed. The supporting discussion of management of data gaps and assumptions is convincing and support of the final criteria development. The explicit identification of data gaps in estuarine/marine species and for plants support the research community in their project proposals to help close those data gaps. As a minor edit, the third line from the bottom on page 30 has an extra space inserted.	

Reviewer	Comments	Response to Comments
	3 Effects Analyses **3.1 Acute Toxicity to Aquatic Animals** The comprehensive data sets of Appendix A and B are introduced. **3.1.1 Freshwater** This section explicitly, and in detail, explores the available freshwater Al toxicity data. The use of tables and figures helps organize the complex available resources. The critical analysis and discussion of the limitations and usefulness of available studies are well done and supported by references to criteria development standards. The clarifying sentence: "That is, EPA's acute freshwater criteria are not fixed values, but equations based on bioavailability and hence toxicity of aluminum under different water chemistry conditions (see Appendix K for additional criteria values)" helps communicate the new Al guidance in the draft document. **3.1.2 Estuarine/Marine** This section discusses the data gaps and limitations that sufficiently support a finding of inability to calculate an estuarine/marine criterion. **3.2 Chronic Toxicity to Aquatic Animals** This section introduces Appendix C that comprehensively lists acceptable chronic toxicity data. **3.2.1 Freshwater** This section explicitly, and in detail, explores the available freshwater Al chronic toxicity data. The use of tables and figures helps organize the complex available resources. The critical analysis and discussion of the limitations and usefulness of available studies are well done and supported by references to criteria development standards. Many of the cited studies are recent and very high quality, addressing the anticipated data needs for developing an Al water quality criterion. The exploration of MDR missing the third family in phylum Chordata is a well-developed, transparent and satisfying approach to missing data. The detailed abstracting of the studies in this section is a useful approach to establishing the scientific basis for consideration of a diverse range of research studies as data resources for criterion development.	

Reviewer	Comments	Response to Comments
	3.2.2 Estuarine/Marine The section contains the explicit data reference of Appendix D for no estuarine/marine chronic toxicity data. **3.3 Bioaccumulation** This is a useful section, especially as it relates to fish accumulation and Al dietary intake potential. While there is absent or limited data, the section is important for completeness. **3.4 Toxicity to Aquatic Plants** This too is a useful section, and while there is absent or limited data, the section is important for completeness. **3.5 Summary of National Criteria** **3.5.1 Freshwater** The summary approach, including example calculations and an example look-up table, with further reference to the full Appendix K look-up tables, is a good approach to introduce the reader to the National Criteria of the draft document. The approach is clear and concise, with good readability. **4 Effects Characterization** **4.1 Effects on Aquatic Animals** **4.1.1 Freshwater Acute Toxicity** This section explicitly, and in detail, explores the available freshwater Al acute toxicity effects data. The critical analysis and discussion of the limitations and usefulness of available studies are well done and supported by references to criteria development standards. The justifications for inclusion or rejection of specific study data appear reasonable and well supported in the document discussion. **4.1.2 Freshwater Chronic Toxicity** This section explicitly, and in detail, explores the available freshwater Al acute toxicity effects data. The critical analysis and discussion of the limitations and usefulness of available studies are well done and supported by references to criteria development standards. The justifications for inclusion or rejection of specific study data, for example, pH or study duration, appear reasonable and well supported in the document discussion.	

Reviewer	Comments	Response to Comments
	4.1.3 Freshwater Field Studies The discussion is useful for completeness of criteria background development and observation of need for Al management, especially in consideration of historical acidification of some surface waters in the eastern US. **4.1.4 Estuarine/Marine Acute Toxicity** The discussion is useful for completeness of criteria background development, and correlation/comparison with other Appendix B and I data. **4.1.5 Estuarine/Marine Chronic Toxicity** The discussion is useful for completeness of criteria background development, and correlation/comparison with other Appendix D and I data. **4.1.6 Bioaccumulation** The discussion of effects of Al bioaccumulation is well executed with good breadth and depth of the available literature. There is limited published work available, and the analysis appears complete. **4.2 Effects on Aquatic Plants** With reference to Appendix E, this section summarizes and reviews toxic effect on plants with available citations. **4.3 Protection of Endangered Species** The summary statement that the 2017 criterion is protective of endangered species is an important early summary statement. **4.3.1 Key acute toxicity data for listed fish species** This section summarizes available data in a complete and clear manner. It is an important observation that there is no acceptable acute toxicity data for endangered or threatened estuarine/ marine aquatic fish species. **4.3.2 Key chronic toxicity data for listed fish species** This section summarizes available data, a single study, in a complete and clear manner. It is an important observation that there is no acceptable chronic toxicity data for endangered or threatened estuarine/marine aquatic fish species.	

Reviewer	Comments	Response to Comments
	4.3.3 Concerns about Federally listed endangered mussels Available studies on Federally listed endangered mussels are reviewed with clarity and completeness. **4.4 Comparison of 1988 and 2016 Criteria Values** The title of this section should be "Comparison of 1988 and 2017 Criteria Values." This comparative analysis is important and useful to the overall document. The assumptions and approaches to setting the data in Table 11 (e.g., hardness at 100 mg/L) are scientifically reasonable and defendable. Statements concerning the relative level of protection, for example, brook trout, are an important qualifier in support of the MLR model approach that is the basis for criteria development. **5 References** The literature supporting the draft document represents a complete and exhaustive review of the available peer-reviewed science.	
Reviewer 4	a) The document is very well prepared. The communication, documentation and clarity of the draft criteria is dramatically improved compared to the 1985 criteria document. b) One criticism though is in the reuse of some material from the 1985 document. This results in some very old references in the initial parts of this new draft criteria document. The initial sections (up to about page 12) draws heavily from the older document and could do with insertion of some of the new references which are detailed in later sections of the draft criteria document. In particular the first paragraph of section 2.2 and the paragraph right before section 2.4.1 are very much like the 1985 document c) There should be an aluminum chemistry section. This can be an appendix and does not necessarily need to be integrated into the text. As it is written now so many "random" aspects of aluminum chemistry are introduced as the document progresses. It would be nice to have one central place discussing aluminum chemistry in some details. Emphasizing aluminum speciation and reaction kinetics. d) In terms of more specific communication issues, detailed comments are included as an appendix to this review.	

Reviewer	Comments	Response to Comments
Reviewer 5	The document very clearly explained the derivation of each criterion, including the toxicity data and the water chemistry parameters used. The Multiple Linear Regression (MLR) model for determining Al criteria uses equations based on pH, hardness, and dissolved organic carbon (DOC). Therefore, the criteria would vary in a site-specific manner. The ways in which these water chemistry variables modify Al toxicity, specifically including changes in Al speciation, is very well explained.	

2.2 Please comment on the technical approach used to derive the draft aluminum criteria; is it logical, does the science support the conclusion, and is it consistent with the protection of freshwater life from acute, chronic, and bioaccumulative effects? Please provide specific comments.

Reviewer 1 had many questions about the technical approach used to derive the draft aluminum criteria: Does bioavailability differ between species due to their physiology or aluminum speciation? Does the model based on *C. dubia* fit other invertebrates? Does the model based on *P. promelas* fit other fish? Do the models fit equally well for acute and chronic toxicity? The reviewer also said that more information is needed about the predicted effects concentrations (EC20s) being within a factor of two of observed values for 95% of the tests and that the model ignores the temperature-solubility issue. Reviewer 1 also responded that "the new criteria values significantly weaken protection of aquatic ecosystems by allowing higher Al concentrations in most situations than the 1988 criteria."

Reviewer 2 responded that the approach is overall technically sound, and that "the MLR method will enable states, tribes, and local governments to assess potential aluminum toxicity at any aquatic site based on its unique aquatic chemistry." The reviewer suggested adding the following:

- A comparison of the 1988 criteria and the 2017 draft criteria to the Executive Summary.
- Details on why the authors of the cited studies thought the dissolved concentrations were relatively level as total aluminum concentrations and toxicity increased to Section 2.2.
- Information about the 1980s acid rain issues to Section 2.3.
- Text that explains that the 10 parameter data requirements were difficult for many state monitoring programs, and that the MLR approach uses data that almost all states already collect to Section 2.4.1.
- A population model, if available, to Section 2.6.2.
- An explanation for why the analysis does not follow a conventional species sensitivity distribution to Section 2.8.
- Bullets on bioconcentration and biomagnification to Section 3.3.
- A discussion about a U.S. habitat where a species would be most at risk to Section 4.3.

Both Reviewer 3 and Reviewer 5 responded that the technical approach is logical. Reviewer 3 also commented that it is "consistent with the protection of freshwater life from acute, chronic, and bioaccumulative effects." Reviewer 5 said the technical approach is scientifically supported.

Reviewer 4 admitted a preference for a biotic ligand model (BLM) approach and commented that the MLR approach should be further justified because it is a compromise compared to a full BLM approach. The reviewer said that aluminum speciation is not fully addressed, and that the selection of total aluminum as the independent variable needs more justification. However, the reviewer concluded the proposed criteria will be protective of aquatic life.

Reviewer	Comments	Response to Comments
Reviewer 1	The criteria are based primarily on multiple linear regression models that use pH, hardness and DOC as modulators of 7-day toxicity in two models – a fish (*Pimephales promelas*) with biomass as the endpoint, and a crustacean (*Ceriodaphnia dubia*) with reproduction as an endpoint. The best fit models for each species were different from each other which poses many questions: a) pH, hardness and DOC are described as important modulators of bioavailability. Do the different models then mean that bioavailability differs between species as a function of their physiology and not as a function of Al speciation? This is extremely important to articulate because the chemistry and bioavailability of Al is complex. Species also vary widely and the physiology of Al uptake is not very well understood comparatively across species. b) Do we know that the model based on *C. dubia* fits other invertebrates well? Do we know that the model based on *P. promelas* fits other fish well? Why are we so willing to make this assumption? Shouldn't this be tested first? c) Do we know that the models fit equally well for acute and chronic toxicity? This is extremely important to articulate the uncertainty of the assumption that the models apply to other species based on gross taxonomy (vertebrates vs invertebrates). P. 15: "In the final MLR model, predicted EC20s were within a factor of two of observed values for 95% of the tests. I assume that this means for the tests performed with *P. promelas* and *C. dubia,* most observations came within a factor of 2 of the model predictions that were built from the same data. More information would be useful here. Are predictions generally biased in an over-protective or under-protective way? If the uncertainty for any specific prediction is a factor of 2, shouldn't this uncertainty	

Reviewer	Comments	Response to Comments
	be built into the criteria? Should an interspecific uncertainty factor be applied as well? P. 3: Al associated with sediments are "unavailable to aquatic organisms"? This is a rather broad sweeping statement to make. P. 9: The temperature-solubility issue is completely ignored in the model. If a 15 degree change is equal to a whole pH unit, and pH drives the model, shouldn't there be a temperature term in the model? P. 10: "aluminum has also been shown to increase respiration, and thereby energy demands"… but an increase in respiration is a cause for excluding data from consideration? P. 30: "Application of water only laboratory toxicity tests to develop water quality criteria to protect aquatic life" fails on many fronts and among the reasons that EPA is considering modernizing the 1985 guidelines. This paragraph is not entirely truthful. P. 30: "…Only chronic data are used in the model, and application to acute data normalization assumes that the same relationships are present". This is a big assumption. Such an assumption would generally fail for BLM models based on chronic data applied to acute data, and to acute BLMs applied to chronic data for other metals. Why should we make this assumption here? P. 60: "Comparison of 1988 and 2016 (sic) criterion values": This paragraph is troubling. The new criteria values significantly weaken protection of aquatic ecosystems by allowing higher Al concentrations in most situations than the 1988 criteria. To suggest otherwise is disingenuous.	
Reviewer 2	Overall, the approach is technically sound. The vast improvement of the MLR method, which includes key water chemistry parameters, over those used for the 1988 aluminum criteria, will enable states, tribes, and local governments to assess potential aluminum toxicity at any aquatic site based on its unique aquatic chemistry. Section 2.2. The justification for the use of total recoverable aluminum is a complex topic which requires a decision to support the criteria development process. Further details on why the authors of the cited toxicity studies thought the dissolved concentrations were relatively level as total aluminum concentrations and toxicity increased would be good. Also, are these studies the exception? Do most toxicity studies show a	

Reviewer	Comments	Response to Comments
	correlation between dissolved concentrations of aluminum and toxicity? The toxic forms of aluminum should be in both of the dissolved and total recoverable values. Section 2.3. At the start of this section the inclusion of text on acid rain issues in the 1980s and the building awareness of aluminum toxicity to aquatic biota would be informative. The discussion and the Figure 2 provide the reader with important information on the environmental chemistry, fate and transport of aluminum. Section 2.4. Increasing the range of pH in toxicity studies included in the criteria to 5.0 – 9.0 is a very good decision. Toxic forms of aluminum increase at a pH lower than 6.5. Section 2.4.1. The discussion on the BLM method and the use of MLR is important for the reader. I think text should be added that the data requirements (ten parameters) were difficult for many state monitoring programs to address. The MLR approach uses data that almost all states already collect, so the MLR would be much more likely to be used and provide increase scientific rigor for the protection of aquatic species from aluminum. A discussion comparing the output of the methods would be good. The 2017 publications cited in this section are still (as of this writing) in review. A text box or text on key points from Brix et al., 2017 and DeForest et al., 2017 and why they support the use of the MLR would help the reader. Some limits of the DeForest study are mentioned on page 30, but It would be good to have that discussion earlier in the document and all in one place. I commend the EPA for their awareness of these studies and the cooperation with these authors to include their research. Once the discussion focuses on the methods and results from DeForest, I would start a new section or subsection of the report. This text is the technical support for your document. Section 2.5. On page 17, when you first mention assessment endpoints, please state that they are listed in Table 1. On page 16, add a web address for Stephan et al., 1985 or as it is also cited, EPA, 1985. Section 2.6.2. This section briefly discusses use of the effects concentration (EC20) to protect populations from long term chronic effects of a toxicant. If population models or a discussion relating the draft criteria to field studies are not going to be included in the document, I would recommend using the wording, “inferred protection of populations… “. Adding a population model as an additional line of evidence would be informative if	

Reviewer	Comments	Response to Comments
	the required biological and effects data are available for a species of interest. Section 2.7. The conceptual model section is well written and important for the user to understand the various exposure pathways and which ones may be most important. Section 2.8. The analysis plan is well presented. The use of a species sensitivity distribution (SSD) for the four most sensitive species and using the 5th percentile of that distribution is conservative. The decision to use a factor of 2 is more conservative and is based on a 1978 Federal Register notice. This process is supposed to protect 95% of the species in an aquatic community. Conventional use of a SSD requires at least 7 species, why is the use of the four most sensitive species proper? A conventional SSD and using an EC5 percentile is predicted to be protective of 95% of a given community (Greenberg et al., 2013). Why not use all of the data? From the text in the section, it seems that the 1978 FR notice data is a method to get to near control test organism survival percentages. I don't think that relates to protection of 95% of the species in an aquatic community. If you want to use the factor of 2, it could be stated as a policy decision. Section 3.1.1. Once again, the factor of 2. Maybe state that it is <u>intended</u> to protect 95% …. Section 3.2. The discussion of the chronic toxicity studies was a very good summary. I would add at the beginning of the section that EC20 values were used for primarily sublethal endpoints. Section 3.3. The authors could consider adding bullets on bioconcentration/biomagnification issues for inorganic metals from the EPA's Framework for Metals Risk Assessment (US EPA, 2007). An example bullet: • Trophic transfer can be an important route of exposure for metals, although biomagnification of inorganic forms of metals in food webs is generally not a concern in metals assessments. Section 3.5. Table 10 is a good presentation of the effects of pH and hardness on the criteria. It would be helpful to the reader to explain why when the pH is above 7.5, and higher, that the criteria values are lower as the hardness concentrations increase. Section 4. The discussion of aquatic toxicity studies that had non-standard experimental designs or results was good to include in this document. The results are discussed and the reasons for the	

Reviewer	Comments	Response to Comments
	lack of inclusion in the datasets are presented. A given regulator could utilize those studies if their inclusion met their needs. Section 4.3. In the discussion on endangered species, it might be good to discuss habitat in the U.S. where a species would be most at risk, e.g., acidic lakes with low hardness and dissolved organic carbon (DOC). An overlap of at risk habitat with an endangered species range would be informative. This suggestion may exceed the objectives of this report, but could be considered in more focused inter-agency studies. Section 4.4. The comparison of 1988 criteria and the 2017 draft criteria should be briefly discussed in the Executive Summary.	
Reviewer 3	Overall the technical approach used to derive the draft Al criteria is logical and supportive of the criteria data found in the look-up tables of Appendix K. Within the constraints of available peer-reviewed data, the technical approach is consistent with the protection of freshwater life from acute, chronic, and bioaccumulative effects. Specific comments related to a critical analysis of each section of the draft document appear above.	
Reviewer 4	Technical comments are itemized below: a) Selection of an MLR approach should be justified. Why not a BLM approach? I have no problem with an MLR approach but it is a bit of a compromise compared to a full BLM approach so the reason for selecting this method should be given. Ideally highlighting how the MLR method is as protective as a "full BLM" would be. If MLR is used for practical reasons I think that is fine as a justification, provided that it is clearly shown that there is no loss of protection. b) I think aluminum speciation is not fully addressed. In particular, solid versus dissolved Al. I am currently reviewing a manuscript by Robert Santore et al. (Windward) that does this very well (submitted to Environ. Toxicol. And Chem. "Development and application of a biotic ligand model for predicting chronic toxicity of dissolved and precipitated aluminum to aquatic organisms"). I suspect that the authors of this report are familiar with that work but if not they should contact the authors to preview a copy. This draft criteria document over simplifies the chemistry of aluminum by lumping particulate and dissolved aluminum	

Reviewer	Comments	Response to Comments
	into a total measure of aluminum. I think this is fine – based on reading the Santore draft paper – but the draft criteria document does not convince me it is fine. Different modes of exposure (particulate versus soluble) and not even really hinted at in the draft criteria document. c) Selection of total Al as the independent variable needs to be more fully justified (see comment above) but also it needs to be made clear if this is in fact what was done. It seems that the discussion on page 3 and 4 about the operational definitions of total and dissolved is used in the rest of the document, in particular, for the numerical criteria. This needs to be made clear. As mentioned in the Appendix (to this review), aluminum speciation needs to be defined more clearly. Total aluminum could include particulate refractory minerals, weathering product minerals (i.e., clay such as kaolinite) or it could include reactive solids such as amorphous gibbsite which can exchange back and forth solid to dissolved phase. Should these really all be included in one "box". This might be ok for lab studies where only amorphous gibbsite occur as solids – but what about field? d) Kinetics of "land scale" processes compared to acute and chronic tests. The criteria is based on lab-based studies but will be implemented with real samples. In acute tests the time scales are very short but in the field longer timescales will control the total aluminum. Not geologic timescales but longer than laboratory tests. e) Bottom-line the proposed criteria will be protective for aquatic life in an acute and chronic sense. The science is logical but needs just a bit more justification (as detailed in my points above). My personal bias is that I would prefer a BLM approach but I completely understand the practical needs that make MLR an attractive and possibly more likely to be adopted option.	
Reviewer 5	The technical approach used to derive the draft Al criteria is logical and scientifically supported. The MLR model has been validated using *Ceriodaphnia dubia* and *Pimephales promelas* with greater than 86% success at predicting Al toxicity values. Biaccumulation of Al is most observed in the lower trophic levels, as specifically stated in the document.	

2.3 Please comment on the data used to derive the revised criteria, including data adequacy/comprehensiveness, and the appropriateness of the data selected and/or excluded from the derivation of the draft criteria. Are there other relevant data that you are aware of that should be included? If so, please provide the reference.

All five reviewers commented favorably about the completeness of the data. None suggested additional studies for inclusion. Reviewer 1 said that the references are thorough and robust, but commented that biologically important endpoints that are non-standard should be retained instead of excluded. Both Reviewer 2 and Reviewer 3 said the criteria for inclusion or exclusion were transparent. Reviewer 2 responded that the aquatic toxicology data was very comprehensive and the bioconcentration discussion in Section 3.4 was well presented. The reviewer suggested further validating the models with similar studies containing data at a pH below 6.0 and explaining the statistical analysis by Cleveland et al. (1989). Reviewer 3 concluded that the data used in the derivation of the revised criteria are the result of a complete and thorough analysis that is comprehensive, linked to formally published guidance, and explicitly discussed. Reviewer 4 replied that the data are comprehensive, the data standards are very high, and the rejected datasets are discussed to present a balanced view. Reviewer 5 commented that there are more available, acceptable data to derive the draft criteria compared to the 1988 document, and that the data for the acute values fulfilled the minimum data requirement. The reviewer agreed with the approach to use “other data” to fill the missing minimum data requirement group for the chronic value. However, this reviewer noted that the database does not meet the minimum data requirements for estuarine/marine toxicity. The reviewer concluded that the data selected for inclusion were appropriate and thorough.

Reviewer	Comments	Response to Comments
Reviewer 1	The compilation of references is thorough and robust. The reasons for excluding data are generally reasonable, though in a data-limited world, it may be more appropriate to look for ways to include data rather than excluding it. Non-standard but biologically important endpoints should probably be retained rather than excluded, but I don’t think they would change criteria values in this case. It is apparent that there was a big push from industry to generate data to change this criteria, and industry is well within their rights to do so. There is no accompanying effort by EPA to ground truth or validate the industry supplied model that drives this criteria.	
Reviewer 2	The presentation of the aquatic toxicology data on aluminum was very comprehensive. The rational for what data were used for acute and chronic calculations for WQC was clear. The presentation and discussion of the data that were not utilized, in appendices I and J, certainly addresses past complaints about transparency and bias in selecting studies to include in analysis. A reader will know why a given study was or was not selected for criteria development. As I stated above, after explaining that there is not enough data to calculate WQC for estuarine/marine biota, moving all of the	

Reviewer	Comments	Response to Comments
	text on that topic from the main body of the report into an appendix would improve the flow of the document. In Figure 2 and in the paragraph on page 9 above the figure, the text states that aluminum in the water column at pH 7.0 is almost all in the insoluble form of aluminum hydroxide. It seems that there should be a more dramatic shift in criteria values moving from pH 7 to pH 5.5. In table 10 at a hardness concentration of 100 ppm and DOC of 1.0 ppm, the criteria at pH 5.5 is six fold lower than the value at pH 7.0. I wonder if the difference could be greater. In section 2.4.1, page 16, the process for normalization of the data to pH 7.0 is explained, however, the models are built from studies with a pH ranging from: pH 6.14 – 8.0, pH 6.3 – 8.1, and pH 6.0 – 8.0. These models extrapolate criteria below pH 6.0. Perhaps further validation of the models would be useful with data from similar studies that have data at a pH below 6.0. The authors, in section 2.9, identified other data needs and resulting uncertainty issues building from the DeForest et al. (2017) manuscript. The chronic brook trout study by Cleveland et al. (1989) is a driver for the chronic aluminum WQC. More discussion of the study could include more information on the statistical analysis of endpoints and the presentation of endpoints that were not significantly different from controls. It is noted that EC20 values were used for primarily sublethal endpoints. On page 17, data requirement from the EPA 1985 report are discussed. It would be informative for the reader to include a table of those data requirement. Since 1985, data quality has become a more studied topic and issues of experimental design, chemical analysis, statistical analysis, replication, etc. have become more important to include in toxicity publications. Section 3.4. The bioconcentration discussion was well presented. I suggested previously that some broader policy statements on the bioconcentration of inorganic metals be included. I would also recommend that the discussion in 4.1.6 be combined with the discussion in this section. Section 4.1.3. The inclusion of field data in the report is a very good step to relate the criteria development process to observed field effects.	
Reviewer 3	The data qualified in the derivation of the revised criteria are the result of a complete and thorough analysis through 2013. The data set is comprehensive, and acceptance or exclusion judgments	

Reviewer	Comments	Response to Comments
	or assumption are transparent, linked to formally published guidance, and explicitly discussed. Exploration of the 47 citations retrieved from a "Web of Science" database search of aluminum aquatic toxicity, for 2013 to the present did not yield any new controlled Al exposure studies. Specific comments related to the data used to derive the revised criteria of each section of the draft document appear above.	
Reviewer 4	a) Data seems very comprehensive and the data standards are very high. The rejected datasets are still discussed in the text so the document has a very balanced view. b) I am not aware of any other data sets that should be included.	
Reviewer 5	There were more available, acceptable data used to derive freshwater acute and chronic criteria in the 2017 Al document as compared to that used in the 1988 document. For acute values, the 2017 Al document included 21 species (12 invertebrates, eight fish species, and one frog species) in 19 genera as opposed to the 15 species (eight invertebrates and seven fish species) in 14 genera that were included in the 1988 Al document. These data fulfilled the minimum data requirements (MDR). For chronic values, the 2017 document included 11 species (seven invertebrates and four fish species), whereas, the 1988 document included only three species (two invertebrates and one fish species). The third family in phylum chordata is missing in the chronic database; however, "Other Data" was used to fulfill the missing MDR group. Data from a study using a wood frog was included in the chronic database for this purpose. I agree that this is the more scientifically defensible route to directly determine the final chronic value rather than using both the final acute value and the final Acute-Chronic Ratio (FACR) to estimate the final chronic value. All of the freshwater data were normalized to a pH of 7, hardness of 100 mg $CaCO_3$/L and DOC of 1 mg/L for comparison. The criteria in the 2017 Al document are determined based on those site-specific parameters (pH, hardness, and DOC). New acceptable, acute estuarine/marine data is available for five families representing five species in the 2017 Al document, as compared to no data before; however, the database still does not meet the minimum of eight families necessary to fulfill the MDR.	

Reviewer	Comments	Response to Comments
	One data gap identified is for estuarine/marine chronic Al toxicity, where there were no acceptable data. The data selected to be included in the databases were, in my opinion, appropriate and thorough. For comparison purposes, studies were also discussed (data provided in Appendices) that were missing some/all water chemistry parameters (i.e., pH) necessary for inclusion in the database.	

2.4 Are the derived criteria appropriately protective of commercially and recreationally important species, and of ecosystems overall?

Reviewer 2, Reviewer 3, and Reviewer 4 stated that the derived criteria are an improvement over the 1988 criteria. Reviewer 2, Reviewer 3, and Reviewer 5 agreed that the derived criteria are protective of ecosystems overall. Reviewer 3 stated that quantifying the effects of aluminum water chemistry helps states, tribes, and the regulated community maintain environmental quality and health aquatic ecosystems. Reviewer 3, Reviewer 4, and Reviewer 5 agreed that the derived criteria are protective of commercially and recreationally important species. Reviewer 4 said that the derived criteria are protective of endangered species. However, Reviewer 5 said that the derived criteria may not be protective of endangered freshwater mussels for which data are lacking.

Reviewer 1 was concerned that cross-species validation and chronic to acute validation were not completed. The reviewer was also concerned that the publication cited as the basis for the criterion is a manuscript that has not completed the peer-review process. The reviewer concluded that "much will depend on the assumptions that state-level entities apply to the stability of pH, DOC and hardness in the systems that these criteria are intended to protect."

Reviewer 2 said that the MLR provides a method that states and tribes can easily use with aquatic monitoring data they already collect. The reviewer suggested a new section that states the major benefits of the draft 2017 criteria and a discussion of the U.S. regions where the aquatic chemistry would indicate aluminum toxicity to be a substantial stressor to biota.

Reviewer	Comments	Response to Comments
Reviewer 1	This is a difficult question to answer because the application of the MLM models to other species has not been validated. Nor has the model transferability between chronic and acute effects been validated. What is particularly concerning, is that the publication cited as being the basis for this criterion is a submitted manuscript that at this time of writing and has not even completed the peer-review process. The model looks promising, but to use it without doing the cross species validation, and without chronic to acute validation is pre-mature.	

Reviewer	Comments	Response to Comments
	If adopted, in practice, much will depend on the assumptions that state-level entities apply to the stability of pH, DOC and hardness in the systems that these criteria are intended to protect. Are they assumed to be stable over time? Are they based on a single measurement in a site specific context? Is a state wide mean to be used for these important toxicity modifiers? Does the normalization of the model to conditions that are minimally toxic leave open the possibility that appropriate conversions based on pH, DOC and hardness may not occur?	
Reviewer 2	I think that the derived criteria are appropriately protective of aquatic biota and aquatic ecosystems. Compared to the 1988 criteria, the 2017 draft criteria are based on a larger dataset which provides more confidence in the criteria development process. The utilization of the aquatic chemistry of aluminum in the development of the criteria is a major improvement which recognizes the complex aspects of aluminum toxicity to aquatic biota. Also, the progression of tools to assess metal aquatic chemistry and toxicity from the BLM to MLR provides a method that states and tribes can easily use with aquatic monitoring data they already collect. At the end of the document, a new section (4.5 or 5.0) entitled conclusions should be added that states these major benefits of the draft 2017 criteria. For most aquatic systems in the U.S., with a pH between 6.5 – 8.0 (or so) aluminum should not be a significant stressor of aquatic biota. I think a broad discussion of the regions of the county where the aquatic chemistry of streams and lakes would indicate where aluminum toxicity could be a significant stressor to biota would be informative to the reader. In addition, issues of background concentrations of aluminum in low pH, hardness, and DOC systems could have aluminum concentrations above the proposed criteria and tolerance or adaptation may have occurred in these systems over time. Section 2.8, page 27. I have concerns on the discussion of the acute criterion where the final acute value (FAV) is divided by 2 to protect 95% of the species in a representative aquatic community from acute effects. The protection of 95% of the species in a community from acute effects is a policy goal and it should be stated as such. The dataset of no observable effects concentration (NOEC) values which are then divided by a safety factor of two is a management decision and I don't see how dividing by two is related to a 95% protection goal. It seems a SSD where 95% of the species are protected would be a more defensible approach.	

Reviewer	Comments	Response to Comments
	On page 45, it appears that the chronic criteria are developed using the MLR with no safety factor. As you move toward finalization of the aluminum WQC, "May The Force be with you."	
Reviewer 3	Yes, the derived criteria appropriately protective of commercially and recreationally important species, and of ecosystems overall. The 2017 draft criteria look-up tables present a research-based improvement on the 1988 criteria. Specific comparative examples such as brook trout in the draft document section 4.4, offer support for this conclusion. Quantifying the effects of water chemistry (pH, DOC, and Hardness) for aluminum risk to aquatic species is an improvement to meeting the goals of the CWA from the 1988 approach, helping states, tribes, and the regulated community maintain environmental quality and healthy aquatic ecosystems.	
Reviewer 4	a) Yes the document clearly demonstrated that criteria will be protective to commercially and recreationally important species and in fact to endangered species. The revised criteria are an improvement over the existing criteria. b) In terms of ecosystem protection overall the field studies that are discussed make it pretty clear that effects were observed for higher aluminum then would be recommended by this draft criteria.	
Reviewer 5	The database for acute and chronic Al data in freshwater and estuarine/marine systems is comprehensive; however, data gaps have been identified. Data for endangered freshwater mussels or closely related species are lacking. For these organisms the derived criteria may not be protective. Using current available data, I believe the derived criteria will be appropriately protective of most commercially and recreationally important species and of ecosystems.	

3.0 ADDITIONAL COMMENTS PROVIDED

Reviewer	Comments	Response to Comments
Reviewer 4	**Appendix: detailed "editorial" comments.** a) Chemistry section is a bit sparse. In particular colloidal and particulate and kinetics. If you measure total Al does it also include minerals like feldspar and weathering minerals like kaolinite? Then how does that compare to freshly precipitated amorphous gibbsite? b) A table of acronyms at the start of the document would be beneficial. Also making sure that each acronym is defined before first use (for example DOC in the executive summary (page vi) is not defined). c) For many comparisons in the text the criteria value for water chemistry of pH 7, DOC=1mg C/L and 100 mg CaCO3 hardness is used. I realize many more possible values are reported in the look up tables at the end of the document but why were these conditions selected for comparisons in the text? (such as the comparisons to the "old" criteria document). pH 7 is not exactly a common pH and DOC of 1 mg C/L is a pretty low value. d) Page vii there is a space missing 4 lines up from the bottom; "1985Guidelines" should be "1985 Guidelines". e) Page 2, "… yet very rarely in the elemental state …". The use of the word "yet" is a strange choice. And I'm not sure Aluminum is ever found in its elemental state in natural systems. f) Paragraph at the start of section 2.1 (page 2) does not mention aluminum in clays or soil or sediment. Just aluminum in rocks and that is complexed by some inorganic and organic ligands. Also, particulate (suspended) aluminum and colloidal and polymeric forms should be mentioned. These forms are referred to later in the document so they should be introduced here in the overview section. g) Paragraph starting at the bottom of page 2 starts with "Aluminum enters the environment …" is a strange opening. Aluminum is everywhere in the environment. The authors point out in the text how abundant aluminum is in the earth's crust. So I think the authors need to be specific here – what forms of aluminum are "entering" the environment? Does this mean dissolved aluminum? The document is	

Reviewer	Comments	Response to Comments
	written as if aluminum is like other metal contaminants and aluminum is very different. h) Paragraph starting at the bottom of page 2 says both natural and anthropogenic sources of aluminum are responsible for a "meaningful amount found in the aquatic environment". I am not clear what meaningful amount means in this context. i) Paragraph starting at the bottom of page 2 might be an appropriate place to mention aluminum solubility and Ksp (unless a separate section on chemistry is adopted per my suggestion above). This is an important concept in nature and a really important concept in the toxicity experiments. j) Page 3 around the middle, aluminum is also used in wastewater treatment to remove phosphorus. k) Page 3 around the middle of the page, it is not clear what is meant by ``diversely abundant". l) Top of page 4, colloidal and particulate aluminum species should be clearly discussed. The current text does explain soluble speciation (i.e., complexation) but fails to recognize solid speciation. There is a large difference between a particle of feldspar or kaolinite from freshly precipitated aluminum hydroxide. Also polymeric species. m) This is a bit of a technical question but why does the minimum detection limit for dissolved concentration differ from the minimum detection limit of total aluminum? (bottom of page 4). n) Figure 1 gives a map of wells but there is no Figure showing a map of the surface water samples discussed. o) Below Figure 1 on page 5 in reference to marine and estuarine systems the recent paper by Angel et al should be cited. It gives specific measured values. This paper is also useful because it points out the dynamics of aluminum in toxicity test solutions – it specifically considers the timescale and chemistry of exposure solutions. [Geochemical controls on aluminum concentrations in coastal waters. Brad M. Angel, Simon C. Apte , Graeme E. Batley and Lisa A. Golding. Environmental Chemistry 13(1) 111-118 http://dx.doi.org/10.1071/EN15029]. p) Bottom of page 6, "higher acid extractable aluminum" should be explained what this term means. Also, this highlights the need for solid (particulate) aluminum	

Reviewer	Comments	Response to Comments
	speciation measurements. There is a difference between geologic minerals and transient hydroxide precipitates. q) Middle of page 7 is the first mention of polymeric and colloidal forms of Al and also the first mention of Al sorbed to clay. These concepts need to be mentioned sooner – again, a separate appendix on Al chemistry should be considered. r) Bottom of page 7 there is a mention of carbonate precipitates of aluminum. I'm not sure these exist. A citation should be given. s) Middle of page 8 is the first mention of measuring aluminum speciation (Driscoll reference). There is a large body of literature from the 1980s on aluminum speciation and although this is old literature it should be reviewed. The authors use this one Driscoll citation to downplay the usefulness of aluminum speciation methods which define fractions such as monomeric and polymeric aluminum. It is true that these methods are operationally defined but perhaps this section could include more substantial literature review before rejecting them out of hand? Lumping all particulate aluminum into a single "box" as is proposed with a single total aluminum measurement has its own limitations. t) Sentence in the middle of the first complete paragraph on page 8 is out of place. "Aluminum toxicity is important in freshwater and marine environments due to both anthropogenic sources …" Just kind of comes out of no where. u) Section 2.3 it is not clear if this section is referring to soluble aluminum transport or also including particulate aluminum transport. v) Page 9 near the top of the page says that at neutral pH aluminum is nearly insoluble – this should be quantified. The Ksp of aluminum hydroxide allows clear estimation of the solubility limits of aluminum. w) Figure 2 is a good figure but another figure showing precipitation equilibria would give a more balanced view of aluminum speciation. Also the final sentence of the figure caption is not in bold face font.	

Reviewer	Comments	Response to Comments
	x) The list of species at the bottom of page 10, top of page 11, is a repetitive list but does not mention the same species as the previous listing of species (i.e., first paragraph of section 2.1.). y) Are all filters created equal? The authors refer to 0.45 µm filtration throughout the document but do not specify filter media. Different media can lead to different answers. Probably beyond the level of detail this document is intended for but I mention this incase the authors want to consider. z) The first complete paragraph on page 12 goes into more details on aluminum chemistry. These details could be included earlier in the more general discussion paragraphs. In particular time is mentioned as a variable here. This is significant in aluminum chemistry and needs to be highlighted earlier. The Angel manuscript mentioned above does a good job of showing aluminum kinetics on the timescale of toxicity testing. aa) The mention of aluminum flocs coprecipitating nutrients in the middle paragraph of page 12 is a well known phenomena. It is in fact how chemically mediated phosphorus removal works in wastewater treatment. Also these examples (and this text) is taken from the 1985 criteria document. Some more recent examples would be better. bb) Page 13 when the Appendices are referred to it would be nice to give the title of the appendix. That would really help the reader (I noticed this in a later section when the title was given in reference to the appendix.). Specifically here though it would be nice to include some of the numbers and not rely on the reader to go and read the Appendix to find the number for themselves. cc) Second to last paragraph on page 13 refers to MLR models as less complex than BLM models I think I would also emphasize that MLR models are not mechanistic. They are referred to as bioavailablity models and it is semantics but since there is no mechanistic aspect to an MLR I would prefer if they were not referred to as bioavailablity models. It is true MLR models take toxicity mitigating factors into account so maybe some reference to that would be better? dd) When MLR models are discussed on page 14 (bottom paragraph) it is not clear if total aluminum (as defined in this	

Reviewer	Comments	Response to Comments
	document) is the independent variable or if other forms of measured aluminum where tested. ee) Last paragraph of page 14, are pH, hardness and DOC really independent variables? More independent than say pH and alkalinity but these aren't really independent variables. Harder water could have more alkalinity (hardness cations are strong base cations after all) and high DOC waters are often acidic because of organic acids. ff) The MLR paragraph at the bottom of page 14 really emphasizes the importance of pH, DOC and hardness. These variables are mentioned in the earlier text but I think could be emphasized more strongly. So many other things like sulfate, phosphate, … are mentioned as well. The reader should be convinced that these are the best variable choices. gg) Page 15 6 lines down, "more simple" should be "simpler". hh) Page 15, only R^2 values are given, p values should also be given. ii) Section 2.7.1 repeats things that have been mentioned before. It almost seems that the earlier sections drew heavily from the 1985 document and then this text was created for this current document. The end result is repetitive though. I prefer how things are presented in section 2.7.1 and I really like Figure 3. jj) A figure similar to Figure 3 could be developed showing aluminum speciation. Or Figure 3 could include some indication of aluminum speciation. kk) Page 30 middle paragraph argues how laboratory toxicity tests can be extrapolated to the field. This is reasonable for other metals but some caveats should be given for aluminum. The slow kinetics and the difference between geological aluminum materials suspended in natural water and the transient precipitates formed during toxicity testing should be discussed. I agree that lab-based toxicity testing is the best we can do for now but longer term field-based aluminum toxicity tests could be very useful. (such as section 4.1.3). ll) Table 3 and elsewhere when SMAV values (or other aluminum values are referred to) it should be made clear if this is referring to total or dissolved aluminum. I think it is	

Reviewer	Comments	Response to Comments
	pretty much always total aluminum but readers may want to just skip to the tables so this should be made clear in the table footnotes. Figure 4 does specifically mention that it is total aluminum. mm) Section 4.1.3 ends by stating that the study exceeded the minimum acceptable guidelines for tests of this duration. It is not clear what this means. nn) Section 4.1.3 are there any field based fish studies? I know that aluminum mobilized by environmental acidification was implicated in fish kills. Are there any systematic studies in this regard? The authors repeat other acidic test results (top of page 50 for example). oo) Throughout section 4 aluminum concentrations are often referred to (i.e., effects concentrations). Are these always total aluminum values? This applies to the summary tables at the end of the document as well – maybe some blanket statement would help avoid any ambiguity here. I am not sure if all accepted papers measured total aluminum so the authors need to clarify this. pp) Page 59 refers to the appendix by name (Appendix J reference). This is a nice way to do this and would make the rest of the document more readable if this convention was adopted throughout. qq) Page 62 says that MLR accounts for the bioavailability of the metal. It is semantics but I think it is overstating things to state that MLR account for bioavailablity. There is no mechanism in an MLR.	

4.0 NEW INFORMATION PROVIDED BY REVIEWERS

This section presents all new information that reviewers provided in addition to or within their specific responses (presented in Section 2, above) to the charge questions.

Reviewer	Comments	Response to Comments
Reviewer 2	REFERENCES Greenberg MS, Schoeters I, Wentsel RS, Charters DW, Mitchell IA, Zajdlik B. Regulatory considerations for the potential	

Reviewer	Comments	Response to Comments
	development and application of metal clean-up values. Integr Environ Assess Manag 2014; 10:401–414. US EPA (2007). Framework for metals risk assessment. EPA 120/R-07/001. March 2007. Office of the Science Advisor. US Environmental Protection Agency. Washington, DC.	

Appendix A

Charge to Reviewers

Technical Charge to External Peer Reviewers

Contract No. EP-C-12-029
Task Order 65
February 2017

External Letter Peer Review of EPA's Draft Aquatic Life Ambient Water Quality Criteria for Aluminum - 2017

BACKGROUND

National Recommended Ambient Water Quality Criteria (AWQC) are established by the United States Environmental Protection Agency (EPA) under the Clean Water Act (CWA). Section 304(a)(1) aquatic life criteria serve as recommendations to states and tribes by defining ambient water concentrations that will protect against unacceptable adverse ecological effects to aquatic life resulting from exposure to pollutants found in water. Aquatic life criteria address the CWA goals of providing for the protection and propagation of fish and shellfish. Once EPA publishes final §304(a) recommended water quality criteria, states and authorized tribes may adopt these criteria into their water quality standards to protect designated uses of water bodies. As required by the CWA, EPA periodically reviews and revises §304(a) AWQC to ensure they are consistent with the latest scientific information. In support of this mission, EPA is working to update water quality criteria to protect aquatic life from the presence of aluminum in freshwater environments.

The current aluminum criteria are from 1988 and applied to pH range 6.5 – 9.0 and across all hardness and DOC ranges. The approach in this peer review is a multi linear regression (MLR) model using pH, hardness, and DOC, with interaction terms between these parameters for freshwater aluminum criteria. This draft includes freshwater acute and chronic mussel tests and additional toxicity studies that have exposure to different combinations of pH, hardness, and DOC ranges.

CHARGE QUESTIONS

The draft document provided is the focus of this review.

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

2. Please comment on the technical approach used to derive the draft aluminum criteria; is it logical, does the science support the conclusion, and is it consistent with the protection of freshwater life from acute, chronic, and bioaccumulative effects? Please provide specific comments.

3. Please comment on the data used to derive the revised criteria, including data adequacy/comprehensiveness, and the appropriateness of the data selected and/or excluded from the derivation of the draft criteria. Are there other relevant data that you are aware of that should be included? If so, please provide the reference.

4. Are the derived criteria appropriately protective of commercially and recreationally important species, and of ecosystems overall?

APPENDIX B

INDIVIDUAL REVIEWER COMMENTS

COMMENTS SUBMITTED BY

Reviewer 1

External Peer Review of Aquatic Life Ambient Water Quality Criteria for Aluminum – 2017

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

Overall, the document is well constructed and reasonably clear in its presentation.

2. Please comment on the technical approach used to derive the draft aluminum criteria; is it logical, does the science support the conclusion, and is it consistent with the protection of freshwater life from acute, chronic, and bioaccumulative effects? Please provide specific comments.

The criteria are based primarily on multiple linear regression models that use pH, hardness and DOC as modulators of 7-day toxicity in two models – a fish (*Pimephales promelas*) with biomass as the endpoint, and a crustacean (*Ceriodaphnia dubia*) with reproduction as an endpoint. The best fit models for each species were different from each other which poses many questions:

a) pH, hardness and DOC are described as important modulators of bioavailability. Do the different models then mean that bioavailability differs between species as a function of their physiology and not as a function of Al speciation? This is extremely important to articulate because the chemistry and bioavailability of Al is complex. Species also vary widely and the physiology of Al uptake is not very well understood comparatively across species.

b) Do we know that the model based on *C. dubia* fits other invertebrates well? Do we know that the model based on *P. promelas* fits other fish well? Why are we so willing to make this assumption? Shouldn't this be tested first?

c) Do we know that the models fit equally well for acute and chronic toxicity?

This is extremely important to articulate the uncertainty of the assumption that the models apply to other species based on gross taxonomy (vertebrates vs invertebrates).

P. 15: "In the final MLR model, predicted EC20s were within a factor of two of observed values for 95% of the tests. I assume that this means for the tests performed with *P. promelas* and *C. dubia,* most observations came within a factor of 2 of the model predictions that were built from the same data. More information would be useful here. Are predictions generally biased in an over-protective or under-protective way? If the uncertainty for any specific prediction is a factor of 2, shouldn't this uncertainty be built into the criteria? Should an interspecific uncertainty factor be applied as well?

P. 3: Al associated with sediments are "unavailable to aquatic organisms"? This is a rather broad sweeping statement to make.

P. 9: The temperature-solubility issue is completely ignored in the model. If a 15 degree change is equal to a whole pH unit, and pH drives the model, shouldn't there be a temperature term in the model?

P. 10: "aluminum has also been shown to increase respiration, and thereby energy demands"… but an increase in respiration is a cause for excluding data from consideration?

P. 30: "Application of water only laboratory toxicity tests to develop water quality criteria to protect aquatic life" fails on many fronts and among the reasons that EPA is considering modernizing the 1985 guidelines. This paragraph is not entirely truthful.

P. 30: "…Only chronic data are used in the model, and application to acute data normalization assumes that the same relationships are present". This is a big assumption. Such an assumption would generally fail for BLM models based on chronic data applied to acute data, and to acute BLMs applied to chronic data for other metals. Why should we make this assumption here?

P. 60: "Comparison of 1988 and 2016 (sic) criterion values": This paragraph is troubling. The new criteria values significantly weaken protection of aquatic ecosystems by allowing higher Al concentrations in most situations than the 1988 criteria. To suggest otherwise is disingenuous.

3. Please comment on the data used to derive the revised criteria, including data adequacy/ comprehensiveness, and the appropriateness of the data selected and/or excluded from the derivation of the draft criteria. Are there other relevant data that you are aware of that should be included? If so, please provide the reference.

The compilation of references is thorough and robust. The reasons for excluding data are generally reasonable, though in a data-limited world, it may be more appropriate to look for ways to include data rather than excluding it. Non-standard but biologically important endpoints should probably be retained rather than excluded, but I don't think they would change criteria values in this case.

It is apparent that there was a big push from industry to generate data to change this criteria, and industry is well within their rights to do so. There is no accompanying effort by EPA to ground truth or validate the industry supplied model that drives this criteria.

4. Are the derived criteria appropriately protective of commercially and recreationally important species, and of ecosystems overall?

This is a difficult question to answer because the application of the MLM models to other species has not been validated. Nor has the model transferability between chronic and acute effects been validated. What is particularly concerning, is that the publication cited as being the basis for this criterion is a submitted manuscript that at this time of writing and has not even completed the peer-review process. The model looks promising, but to use it without doing the cross species validation, and without chronic to acute validation is pre-mature.

If adopted, in practice, much will depend on the assumptions that state-level entities apply to the stability of pH, DOC and hardness in the systems that these criteria are intended to protect. Are they assumed to be stable over time? Are they based on a single measurement in a site specific context? Is a state wide mean to be used for these important toxicity modifiers? Does the normalization of the model to conditions that are minimally toxic leave open the possibility that appropriate conversions based on pH, DOC and hardness may not occur?

COMMENTS SUBMITTED BY

Reviewer 2

External Peer Review of Aquatic Life Ambient Water Quality Criteria for Aluminum – 2017

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

I would first like to commend the authors for producing this draft report on the Water Quality Criteria (WQC) for aluminum. It is a complex topic and the authors incorporate very recent scientific approaches which improves the reports technical basis and its application to states and other users of the criteria. The Environmental Protection Agency and the Office of Water, in particular, have a successful history of improving the quality and utilization of our nation's waters and while others may forget turbid, oily, lakes and streams with detergent bubbles floating on them, I do not, and I thank you for continuing your important role in the protection of people and the environment in which they live.

The purpose of the document, in the Executive Summary, is listed as improving scientific knowledge. I would suggest adding – and to increase the utilization by states, tribes, and other users. Developing the criteria using the multiple linear regression (MLR) approach will use water chemistry parameters that the vast majority of the states currently measure in their water monitoring programs. The relevance of these draft criteria will be very high because the states and tribes can easily apply them in their water quality programs.

In the Executive Summary, I would suggest a paragraph on the history of utilization of the previous aluminum WQC (or others) by the states to show how useful the documents are at protecting our Nation's water.

Overall the clarity of the writing is very good. The document organization follows the format of the ecological risk assessment framework and that process is familiar to the readers of the document.

There is not enough data to conduct WQC for estuaries and marine environments, however discussion sections continue through the document. I would suggest stating that there is not enough data and then combine the further text and data on estuary/marine systems into an appendix.

I think the reader needs to brought along with the progression of the use of the biotic ligand model (BLM) in the copper WQC to the current selection of the MLR approach. An expansion of the text on page 13 would benefit the report. A discussion on the development and application of the full BLM, which could include text on the collaborative effort to produce the BLM approach and its utilization of aquatic chemistry data. Implementation issues concerned the amount of water quality data required to run the full BLM has been a significant for the states and tribes. Then text on potential solutions to the implementation issues could include a discussion of analysis of the partial BLM approach and analysis using the MLR. This text would help the reader understand the amount of effort, in coordination with the scientific community, that went into the development of a more relevant approach.

2. **Please comment on the technical approach used to derive the draft aluminum criteria; is it logical, does the science support the conclusion, and is it consistent with the protection of freshwater life from acute, chronic, and bioaccumulative effects? Please provide specific comments.**

Overall, the approach is technically sound. The vast improvement of the MLR method, which includes key water chemistry parameters, over those used for the 1988 aluminum criteria, will enable states, tribes, and local governments to assess potential aluminum toxicity at any aquatic site based on its unique aquatic chemistry.

Section 2.2. The justification for the use of total recoverable aluminum is a complex topic which requires a decision to support the criteria development process. Further details on why the authors of the cited toxicity studies thought the dissolved concentrations were relatively level as total aluminum concentrations and toxicity increased would be good. Also, are these studies the exception? Do most toxicity studies show a correlation between dissolved concentrations of aluminum and toxicity? The toxic forms of aluminum should be in both of the dissolved and total recoverable values.

Section 2.3. At the start of this section the inclusion of text on acid rain issues in the 1980s and the building awareness of aluminum toxicity to aquatic biota would be informative. The discussion and the Figure 2 provide the reader with important information on the environmental chemistry, fate and transport of aluminum.

Section 2.4. Increasing the range of pH in toxicity studies included in the criteria to 5.0 – 9.0 is a very good decision. Toxic forms of aluminum increase at a pH lower than 6.5.

Section 2.4.1. The discussion on the BLM method and the use of MLR is important for the reader. I think text should be added that the data requirements (ten parameters) were difficult for many state monitoring programs to address. The MLR approach uses data that almost all states already collect, so the MLR would be much more likely to be used and provide increase scientific rigor for the protection of aquatic species from aluminum. A discussion comparing the output of the methods would be good. The 2017 publications cited in this section are still (as of this writing) in review. A text box or text on key points from Brix et al., 2017 and DeForest et al., 2017 and why they support the use of the MLR would help the reader. Some limits of the DeForest study are mentioned on page 30, but It would be good to have that discussion earlier in the document and all in one place. I commend the EPA for their awareness of these studies and the cooperation with these authors to include their research.

Once the discussion focuses on the methods and results from DeForest, I would start a new section or subsection of the report. This text is the technical support for your document.

Section 2.5. On page 17, when you first mention assessment endpoints, please state that they are listed in Table 1. On page 16, add a web address for Stephan et al., 1985 or as it is also cited, EPA, 1985.

Section 2.6.2. This section briefly discusses use of the effects concentration (EC20) to protect populations from long term chronic effects of a toxicant. If population models or a discussion relating the draft criteria to field studies are not going to be included in the document, I would recommend using the wording, "inferred protection of populations… ". Adding a population model as an additional line of evidence would be informative if the required biological and effects data are available for a species of interest.

Section 2.7. The conceptual model section is well written and important for the user to understand the various exposure pathways and which ones may be most important.

Section 2.8. The analysis plan is well presented. The use of a species sensitivity distribution (SSD) for the four most sensitive species and using the 5th percentile of that distribution is conservative. The decision to use a factor of 2 is more conservative and is based on a 1978 Federal Register notice. This process is supposed to protect 95% of the species in an aquatic community. Conventional use of a SSD requires at least 7 species, why is the use of the four most sensitive species proper? A conventional SSD and using an EC5 percentile is predicted to be protective of 95% of a given community (Greenberg et al., 2013). Why not use all of the data? From the text in the section, it seems that the 1978 FR notice data is a method to get to near control test organism survival percentages. I don't think that relates to protection of 95% of the species in an aquatic community. If you want to use the factor of 2, it could be stated as a policy decision.

Section 3.1.1. Once again, the factor of 2. Maybe state that it is intended to protect 95% ….

Section 3.2. The discussion of the chronic toxicity studies was a very good summary. I would add at the beginning of the section that EC20 values were used for primarily sublethal endpoints.

Section 3.3. The authors could consider adding bullets on bioconcentration/biomagnification issues for inorganic metals from the EPA's Framework for Metals Risk Assessment (US EPA, 2007). An example bullet:

- Trophic transfer can be an important route of exposure for metals, although biomagnification of inorganic forms of metals in food webs is generally not a concern in metals assessments.

Section 3.5. Table 10 is a good presentation of the effects of pH and hardness on the criteria. It would be helpful to the reader to explain why when the pH is above 7.5, and higher, that the criteria values are lower as the hardness concentrations increase.

Section 4. The discussion of aquatic toxicity studies that had non-standard experimental designs or results was good to include in this document. The results are discussed and the reasons for the lack of inclusion in the datasets are presented. A given regulator could utilize those studies if their inclusion met their needs.

Section 4.3. In the discussion on endangered species, it might be good to discuss habitat in the U.S. where a species would be most at risk, e.g., acidic lakes with low hardness and dissolved organic carbon (DOC). An overlap of at risk habitat with an endangered species range would be informative. This suggestion may exceed the objectives of this report, but could be considered in more focused inter-agency studies.

Section 4.4. The comparison of 1988 criteria and the 2017 draft criteria should be briefly discussed in the Executive Summary.

3. **Please comment on the data used to derive the revised criteria, including data adequacy/ comprehensiveness, and the appropriateness of the data selected and/or excluded from the derivation of the draft criteria. Are there other relevant data that you are aware of that should be included? If so, please provide the reference.**

The presentation of the aquatic toxicology data on aluminum was very comprehensive. The rational for what data were used for acute and chronic calculations for WQC was clear. The presentation and discussion of the data that were not utilized, in appendices I and J, certainly addresses past complaints about transparency and bias in selecting studies to include in analysis. A reader will know why a given study was or was not selected for criteria development.

As I stated above, after explaining that there is not enough data to calculate WQC for estuarine/marine biota, moving all of the text on that topic from the main body of the report into an appendix would improve the flow of the document.

In Figure 2 and in the paragraph on page 9 above the figure, the text states that aluminum in the water column at pH 7.0 is almost all in the insoluble form of aluminum hydroxide. It seems that there should be a more dramatic shift in criteria values moving from pH 7 to pH 5.5. In table 10 at a hardness concentration of 100 ppm and DOC of 1.0 ppm, the criteria at pH 5.5 is six fold lower than the value at pH 7.0. I wonder if the difference could be greater. In section 2.4.1, page 16, the process for normalization of the data to pH 7.0 is explained, however, the models are built from studies with a pH ranging from: pH 6.14 – 8.0, pH 6.3 – 8.1, and pH 6.0 – 8.0. These models extrapolate criteria below pH 6.0. Perhaps further validation of the models would be useful with data from similar studies that have data at a pH below 6.0. The authors, in section 2.9, identified other data needs and resulting uncertainty issues building from the DeForest et al. (2017) manuscript.

The chronic brook trout study by Cleveland et al. (1989) is a driver for the chronic aluminum WQC. More discussion of the study could include more information on the statistical analysis of endpoints and the presentation of endpoints that were not significantly different from controls. It is noted that EC20 values were used for primarily sublethal endpoints.

On page 17, data requirement from the EPA 1985 report are discussed. It would be informative for the reader to include a table of those data requirement. Since 1985, data quality has become a more studied topic and issues of experimental design, chemical analysis, statistical analysis, replication, etc. have become more important to include in toxicity publications.

Section 3.4. The bioconcentration discussion was well presented. I suggested previously that some broader policy statements on the bioconcentration of inorganic metals be included. I would also recommend that the discussion in 4.1.6 be combined with the discussion in this section.

Section 4.1.3. The inclusion of field data in the report is a very good step to relate the criteria development process to observed field effects.

4. **Are the derived criteria appropriately protective of commercially and recreationally important species, and of ecosystems overall?**

I think that the derived criteria are appropriately protective of aquatic biota and aquatic ecosystems.

Compared to the 1988 criteria, the 2017 draft criteria are based on a larger dataset which provides more confidence in the criteria development process. The utilization of the aquatic chemistry of aluminum in the development of the criteria is a major improvement which recognizes the complex aspects of aluminum toxicity to aquatic biota. Also, the progression of tools to assess metal aquatic chemistry and toxicity from the BLM to MLR provides a method that states and tribes can easily use with aquatic monitoring data they already collect. At the end of the document, a new section (4.5 or 5.0) entitled conclusions should be added that states these major benefits of the draft 2017 criteria.

For most aquatic systems in the U.S., with a pH between 6.5 – 8.0 (or so) aluminum should not be a significant stressor of aquatic biota. I think a broad discussion of the regions of the county where the aquatic chemistry of streams and lakes would indicate where aluminum toxicity could be a significant stressor to biota would be informative to the reader. In addition, issues of background concentrations of aluminum in low pH, hardness, and DOC systems could have aluminum concentrations above the proposed criteria and tolerance or adaptation may have occurred in these systems over time.

Section 2.8, page 27. I have concerns on the discussion of the acute criterion where the final acute value (FAV) is divided by 2 to protect 95% of the species in a representative aquatic community from acute effects. The protection of 95% of the species in a community from acute effects is a policy goal and it should be stated as such. The dataset of no observable effects concentration (NOEC) values which are then divided by a safety factor of two is a management decision and I don't see how dividing by two is related to a 95% protection goal. It seems a SSD where 95% of the species are protected would be a more defensible approach. On page 45, it appears that the chronic criteria are developed using the MLR with no safety factor.

As you move toward finalization of the aluminum WQC, "May The Force be with you."

REFERENCES

Greenberg MS, Schoeters I, Wentsel RS, Charters DW, Mitchell IA, Zajdlik B. Regulatory considerations for the potential development and application of metal clean-up values. Integr Environ Assess Manag 2014; 10:401–414.

US EPA (2007). Framework for metals risk assessment. EPA 120/R-07/001. March 2007. Office of the Science Advisor. US Environmental Protection Agency. Washington, DC.

COMMENTS SUBMITTED BY

Reviewer 3

External Peer Review of Aquatic Life Ambient Water Quality Criteria for Aluminum – 2017

1. Comments on the overall clarity of the document and construction as it relates to the derivation of each criterion.

Overall the draft criteria document is well organized and clearly written. It contains a satisfying level of breadth and depth in analyzing the results of a comprehensive literature review and successfully performing a critical analysis of a large peer-reviewed data set concerning aluminum risk to aquatic ecosystems. The MLR model approach and the thoroughness of the input data selection/rejection are strengths of this criteria development document.

Specific Document Section Comments

Executive Summary

The Executive Summary has sufficient detail to provide readers with the regulatory premise, background, critical concepts, derivation and modeling approaches and the criteria conclusions presented in the document. I would suggest "getting to the point" in the opening paragraph of the Executive Summary with the following edit: "This update establishes a freshwater criteria magnitude that is affected by total hardness, pH, and DOC and expands on the toxicity database to include those studies beyond the pH range of 6.5-9.0. The criteria results are presented in a series of look-up tables in Appendix K."

Introduction and Background

This is a well-developed and informative section. The regulatory premise and application of criteria are discussed in clear exposition. The active web link to National Recommended Water Quality Criteria - Aquatic Life Criteria Table will be helpful in establishing context to some readers, especially non-experts, who may access this document with limited background knowledge or as their first introduction to water quality criteria.

2 Problem Formulation

This section establishes cause and framework for the document in a concise manner.

2.1 Overview of Stressor Sources and Occurrence

This section is a well-cited review. The analysis is complete and representative. The supporting information and in context explanations (i.e., dissolved Al samples) are helpful without being pedantic.

2.2 Stressors of Concern

The treatment of 2009 total recoverable metal differences of aquatic Al concentrations and the potential for "conservative" risk assessment is balanced and offers a fair treatment of potential for bias from a change in sampling and analysis methodology. The potential for bias is managed in the selection of published controlled exposure studies cited in the draft document that uses the same dose construction, sampling, and analysis approaches.

2.3 Environmental Fate and Transport

Aluminum environmental chemistry is complex, and Figure 2 is very helpful in demonstrating this complexity to the reader.

2.4 Mode of Action and Toxicity

This section is a well-developed analysis and introduces the reader to the aquatic chemistry challenges of Al, as it relates to exposure studies. It establishes the context for study criteria to be included in model inputs. The section is comprehensive in exploring the practical challenges of performing controlled Al dosing studies. The discussion of pH effects and limits for study inclusion is a strength.

2.4.1 Water Quality Parameters Affecting Toxicity

This section introduces BLM and MLR modeling approaches with sufficient depth and breadth and is well cited. The MLR approach is well documented and clearly discussed in this section. The in-review manuscripts of Cardwell 2017, Gensemer 2017, Brix 2017 and DeForest 2017 are a critical foundation needed to support the draft document. Substantive changes to the final peer-reviewed papers will need to be reflected in changes to this draft criteria document. Because of the entwined fate of those papers and this draft document, acceptance for publication with minor changes is a hopeful outcome of that review process. My read of the DeForest 2017 manuscript suggests it is a strong work.

2.5 Measurement Endpoints

This section establishes a context for the approach. The outline of process and procedures is clearly developed and clearly written with a sufficient and satisfying level of explanation for a wide range of readers.

2.5.1 Overview of Toxicity Data Requirements

The use of bullets in this section is effective to outline an overview of data requirements.

2.6 Measures of Effect

The section does a solid job in explaining the data and study research search methodology that is used to determine test acceptability the development of the criteria.

2.6.1 Acute measures of effect

The section solidly introduces the reader to fundamental terms and resources found in the draft document and appendices.

2.6.2 Chronic measures of effect

The section solidly introduces the reader to fundamental terms and resources found in the draft document and appendices. Table 1 is a useful organization and summary of concepts.

2.6.3 Aluminum Toxicity Data Fulfilling Minimum Data Requirements

The section introduces the summary of acceptable toxicity data in Table 2 and the insufficient minimum dataset requirements for estuarine/marine criterion value determination with sufficient robustness. Communicating the need for additional acute and chronic toxicity testing on estuarine/marine taxa in this section is helpful in supporting research proposals to develop that data gap.

2.7 Conceptual Model

The approach of conceptual models is introduced with clarity and conciseness.

2.7.1 Conceptual Diagram

Figure 3 and the related discussion is an excellent resource for organizing the sources, pathways, receptors and controls of environmental Al. This section and figure do not have citations, and this is a curious omission as Al is a primary resource in common use. The authors may want to explore, review and cite similar conceptual diagrams and models that may exist in other quality resources.

2.8 Analysis Plan

This section is a thorough and comprehensive discussion of the background and approach for developing the data set used in modeling input and criteria development. The complexity of the topic is well treated in this exposition, and the authors present clear explanations with sufficient detail and definitions. The section is supportive of the criteria development goals of the draft document.

2.9 Identification of Data Gaps and Uncertainties for Aquatic Organisms

This section presents a fair and transparent treatment of data gaps and uncertainties in Al criterion development. Assumptions are clearly and openly discussed. The supporting discussion of management of data gaps and assumptions is convincing and support of the final criteria development. The explicit identification of data gaps in estuarine/marine species and for plants support the research community in their project proposals to help close those data gaps. As a minor edit, the third line from the bottom on page 30 has an extra space inserted.

3 Effects Analyses

3.1 Acute Toxicity to Aquatic Animals

The comprehensive data sets of Appendix A and B are introduced.

3.1.1 Freshwater

This section explicitly, and in detail, explores the available freshwater Al toxicity data. The use of tables and figures helps organize the complex available resources. The critical analysis and discussion of the limitations and usefulness of available studies are well done and supported by references to criteria development standards. The clarifying sentence: “That is, EPA’s acute freshwater criteria are not fixed values, but equations based on bioavailability and hence toxicity of aluminum under different water chemistry conditions (see Appendix K for additional criteria values)" helps communicate the new Al guidance in the draft document.

3.1.2 Estuarine/Marine

This section discusses the data gaps and limitations that sufficiently support a finding of inability to calculate an estuarine/marine criterion.

3.2 Chronic Toxicity to Aquatic Animals

This section introduces Appendix C that comprehensively lists acceptable chronic toxicity data.

3.2.1 Freshwater

This section explicitly, and in detail, explores the available freshwater Al chronic toxicity data. The use of tables and figures helps organize the complex available resources. The critical analysis and discussion of the limitations and usefulness of available studies are well done and supported by references to criteria development standards. Many of the cited studies are recent and very high quality, addressing the anticipated data needs for developing an Al water quality criterion. The exploration of MDR missing the third family in phylum Chordata is a well-developed, transparent and satisfying approach to missing data. The detailed abstracting of the studies in this section is a useful approach to establishing the scientific basis for consideration of a diverse range of research studies as data resources for criterion development.

3.2.2 Estuarine/Marine

The section contains the explicit data reference of Appendix D for no estuarine/marine chronic toxicity data.

3.3 Bioaccumulation

This is a useful section, especially as it relates to fish accumulation and Al dietary intake potential. While there is absent or limited data, the section is important for completeness.

3.4 Toxicity to Aquatic Plants

This too is a useful section, and while there is absent or limited data, the section is important for completeness.

3.5 Summary of National Criteria

3.5.1 Freshwater

The summary approach, including example calculations and an example look-up table, with further reference to the full Appendix K look-up tables, is a good approach to introduce the reader to the National Criteria of the draft document. The approach is clear and concise, with good readability.

4 Effects Characterization

4.1 Effects on Aquatic Animals

4.1.1 Freshwater Acute Toxicity

This section explicitly, and in detail, explores the available freshwater Al acute toxicity effects data. The critical analysis and discussion of the limitations and usefulness of available studies are well done and supported by references to criteria development standards. The justifications for inclusion or rejection of specific study data appear reasonable and well supported in the document discussion.

4.1.2 Freshwater Chronic Toxicity

This section explicitly, and in detail, explores the available freshwater Al acute toxicity effects data. The critical analysis and discussion of the limitations and usefulness of available studies are well done and supported by references to criteria development standards. The justifications for inclusion or rejection of specific study data, for example, pH or study duration, appear reasonable and well supported in the document discussion.

4.1.3 Freshwater Field Studies

The discussion is useful for completeness of criteria background development and observation of need for Al management, especially in consideration of historical acidification of some surface waters in the eastern US.

4.1.4 Estuarine/Marine Acute Toxicity

The discussion is useful for completeness of criteria background development, and correlation/comparison with other Appendix B and I data.

4.1.5 Estuarine/Marine Chronic Toxicity

The discussion is useful for completeness of criteria background development, and correlation/comparison with other Appendix D and I data.

4.1.6 Bioaccumulation

The discussion of effects of Al bioaccumulation is well executed with good breadth and depth of the available literature. There is limited published work available, and the analysis appears complete.

4.2 Effects on Aquatic Plants

With reference to Appendix E, this section summarizes and reviews toxic effect on plants with available citations.

4.3 Protection of Endangered Species

The summary statement that the 2017 criterion is protective of endangered species is an important early summary statement.

4.3.1 Key acute toxicity data for listed fish species

This section summarizes available data in a complete and clear manner. It is an important observation that there is no acceptable acute toxicity data for endangered or threatened estuarine/marine aquatic fish species.

4.3.2 Key chronic toxicity data for listed fish species

This section summarizes available data, a single study, in a complete and clear manner. It is an important observation that there is no acceptable chronic toxicity data for endangered or threatened estuarine/marine aquatic fish species.

4.3.3 Concerns about Federally listed endangered mussels

Available studies on Federally listed endangered mussels are reviewed with clarity and completeness.

4.4 Comparison of 1988 and 2016 Criteria Values

The title of this section should be "Comparison of 1988 and 2017 Criteria Values." This comparative analysis is important and useful to the overall document. The assumptions and approaches to setting the data in Table 11 (e.g., hardness at 100 mg/L) are scientifically reasonable and defendable. Statements concerning the relative level of protection, for example, brook trout, are an important qualifier in support of the MLR model approach that is the basis for criteria development.

5 References

The literature supporting the draft document represents a complete and exhaustive review of the available peer-reviewed science.

2. **Please comment on the technical approach used to derive the draft aluminum criteria; is it logical, does the science support the conclusion, and is it consistent with the protection of freshwater life from acute, chronic, and bioaccumulative effects? Please provide specific comments.**

Overall the technical approach used to derive the draft Al criteria is logical and supportive of the criteria data found in the look-up tables of Appendix K. Within the constraints of available peer-reviewed data, the technical approach is consistent with the protection of freshwater life from acute, chronic, and bioaccumulative effects. Specific comments related to a critical analysis of each section of the draft document appear above.

3. **Please comment on the data used to derive the revised criteria, including data adequacy/comprehensiveness, and the appropriateness of the data selected and/or excluded from the derivation of the draft criteria. Are there other relevant data that you are aware of that should be included? If so, please provide the reference.**

The data qualified in the derivation of the revised criteria are the result of a complete and thorough analysis through 2013. The data set is comprehensive, and acceptance or exclusion judgments or assumption are transparent, linked to formally published guidance, and explicitly discussed. Exploration of the 47 citations retrieved from a "Web of Science" database search of aluminum aquatic toxicity, for 2013 to the present did not yield any new controlled Al exposure studies. Specific comments related to the data used to derive the revised criteria of each section of the draft document appear above.

4. Are the derived criteria appropriately protective of commercially and recreationally important species, and of ecosystems overall?

Yes, the derived criteria appropriately protective of commercially and recreationally important species, and of ecosystems overall. The 2017 draft criteria look-up tables present a research-based improvement on the 1988 criteria. Specific comparative examples such as brook trout in the draft document section 4.4, offer support for this conclusion. Quantifying the effects of water chemistry (pH, DOC, and Hardness) for aluminum risk to aquatic species is an improvement to meeting the goals of the CWA from the 1988 approach, helping states, tribes, and the regulated community maintain environmental quality and healthy aquatic ecosystems.

COMMENTS SUBMITTED BY

Reviewer 4

External Peer Review of Aquatic Life Ambient Water Quality Criteria for Aluminum – 2017

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

    a) The document is very well prepared. The communication, documentation and clarity of the draft criteria is dramatically improved compared to the 1985 criteria document.

    b) One criticism though is in the reuse of some material from the 1985 document. This results in some very old references in the initial parts of this new draft criteria document. The initial sections (up to about page 12) draws heavily from the older document and could do with insertion of some of the new references which are detailed in later sections of the draft criteria document. In particular the first paragraph of section 2.2 and the paragraph right before section 2.4.1 are very much like the 1985 document

    c) There should be an aluminum chemistry section. This can be an appendix and does not necessarily need to be integrated into the text. As it is written now so many "random" aspects of aluminum chemistry are introduced as the document progresses. It would be nice to have one central place discussing aluminum chemistry in some details. Emphasizing aluminum speciation and reaction kinetics.

    d) In terms of more specific communication issues, detailed comments are included as an appendix to this review.

2. **Please comment on the technical approach used to derive the draft aluminum criteria; is it logical, does the science support the conclusion, and is it consistent with the protection of freshwater life from acute, chronic, and bioaccumulative effects? Please provide specific comments.**

Technical comments are itemized below:

a) Selection of an MLR approach should be justified. Why not a BLM approach? I have no problem with an MLR approach but it is a bit of a compromise compared to a full BLM approach so the reason for selecting this method should be given. Ideally highlighting how the MLR method is as protective as a "full BLM" would be.

    If MLR is used for practical reasons I think that is fine as a justification, provided that it is clearly shown that there is no loss of protection.

b) I think aluminum speciation is not fully addressed. In particular, solid versus dissolved Al. I am currently reviewing a manuscript by Robert Santore et al. (Windward) that does this very well (submitted to Environ. Toxicol. And Chem. "Development and application of a biotic ligand model for predicting chronic toxicity of dissolved and precipitated aluminum to aquatic organisms"). I suspect that the authors of this report are familiar with that work but if not they should contact the authors to preview a copy.

This draft criteria document over simplifies the chemistry of aluminum by lumping particulate and dissolved aluminum into a total measure of aluminum. I think this is fine – based on reading the Santore draft paper – but the draft criteria document does not convince me it is fine. Different modes of exposure (particulate versus soluble) and not even really hinted at in the draft criteria document.

c) Selection of total Al as the independent variable needs to be more fully justified (see comment above) but also it needs to be made clear if this is in fact what was done. It seems that the discussion on page 3 and 4 about the operational definitions of total and dissolved is used in the rest of the document, in particular, for the numerical criteria. This needs to be made clear.

As mentioned in the Appendix (to this review), aluminum speciation needs to be defined more clearly. Total aluminum could include particulate refractory minerals, weathering product minerals (i.e., clay such as kaolinite) or it could include reactive solids such as amorphous gibbsite which can exchange back and forth solid to dissolved phase. Should these really all be included in one “box”. This might be ok for lab studies where only amorphous gibbsite occur as solids – but what about field?

d) Kinetics of “land scale” processes compared to acute and chronic tests. The criteria is based on lab-based studies but will be implemented with real samples. In acute tests the time scales are very short but in the field longer timescales will control the total aluminum. Not geologic timescales but longer than laboratory tests.

e) Bottom-line the proposed criteria will be protective for aquatic life in an acute and chronic sense. The science is logical but needs just a bit more justification (as detailed in my points above). My personal bias is that I would prefer a BLM approach but I completely understand the practical needs that make MLR an attractive and possibly more likely to be adopted option.

3. Please comment on the data used to derive the revised criteria, including data adequacy/comprehensiveness, and the appropriateness of the data selected and/or excluded from the derivation of the draft criteria. Are there other relevant data that you are aware of that should be included? If so, please provide the reference.

a) Data seems very comprehensive and the data standards are very high. The rejected datasets are still discussed in the text so the document has a very balanced view.

b) I am not aware of any other data sets that should be included.

4. Are the derived criteria appropriately protective of commercially and recreationally important species, and of ecosystems overall?

a) Yes the document clearly demonstrated that criteria will be protective to commercially and recreationally important species and in fact to endangered species. The revised criteria are an improvement over the existing criteria.

b) In terms of ecosystem protection overall the field studies that are discussed make it pretty clear that effects were observed for higher aluminum then would be recommended by this draft criteria.

Appendix: detailed "editorial" comments.

a) Chemistry section is a bit sparse. In particular colloidal and particulate and kinetics. If you measure total Al does it also include minerals like feldspar and weathering minerals like kaolinite? Then how does that compare to freshly precipitated amorphous gibbsite?

b) A table of acronyms at the start of the document would be beneficial. Also making sure that each acronym is defined before first use (for example DOC in the executive summary (page vi) is not defined).

c) For many comparisons in the text the criteria value for water chemistry of pH 7, DOC=1mg C/L and 100 mg CaCO3 hardness is used. I realize many more possible values are reported in the look up tables at the end of the document but why were these conditions selected for comparisons in the text? (such as the comparisons to the "old" criteria document). pH 7 is not exactly a common pH and DOC of 1 mg C/L is a pretty low value.

d) Page vii there is a space missing 4 lines up from the bottom; "1985Guidelines" should be "1985 Guidelines".

e) Page 2, "… yet very rarely in the elemental state …". The use of the word "yet" is a strange choice. And I'm not sure Aluminum is ever found in its elemental state in natural systems.

f) Paragraph at the start of section 2.1 (page 2) does not mention aluminum in clays or soil or sediment. Just aluminum in rocks and that is complexed by some inorganic and organic ligands. Also, particulate (suspended) aluminum and colloidal and polymeric forms should be mentioned. These forms are referred to later in the document so they should be introduced here in the overview section.

g) Paragraph starting at the bottom of page 2 starts with ``Aluminum enters the environment …" is a strange opening. Aluminum is everywhere in the environment. The authors point out in the text how abundant aluminum is in the earth's crust. So I think the authors need to be specific here – what forms of aluminum are "entering" the environment? Does this mean dissolved aluminum? The document is written as if aluminum is like other metal contaminants and aluminum is very different.

h) Paragraph starting at the bottom of page 2 says both natural and anthropogenic sources of aluminum are responsible for a "meaningful amount found in the aquatic environment". I am not clear what meaningful amount means in this context.

i) Paragraph starting at the bottom of page 2 might be an appropriate place to mention aluminum solubility and Ksp (unless a separate section on chemistry is adopted per my suggestion above). This is an important concept in nature and a really important concept in the toxicity experiments.

j) Page 3 around the middle, aluminum is also used in wastewater treatment to remove phosphorus.

k) Page 3 around the middle of the page, it is not clear what is meant by ``diversely abundant".

l) Top of page 4, colloidal and particulate aluminum species should be clearly discussed. The current text does explain soluble speciation (i.e., complexation) but fails to recognize solid speciation. There is a large difference between a particle of feldspar or kaolinite from freshly precipitated aluminum hydroxide. Also polymeric species.

m) This is a bit of a technical question but why does the minimum detection limit for dissolved concentration differ from the minimum detection limit of total aluminum? (bottom of page 4).

n) Figure 1 gives a map of wells but there is no Figure showing a map of the surface water samples discussed.

o) Below Figure 1 on page 5 in reference to marine and estuarine systems the recent paper by Angel et al should be cited. It gives specific measured values. This paper is also useful because it points out the dynamics of aluminum in toxicity test solutions – it specifically considers the timescale and chemistry of exposure solutions. [Geochemical controls on aluminum concentrations in coastal waters. Brad M. Angel , Simon C. Apte , Graeme E. Batley and Lisa A. Golding. Environmental Chemistry 13(1) 111-118 http://dx.doi.org/10.1071/EN15029].

p) Bottom of page 6, "higher acid extractable aluminum" should be explained what this term means. Also, this highlights the need for solid (particulate) aluminum speciation measurements. There is a difference between geologic minerals and transient hydroxide precipitates.

q) Middle of page 7 is the first mention of polymeric and colloidal forms of Al and also the first mention of Al sorbed to clay. These concepts need to be mentioned sooner – again, a separate appendix on Al chemistry should be considered.

r) Bottom of page 7 there is a mention of carbonate precipitates of aluminum. I'm not sure these exist. A citation should be given.

s) Middle of page 8 is the first mention of measuring aluminum speciation (Driscoll reference). There is a large body of literature from the 1980s on aluminum speciation and although this is old literature it should be reviewed. The authors use this one Driscoll citation to downplay the usefulness of aluminum speciation methods which define fractions such as monomeric and polymeric aluminum. It is true that these methods are operationally defined but perhaps this section could include more substantial literature review before rejecting them out of hand? Lumping all particulate aluminum into a single "box" as is proposed with a single total aluminum measurement has its own limitations.

t) Sentence in the middle of the first complete paragraph on page 8 is out of place. "Aluminum toxicity is important in freshwater and marine environments due to both anthropogenic sources …" Just kind of comes out of no where.

u) Section 2.3 it is not clear if this section is referring to soluble aluminum transport or also including particulate aluminum transport.

v) Page 9 near the top of the page says that at neutral pH aluminum is nearly insoluble – this should be quantified. The Ksp of aluminum hydroxide allows clear estimation of the solubility limits of aluminum.

w) Figure 2 is a good figure but another figure showing precipitation equilibria would give a more balanced view of aluminum speciation. Also the final sentence of the figure caption is not in bold face font.

x) The list of species at the bottom of page 10, top of page 11, is a repetitive list but does not mention the same species as the previous listing of species (i.e., first paragraph of section 2.1.).

y) Are all filters created equal? The authors refer to 0.45 µm filtration throughout the document but do not specify filter media. Different media can lead to different answers. Probably beyond the level of detail this document is intended for but I mention this incase the authors want to consider.

z) The first complete paragraph on page 12 goes into more details on aluminum chemistry. These details could be included earlier in the more general discussion paragraphs. In particular time is mentioned as a variable here. This is significant in aluminum chemistry and needs to be highlighted earlier. The Angel manuscript mentioned above does a good job of showing aluminum kinetics on the timescale of toxicity testing.

aa) The mention of aluminum flocs coprecipitating nutrients in the middle paragraph of page 12 is a well known phenomena. It is in fact how chemically mediated phosphorus removal works in wastewater treatment. Also these examples (and this text) is taken from the 1985 criteria document. Some more recent examples would be better.

bb) Page 13 when the Appendices are referred to it would be nice to give the title of the appendix. That would really help the reader (I noticed this in a later section when the title was given in reference to the appendix.). Specifically here though it would be nice to include some of the numbers and not rely on the reader to go and read the Appendix to find the number for themselves.

cc) Second to last paragraph on page 13 refers to MLR models as less complex than BLM models I think I would also emphasize that MLR models are not mechanistic. They are referred to as bioavailablity models and it is semantics but since there is no mechanistic aspect to an MLR I would prefer if they were not referred to as bioavailablity models. It is true MLR models take toxicity mitigating factors into account so maybe some reference to that would be better?

dd) When MLR models are discussed on page 14 (bottom paragraph) it is not clear if total aluminum (as defined in this document) is the independent variable or if other forms of measured aluminum where tested.

ee) Last paragraph of page 14, are pH, hardness and DOC really independent variables? More independent than say pH and alkalinity but these aren't really independent variables. Harder water could have more alkalinity (hardness cations are strong base cations after all) and high DOC waters are often acidic because of organic acids.

ff) The MLR paragraph at the bottom of page 14 really emphasizes the importance of pH, DOC and hardness. These variables are mentioned in the earlier text but I think could be emphasized more strongly. So many other things like sulfate, phosphate, … are mentioned as well. The reader should be convinced that these are the best variable choices.

gg) Page 15 6 lines down, "more simple" should be "simpler"

hh) Page 15, only R^2 values are given, p values should also be given.

ii) Section 2.7.1 repeats things that have been mentioned before. It almost seems that the earlier sections drew heavily from the 1985 document and then this text was created for this current document. The end result is repetitive though. I prefer how things are presented in section 2.7.1 and I really like Figure 3.

jj) A figure similar to Figure 3 could be developed showing aluminum speciation. Or Figure 3 could include some indication of aluminum speciation.

kk) Page 30 middle paragraph argues how laboratory toxicity tests can be extrapolated to the field. This is reasonable for other metals but some caveats should be given for aluminum. The slow kinetics and the difference between geological aluminum materials suspended in natural water and the transient precipitates formed during toxicity testing should be discussed. I agree that lab-based toxicity testing is the best we can do for now but longer term field-based aluminum toxicity tests could be very useful. (such as section 4.1.3).

ll) Table 3 and elsewhere when SMAV values (or other aluminum values are referred to) it should be made clear if this is referring to total or dissolved aluminum. I think it is pretty much always total aluminum but readers may want to just skip to the tables so this should be made clear in the table footnotes. Figure 4 does specifically mention that it is total aluminum.

mm) Section 4.1.3 ends by stating that the study exceeded the minimum acceptable guidelines for tests of this duration. It is not clear what this means.

nn) Section 4.1.3 are there any field based fish studies? I know that aluminum mobilized by environmental acidification was implicated in fish kills. Are there any systematic studies in this regard? The authors repeat other acidic test results (top of page 50 for example).

oo) Throughout section 4 aluminum concentrations are often referred to (i.e., effects concentrations). Are these always total aluminum values? This applies to the summary tables at the end of the document as well – maybe some blanket statement would help avoid any ambiguity here. I am not sure if all accepted papers measured total aluminum so the authors need to clarify this.

pp) Page 59 refers to the appendix by name (Appendix J reference). This is a nice way to do this and would make the rest of the document more readable if this convention was adopted throughout.

qq) Page 62 says that MLR accounts for the bioavailability of the metal. It is semantics but I think it is overstating things to state that MLR account for bioavailablity. There is no mechanism in an MLR.

COMMENTS SUBMITTED BY

Reviewer 5

External Peer Review of Aquatic Life Ambient Water Quality Criteria for Aluminum – 2017

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

The document very clearly explained the derivation of each criterion, including the toxicity data and the water chemistry parameters used. The Multiple Linear Regression (MLR) model for determining Al criteria uses equations based on pH, hardness, and dissolved organic carbon (DOC). Therefore, the criteria would vary in a site-specific manner. The ways in which these water chemistry variables modify Al toxicity, specifically including changes in Al speciation, is very well explained.

2. **Please comment on the technical approach used to derive the draft aluminum criteria; is it logical, does the science support the conclusion, and is it consistent with the protection of freshwater life from acute, chronic, and bioaccumulative effects? Please provide specific comments.**

The technical approach used to derive the draft Al criteria is logical and scientifically supported. The MLR model has been validated using *Ceriodaphnia dubia* and *Pimephales promelas* with greater than 86% success at predicting Al toxicity values. Biaccumulation of Al is most observed in the lower trophic levels, as specifically stated in the document.

3. **Please comment on the data used to derive the revised criteria, including data adequacy/comprehensiveness, and the appropriateness of the data selected and/or excluded from the derivation of the draft criteria. Are there other relevant data that you are aware of that should be included? If so, please provide the reference.**

There were more available, acceptable data used to derive freshwater acute and chronic criteria in the 2017 Al document as compared to that used in the 1988 document.

For acute values, the 2017 Al document included 21 species (12 invertebrates, eight fish species, and one frog species) in 19 genera as opposed to the 15 species (eight invertebrates and seven fish species) in 14 genera that were included in the 1988 Al document. These data fulfilled the minimum data requirements (MDR).

For chronic values, the 2017 document included 11 species (seven invertebrates and four fish species), whereas, the 1988 document included only three species (two invertebrates and one fish species). The third family in phylum chordata is missing in the chronic database; however, "Other Data" was used to fulfill the missing MDR group. Data from a study using a wood frog was included in the chronic database for this purpose. I agree that this is the more scientifically defensible route to directly determine the final chronic value rather than using both the final acute value and the final Acute-Chronic Ratio (FACR) to estimate the final chronic value.

All of the freshwater data were normalized to a pH of 7, hardness of 100 mg $CaCO_3$/L and DOC of 1 mg/L for comparison. The criteria in the 2017 Al document are determined based on those site-specific parameters (pH, hardness, and DOC).

New acceptable, acute estuarine/marine data is available for five families representing five species in the 2017 Al document, as compared to no data before; however, the database still does not meet the minimum of eight families necessary to fulfill the MDR. One data gap identified is for estuarine/marine chronic Al toxicity, where there were no acceptable data.

The data selected to be included in the databases were, in my opinion, appropriate and thorough. For comparison purposes, studies were also discussed (data provided in Appendices) that were missing some/all water chemistry parameters (i.e., pH) necessary for inclusion in the database.

4. Are the derived criteria appropriately protective of commercially and recreationally important species, and of ecosystems overall?

The database for acute and chronic Al data in freshwater and estuarine/marine systems is comprehensive; however, data gaps have been identified. Data for endangered freshwater mussels or closely related species are lacking. For these organisms the derived criteria may not be protective. Using current available data, I believe the derived criteria will be appropriately protective of most commercially and recreationally important species and of ecosystems.