



**EPA RESPONSE TO THE
EXTERNAL PEER REVIEW
REPORT ON THE CHRONIC
TOXICITY OF ALUMINUM
TO THE CLADOCERAN,
CERIODAPHNIA DUBIA:
EXPANSION OF THE EMPIRICAL
DATABASE FOR
BIOAVAILABILITY MODELING
(2018)**

EPA RESPONSE TO THE EXTERNAL PEER REVIEW REPORT ON THE CHRONIC
TOXICITY OF ALUMINUM TO THE CLADOCERAN, *CERIODAPHNIA DUBIA*:
EXPANSION OF THE EMPIRICAL DATABASE FOR BIOAVAILABILITY MODELING
(2018)

January 2020

U.S. ENVIRONMENTAL PROTECTION AGENCY
OFFICE OF WATER
OFFICE OF SCIENCE AND TECHNOLOGY
HEALTH AND ECOLOGICAL CRITERIA DIVISION
WASHINGTON, D.C.

TABLE OF CONTENTS

1	Introduction	1
1.1	Background	1
1.2	Peer Reviewers.....	1
1.3	Review Materials Provided	1
1.4	Charge Questions	1
2	External Peer Reviewer Comments and EPA Responses, Organized by Charge Question	2
2.1	Charge Question 1	3
2.2	Charge Question 2.....	5
2.3	Charge Question 3.....	6
2.4	Charge Question 4.....	8
2.5	Charge Question 5.....	9
2.6	Charge Question 6.....	11
2.7	Charge Question 7.....	13
2.8	Charge Question 8.....	15
2.9	Charge Question 9.....	17
2.10	Charge Question 10	19
2.11	Charge Question 11	21
2.12	Charge Question 12	23
3	Additional Comments Provided	26
4	References Cited by Reviewers and EPA Responses.....	28

1 INTRODUCTION

EPA organized a contractor-led independent, external peer review of an aquatic life toxicity test report entitled “Chronic Toxicity of Aluminum to the cladoceran, *Ceriodaphnia dubia*: Expansion of the empirical database for bioavailability modeling” (OSU 2018). Oregon State University (OSU) conducted the invertebrate toxicity tests for aluminum to expand the toxicity test dataset that may be used for bioavailability model development to estimate the effects of aluminum on aquatic organisms.

The external peer review was completed on July 31, 2018. The external peer reviewers provided their independent responses to EPA’s charge questions. This report documents EPA’s response to the external peer review comments provided to EPA.

This report presents the 12 peer review charge questions and five individual external peer reviewer comments (verbatim) on the charge questions in Sections 2.1 through 2.12. Additional comments outside of the charge questions are presented in Section 3. New information (e.g., references) provided by reviewers is presented in Section 4. Each reviewer’s comments were separated by charge question into distinct topics and EPA responded to each topic individually.

1.1 BACKGROUND

Section 304(a) (1) of the Clean Water Act, 33 U.S.C. § 1314(a)(1), directs the Administrator of EPA to publish water quality criteria that accurately reflecting the latest scientific knowledge on the kind and extent of all identifiable effects on health and welfare that might be expected from the presence of pollutants in any body of water. In support of this mission, EPA is working to update water quality criteria to protect aquatic life from the potential effects of aluminum in freshwater environments. Invertebrate toxicity tests for aluminum have been conducted by Oregon State University and are yet unpublished in the peer-reviewed literature. EPA thus funded a contractor-led focused, objective evaluation of these invertebrate toxicity tests, to determine if their quality was sufficient for EPA to include them in the development of a bioavailability model to calculate the effects of aluminum on aquatic organisms under a range of water chemistry conditions.

1.2 PEER REVIEWERS

An EPA contractor identified and selected five expert external reviewers who met the technical expertise criteria provided by EPA and who had no conflict of interest in performing this review.

The EPA contractor provided reviewers with instructions, the final report, and the charge to reviewers prepared by EPA. Reviewers worked individually to develop written comments in response to the charge questions.

1.3 REVIEW MATERIALS PROVIDED

- OSU 2018 Final Report and Appendices

1.4 CHARGE QUESTIONS

1. Were an adequate number of concentrations tested to fully-characterize concentration-response and determine an accurate and scientifically-defensible chronic effect concentration (e.g., EC₂₀)?

2. Was there a sufficient number of replicates for each test concentration and control to pass statistical rigor for the type of test and test conditions?
3. Was the source, maintenance, and husbandry of test organisms well described?
4. Were the control's survival rates acceptable?
5. Were test organisms appropriately acclimated for the type of test and test water conditions to represent their chronic sensitivity under those conditions?
6. Were test endpoints and data acceptability criteria well defined and explained?
7. Was preparation of test solutions fully described and target test concentrations verified prior to testing?
8. Were manipulated test water quality variables (e.g., pH, DOC, water hardness) measured with sufficient frequency and accuracy to represent intended levels?
9. Was the frequency and accuracy of chemical concentrations measured in test solutions sufficient to represent intended exposure levels throughout the duration of the test(s)?
10. Were any anomalies in the test explained or justified with additional information or testing?
11. Do the reported test results meet or exceed expectations for use in model development for the derivation of ambient water quality criteria for the protection of aquatic life?
12. Is there any reason to be concerned with the use of the test results in the criteria derivation process?

2 EXTERNAL PEER REVIEWER COMMENTS AND EPA RESPONSES, ORGANIZED BY CHARGE QUESTION

The following tables list the charge questions submitted to the external peer reviewers, the external peer reviewers' comments regarding those questions (broken into distinct topics), and EPA's responses to the peer reviewers' comments.

2.1 CHARGE QUESTION 1

1. Were an adequate number of concentrations tested to fully-characterize concentration-response and determine an accurate and scientifically-defensible chronic effect concentration (e.g., EC₂₀)?

Reviewer	Comments	EPA Response to Comment
Reviewer 1	Yes. The test was conducted following standard US EPA chronic testing methodology according to US EPA (2002). This reference is not provided in the reference list (it should be), but presumably refers to EPA-821-R-02-013. According to this guidance, a minimum of 5 test concentrations and a control should be used in a definitive test. As each test in this study included 5 exposure concentrations and a dilution water control (p. 2-2), it is judged to be adequate for the test purpose. The range of concentrations chosen was also deemed adequate to achieve estimates of the desired effect levels for reproduction (10, 20, and 50% effect; Table 3-13). With the exception of one test in which effects on survival occurred, all test concentrations could be used to estimate reproductive effects.	Thank you for your suggestion. The reference is not in the main body of the report but is cited in Appendix A of the report.
Reviewer 2	A total of nine different tests were conducted under different pH, hardness and DOC conditions. Five total Al concentrations plus controls were generally used in the various tests. This number of concentrations is generally considered adequate.	Thank you for your comment.
Reviewer 3	Yes, 5 concentrations of Al and a negative control were used for each test. This design appeared to follow the EPA guidelines for toxicology testing with freshwater organisms. The concentrations used were low that did not result in complete mortality at the highest concentration of each test. Therefore, lethal effect concentrations (LCs) could not be calculated.	Thank you for your comment. While lethal concentrations were not observed in all tests, the chronic endpoint for reproductive effects occurred in all tests.
Reviewer 4	<p><u>Comment:</u></p> <p>In my opinion, an adequate number of concentrations were tested to allow full characterization of the concentration response and allow determination of a scientifically-defensible chronic effect concentration.</p> <p><u>Rationale:</u></p> <p>This research project evaluated the effects of multiple water quality variables on the toxicity of Aluminum (Al) to the cladoceran <i>Ceriodaphnia dubia</i>. The goal of the study was to increase the range of water quality variables under which a reasonable prediction of invertebrate toxicity could be performed under a given set of water quality variables. The test followed standard USEPA methodology (US EPA 2002). The methods included in this</p>	Thank you for your comment.

Reviewer	Comments	EPA Response to Comment
	<p>manual are referenced in Table IA, 40 CFR Part 136 regulations and, therefore, constitute approved methods for acute toxicity tests. These methods were used in the present study with modifications to address different water types and pH levels. For example, concentrations were based on previous studies shown to cause a negative impact on <i>C. dubia</i> survival and reproduction. The standard EPA protocol calls for five test concentrations and a control and this was mostly followed in the present study. For one test (Test #: Al 1185 CDC; p. 12, Appendices (page 1, Appendix B) six concentrations of Al were used, plus a treatment labeled “non pH”). This was apparently a confirmatory test for comparison to results obtained at the Chilean Mining and Metallurgy Research Center (CIMM; Santiago, Chile) and Universidad Adolfo Ibañez (UAI; Santiago, Chile) and reported in Gensemer et al. (2018) as indicated on p. 29, paragraph 3. Five concentrations is the number usually followed by most toxicity testing laboratories including those administered by the US EPA (such as the EPA facility in Cincinnati, OH with which I am familiar). This allows the present study to be compared to the results of other laboratories and have such results be incorporated into the statistical model developed by the authors. This regression model can be used to develop a scientifically defensible chronic effect concentration such as the EC20 (dose which causes a 20% change from control response of the test organisms).</p>	
Reviewer 5	<p>The study was performed following the agreed to protocol. However, one study used a 45% bisection of the test concentrations rather than the protocol specified 50% bisection. While I do not believe that this is a fatal flaw in the analysis, I believe that it does warrant a section in the report for protocol deviations (rather than as only noted in Section 2.5 [page 2-2]). This would also provide an opportunity to offer the analytical issues (as identified in Section 3.2 [page 3-4]). I also believe the authors should assess whether the analytical anomalies bias the results high, low, or neutral. This is very helpful in the use of these results. In my overall opinion, all test concentrations were sufficiently characterized to provide a meaningful and accurate description of the test results and the chronic toxicity of aluminum.</p>	<p>Thank you for your comment. Section 3.5 “Protocol Deviations and Amendments” provides a statement that the authors noted that no protocol deviations occurred during the toxicity tests which would affect the study outcomes.</p>

2.2 CHARGE QUESTION 2

2. Was there a sufficient number of replicates for each test concentration and control to pass statistical rigor for the type of test and test conditions?

Reviewer	Comments	Response to Comments
Reviewer 1	Yes. There were 10 replicate chambers for each exposure concentration and control, each containing one cladoceran. This is consistent with US EPA guidance (EPA-821-R-02-013).	Thank you for your comment.
Reviewer 2	Yes. Ten replicates per treatment is adequate.	Thank you for your comment.
Reviewer 3	Yes, 10 replicates per treatment were usually used for this type of test. The report (section 2.9) did not clearly say the number of organisms used per replicate chamber.	Thank you for your comment. The number of replicates, as stated in Section 2.5 is ten.
Reviewer 4	<p><u>Comment:</u></p> <p>Yes, the number of replicates (10 per AI treatment concentration and 10 in the non-treated control) was sufficient to allow sufficient statistical rigor for a <i>C. dubia</i> chronic toxicity evaluation under the stated test conditions.</p> <p><u>Rationale:</u></p> <p>Ten replicates of each toxicant concentration and the control is the number recommended by the US EPA (2002). This number of replicates is used by most toxicity testing laboratories, allowing comparison of the results of the present study with previous (and likely future) results from other laboratories. Statistical dogma suggests that ≈ 30 replicates is the optimal number when evaluating biological data. However, in this (and most other toxicity testing laboratories) the test conditions were carefully controlled, using 1) moderately hard diluent water prepared in-house (please see question 7 below), 2) environmental chambers controlled for pH and light regimen, and 3) neonates that were all less than 24 hours old. All of these conditions will serve to reduce variability in organism response to exposure, which will support rigorous statistical testing using 10 replicates.</p>	Thank you for your comment.
Reviewer 5	The number of replicates (10) and test concentrations (minimally 5 plus a control) were standard with in ecotoxicity testing with <i>Ceriodaphnia dubia</i> . These are acceptable.	Thank you for your comment.

2.3 CHARGE QUESTION 3

3. Was the source, maintenance, and husbandry of test organisms well described?

Reviewer	Comments	Response to Comments
Reviewer 1	Partially. The source of the organisms was well described. They were obtained from in-house cultures that had been maintained for over 10 years and originally obtained from Aquatic BioSystems (Fort Collins, CO, USA) (p. 2-1). Maintenance and husbandry of the test organisms were not described in the report, although the authors did indicate that they conducted monthly tests with a reference toxicant (NaCl) to confirm that the organisms were in good condition (p. 2-1).	Thank you for your comment. EPA confirmed with the authors that <i>C. dubia</i> were cultured in-house on brood boards according to standard methodology (USEPA 2002). Cultures underwent 100% water renewals (moderately hard reconstituted water) five times per week, were fed daily, and reproduction was tracked to ensure health acceptability for testing.
Reviewer 2	Not particularly. This section was remarkably brief and lacking details of animal performance for the reference toxicant tests. The reporting of volumes of algal suspensions used for feeding are not useful unless cell densities are reported.	Thank you for your comment. EPA confirmed with the authors that <i>C. dubia</i> were cultured in-house according to standard methodology (USEPA 2002). Each test chamber was fed 0.3 mL of an algal (<i>Pseudokirchneriella subcapitata</i>) and yeast/trout chow/cereal leaf (YTC) suspension (1:1) at test initiation (prior to test organism introduction) and once daily prior to water renewal. The algal density was 3×10^7 cells/mL used in the food suspension.
Reviewer 3	Organisms were originally from Aquatic Biosystems and cultured at OSU for more than 10 years. Organisms were cultured in moderately hard water. Other environmental conditions and maintenance procedures were not described, such as temperature, photoperiod (light:dark hours), food, feeding rates, biomass/water volume, water change, etc.	Thank you for your comment. EPA confirmed with the authors that <i>C. dubia</i> were cultured in-house according to standard methodology (USEPA 2002). By following this protocol all maintenance and husbandry conditions are deemed to be appropriate.
Reviewer 4	<p><u>Comment:</u></p> <p>No, an adequate description of the source, maintenance, and husbandry of the <i>C. daphnia</i> test organism was not provided.</p> <p><u>Rationale:</u></p> <p>In the report, section 2.3.2 SOURCE, the authors state that the <24 hour old neonates were obtained from in-house cultures which have been maintained successfully at the Aquatic Toxicology laboratory at Oregon State University (Corvallis) for >10 years. In Appendix A, section 2.2 and 2.3, feeding diet and feeding regimen during toxicity testing were described.</p>	Thank you for your comment. EPA confirmed with the authors that <i>C. dubia</i> were cultured in-house according to standard methodology (USEPA 2002). By following this protocol all maintenance and husbandry conditions are deemed to be appropriate.

Reviewer	Comments	Response to Comments
	<p>However, nowhere that I could find in the report was it explicitly stated that the test organisms were cultured and maintained under these same conditions. I believe this is an oversight in reporting, not a failure of procedure, and this oversight can be readily remedied by the authors by providing the missing information. Husbandry of the test organisms during culture and testing as described appeared to be adequate.</p>	
Reviewer 5	<p>The description of the test animals was adequately presented in the report. Reference toxicant testing was regularly performed as part of the quality assurance program.</p>	<p>Thank you for your comment.</p>

2.4 CHARGE QUESTION 4

4. Were the control's survival rates acceptable?

Reviewer	Comments	Response to Comments
Reviewer 1	Yes. The authors report that in all tests, control acceptability criteria (> 80 % survival and > 60% surviving females having 15 or more neonates) were met (p. 3-14). These fulfill the criteria for test acceptability outlined in EPA-821-R-02-013.	Thank you for your comment.
Reviewer 2	The average number of neonates/female in controls ranged from 22 to 37 with 42.5 reported from a “concurrent control”. The test with the poor control reproductive output (A11199 CDC) should not be used.	According to USEPA 2002, “In <i>Ceriodaphnia dubia</i> controls, 60% or more of the surviving females must have produced their third brood in 7±1 days, and the number of young per surviving female must be 15 or greater.” Since the control group in A1 1199 CDC met these conditions, EPA disagrees that it should not be used.
Reviewer 3	The survival of the control organisms of each test was 100%. This meets the test acceptability criteria of the test method (80-100%).	Thank you for your comment.
Reviewer 4	<p><u>Response:</u></p> <p>Yes, it appears that the survival rate of <i>C. dubia</i> used in the control (no aluminum) treatments met the accepted survival rate for this type of toxicity testing.</p> <p><u>Rationale:</u></p> <p>The standard methodology as developed by the US EPA (1982) calls for at least 80% survival of the control test organisms for the test to be considered valid. On p. 29, paragraph 2, the authors state that, in all tests, control acceptability criteria (> 80 % survival and > 60% surviving females having 15 or more neonates) were met. Table 3-12 (p. 30 of report) and Appendix D Raw Data both indicate that control survival was uniformly 100%, clearly meeting the EPA (2002) control standard for test acceptability.</p>	Thank you for your comment.
Reviewer 5	Control survival rates were acceptable.	Thank you for your comment.

2.5 CHARGE QUESTION 5

5. Were test organisms appropriately acclimated for the type of test and test water conditions to represent their chronic sensitivity under those conditions?

Reviewer	Comments	Response to Comments
Reviewer 1	Yes, as far as hardness is concerned. Organisms cultured under standard conditions (100 mg/L as CaCO ₃) were used in the moderately hard water tests (120 mg/L as CaCO ₃). Organisms were acclimated to the soft (60 mg/L as CaCO ₃) and hard water (250 and 400 mg/L as CaCO ₃) conditions for multiple generations (i.e., over two months), and survival and reproduction were reported to be excellent (p. 2-2). As far as indicated in the report, there was no acclimation for different pH (tested range: 6.3 – 8.8; standard culture at 7.8-8.0) or DOC (tested range: 1-14 mg/L; standard culture unknown) conditions.	Thank you for your comment. EPA confirmed with the authors that <i>C. dubia</i> cultures were not acclimated to pH or DOC test conditions. However all control exposures met the data quality criteria according to USEPA (2002). Additionally, OSU lab data quality conditions (Appendix A Section 4.9) were also met in all tests.
Reviewer 2	The report only mentions acclimation of cultures to different hardness levels, but not pH and DOC or buffers.	Thank you for your comment. EPA confirmed with the authors that <i>C. dubia</i> cultures were not acclimated to pH or DOC test conditions. However all control exposures met the data quality criteria according to USEPA (2002). Additional, OSU lab data quality conditions (Appendix A Section 4.9) were also met in all tests.
Reviewer 3	Yes, the acclimation of the organisms to the hardness of test waters (250 and 400 mg/L as CaCO ₃) for multiple generations and over more than 2 months should be adequate.	Thank you for your comment.
Reviewer 4	<p><u>Comment:</u></p> <p>It would appear that the <i>C. dubia</i> used in these toxicity tests were appropriately acclimated for the stated test type and described test water conditions at the time the chronic toxicity testing was performed.</p> <p><u>Rationale:</u></p> <p>The <i>C. dubia</i> used for the present study were reported (Section 2.3.4 ACCLIMATION p. 2-2;) as being cultured at the Ohio State University AquaTox laboratory, in a “moderately hard” reconstituted water that was prepared as detailed in standard USEPA methods (USEPA 2002). This diluent was reported to have a measured hardness of 100 mg/L as CaCO₃ and pH of 7.8 – 8.0, p. 2-2). All acclimated cultures for all of the toxicity tests were successfully maintained in their respective laboratory water for multiple generations (2+ months).</p>	Thank you for your comment.

Reviewer	Comments	Response to Comments
	Organism survival and reproduction were reported as excellent and organism health was maintained over the period of acclimation. Note: In section 2.3.4, ACCLIMATION is erroneously labeled, in section 2.3.2 SOURCE, as section 2.4.3).	
Reviewer 5	I was quite impressed with the acclimation process used in this study. In many instances, researchers do not go to the length of details used for the acclimation protocol performed in this study. The researches should be commended on this practice.	Thank you for your comment.

2.6 CHARGE QUESTION 6

6. Were test endpoints and data acceptability criteria well defined and explained?

Reviewer	Comments	Response to Comments
Reviewer 1	Yes. Test endpoints included NOEC and LOEC for survival and reproduction (if data met assumptions of normality and homogeneity), as well as effect concentrations (i.e., LC10/LC20/LC50 for survival and ECx10/EC20/EC50 for reproduction). The authors mentioned that any concentrations for which significant survival effects occurred were not included in the analysis of reproductive effects. Acceptability criteria for temperature (25 +/- 2°C) and dissolved oxygen (>60%) were indicated (p. 3-1) and met. The authors documented the range of measured pH and DOC measurements (p. 3-1), but did not indicate what was considered an acceptable range (Note: there are no acceptability criteria defined in EPA guidance EPA-821-R-02-013 for these parameters). The authors report that Al concentrations among all quality control samples were within acceptability criteria of 85-115%, whereas the standard addition recoveries were within acceptability criteria of 116-102% with a few exceptions (n=7) (p. 3-4).	Thank you for your comment.
Reviewer 2	Data acceptability criteria were not explicitly discussed but the software packages used to assess data have built in tests for homogeneity of variance, etc. Control performance should be explicitly discussed however.	Thank you for your comment. While control performance is not discussed, all control information (i.e., survival, reproduction) is reported.
Reviewer 3	Determination of NOEC, LOEC, LCs, and ECs were described in the statistical analysis section. However, a separate section to define the measured endpoints of the test is recommended.	Section 2.10.2 does state that live and dead counts (i.e., survival), and the number of young (i.e., reproduction) was counted daily.
Reviewer 4	<p><u>Comment:</u></p> <p>Test endpoints were sufficiently defined and explained. Data acceptability criteria were not well defined and explained.</p> <p><u>Rationale:</u></p> <p>Although rather brief, the authors state under section 2.10.2 BIOLOGICAL MONITORING p. 2-5 that observations of live and dead organisms were conducted on a daily basis from initiation to termination, and that the numbers of young were counted daily. This is sufficient to understand the test endpoints used, but it would be useful to know under what conditions the organisms were observed (light table? microscope? visual inspection only? time of day?) and how the test organisms were determined to be either dead or alive. Data acceptability criteria for this project were not offered. Most uses of data acceptance criteria involve some</p>	Appendix A, Section 4.9 of the report does discuss data acceptability. Data analysis followed the statistical decision tree/flow chart according to methodology described in USEPA 2002 and is detailed in Appendix D of the study report.

Reviewer	Comments	Response to Comments
	<p>type of comparison among the data groups to determine if variability falls within a predetermined acceptable range but the predetermined acceptable range for normality and homogeneity for these tests were not stated by the authors. The only data acceptability evaluation offered was that if the data met the assumptions of normality and homogeneity, the NOEC and LOEC were estimated using an analysis of variance to compare (p. 2-6, the authors use “$p = 0.05$” as the threshold for accepting a significant effect but the correct variable here would be “$\alpha = 0.05$”). There was no explanation offered on how the data were handled when the data did not meet assumptions of normality and homogeneity. If all data met those assumptions it should be stated in the report.</p>	
Reviewer 5	<p>The test endpoints and data acceptability criteria were well defined and explained in the text. I would like the authors to further evaluate the pH 6.3, hardness 60, DOC 2 treatment as to the appropriateness of the results. The 529 Al treatment had slightly better reproduction average than the next lower concentration (264.5 Al treatment). While I know that this sometimes happens, the control through the 529 Al treatment (represents 5 of the treatments) ranged in reproduction from 32.6 to 26.0 neonates (Table 3-12, page 3-15). This represents a wide range of treatment concentrations, with minimal change in neonate average production. I couldn't further evaluate whether there was something in this test that might explain this effect? All other tests looked adequate and were well defined and explained.</p>	<p>The concentration-response data for test Al 1185 CDC does not appear to be abnormal. It is not uncommon to see response data where the higher test concentrations may vary in the measured test endpoint (i.e., in this case the average number of neonates). Furthermore this test was replicated based on previous work (Gensemer et al. 2018) and the t-test analysis between the reported EC_{20}s had no significant difference (see Section 3.3 of the report). Additionally while the reported NOEC-LOEC of the test was 264.5-529.0 $\mu\text{g/L}$ total aluminum with very similar average reproduction (25.8 vs. 26.0), the reported EC_{20} for the test was 828.6 $\mu\text{g/L}$. This demonstrates that while that 529 Al treatment may be significant as the LOEC, the 20% reduction in reproduction is modeled to occur at a higher aluminum concentration.</p>

2.7 CHARGE QUESTION 7

7. Was preparation of test solutions fully described and target test concentrations verified prior to testing?

Reviewer	Comments	Response to Comments
Reviewer 1	Yes. Preparation of the test solutions is described in detail at the top of p. 2-3. Analytical samples from each treatment were collected for total Al and dissolved Al (<45 µm) analysis from newly prepared waters (after the 3-hr equilibrium period) at test initiation, during the tests, and from a composite of replicates at test termination (p. 2-5). Total Al concentrations prior to addition to test chambers were between 93 and 115% of nominal spiked concentrations, with four measurements outside of this range (with measurements of 75, 117, 120, and 130% of nominal). Total Al concentrations in test solutions measured in the replicate chambers at the end of the tests were more variable and the authors explained that it was more difficult to obtain homogeneous samples from the chambers and that these measurements were therefore less reliable (p. 3-4). In addition, dissolved Al concentrations were found to be highly variable, ranging from 0.1 to 111% of total Al. The authors explained that this was expected because the majority of solutions were well above solubility limits. There was some variability in the background levels of Al in the control water, presumably due to differences in natural organic matter.	Thank you for your comment.
Reviewer 2	Test solutions that were aged 3 hours were taken on day 0 for both total and dissolved Al concentrations. All tests except Al 1185 CDC also had test solutions measured on days 3 and 6. The Al1185 tests did not have a day 3 sample reported.	Thank you for your comment. EPA agrees that this is the only test where aluminum concentrations were not measured on Day 3.
Reviewer 3	Yes, the preparation of the test solutions was fully described. The measured total Al were closed to the nominal concentrations. Usually stock concentrations are verified prior to use. However, it was not mentioned in the report.	Thank you for your comment. EPA confirmed with the authors that stock solutions were not measured. However concentrations were measured in the test chambers at appropriate intervals to verify appropriate dosing.
Reviewer 4	<u>Comment:</u> Yes, the methods of test solution preparation were fully described. The target test concentrations (both of the treatment chemical, aluminum, and the evaluated water quality variables) appears to have been extensively tested and verified during the study but there is no indication that this occurred prior to the study. <u>Rationale:</u> It appears that great attention was paid to chemical analyses in this project. The report	Thank you for your comment.

Reviewer	Comments	Response to Comments
	<p>provides an extensive description of the analytical methodology used, including composition of sampling containers, commercial source, preparation, and storage of test substance (p. 1-2), preparation and distribution of test concentrations (p. 2-1), method of pH control (p. 2-3), timing of collection, treatment and holding time of samples after collection, calibration of analytical instrumentation, use of blanks (p. 2-5), chain of custody documentation for samples analyzed, and data handling and storage of results. Analytical samples for each treatment were obtained from the newly prepared and equilibrated (3 hrs) test concentration prior to the start of the test but there is no indication that concentrations were verified before testing. Samples were taken for chemical analysis just prior to introduction of test organisms to the test chambers. According to Section 2.11 ANALYTICAL CONFIRMATION samples were analyzed for total and dissolved (defined as sample water that has passed through a 0.45 µM filter) using a Spectro Arcos ICP-OE according to US EPA Method 200.7. with quality control samples and spiked samples to determine % recovery. Appendix A (Protocol) indicates that this was a standard procedure for metal analysis to determine Al concentrations using an Inductively Coupled Plasma with either Optical Emission Spectrometry or Mass Spectrometry (p.7). The raw data for these analyses are provided in APPENDIX B – Metals Analytical Data and comprise the majority of the 405 pages of the appendices. Spiked samples were used to determine accuracy of analyses by calculating metal recovery and were shown to be within acceptable analytical limits.</p>	
Reviewer 5	The test solutions were well described and were sufficiently verified prior to testing.	Thank you for your comment.

2.8 CHARGE QUESTION 8

8. Were manipulated test water quality variables (e.g., pH, DOC, water hardness) measured with sufficient frequency and accuracy to represent intended levels?

Reviewer	Comments	Response to Comments
Reviewer 1	Yes. Temperature, pH, conductivity, and dissolved oxygen (DO) were measured in each concentration at test initiation, once daily, and at test termination. Hardness, alkalinity, ammonia, and total residual chlorine (TRC) were measured in the control water of each test at test initiation (p. 2-4). Other parameters (i.e., Calcium, magnesium, sodium, potassium, chloride, sulfate, cations, anions, and DOC) were measured by outside labs using accepted methods, but it is not entirely clear from the report how often these measurements were done.	Thank you for your comment. EPA confirmed with the authors that analytes and DOC were measured in the dilution water at test initiation and are reported in Section 3.1 and Appendix C.
Reviewer 2	Temperature, pH, conductivity and DO were measured daily. Details of the frequency of verification for DOC concentrations were not found.	<p>According to Appendix A, Section 4.5 and verified by the authors to EPA:</p> <ol style="list-style-type: none"> 1. Hardness, alkalinity, total ammonia, and total residual chlorine were measured in the dilution water control at test initiation. 2. A sample of each control/dilution water (prior to addition of buffer or pH adjustment) was sent to an outside analytical laboratory for analysis of calcium, magnesium, sodium, potassium, chloride, sulfate, and dissolved organic carbon at test initiation. 3. Dissolved oxygen, temperature, conductivity, and pH were measured and recorded daily in the new waters of each treatment. Dissolved oxygen, temperature, and pH was measured daily in the old waters of each treatment.
Reviewer 3	The procedure for controlling test water quality, such as pH was clearly described. It was conducted carefully. Measurement of pH, DO, conductivity, and temperature were sufficient. The measured values represent the target values. However, hardness and alkalinity were measured only in the control water of each test at test initiation. This is weak rather than sufficient. These parameters are usually measured at least in control, the lowest and highest treatment concentrations at test initiation and termination to make sure the addition of toxicant into the test treatments does not change the water quality of the test water.	Thank you for your comment. EPA determined that measured hardness and alkalinity would not be expected to vary greatly during a test exposure and thus measurement only at the beginning of this test would be sufficient.

Reviewer	Comments	Response to Comments
Reviewer 4	<p><u>Comment:</u></p> <p>Yes, it appears that the manipulated test water quality variables (pH, hardness, and DOC; incorrectly called parameters in the report) were measured with sufficient frequency and accuracy to represent intended levels and allow incorporation into an updated predictive model of aluminum toxicity under varying water quality conditions.</p> <p><u>Rationale:</u></p> <p>Under Section 2.10 TEST MONITORING, subsection 2.10.1 WATER QUALITY the authors indicate that pH, hardness, and dissolved organic carbon (DOC) were measured during toxicity testing. pH was measured in each concentration at test initiation, once daily, and at test termination using a HACH HQ3od pH meter. Water hardness was measured in the control water of each test at test initiation using a colorimetric titration method following Standard Methods 2340B/C (APHA 2012). DOC was measured by an outside laboratory (Oregon State University Cooperative Chemical Analytical Laboratory (Corvallis, OR, USA) using a Shimadzu TOC-VCNS total organic carbon analyzer (Shimadzu Scientific Instruments, Columbia, Maryland) following a Combustion method ((Standard Methods 5310B APHA 2012). All of the analytical instrumentation used are of sufficient quality to provide accurate, reproducible data results. Both water hardness and DOC would not be expected to vary greatly during a test exposure and thus measurement only at the beginning of the test would be sufficient. The mean and raw values for the data from these analyses are presented in Tables 3-1 and 3-1 in the report, and the Appendices C and D, respectively.</p>	Thank you for your comment.
Reviewer 5	Water quality variables were adequately manipulated. I believe that the use of the buffers as well as CO2 headspace was warranted for keeping these tight conditions with regards to the challenging pH parameter.	Thank you for your comment.

2.9 CHARGE QUESTION 9

9. Was the frequency and accuracy of chemical concentrations measured in test solutions sufficient to represent intended exposure levels throughout the duration of the test(s)?

Reviewer	Comments	Response to Comments
Reviewer 1	Yes. Al concentrations were measured at test initiation and once during each test, and from a composite of replicates at test termination. Samples were analyzed for total and dissolved (<45 µm) Al using standard US EPA methods. Blanks and quality control samples were also run (p. 2-5).	Thank you for your comment.
Reviewer 2	Generally, yes for total Al concentrations. Test A11199 CDC reported considerable variation in total Al concentrations among days for a given nominal concentration. Dissolved Al concentrations were all over the map and incredibly inconsistent.	Thank you for your comment. EPA notes that total aluminum concentrations will be used for determining toxicity effect concentrations, not dissolved concentrations.
Reviewer 3	Total and dissolved Al were measured in new and old waters at test initiation and termination and during the test period. This is sufficient. In addition, the measured concentrations of total Al were close to the nominal concentrations, presenting an accuracy of preparation and measurement of the test solutions. However, the measured dissolved Al concentrations were far away from the total concentrations. This weakens the confidence of this study.	Thank you for your comment. EPA notes that total aluminum concentrations will be used for determining toxicity effect concentrations, not dissolved concentrations. The results from this study are similar to tests from other laboratories.
Reviewer 4	<p><u>Comment:</u></p> <p>The frequency and accuracy of chemical concentrations of the non-manipulated water quality variables measured in test solutions appeared to be sufficient to represent intended exposure levels throughout the duration of the tests.</p> <p><u>Rationale:</u></p> <p>Temperature, conductivity, and dissolved oxygen (DO) were measured in each concentration at test initiation, once daily from one of the test chambers at each concentration of aluminum, and at test termination. This frequency is standard protocol for water quality variables that may exhibit some variation in concentration over the duration of a test exposure. They were also measured in the renewal water prior to changing out the adult daphnids. These were reported to be calibrated prior to starting a measurement in Appendix A Protocol following Oregon State University Aquatic Toxicology Laboratory Standard Operating Procedures. These were measured using calibrated digital instrumentation as described in Section 2.4 DILUTION WATERS and reported in Table 2-1. Alkalinity, ammonia, and total residual</p>	Thank you for your comment.

Reviewer	Comments	Response to Comments
	chlorine (TRC), were measured in the control water of each test at test initiation using digital meters. Temperature was measured with a standard laboratory thermometer. Test solution pH was measured using a HACH (Loveland, CO, USA) HQ30d pH meter. These methods of measurement usually provide highly accurate and reproducible results sufficient to ensure determination of intended exposure levels.	
Reviewer 5	I believe that the frequency and accuracy of the chemical concentrations were sufficiently performed through the duration of the test. (see next charge question for additional input to this charge question).	Thank you for your comment.

2.10 CHARGE QUESTION 10

10. Were any anomalies in the test explained or justified with additional information or testing?

Reviewer	Comments	Response to Comments
Reviewer 1	Yes. The only anomalies were variability in the total Al concentrations measured in the chambers at the end of the test and in dissolved Al measurements. The authors explained these results (see answer to question 7). There was one test in which significant effects on reproduction occurred, and the authors addressed this by omitting the affected test concentrations from the reproductive effects analysis.	Thank you for your comment. Since the highest tested concentration in Test Al 1196R CDC exhibited a significant effect on survival, EPA (2002) recommends not including the concentration in the analysis of reproduction effects. The authors did not use this test in the reproductive effects analysis.
Reviewer 2	No. Anomalies (see control reproduction in Al1199 CDC) were not explained or justified with additional testing.	The control group in Al 1199 CDC met data acceptability conditions outlined in USEPA 2002. EPA disagrees that it is an anomaly.
Reviewer 3	Not really, except for the procedure for controlling the pH of the test waters.	Thank you for your comment, additional details about testing were provided in Appendix A.
Reviewer 4	<p><u>Comment:</u></p> <p>The relatively few anomalous data were explained/justified without the need for additional data or testing.</p> <p><u>Rationale:</u></p> <p>In Section 3. RESULTS AND CONCLUSIONS, subsection 3.1 TEST CONDITIONS the authors observed some variability in measured DOC. This has been observed in their testing laboratory previously and they believe it is due to using multiple batches of Suwanee Natural Organic Matter (NOM) which shows some variation in % DOC among batches. They also acknowledge that observed differences may be due to variability in analytical measurements. Because the DOC concentrations are reported as measured and not nominal, they should be acceptable for this project's goals of incorporation and expansion into the previously established predictive model. pH was maintained within 0.2 SU of the target pH in freshly prepared ("new") solutions after the equilibrium period. However, in some studies, an increase in pH occurred in the "old" waters (pH up to 0.3 – 0.4 SU above the "new" waters) between each 24-hr water renewal. Both the use of the buffer to control pH, and also slightly adjusting the CO2 atmosphere, limited observed pH drift within limits that allowed incorporation of mean pH values into the predictive model. Mean conductivity values</p>	Thank you for your comment.

Reviewer	Comments	Response to Comments
	<p>remained consistent over the 24-hr period between water renewals. But in certain cases the range in conductivity was wide, primarily in the higher DOC tests (Table 3-2, p. 3-2). This is likely due to the higher DOC and cannot be eliminated as a (slightly) confounding factor. The authors also speculate that some increase in conductivity in the “old” water may be due to addition of food to the test chambers. The authors observed some variability in total Al recovery from “old” solutions and suggest this was primarily due to the difficulty in removing the entire homogenized aliquot because it has been altered during final enumeration of neonates by removing the organisms during counting (to prevent double counting). They believe this may have resulted in the accidental removal of precipitates from the non-homogeneous solution, potentially resulting in a misrepresentation of the entire fraction in the test chamber. Therefore, they feel that the “new” solutions are the most appropriate measurements for average exposure determination of Al. When comparing total Al to dissolved Al in the same sample, dissolved Al was much more variable than total Al, ranging from 0.1 to 111% of total Al. The author’s expected this as the majority of solutions were well above solubility limits. The observed trend in dissolved concentrations was that higher percentages of dissolved/total were apparent in the lower exposure concentrations and percentages decreased as total Al increased. A few dissolved Al measurements were elevated and unexpected (and did not correspond to total dissolved Al samples from the identical concentration). The authors feel this is most likely associated with breaching of the 0.45 µM filter by insoluble Al clogging the filter and requiring additional pressure on the filter to obtain sufficient sample volume. The authors addressed this by keeping pressure on the filter at a minimum. Because (unlike most metals) the dissolved/free ion species of Al has relatively less effect on toxicity than the Al hydroxide species at circumneutral pH (6–8), and Al concentration–toxicity relationships correspond to total Al (Cardwell et al., 2017), total Al was incorporated into the predictive model.</p>	
Reviewer 5	<p>I believe that the anomalies observed during testing were well explained and the justification was sufficiently presented and plausible (page 3-4). However, these anomalies can be classified as deviations from protocol. I think this report would benefit from a section in the report presenting these identified anomalies and also the researchers should attempt to assess whether these anomalies potentially bias the results high, low, or neutral. I think that this section will help strengthen the report and further demonstrate a transparent process.</p>	<p>Thank you for your comment. Section 3.5 “Protocol Deviations and Amendments” provides a statement that the authors noted that no protocol deviations occurred during the toxicity tests which would affect the study outcomes.</p>

2.11 CHARGE QUESTION 11

11. Do the reported test results meet or exceed expectations for use in model development for the derivation of ambient water quality criteria for the protection of aquatic life?

Reviewer	Comments	Response to Comments
Reviewer 1	As far as I can tell. The authors followed standard US EPA guidance for conducting chronic toxicity tests with <i>Ceriodaphnia dubia</i> with some modifications to account for specific water types and to achieve effective pH control. The general US EPA criteria for test design and test acceptability were met, and the authors applied principles consistent with Good Laboratory Practice (GLP). Although documentation on culture maintenance and husbandry were not included in the report, the fact that the laboratory has been culturing this species successfully for over a decade and that control organisms showed acceptable performance, give little cause for concern related to maintenance and husbandry.	Thank you for your comment. EPA confirmed with the authors that <i>C. dubia</i> were cultured in-house on brood boards according to standard methodology (USEPA 2002). Cultures underwent 100% water renewals (moderately hard reconstituted water) five times per week, were fed daily, and reproduction was tracked to ensure health acceptability for testing.
Reviewer 2	Without seeing the entire package of how water chemistry parameters are going to be used to model both dissolved and particulate/precipitate concentrations and link these to toxicity, it is impossible to answer this question. The use of total recoverable Al as a descriptor for toxicity seems to run counter to BLM principles. Without direct evidence and mechanistic understanding of how Al precipitates are toxic to daphnids, it is going to be very difficult to convince people that the dissolved concentrations reported in these tests can be predictive of toxicity.	EPA agrees that dissolved aluminum concentrations are not appropriate for use in criteria derivation. EPA notes that total aluminum for toxicity test effect concentrations will be used in model development. The use of total recoverable aluminum does not run counter to BLM principles, in fact, the aluminum BLM also uses total aluminum concentrations (Santore et al. 2018).
Reviewer 3	This study covered a wide range of water quality parameters that are suitable for BLM development and calibration. Reproductive results showed concentration-response relationships that are useful for determination of effect concentrations based on total concentration basis but not for dissolved concentration basis.	Thank you for your comment.
Reviewer 4	<p><u>Comment:</u></p> <p>The reported test results do meet or exceed expectations for use in model development for the derivation of ambient water quality criteria for the protection of aquatic life.</p> <p><u>Rationale:</u></p> <p>This study appears to have been carefully planned and executed and seems to compare well with the results of other similar studies and laboratories. For instance, the authors compared their (EC10/EC20 with 95% confidence interval results with Gensemer et al. (2018) using a</p>	Thank you for your comment.

Reviewer	Comments	Response to Comments
	<p>one-sample paired-comparison t-test and found that the values were not statistically different between laboratories. The authors also endeavored to make the study results appropriate for inclusion in previously developed models. For example, the Biotic Ligand Model (BLM) uses Ca and Mg (in mg/L) as input variables to calculate hardness values and the multiple linear regression (MLR) for the Al toxicity prediction model on which the Water Quality Criterion is based uses hardness (as mg/L CaCO₃). The calculated hardness values in Table 3-1 were used in the MLR analysis to maintain consistency between model input values derived from other studies. The results of this study are directly applicable to the EPA-developed WQC because that value is derived using an MLR model based on a site's pH, DOC, and hardness (EPA 2017). These water quality variables are precisely those evaluated by manipulation in this study and thus the datasets can be included as part of the model refinement effort.</p>	
Reviewer 5	<p>I believe that these test results will strengthen the aluminum water quality criteria, however, I am not sure the results were meant to meet all of this charge question the way it was described. I am confident that these results will be very useful to the application of the BLM model and MLR model, however, the results presented in the report do not provide the details to make this assessment.</p>	Thank you for your comment.

2.12 CHARGE QUESTION 12

12. Is there any reason to be concerned with the use of the test results in the criteria derivation process?

Reviewer	Comments	Response to Comments
Reviewer 1	<p>Three of the tests had very steep concentration-response relationships and were flagged by the TRAP model as being useful for exploratory analysis only due to an inadequate number of partial effects. It is difficult to judge what the effect of including these test results in the Biotic Ligand Model and Multiple Linear Regression Model would be. Certainly the models could be run with and without these data and a judgement made as to whether their precision was sufficient for inclusion in the model refinements.</p>	<p>According to the authors, “As shown in Table 3-12 and Appendix D, modeling of reproduction data resulted in qualifiers identified by the TRAP model, in addition to undeterminable confidence intervals. These were identified due to the lack of partial effects concentrations in the datasets and associated steep slopes between the concentration with no effect (NOEC) and the concentration with a reproductive effect (LOEC). According to available TRAP guidance, datasets identified as “exploratory” should be examined on a case-by-case basis to assess the confidence around the result based upon the exposure-response relationship. As the three tests did show a quite significant reproductive effect at the highest concentration, it is believed that these datasets can be included as part of the model refinement effort.” EPA concurs with this assessment.</p>

Reviewer	Comments	Response to Comments
Reviewer 2	<p>The complexity of Al chemistry makes this very challenging. We do not appear to be closer to understanding the effects of dissolved Al and its speciation on <i>C. dubia</i> as a result of these studies, because the dissolved concentrations are not tractable due to precipitation issues. The uncertainty of the kinetics of precipitate formation and the effects of those precipitates on different forms of aquatic life bring a large amount of uncertainty into the equation. How does a 3 hour equilibration period in the laboratory (with high buffer concentrations) translate to animal exposures in nature? It is interesting that EPA is willing to consider Al solid phases in toxicity characterization, but generally refuses to consider the effects of dietary exposures of metals – which are known to cause deleterious effects in aquatic life.</p> <p>Thus, there appears to be considerable uncertainty with respect to both dissolved and particulate Al forms. It would appear that both dissolved criteria based on BLB type principles and particulate criteria would be needed – or that a considerably large uncertainty factor would be applied to a total Al measurement.</p>	<p>EPA used total aluminum for toxicity test effect concentrations to account for these precipitated forms and their potential presence in the environment. Total aluminum effects are used in criteria derivation; dissolved aluminum concentrations are not appropriate for use. The aluminum BLM also uses total aluminum (Santore et al. 2018). EPA assumes organisms are exposed to both dissolved and particulate aluminum in the treatment concentrations and in the environment.</p>
Reviewer 3	<p>The concern about this study is the measured dissolved Al concentrations. Dissolved Al concentrations were totally off the total concentrations, especially at high concentrations. A few examples are the measured dissolved concentrations were below the detection limit or 7 or 45 ug/L at the total Al concentrations of 5000 and 10000 ug/L (Table 3-6, Test Al 1205CDC), or 80-217 ug/L at the total Al concentrations of 30012000ug/L (Table 3-8, Test Al 1198CDC). Dissolved metal concentration has been using for evaluating metal bioavailability, especially using the BLM approach. Given that said, I don't know how the BLM can be applied to the dissolved concentration data set in this report.</p>	<p>EPA agrees that dissolved aluminum concentrations are not appropriate for use in aluminum criteria derivation and will use total aluminum for toxicity test effect concentrations. Dissolved aluminum concentrations have not been used for evaluating bioavailability. The aluminum BLM also uses total aluminum (Santore et al. 2018).</p> <p>In numerous studies where both dissolved and total concentrations were reported, the relationship between total and dissolved aluminum varies. When the total aluminum concentrations increase, the dissolved aluminum concentrations do not increase as expected. The total aluminum concentration is used because it includes dissolved and particulate aluminum and the dissolved aluminum fraction varies and is not reliable.</p>

Reviewer	Comments	Response to Comments
Reviewer 4	<p><u>Comment:</u></p> <p>I do not believe there is any significant reason to be concerned with using the test results from this report in the water quality criterion derivation process.</p> <p><u>Rationale:</u></p> <p>The main goal of this project was to increase understanding of the bioavailability and toxicity of Al to aquatic organisms. To reach this goal, the main objectives of this project were 1) to quantify the effects of water quality on Al toxicity and 2) to use the results to develop a bioavailability-based model to predict Al toxicity across a wider range of certain water quality variables (specifically pH, hardness, and dissolved organic carbon). I believe this study has achieved these objectives and has increased the applicable range of previous predictive models used to derive an Al WQC. The expansion included increasing pH from 8.10 up to 8.70, hardness (as CaCO₃) up to 428 mg/L from 123 mg/L, and dissolved organic carbon from 4.0 mg/L up to 12.30 mg/L. Comparison of the current model predicted effect concentrations with observed effect concentrations, for water types outside the previous range of model development, suggests very good predictive capabilities of this new model (Table 3 – 13) and thus may be confidently used in the water quality criterion derivation process.</p> <p>In terms of future Al toxicity testing with the goal of developing a new WQC, I would like to see the following suggestions to be considered:</p> <ol style="list-style-type: none"> 1) Al toxicity tests performed with sodium aluminum sulfate (probably as NaAl(SO₄)₂·12H₂O. This would help address the massive problem with sulfuric acid-derived acid mine drainage (AMD), of which elevated Al is often a constituent. There are more than 500,000 abandoned and inactive mines in 32 states and AMD has degraded more than 8,000 miles of streams in Appalachia alone. 2) I would have preferred to see pH controlled in a flow-thru set-up, perhaps using a digital controller (Grippe 1997) rather than by buffers, which introduce a possibly confounding effect on the results. A flow-through protocol has not yet been developed for fecundity of <i>Ceriodaphnia dubia</i> but development of such a protocol would significantly increase environmental realism. 	Thank you for your comment.
Reviewer 5	I have not [<i>sic</i>] concerns with regards to the use of the test results in the criteria derivation process.	Thank you for your comment.

3 ADDITIONAL COMMENTS PROVIDED

Reviewer	Comments	Response to Comments
Reviewer 4	<p><u>Suggestions to authors</u></p> <p>-Authors frequently use the phrase “In order to”. Reducing this phrase to simply “To” will convey the same meaning with fewer words, enhancing the goal of preparing scientific prose that exhibits clarity and brevity.</p>	<p>This comment is on the toxicity report for invertebrates that was conducted by OSU, not on an EPA document. The authors of the OSU report can access these comments on our website.</p>
Reviewer 4	<p>-In Part 3.3 BIOLOGICAL RESULTS, paragraph 3 the authors state “The results were quite comparable to those reported in Gensemer et al. (2018) (EC10/EC20 with 95% confidence intervals: 504.4 (226 – 1126) µg/L total Al and 631.3 (362 1101) µg/L total Al, respectively). A one sample t-test was performed and the values were not statistically different between laboratories. Because the comparison was between two independent populations of test results (ration of EC10/EC20 a two – sample t-test may have been more appropriate.</p>	<p>This comment is on the toxicity report for invertebrates that was conducted by OSU, not on an EPA document. The authors of the OSU report can access these comments on our website.</p>
Reviewer 4	<p>-Table 3-12. Some of the data are set off by both asterisks and bold-type. In the text it is stated that this indicates significant differences. I suggest including an explanation of what the bold-face and asterisks denote in the table heading, rather than the text, so the reader does not have to go searching in the text to determine the meaning of these highlighted results.</p>	<p>This comment is on the toxicity report for invertebrates that was conducted by OSU, not on an EPA document. The authors of the OSU report can access these comments on our website.</p>
Reviewer 5	<p>General Comments:</p> <p>I found this report to be well written and supported using the information in the appendices. I support the use of these results for the derivation of the aluminum ambient water quality criteria.</p>	<p>This comment is on the toxicity report for invertebrates that was conducted by OSU, not on an EPA document. The authors of the OSU report can access these comments on our website.</p>
Reviewer 5	<p>Specific comments from reviewer:</p> <ul style="list-style-type: none"> • While the <i>Ceriodaphnia</i> tests followed the protocols as presented in Appendix A, the test as described by US EPA is a 3-brood test. However as specified in the protocol, the tests were carried out with 7-days of exposure (and potentially extended another day if 3-broods did not occur) rather than as a 3-brood test. Thus, the average neonates were considerably higher than normal 3-brood tests. I think that this should be mentioned in the results. Also, some of the variability during testing might also be explained because the protocol did not specify that the neonates are <24 hours old (from an 8-hour window). While the researchers followed the protocol, these two issues are outside of the US EPA methods that were reported in the Methods and Materials section (page 2-1). 	<p>Thank you for comment. While the protocol mentions this caveat, the raw data in Appendix D verifies that all tests were indeed seven-day three-brood tests. Additionally, both the protocol in Appendix A and the final report states that the neonates are less than 24 hours old.</p>

Reviewer	Comments	Response to Comments
Reviewer 5	<ul style="list-style-type: none"> • What was the normality of the dilute NaOH and HCl? (Section 2.5, page 2-3) 	<p>Thank you for your comment. EPA confirmed with the authors that the acids/bases used in the studies are reported as molarity in the appendices for each study (Appendix D). The molarity of the acids and bases used for pH adjustment were: 0.01 M, 0.1 M, 1 M, 5 M HCl and 0.1 M, 1 M, 5 M NaOH.</p>
Reviewer 5	<ul style="list-style-type: none"> • Section 2.8 it should be pH rather than all capital letters (page 2-3). 	<p>This comment is on the toxicity report for invertebrates that was conducted by OSU, not on an EPA document. The authors of the OSU report can access these comments on our website.</p>
Reviewer 5	<ul style="list-style-type: none"> • Good spike response, however, I think the dissolved Al observation needs its own paragraph. It is buried in the middle of the second paragraph on page 3-4. 	<p>This comment is on the toxicity report for invertebrates that was conducted by OSU, not on an EPA document. The authors of the OSU report can access these comments on our website. No change is needed.</p>
Reviewer 5	<ul style="list-style-type: none"> • The report states that there was no protocol deviations and amendments, however, there were several deviations that were noted in the text (i.e., 45% bisections rather than 50% bisections). This section needs revised as well as I recommend, as stated above, the researchers should assess whether the deviations bias the results potentially high, low, or neutral. 	<p>This comment is on the toxicity report for invertebrates that was conducted by OSU, not on an EPA document. Section 3.5 “Protocol Deviations and Amendments” provides a statement that the authors noted that no protocol deviations occurred during the toxicity tests which would affect the study outcomes. The authors of the OSU report can access these comments on our website.</p>

4 REFERENCES CITED BY REVIEWERS AND EPA RESPONSES

American Public Health Association (APHA). 2012. Standard Methods for the Examination of Water and Wastewater, 22nd edition. Washington, D.C.

Cardwell, A.S., W.J. Adams, R.W. Gensemer, E. Nordheim, R.C. Santore, A.C. Ryan and W.A. Stubblefield. 2017. Chronic toxicity of aluminum, at a pH of 6, to freshwater organisms: empirical data for the development of international regulatory standards/criteria. *Environ. Toxicol. Chem.* 37: 36-48.

Gensemer, R., J. Gondek, P. Rodriguez, J.J. Arbildua, W.A. Stubblefield, A.S. Cardwell, R.C. Santore, A. Ryan, W.J. Adams and E. Nordheim. 2018. Evaluating the effects of pH, hardness, and dissolved organic carbon on the toxicity of aluminum to freshwater aquatic organisms under circumneutral conditions. *Environ. Toxicol. Chem.* 37: 49-60.

Grippio, R.S. 1997. A gravity-based system for controlling pH in flow-through aquatic toxicity experiments. *Environ. Technol.* 18: 763-768.

Santore, R., A.C. Ryan, F. Kroglund, P.H. Rodriguez, W.A. Stubblefield, A.S. Cardwell, W.J. Adams, E. Nordheim. 2018. Development and application of a biotic ligand model for predicting the chronic toxicity of dissolved and precipitated aluminum to aquatic organisms. *Environ. Toxicol. Chem.* 37:70-79.

USEPA 2002. USEPA. 2002. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms. Fourth edition. Office of Water, U.S. Environmental Protection Agency, Washington, DC 20460. EPA-821-R-02-013.

USEPA 2017. United States Environmental Protection Agency. Fact Sheet: Draft Aquatic Life Ambient Water Quality Criteria for Aluminum in Freshwaters Office of Water EPA 820-F-17-002 July 2017.