BEFORE THE
UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

COMMENTS OF THE NORTH AMERICAN METALS COUNCIL-SELENIUM WORK GROUP IN RESPONSE TO EPA'S DRAFT NATIONAL RECOMMENDED AQUATIC LIFE CRITERION FOR THE POLLUTANT SELENIUM

External Peer Review Draft Aquatic Life Ambient Water Quality Criterion for Selenium -- Freshwater 2014 (79 Fed. Reg. 27601) (May 14, 2014)	)))))	EPA–HQ–OW–2004–0019

William J. Adams, Ph.D.
Chairman, North American Metals Council
Selenium Work Group
2200 Pennsylvania Avenue, N.W.
Suite 100W
Washington, D.C. 20037

Kathleen M. Roberts
Executive Director
North American Metals Council
2200 Pennsylvania Avenue, N.W.
Suite 100W
Washington, D.C. 20037

Peter Chapman, Ph.D.
North American Metals Council
Selenium Work Group
2200 Pennsylvania Avenue, N.W.
Suite 100W
Washington, D.C. 20037

Dated: July 28, 2014

EXECUTIVE SUMMARY

The North American Metals Council (NAMC) submits these comments in response to the U.S. Environmental Protection Agency (EPA) request for input on the 2014 Draft Selenium Criterion Document. NAMC is an unincorporated, not-for-profit organization serving as a collective voice for North American metals producers and users. NAMC has been a leading voice for the metals industry on science and policy-based issues affecting metals. Our organization has worked closely with U.S. federal and international agencies to address risk assessment issues that are unique to metals at various stages of their lifecycle -- sourcing, production, engineering, use, recycling, and recovery. We advocate policy based on good science.

This document comprises integrated comments provided by individual Members and Associates of the North American Metals Council-Selenium Work Group (NAMC-SWG). The NAMC-SWG is engaged in technical research on issues pertaining to selenium. Activities include the development of water quality tissue-based standards for selenium, the implementation of such standards, development of effects thresholds, and the identification of analytical methods pertinent to such standards. As part of its ongoing efforts, the NAMC-SWG develops papers on these topics and shares them publicly.

NAMC commends EPA on key technical improvements to the proposed document, including: the focus on chronic not acute effects; the recognition that reproductive effects are of greater ecological concern and significance than non-reproductive effects; the reliance on diet as the primary pathway of selenium exposure for both invertebrates and vertebrates; the decision, supported by the fact that the winter stress syndrome has not been shown to occur in field studies and not to include overwinter survival or juvenile survival in the development of the proposed criterion; and, the provision of the option for development of site-specific criteria based on the principle that toxicity of selenium in aquatic systems is highly dependent upon site-specific factors, including food web structure and hydrology. We are highly supportive of the tissue-based approach and believe this is the most credible and scientific approach to assessing potential environmental effects from selenium and protecting aquatic resources in the future.

We urge EPA and the peer reviewers to consider the appended report recently completed by the NAMC-SWG, *Selenium Partitioning between Water and Fish Tissue in Freshwater Systems: Development of Water-based Selenium Screening Guidelines*, in the review process. The report contains new information that fills existing data gaps on the translation of tissue-based criteria to water-based selenium screening levels.

We also highlight the need for a tiered approach beginning with screening based on water selenium concentrations then proceeding, if necessary, to generic tissue concentrations and to site-specific tissue concentrations. We also note the need for a clear definition of lentic and lotic waters.

NAMC is concerned that the lentic water criteria concentrations and the tissue criteria concentrations in the Draft Selenium Criterion Document are not technically defensible

and are demonstrably overly conservative. We provide evidence for higher but still environmentally protective concentrations that would not unnecessarily expend limited regulatory resources to the detriment of genuine environmental issues, nor unduly penalize human industrial or other activities. We are concerned that the low lentic criteria concentrations will result in a serious misallocation of resources, thereby reducing rather than enhancing the nation's ability to address environmental problems.

We are also concerned, relative to addressing genuine environmental problems nationally, that the tissue criterion is inappropriately expressed as "never to be exceeded" on an instantaneous basis, which is inconsistent with EPA Guidelines. We recommend that, in accord with EPA Guidelines, "instantaneous" be replaced with "seasonal average," and "never to be exceeded" be replaced with "not to be exceeded more than once in three years on average" applicable to the seasonal average concentration.

Finally, while we thank EPA for the extension of the 30-day public comment period to 60 days, we note that this extension came only hours before the original deadline for public comment and several weeks after we had been told that no extension was possible. We respectfully request that the planned second public comment time period be extended *a priori* to at least 60 days to allow adequate time for public review and comment.

TABLE OF CONTENTS

FIGURES

TABLES

1.0 INTRODUCTION

This document, submitted by the North American Metals Council (NAMC), comprises integrated comments provided by individual Members and Associates of the North American Metals Council-Selenium Work Group (NAMC-SWG). It is provided in response to the U.S. Environmental Protection Agency's (EPA) May 14, 2014, request for public comment on a draft updated national recommended aquatic life criterion for the pollutant selenium.[1]

We understand that there will be a second 30-day opportunity to comment on the draft criterion, following the external peer review, and that this second comment period on the Draft Selenium Criterion Document is planned for late 2014. NAMC intends to provide comments during this second opportunity. We respectfully request that this second public review comment period be extended to 60 days ahead of time. If this first opportunity for comment had been longer from the start, as requested by several organizations, including NAMC, we would likely have had additional comments.

We thank EPA for the opportunity to provide the comments below. We applaud this approach and we look forward to the eventual adoption of a technically defensible selenium aquatic life criterion that is appropriately, but not unnecessarily protective.

2.0 KEY TECHNICAL IMPROVEMENTS IN THE PROPOSED CRITERION DOCUMENT

We are pleased to highlight key technical improvements in the current proposed criteria in comparison with previous criteria. The focus on chronic as opposed to acute effects is appropriate and technically defensible, and is supported by extensive scientific evidence, as provided in the Draft Selenium Criterion Document. Similarly, we fully agree with EPA that reproductive effects, linked to the magnitude of fish egg-ovary selenium concentrations, are of greater ecological concern and provide a more reliable basis for the criterion than non-reproductive endpoints (*e.g.,* survivorship, growth). Again, this is supported by extensive scientific evidence, as provided in the Draft Selenium Criterion Document.

We strongly support EPA's decision to not use juvenile survival, in particular juvenile overwinter survival, as an endpoint in developing the proposed criterion; again, the scientific evidence strongly supports this decision. The juvenile bluegill overwinter survival testing conducted by Lemly (1993) to demonstrate increased overwinter mortality due to selenium (the "winter stress syndrome" -- Lemly 1996) is technically questionable. With only one treatment, there is no supporting information regarding a dose-response relationship and a partial response with only one treatment, although possible, is surprising. More importantly, the relatively high selenium toxicity in Lemly's (1993) study was not found by McIntyre *et al*. (2008). Although the McIntyre *et al*. (2008) study has been criticized for a possible slight

[1] EPA Office of Water, Office Science and Technology, "External Peer Review Draft Aquatic Life Ambient Water Quality Criterion for Selenium – Freshwater 2014" (May 2014) (Draft Selenium Criterion Document).

difference in light-dark test conditions, this relatively minor difference from Lemly's (1993) study should not have obviated replicating that study's findings of relatively high selenium toxicity if, in fact, the Lemly (1993) study results were robust and repeatable. The McIntyre *et al.* (2008) study should be viewed as the more reliable result because of: the greater number of concentrations tested; the more realistic exposure pathway (yeast to oligochaetes to fish); and the initial exposure period at a warm temperature. The latter two improvements over the Lemly (1993) experimental design were recommended by John Besser of the U.S. Geological Survey.

The critical aspect of Lemly's (1993, 1996) winter stress syndrome hypothesis is increased metabolism resulting in an energy deficit as fish rely on stored energy to survive winter months due to low food availability. In fact, fish do not generally rely on stored energy over the winter months; active feeding occurs (Sogard and Olla, 2000; McCollum *et al.*, 2003; Biro *et al.*, 2004; Parrish *et al*., 2004; Eckmann, 2004; Bennett and Janz, 2007a,b).

Field studies have provided no support for Lemly's (1993, 1996) hypothesis of a winter stress syndrome related to selenium exposures. Hermanutz *et al.* (1992, 1996) exposed bluegills to selenium in outdoor experimental streams over winter but did not report increased overwinter mortality. Aspects of the winter stress syndrome hypothesis have been investigated in field studies of juvenile fish inhabiting areas receiving complex metal mine effluents containing elevated selenium concentrations (Bennett and Janz, 2007a,b; Kelly and Janz, 2008; Drieger *et al*., 2009). Support for the winter stress syndrome hypothesis would have come from decreased growth and energy storage over winter, more so with elevated selenium concentrations than in reference areas. The opposite was found to occur in juvenile northern pike (*Esox lucius*), burbot (*Lota lota*), fathead minnows (*Pimephales promelas*), creek chubs (*Semotilus atromaculatus*), and white suckers (*Catostomus commersoni*). Slimy sculpins (*Cottus cognatus*) exhibited changes in whole body triglycerides consistent with the winter stress syndrome hypothesis, but this occurred at both sites with elevated selenium and reference sites, thus also providing no support for the winter stress hypothesis related to elevated selenium concentrations.

It is misleading for the draft criteria document to average together the Lemly (1993) and McIntyre *et al.* (2008) results to create a *Lepomis* non-reproductive, cold season genus mean chronic value (GMCV), against which the document compares its reproduction-based criterion. A *Lepomis* non-reproductive, cold season GMCV should not be provided; rather, it should be acknowledged that uncertainty remains regarding juvenile overwinter sensitivity while noting that the weight of field and laboratory evidence indicates that the reproductive endpoint is the critical endpoint.

Finally, we fully support the option for the development of site-specific criteria based on appropriate scientific studies, where appropriate and practical. As noted in the Draft Selenium Criterion Document and in the extensive references provided with that document, the fate and effects of selenium introduced to aquatic environments is complex and there will certainly be cases where the generic national criteria are unnecessarily overprotective. Cases of naturally elevated selenium concentrations in water have been documented, as have elevated selenium concentrations in fish without adverse effects (Chapman *et al*., 2010).

3.0 CONSTRUCTIVE CRITICISMS

3.1 Need for Tiered Approach

The national criterion proposed for selenium in fresh waters has four-parts:[2]

1. The concentration of selenium in the eggs or ovaries of fish does not exceed 15.2 mg/kg, dry weight;

2. The concentration of selenium (a) in whole-body of fish does not exceed 8.1 mg/kg dry weight, or (b) in muscle tissue of fish (skinless, boneless fillet) does not exceed 11.8 mg/kg dry weight;

3. The 30-day average concentration of selenium in water does not exceed 4.8 µg/L in lotic (flowing) waters and 1.3 µg/L in lentic (standing) waters more than once in three years on average; and,

4. The intermittent concentration of selenium in either a lentic or lotic water, as appropriate, does not exceed an intermittent exposure calculated value:

$$WQC_{int} = \frac{WQC_{30\text{-}day} - C_{bkground}\,(1 - f_{int})}{f_{int}}$$

Research has shown for quite some time that chronic toxicity from selenium cannot be determined solely by exposure to selenium in the water column (Canton and Van Derveer, 1997). Toll *et al.* (2005) and Brix *et al.* (2005) proposed the use of tissue residues to establish a site-specific water quality standard for selenium. NAMC supports the use of an egg or ovary tissue value as the criterion.

As described above, EPA has proposed a four-part criterion for selenium. The proposal does provide that "(egg/ovary) overrides any whole-body, muscle, or water column elements when fish egg/ovary concentrations are measured." The four-part criterion as proposed, however, add complexity and potential confusion in regard to implementation. NAMC recommends that EPA make it clear that the official criterion is the egg-ovary value.

It is appropriate and, in fact, necessary for EPA to specify a tiered (step-wise) approach to applying selenium criteria in aquatic systems. Thus, NAMC further recommends that EPA use a tiered approach in which the water column selenium lentic or lotic value is a screening value. Exceedance of the screening value would trigger monitoring of the fish community to determine status in relation to the criterion, possibly followed by development of a site-specific tissue benchmark.

---

[2] Draft Selenium Criterion Document at 96-97.

In this regard, we recommend the use of the following conceptual tiered (step-wise) system:

1. The water selenium concentration threshold (lotic or lentic as appropriate, *see* Section 3.3) should be used as an initial screening value (Tier 1);

2. Exceedance of Tier 1 would indicate a potential risk, triggering a more robust line of evidence, specifically the generic fish egg-ovary selenium concentration to further evaluate potential risk (Tier 2); and,

3. Where appropriate, in particular due to confounding or modifying site conditions and/or limited or seasonal species distribution, it may also be necessary for a site-specific fish egg-ovary to be derived and applied at a site in question to provide a final evaluation of potential risk (Tier 3).

The tiered approach described above applies step-wise lines of evidence (tiers) with increasing certainty, when thresholds are exceeded. EPA has approved such an approach for the State of Kentucky (USEPA, 2013a). Moreover, this approach provides users with an opportunity to develop practical, cost-effective, and minimally invasive monitoring programs (*i.e.,* reduce the need to sample and analyze fish tissue, if water can be monitored), while also ensuring environmental protection.

3.2 Additional Recent Scientific Document for Consideration

We recognize that it is not always possible for a document such as the proposed selenium criterion document to incorporate and integrate all of the emerging literature at the time of release, particularly if there is a large volume of research and publications being generated in the field. We note, however, there is a specific, recently-completed document that needs to be considered by EPA, and integrated into the next version of the selenium criterion document. Specifically, the NAMC-SWG has undertaken a three-year research effort, the product of which is a recently-completed final report entitled *Selenium Partitioning between Water and Fish Tissue in Freshwater Systems: Development of Water-based Selenium Screening Guidelines*."[3] Peer-reviewed publication of this report in an international scientific journal is anticipated in the near future.

In addition to other new information, DeForest *et al.* derives and recommends technically- and statistically-defensible lentic and lotic water selenium thresholds, based on regression relationships using tissue thresholds previously determined using a species sensitivity distribution (SSD) from literature values for reproductive effects on fish, due to dietary exposure

[3] DeForest *et. al.* "Selenium Partitioning between Water and Fish Tissue in Freshwater Systems: Development of Water-based Selenium Screening Guidelines" (May 2014). *See* http://www.namc.org/docs/Selenium%20Integrated%20Report%20-%20Final%20(2014-05-20).pdf.

to selenium. We urge EPA and its peer reviewers to consider the appended report and incorporate its findings in the next revision of the selenium water criterion document.

We note that the draft EPA document mentions that selenate (SeO_4^{2-}) in the water column is taken up only slowly by bacteria, especially if competition with sulfate is involved.[4] There was no further attempt to quantify this sulfate parameter as an important moderator of selenium toxicity. The NAMC-SWG research studies reviewed in the appended report modeled and parameterized the influence of sulphate on selenium toxicity. We encourage EPA to consider applying this sulphate modification model to selenium water screening thresholds.

3.3 Clear Definition of Lentic and Lotic Waters

The Draft Selenium Criterion Document appropriately differentiates lentic and lotic waters relative to potential selenium bioavailability and effects. The definition of lentic and lotic is not clear, however. We strongly recommend clarity and completeness in this definition as there is great potential for confusion and misapplication of criteria if water bodies are misidentified. Sources of our concern are further detailed below.

3.3.1 Definition of Residence Times for Lentic and Lotic Water Bodies

EPA identifies in the Draft Selenium Criterion Document that the build-up of potentially reactive forms of selenium in aquatic environments is higher in environments where water residence times are extended.[5] This was the premise in developing a separate water criterion for lentic and lotic environments. While we agree with EPA on the need for different selenium criteria based on the aquatic system conditions, EPA needs to further clarify or parameterize this distinction between lentic and lotic environments.

In the Draft Selenium Criterion Document, EPA identifies environments that have extended or long "water residence times" as wetlands, estuaries, lakes, and reservoirs.[6] The definition of extended or long "water residence times," however, is not described in the Draft Selenium Criterion Document beyond water body type labels (lake, reservoir, etc.). If residence time is to be used to categorize aquatic environments, then guidance on residence times must be provided. There was no attempt to provide any technical guidance or correlation analysis on actual residence times in these water bodies and their enrichment factors (EF). NAMC recommends that such be provided. This is an area where the scientific literature is lacking and there may be a need for some site assessments that accompany selenium bioconcentration (enrichment) factors.

---

4 Draft Selenium Criterion Document at 14.

5 *Id.* at 15.

6 *Id.* at 15, 16, and 82.

3.3.2 Statistical Grouping of Lentic and Lotic Studies

In Figure 9 of the Draft Selenium Criterion Document, different water bodies are grouped by "residence times" (with no residence time actually provided): (1) lakes and reservoirs; (2) ponds and marshes; (3) rivers; and, (4) streams, creeks, drains, and washes. Non-parametric statistics were conducted on these four groups and no significant differences were found between the water body groupings and EFs except between the lakes and reservoirs group (group 1) and the streams, creeks, drains, and washes group (group 4). The other two groups that were not significantly different than lakes and reservoirs were allocated between each of the two groups significantly different than each other (groups 1 and 4), and then re-labeled overall as "lentic" and "lotic" in Figure 10 of the Draft Selenium Criterion Document. Ponds and marshes (group 2) were grouped with lakes and reservoirs (group 1) and classified as "lentic." Rivers (group 3) were grouped with streams, creeks, drains, and washes (group 4) and classified as "lotic" in Figure 10 of the Draft Selenium Criterion Document.

This is an inappropriate use of statistics. Statistics cannot be applied continually until a significant result is obtained on desired groupings as in the case here of the selenium lentic and lotic groupings. If all statistical combinations were studied and presented, this approach could be acceptable, but this was not done. Two groups not statistically significant from each other cannot be divided up and allocated to other groups arbitrarily without an *a priori* prescribed statistical design and evidence that they belong to the two larger groups.

NAMC recommends that EPA appropriately and fully categorize and define lentic and lotic water bodies. It is clear from a science-based viewpoint that the key is the EF. From a regulatory viewpoint, there is an issue to resolve around lentic and lotic definitions in terms of the water concentration used to trigger tissue investigations. We would be pleased to discuss this further with the Agency.

3.3.3 Use and Definition of Estuaries in Freshwater Criteria

EPA includes estuaries among aquatic environments having extended or long water residence times.[7] Estuaries and lakes, which the EPA links in the Draft Selenium Criterion Document, are very different from each other, however. Estuaries, in contrast to lakes, are not wholly freshwater environments. Estuaries were not included in EPA's analysis of residence times and EFs in Section 4.5.2 of the Draft Selenium Criterion Document. NAMC recommends further clarification of the use and definition of estuaries in the Draft Selenium Criterion Document. Specifically, NAMC recommends that, since the draft criteria are based on freshwater data, EPA should clearly state that they do not apply to estuaries or other transitional waters (*e.g.*, coastal lagoons).

---

[7] *Id.* at 15, 16, and 82.

3.4 Too Low Lentic Water Criterion

As detailed below, the proposed lentic water criterion is unnecessarily overly conservative. A higher criterion would provide environmental protection without unduly penalizing human industrial or other activities, and without unnecessarily expending limited regulatory resources to the detriment of genuine environmental issues.

3.4.1 Influence of Sampling Error on EPA's Calculated Sensitivity of the Water Criterion

We have examined the Draft Selenium Criterion Document's novel application of the type of statistics that are often applied to medical diagnostic tests. We find its explanation in Appendix H and its application in Section 7.2.2 to be interesting and believe the concepts might have other useful applications in environmental analysis. Although we endorse the concept that the measurement of water concentrations can be viewed as a diagnostic test for the possibility that tissue concentrations are elevated, we point out one significant problem -- as few as two water samples were often coupled with a tissue concentration measurement to ascertain whether exceedance of the tissue criterion was accompanied by an exceedance of the water criterion (thereby determining diagnostic test sensitivity).

We recognize the limitations in the data available for this analysis of binary statistics, and we recognize that EPA has incorporated adjustments into the analysis in an attempt to estimate from as few as two samples whether the 30-day, once-in-three-year concentration would exceed the water criterion if more samples had been taken. Nevertheless, the problem remains that the use of two samples involves a great deal of sampling error. In real-world site decision-making, two samples do not generally serve as the basis for an assessment. Relative to how site monitoring is actually done, using only two samples might be viewed as a rather haphazard application of the water-measurement diagnostic test. As with any diagnostic test, haphazard application increases random noise, and the influence of random noise on sensitivity is not unbiased. As is apparent, random noise depresses sensitivity, increases uncertainty, and results in more conservatism since randomness is not predictive of anything. As a consequence of the resulting depression in sensitivity caused by the high level of sampling error in EPA's analysis, we believe EPA has mistakenly set the water concentrations unnecessarily low in order to achieve high sensitivity.

3.4.2 Concerns Regarding "False Alarms" Caused by Setting the Criteria Too Low

Beyond the above problem, we are concerned about the high number of false positives that are likely to result from setting the lentic water criteria concentrations as low as EPA has proposed. In particular, we are concerned about the poor Positive Predictive Values (PPV) of the water criteria. As EPA notes in Appendix H of the Draft Selenium Criterion Document, given that the water criterion is exceeded, PPV represents the probability that the tissue criterion is exceeded. On page H-4, EPA correctly presents the standard equation for adjusting PPV for the difference in prevalence of tissue criteria exceedances in (a) the high-risk site population represented by Tables 18 and 19, and (b) the waters of the U.S. as a whole, as

intended to be represented by EPA's National Rivers and Streams Assessment (NRSA; USEPA, 2013b):

$$PPV = Sens \cdot P / ((Sens \cdot P) + (1 - Spec)(1 - P))$$

where *Sens* is sensitivity, *Spec* is specificity, and *P* is the nationwide prevalence of waters with fish-tissue criteria exceedances.

The Draft Selenium Criterion Document does not show an application of the above equation. Nor does it present the NRSA results in terms of the prevalence of fish muscle selenium concentrations above the muscle criterion. Nevertheless, we estimate that less than 1% of U.S. waters have fish muscle selenium concentrations exceeding EPA's muscle criterion.

In such case, applying the above EPA equation using EPA's tabulated sensitivity and specificity for lentic and lotic waters, we calculate PPV to be less than 2% when applying the water criteria to all waters of the U.S. This means that more than 98% of water criteria exceedances would be false alarms. That is, they would not be accompanied by tissue criterion exceedances. This calculation can be checked against the online calculator at http://vassarstats.net/clin2.html.

We are extremely concerned by this apparent high rate of false alarms. State pollution control agency budgets are fixed. Each time there is a false alarm, state government resources that had been allocated to solving genuine environmental problems must be reallocated to resolve the false alarm, even if that only means overseeing a tissue monitoring study performed by the discharger and arriving at a conclusion after examining the results. Since we are supportive of the mission of the state pollution control agencies and of EPA, we are concerned that setting the water criteria too low will result in a serious misallocation of resources, thereby reducing rather than enhancing the nation's ability to address environmental problems.

3.4.3 30-Day Duration for Water Concentrations

We commend EPA for evaluating data relevant to setting the appropriate averaging period for the water criterion. With the recent ammonia criteria document also evaluating such data, we are hopeful that EPA is setting a precedent to evaluate the relevant underlying data and derive a pollutant-specific averaging period in every future criterion document.

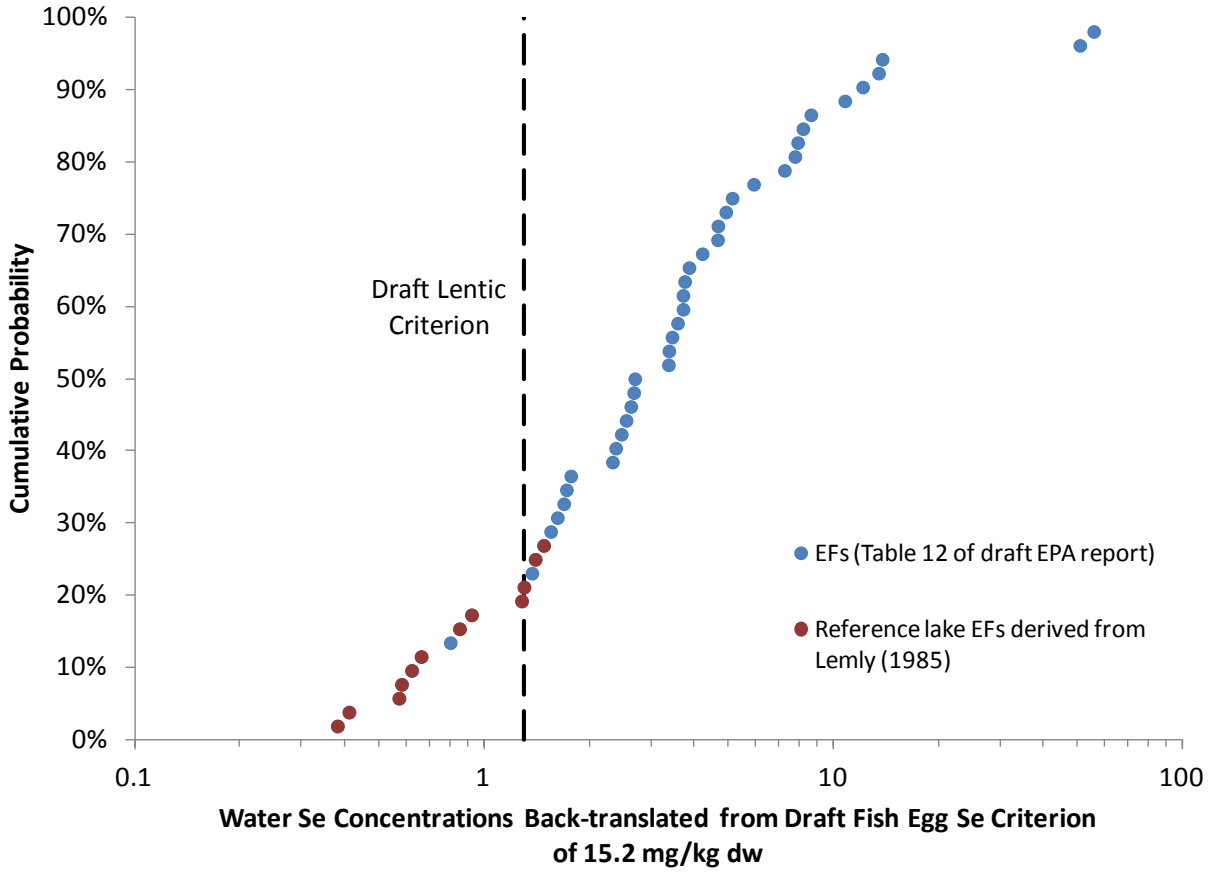
We find the analysis that EPA has done to be appropriate but question why the averaging period is not 60 days rather than 30 days. EPA has presented a characteristic time of 60 days. USEPA (1995) indicates that the averaging period should equal the characteristic time. We cannot find a reason why the averaging period should be shortened to one-half the characteristic time, or 30 days. Such shortening of the averaging period to 30 days appears to be arbitrary.

Because EPA's characteristic time derivation incorporated the environmentally conservative assumption of instantaneous kinetics at the sediment and primary producer level, we believe that a 60-day averaging period would be highly protective.

3.4.4 Underlying Database Should Be Closely Examined

The draft waterborne selenium criterion of 1.3 µg/L for lentic waters is strongly influenced by a small number of uncertain EF derived from a single study. This draft criterion was derived based on translation of the draft fish egg selenium criterion of 15.2 mg/kg dry weight (dw) back to a water selenium concentration, using EFs, egg-to-whole body or muscle conversion factors (CF), and composited trophic transfer factors (TTF) for fish and their prey. A total of 51 back-translated water selenium concentrations were derived for lentic waters, which ranged from 0.38 to 55.6 µg/L (Table 12 in the Draft Selenium Criterion Report). Of these, 18 of the 26 lowest were derived from data presented in Lemly (1985). Further, of the 11 back-calculated water selenium concentrations that were ≤1.3 µg/L, ten were from Lemly (1985) (Figure 1). Accordingly, this single study has a profound influence on the draft selenium criterion of 1.3 µg/L for lentic waters, and is cause for concern.

Figure 1. Cumulative distribution of water selenium concentrations back-translated from draft fish egg selenium criterion of 15.2 mg/kg dw (from Table 12 of the Draft Selenium Criterion Document). Red values highlight the values derived from the data for two reference lakes reported in Lemly (1985).



3.4.4.1 Uncertainty in the Selenium Enrichment Factors Derived from Lemly (1985)

The 18 values derived from Lemly (1985) were from three reservoirs (six fish species each). The lowest values were based on data collected from Badin Lake, an uncontaminated lake with a water selenium concentration of 0.32 µg/L (Lemly, 1985), for which an EF of 8.54 L/g (8,540 L/kg) was derived.

Tables showing derivation of the EFs do not appear to be included in the Draft Selenium Criterion Document, however, in Appendix C, the selenium concentrations in algae and sediment were reported as 8.20 and 0.91 mg/kg dw, and the arithmetic mean of these two particulate concentrations was 4.56 mg/kg dw (the mean particulate selenium concentration of 4.56 mg/kg dw was used to derive the invertebrate TTF in Appendix C of the Draft Selenium Criterion Document).

Lemly (1985) reported selenium concentrations of 0.77, 0.87, and 0.91 mg/kg wet weight (ww) in periphyton, plankton, and sediment, respectively (all quite similar); thus, it appears that the algae selenium concentration reported in Appendix C was based on conversion

of periphyton and plankton concentrations to dw concentrations assuming 90% moisture; while the sediment concentration was not converted to dw. This algal conversion is highly uncertain, as the moisture content of periphyton and plankton is highly variable. If the mean particulate selenium concentration of 4.56 mg/kg dw was used to derive the EF (as it was used in derivation of the invertebrate TTF in Appendix C), the resulting EF would actually be 14,250 L/kg (*i.e.*, 4.56 mg/kg dw ÷ 0.32 µg/L × 1,000 µg/mg = 14,250 L/kg). This even higher EF only results in further questions regarding the relevance and application of these data for water selenium criteria development. In summary, we question the disparity between the values 8.2 µg/g (algae) and 0.91 µg/g (sediment) as it appears there is a unit conversion that is wrong and it is unclear why algae and sediment values should be averaged, especially if bedded sediment samples are used.

In addition to uncertainties in how the selenium EF of 8,540 L/kg was derived for Badin Lake, it also appears that the sediment selenium concentration of 0.91 mg/kg dw used in the evaluation is erroneous as this was reported as a ww concentration in Lemly (1985). According to Lemly (1985), the mean moisture content in the sediment was 56%, which results in a sediment selenium concentration of 2.1 mg/kg dw. Consequently, the arithmetic mean of the algae and sediment selenium concentrations of 8.20 and 2.1 mg/kg dw is 5.2 mg/kg dw. Using the water selenium concentration of 0.32 µg/L for Badin Lake, as reported in Lemly (1985), the resulting EF of 16,250 L/kg would yet again be even larger than that reported in Table 12 of the Draft Selenium Criterion Document. Increasing the EF even further results in further questions regarding these questionable data. The same issue exists with the ww sediment selenium concentration for High Rock Lake and Belews Lake reported in Lemly (1985).

3.4.4.2 Uncertain Enrichment Factors Result in Questioning of Overall Approach for Developing Water Selenium Criteria

Aside from the issue that the selenium EFs derived from Lemly (1985) are uncertain and potentially erroneous, selenium EFs of these magnitudes (whether 8,540, 14,250, or 16,250 L/kg in the case of Badin Lake) are all anomalously large. This anomaly can be evaluated by considering the fish tissue selenium data that were also reported in Lemly (1985). For example, if we apply the EFs, CFs, and TTFs from Table 12 of the Draft Selenium Criterion Document to the Badin Lake water selenium concentration of 0.32 µg/L (reported in Lemly (1985)), the egg selenium concentrations for the six fish species can be estimated. The predicted egg selenium concentrations for the six species, using the EPA model parameters in Table 12 of the Draft Selenium Criterion Document, range from 8.7 to 15.3 mg/kg dw (Table 1). The latter predicted egg selenium concentration exceeds the draft fish egg selenium criterion of 15.2 mg/kg dw, despite the fact that Badin Lake was defined as uncontaminated with a water selenium concentration of 0.32 µg/L (*i.e.*, a false positive). If the higher calculated EF for Badin Lake (16,250 L/kg) is used to predict egg selenium concentrations for these six fish species, predicted egg selenium concentrations range from 16.6 to 29.0 mg/kg dw (Table 1). All of these predicted values are false positives, exceeding the draft fish egg selenium criterion of 15.2 mg/kg dw. For these reasons, we question EPA's reliance on the Lemly (1985) data.

For comparison, Lemly (1985) reported muscle selenium concentrations in the six fish species. These concentrations were reported on a ww basis and converted to a dw basis

assuming 75% moisture. The muscle selenium concentrations were used to calculate egg selenium concentrations using the CFs used in Table 12 of the Draft Selenium Criterion Document. On average, across all six fish species, the predicted egg selenium concentrations based on the EPA model were a factor of 1.9 greater than those calculated from actual measured muscle selenium concentrations. A similar pattern of over-predicted egg selenium concentrations is observed for High Rock Lake (another uncontaminated lake; water 0.67 µg/L selenium) and Belews Lake (contaminated lake; water 10.91 µg/L selenium) (Table 1). If the higher calculated EF for Badin Lake (16,250 L/kg) is used to predict egg selenium concentrations for these six fish species, predicted values are on average a factor of 3.5 greater than those calculated from actual measured muscle selenium concentrations (Table 1).

Table 1. Predicted fish egg selenium concentrations in two uncontaminated reservoirs based on EF, CF, and TTF assumptions from Table 12 of the draft EPA report. Water selenium and measured fish muscle selenium concentrations from Lemly (1985).

	Egg Se Predicted from Measured Water Se					Egg Se Calculated from Measured Muscle Se		
Fish Species	Water Se (µg/L)	EF (L/kg)	CF	TTF	Predicted Egg Se (mg/kg dw)	Measured Muscle Se (mg/kg ww)	Muscle Se (mg/kg dw)	Calculated Egg Se (mg/kg dw)
Badin Lake (Uncontaminated) – with lower EF estimate								
Black bullhead	0.38	8540	1.71	1.66	9.2	0.94	3.76	6.4
Carp	0.38	8540	1.92	1.63	10.2	1.46	5.84	11.2
Fathead minnow	0.38	8540	2.00	2.35	15.3	0.57	2.28	4.6
Green sunfish	0.38	8540	1.45	2.11	9.9	0.77	3.08	4.5
Mosquitofish	0.38	8540	1.71	1.57	8.7	1.22	4.88	8.3
Red shiners	0.38	8540	2.00	2.16	14.0	0.80	3.20	6.4
High Rock Lake (Uncontaminated)								
Black bullhead	0.67	3810	1.71	1.66	7.2	1.18	4.72	8.1
Carp	0.67	3810	1.92	1.63	8.0	1.13	4.52	8.7
Fathead minnow	0.67	3810	2.00	2.35	12.0	0.72	2.88	5.8
Green sunfish	0.67	3810	1.45	2.11	7.8	0.74	2.96	4.3
Mosquitofish	0.67	3810	1.71	1.57	6.9	1.29	5.16	8.8
Red shiners	0.67	3810	2.00	2.16	11.0	0.83	3.32	6.6
Belews Lake (Contaminated)								
Black bullhead	10.91	2090	1.71	1.66	64.7	6.31	25.2	43.2
Carp	10.91	2090	1.92	1.63	71.4	9.80	39.2	75.3
Fathead minnow	10.91	2090	2.00	2.35	107.2	5.17	20.7	41.4
Green sunfish	10.91	2090	1.45	2.11	69.8	4.93	19.7	28.6
Mosquitofish	10.91	2090	1.71	1.57	61.2	9.91	39.6	67.8
Red shiners	10.91	2090	2.00	2.16	98.5	6.94	27.8	55.5
Badin Lake (Uncontaminated) – with higher EF estimate								
Black bullhead	0.38	16250	1.71	1.66	17.5	0.94	3.76	6.4
Carp	0.38	16250	1.92	1.63	19.3	1.46	5.84	11.2
Fathead minnow	0.38	16250	2.00	2.35	29.0	0.57	2.28	4.6
Green sunfish	0.38	16250	1.45	2.11	18.9	0.77	3.08	4.5
Mosquitofish	0.38	16250	1.71	1.57	16.6	1.22	4.88	8.3
Red shiners	0.38	16250	2.00	2.16	26.7	0.80	3.20	6.4

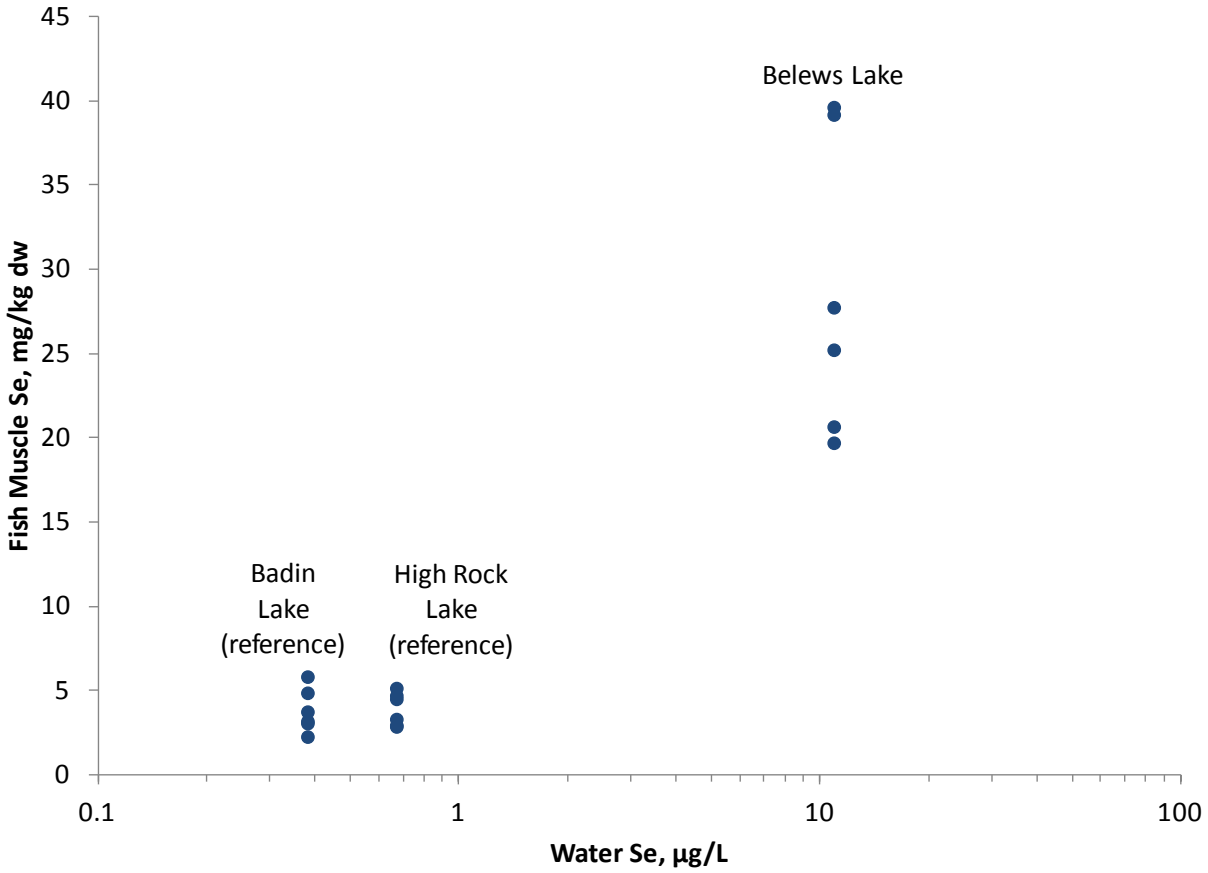
3.4.4.3 Selenium Criteria Should Not Be Driven by Reference Area Data

The above evaluation demonstrates that application of the selenium model parameters to reference areas such as Badin Lake and High Rock Lake results in over-prediction of egg selenium concentrations, which in turn results in translation to inappropriately low water selenium criteria concentrations. This assumes the data reported by Lemly (1985) are accurate or appropriately interpreted. As also noted above, the draft water selenium criterion of 1.3 µg/L is driven almost entirely by the data for these two reference lakes (Figure 1).

The relevance of reference area data for derivation of water selenium criteria is questionable. First, an inverse relationship is consistently observed between EFs and TTFs and exposure concentrations. Thus, higher EFs and TTFs are almost always observed in waters with low water selenium concentrations. Not properly accounting for this can result in an erroneous application of these values. Should the Agency have data for sites where the selenium concentration in water is at 0.01 µg/L, the EF will be significantly larger but not representative of most aquatic environments that are typically in the range of 1 µg/L. Choice of reference area is important. The way around this is to model EFs as a function of concentration in the water.

A second related issue is that a "hockey stick" relationship may be observed between tissue selenium concentrations and corresponding water selenium concentrations (*e.g.*, Brix *et al.*, 2005). There is also evidence of this relationship in the co-located fish and water selenium data for Badin Lake, High Rock Lake, and Belews Lake, albeit limited due to the narrow range of water selenium concentrations for the two reference lakes (Figure 2). The possibility of such a "blade" further emphasizes that EFs and TTFs are not constants across a range of exposure concentrations (*i.e.*, it is along the blade where these factors vary the greatest). It was for this reason that reference area and laboratory control data were excluded from the recent analysis sponsored by the NAMC-SWG in which proposed water selenium screening criteria were developed (DeForest *et al.*, 2014).

Figure 2. Relationship between muscle selenium concentrations for six fish species and co-located water selenium concentrations for Badin Lake, High Rock Lake, and Belews Lake. Data from Lemly (1985).



Finally, because the draft lentic water selenium criterion was derived as the 20th percentile of the cumulative distribution, we question whether it is appropriate that the distribution represents individual species values or whether it would be more appropriate to define the distribution of sites (*i.e.*, each data point would represent a single site). Based on the approach used by EPA, the two reference areas, along with the uncertainties and concerns in using these data, are being over-represented in the data set since data for six species in each lake are represented (*i.e.*, 24% of the data are based on just two reference lakes).

3.5 Uncertainties in Conversions to Whole Body and to Muscle

Appendix B of the Draft Selenium Criterion Document calculates species-specific CFs relating egg-ovary selenium concentrations to muscle or whole-body selenium concentrations. These CFs are used to convert reproductive effect egg-ovary concentrations to associated whole-body (Table 7a of the Draft Selenium Criterion Document) and muscle concentrations (Table 8a of the Draft Selenium Criterion Document) to support calculation of the whole-body and muscle criteria. These CFs are also used to back-calculate water criterion concentrations from the egg-ovary criterion (Equation 18 in the Draft Selenium Criterion Document). Thus, uncertainty in species-specific CFs translates directly into uncertainty in whole-body, muscle, and water criteria.

Species-specific CFs were calculated in Appendix B from empirical ratios of egg-ovary to whole-body or muscle selenium concentrations. If linear regression of egg-ovary vs. whole-body or muscle selenium concentrations resulted in a significant fit ($p \leq 0.05$) with a positive regression coefficient, the ratio of the egg-ovary to whole-body or muscle selenium concentration of each matched pair was calculated and the median ratio was used as the CF for the species. For nearly all species tested, linear regression resulted in a significant fit with a positive regression coefficient. It is, however, apparent from the plots presented in Appendix B of the Draft Selenium Criterion Document that in many cases, the assumptions of linear regression were violated, and thus the results are unreliable. It is also apparent from these plots that approximately half of these linear regressions had intercepts different from zero and thus do not, strictly speaking, support the calculation of a single median CF ratio. (Calculating a single ratio is analogous to fitting a line with an intercept of zero.)

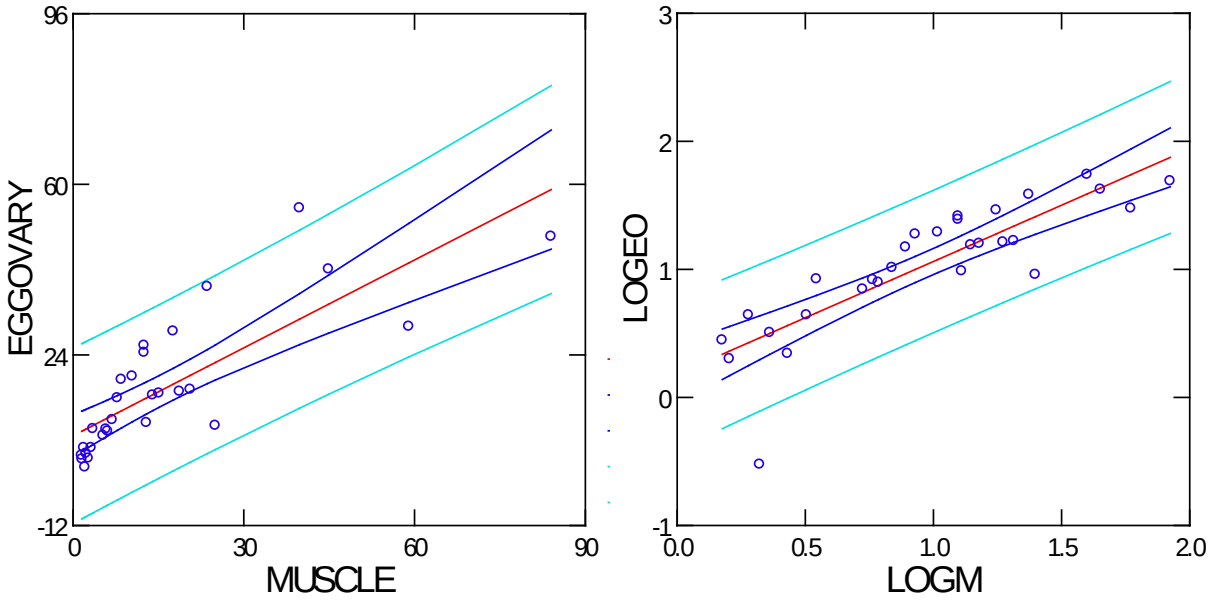
To illustrate this point, bluegill egg-ovary and muscle selenium concentration data were extracted from Appendix B of the Draft Selenium Criterion Document and are plotted below as presented in Appendix B (left panel of Figure 3) and log-transformed (right panel of Figure 3). The untransformed bluegill regression exhibits heteroscedasticity and structured residuals. The regression is significant and positive, but unreliable. Furthermore, the pattern of data does not suggest a constant ratio of egg-ovary to muscle selenium concentrations, but rather a steep relationship with relatively high ratios (most between 1.5 and 2.5) at muscle selenium concentrations less than 10 mg/kg dw, and a shallower relationship with relatively lower ratios (most between 0.5 and 1.5) at muscle selenium concentrations greater than approximately 10 mg/kg dw. The median CF from these data does not reliably reflect the actual CF across the range of data. Bluegill is the most sensitive species in the whole-body criterion calculation and the second most sensitive in the egg-ovary and muscle criterion calculations. Thus, uncertainty in the bluegill CF bears directly on calculated criterion concentrations.

The log-transformed analysis shown in the right panel of Figure 3 has improved heterogeneity of residuals and less residual structure, indicating that this is a more reliable regression than using untransformed data. The fitted slope of the line is 0.88, indicating that the best-fit curve is not a straight line, but rather a concave-down power function. A non-linear function such as this cannot reliably be represented by a single CF ratio.

The median egg-ovary/muscle ratio for bluegill was calculated in Appendix B of the Draft Selenium Criterion Document to be 1.375. Using the log-log relationship shown in the right panel of Figure 3, the CF varies across the range of data. At an egg-ovary concentration of 15.2 mg/kg dw, the CF is calculated to be 1.108. The difference between the concentration-specific CF (1.108) and the median-ratio CF (1.375) is 24%, a difference that would translate directly into calculated muscle or water criterion concentrations.

Ultimately, the relevant CF that is needed for calculating criterion concentrations is the CF that occurs at the criterion concentrations in egg or ovary. The analysis discussed above and illustrated in Figure 3 indicates that the median empirical ratio for any given species may or may not be a good estimate of this relevant CF value.

Figure 3. Relationship between egg-ovary and muscle selenium concentrations for bluegill. Left panel is untransformed data, right panel is log-transformed data. Data from EPA Draft Selenium Criterion Document (Appendix B). Lines are best-fit (red), 95% confidence limits (blue), and 95% prediction limits (cyan).



3.6 Too Low Tissue Criteria Concentrations

The proposed tissue criterion based on brown trout data is overly conservative and not technically defensible. A higher criterion would provide environmental protection without unduly penalizing human industrial or other activities, and without unnecessarily expending limited regulatory resources to the detriment of genuine environmental issues.

3.6.1 Brown Trout EC10

We commend the exhaustive effort that EPA has put into the analysis of this dataset. Nevertheless, we question the scientific defensibility of the brown trout EC10 of 15.91 mg/kg dw. EPA has selected the lowest of the six EC10 values presented.[8] The value EPA favors is based on deformities, assuming that individuals lost in the laboratory accident were deformed, which is questionable. The scientific defensibility of this EC10 is problematic in four respects, which are discussed in more detail below: (1) the model fit obtained by EPA does not fit the data well; (2) the correct best-fit Toxicity Relationship Analysis Program (TRAP) calculation for the deformities worst-case endpoint is 21.58 mg/kg dw, not 15.91 mg/kg dw; (3) there is no sound basis for using either the survival endpoint alone or deformities endpoint alone when the combined survival and deformities endpoint is available; and, (4) there is no basis for the assumption that organisms lost in the laboratory accident were dead, dying, or deformed.

---

[8] *Id.* at 110.

1. *The model does not fit the data well.* Figure 13(b) on page 110 of the Draft Selenium Criterion Document demonstrates that the curve begins its downward bend too early, missing the points at 17.7 and 20.5 mg/kg dw, a concentration range critical for defining the threshold for any fish species thought to be sensitive to selenium. Visual inspection of the data points in Figure 13(b) does not suggest that the percent normal is reduced at those concentrations.

To explore the issue quantitatively, we have presented EPA's count percentages for survival, deformities, and combined survival and deformities for the field samples in Table 2. We have excluded all the hatchery data to avoid the confound results inherent when including the eyed-embryo SPC hatchery samples, which performed differently from all other samples for reasons that cannot be related to selenium exposure.

Table 2 Brown trout results (EPA's accounting) for field stations, arranged in increasing order of egg concentration. The 17.7 and 20.5 mg/kg concentrations are in bold, highlighted.

Field station	Egg conc, mg/kg dw	% Normal		% Survival		% Combined: normal and surviving	
		"optimistic"	worst case	"optimistic"	worst case	"optimistic"	worst case
CC-150-020	6.2	81.2	81.2	99	99	80.3	80.3
CC-150-017	6.6	60.7	50.5	96.9	65	58.2	48.8
CC-150-018	6.9	40.1	40.1	96.9	96.9	38.8	38.8
CC-350-007	6.9	51.9	45	93.9	75.1	48.2	42.2
CC-150-016	7.5	56.5	19.7	96.6	48.3	50	18.8
CC-150-011	8.4	32.7	32.7	99.3	99.3	32.5	32.5
CC-150-013	8.4	44.2	40.8	86.6	64.2	38.5	35.9
CC-150-012	8.5	55.3	55.3	95.5	95.5	53.1	53.1
CC-150-015	9.1	47.2	47.2	97.1	97.1	45.8	45.8
CC-350-008	9.5	79.6	73.5	93.9	67.6	74.9	69.5
LSV2C-020	11.2	81.1	72.3	98.6	63	79.7	71.2
LSV2C-019	12.5	68.5	62.2	95.3	58.2	64.8	59.2
CC-150-009	12.8	74.6	74.6	93.1	93.1	69.3	69.3
LSV2C-002	12.8	88.8	86.3	99.3	83.4	88.1	85.6
LSV2C-012	13.2	39.2	39.2	97.1	97.1	38	38
LSV2C-016	13.4	83	83	96.5	96.5	80	80
CC-350-006	14	31.1	31.1	92.6	92.6	29.1	29.1
LSV2C-008	**17.7**	**75.8**	**61.5**	**97**	**53.2**	**74.2**	**60.5**
LSV2C-017	**20.5**	**73.3**	**65.1**	**85.3**	**69.1**	**61.8**	**55.8**
LSV2C-005	26.8	5.4	5.4	0	0	0	0
LSV2C-021	28.1	4.7	4.7	0	0	0	0
LSV2C-004	36	11.3	11.3	0	0	0	0
LSV2C-010	38.8	6.3	6.3	0	0	0	0
LSV2C-003	40.3	2	2	0	0	0	0

Having ranked the stations by increasing concentration, we now ask the question: how do results for station LSV2C-008 (17.7 mg/kg dw) and station LSV2C-017 (20.5 mg/kg

dw) compare with the averages for all field stations with lower concentrations? The field samples at or below 20.5 mg/kg dw span a 3.3-fold concentration range (6.5 to 20.5 mg/kg dw), which would easily be enough to find differences in performance if selenium were having any effects within this range.

At 17.7 mg/kg dw:

Worst-case percent normal (61.5% normal): 1.12-fold better performance (*higher percent normal, fewer deformities*) than the average of field samples at lower concentrations (55.0% normal).

"Optimistic" case percent normal (75.8% normal): 1.27-fold better performance than the average of field samples at lower concentrations (59.7% normal).

At 20.5 mg/kg dw:

Worst-case percent normal (65.1% normal): 1.18-fold better performance than the average of field samples at lower concentrations (55.3% normal).

"Optimistic" case percent normal (73.3% normal): 1.21-fold better performance than the average of field samples at lower concentrations (60.6% normal).

We believe that the above close examination of the data demonstrates that the calculated EC10 of 15.9 mg/kg dw is definitely too low. The observations at and below 20.5 mg/kg dw do not indicate an effect of selenium on deformities.

As EPA notes in its Draft Selenium Criterion Document, the model's inability to fit the key data at 17.7 and 20.5 mg/kg dw is an artifact of the influence of the data configuration at >36 mg/kg dw.[9] We do not believe that oddities in the responses at >36 mg/kg dw provide a scientifically defensible basis for a brown trout EC10 of 15.9 mg/kg dw.

2. *The correct best-fit TRAP calculation for the deformities worst-case endpoint is 21.58 mg/kg, not 15.91 mg/kg.* To obtain its solution, the computer program TRAP (cited in the Draft Selenium Criterion Document as USEPA 2011) starts with an initial parameter estimate. We have discovered that the tolerance distribution solution that TRAP version 1.21 obtains for the deformities worst-case endpoint is dependent on the initial estimate provided to TRAP. Table 3 shows the two results obtained by different initial estimates of the model parameter values.

[9] *Id*. at 108.

Table 3 TRAP output for the deformities, worst-case dataset, given two different initial estimates for the model parameters

	Set 1		Set 2	
	Initial Guess	Final Estimate	Initial Guess	Final Estimate
Log X50	1.4	1.3716	1.4	1.3849
Std Dev	0.15	0.12555	0.05	0.037545
Y0	0.6	0.63588	0.6	0.63138
Resulting EC10		15.908		21.579
Error Sum of Squares		1.1518		1.0643

Table 3 shows the error sum of squares for the model fit. For the tolerance distribution, TRAP does not provide this sum, but its "Model Fit Summary" provides everything needed to calculate it. Table 4 provides the information provided by TRAP, along with the calculated difference between prediction and observation, and its square. It can be seen that the model error is lower for the fit that yields EC10=21.58 mg/kg dw. Table 4 shows the error for each data point.

Table 4 TRAP Model Fit Summary and calculated fitting errors for the alternative EC10s obtained for the deformities, worst-case dataset.

For EC10=15.91 mg/kg					For EC10=21.58 mg/kg				
TRAP's "Model Fit Summary"			Error Magnitude		TRAP's "Model Fit Summary"			Error Magnitude	
Log X	Y	Predictn Y	Predictn Error	Error squared	Log X	Y	Predictn Y	Predictn Error	Error squared
-0.1367	0.7680	0.6359	-0.1321	0.0175	-0.1367	0.7680	0.6314	-0.1366	0.0187
-0.1367	0.7491	0.6359	-0.1132	0.0128	-0.1367	0.7491	0.6314	-0.1177	0.0139
-0.1367	0.7130	0.6359	-0.0771	0.0059	-0.1367	0.7130	0.6314	-0.0816	0.0067
-0.1367	0.6807	0.6359	-0.0448	0.0020	-0.1367	0.6807	0.6314	-0.0493	0.0024
-0.1192	0.5478	0.6359	0.0881	0.0078	-0.1192	0.5478	0.6314	0.0836	0.0070
-0.0809	0.4212	0.6359	0.2147	0.0461	-0.0809	0.4212	0.6314	0.2102	0.0442
-0.0362	0.3286	0.6359	0.3073	0.0944	-0.0362	0.3286	0.6314	0.3028	0.0917
-0.0269	0.6372	0.6359	-0.0013	0.0000	-0.0269	0.6372	0.6314	-0.0058	0.0000
-0.0177	0.7669	0.6359	-0.1311	0.0172	-0.0177	0.7669	0.6314	-0.1356	0.0184
0.0000	0.6788	0.6359	-0.0430	0.0018	0.0000	0.6788	0.6314	-0.0475	0.0023
0.0792	0.8542	0.6359	-0.2183	0.0477	0.0792	0.8542	0.6314	-0.2228	0.0497
0.0792	0.5476	0.6359	0.0883	0.0078	0.0792	0.5476	0.6314	0.0838	0.0070
0.7924	0.8118	0.6359	-0.1760	0.0310	0.7924	0.8118	0.6314	-0.1805	0.0326
0.8195	0.5051	0.6359	0.1308	0.0171	0.8195	0.5051	0.6314	0.1263	0.0159
0.8389	0.4503	0.6359	0.1856	0.0344	0.8389	0.4503	0.6314	0.1811	0.0328
0.8389	0.4012	0.6359	0.2347	0.0551	0.8389	0.4012	0.6314	0.2302	0.0530
0.8751	0.1970	0.6359	0.4389	0.1926	0.8751	0.1970	0.6314	0.4344	0.1887
0.9243	0.4077	0.6359	0.2281	0.0520	0.9243	0.4077	0.6314	0.2236	0.0500

Table 4 TRAP Model Fit Summary and calculated fitting errors for the alternative EC10s obtained for the deformities, worst-case dataset.

For EC10=15.91 mg/kg					For EC10=21.58 mg/kg				
TRAP's "Model Fit Summary"			Error Magnitude		TRAP's "Model Fit Summary"			Error Magnitude	
Log X	Y	Predictn Y	Predictn Error	Error squared	Log X	Y	Predictn Y	Predictn Error	Error squared
0.9243	0.3271	0.6359	0.3088	0.0954	0.9243	0.3271	0.6314	0.3043	0.0926
0.9294	0.5532	0.6359	0.0827	0.0068	0.9294	0.5532	0.6314	0.0782	0.0061
0.9590	0.4719	0.6359	0.1640	0.0269	0.9590	0.4719	0.6314	0.1595	0.0254
0.9777	0.7350	0.6359	-0.0991	0.0098	0.9777	0.7350	0.6314	-0.1036	0.0107
1.0492	0.7229	0.6359	-0.0870	0.0076	1.0492	0.7229	0.6314	-0.0915	0.0084
1.0969	0.6224	0.6323	0.0099	0.0001	1.0969	0.6224	0.6314	0.0090	0.0001
1.1072	0.8625	0.6296	-0.2329	0.0542	1.1072	0.8625	0.6314	-0.2311	0.0534
1.1072	0.7465	0.6296	-0.1169	0.0137	1.1072	0.7465	0.6314	-0.1151	0.0132
1.1206	0.3917	0.6252	0.2335	0.0545	1.1206	0.3917	0.6314	0.2397	0.0574
1.1271	0.8302	0.6225	-0.2077	0.0431	1.1271	0.8302	0.6314	-0.1988	0.0395
1.1461	0.3109	0.6132	0.3024	0.0914	1.1461	0.3109	0.6314	0.3205	0.1027
1.2480	0.6151	0.5222	-0.0929	0.0086	1.2480	0.6151	0.6314	0.0163	0.0003
1.3118	0.6509	0.4297	-0.2212	0.0489	1.3118	0.6509	0.6181	-0.0328	0.0011
1.4281	0.0537	0.2118	0.1581	0.0250	1.4281	0.0537	0.0886	0.0349	0.0012
1.4487	0.0465	0.1785	0.1320	0.0174	1.4487	0.0465	0.0296	-0.0169	0.0003
1.5563	0.1127	0.0508	-0.0619	0.0038	1.5563	0.1127	0.0001	-0.1126	0.0127
1.5888	0.0625	0.0275	-0.0350	0.0012	1.5888	0.0625	0.0001	-0.0624	0.0039
1.6053	0.0200	0.0184	-0.0016	0.0000	1.6053	0.0200	0.0001	-0.0199	0.0004
Sum of Squares:				1.1518	Sum of Squares:				1.0643

It appears that the two alternative results represent local minima for errors, and to which of these minima TRAP converges depends on where the initial estimate tells it to start. As shown by Figure 4, the fit that yields an EC10 of 21.58 mg/kg dw is a natural one, and does not suffer any of the shortcomings that we present in our first constructive criticism of the 15.91 mg/kg dw EC10 above. Table 4 demonstrates that an EC10 of 15.91 mg/kg dw is not defensible; however, we believe our first argument, regarding the lack of apparent effects at key observed concentrations of 17.7 and 20.5 mg/kg dw, is actually the more important consideration.

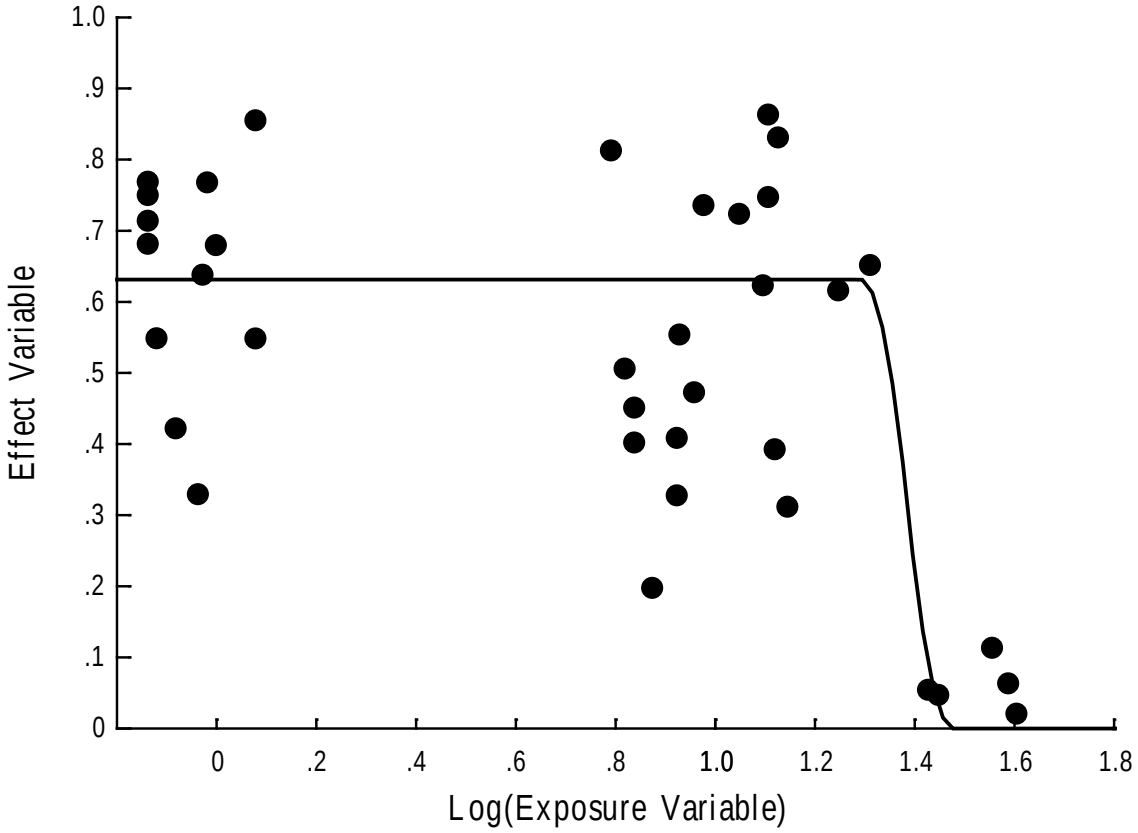


Figure 4 TRAP tolerance distribution graphical output for the model fit yielding the EC10 of 21.58 mg/kg dw for the deformities worst-case endpoint.

3. *There is no basis for using either the survival endpoint alone or the deformities endpoint alone when the combined survival and deformities endpoint is available.* EPA's first application of regression analysis for estimating ECx values was in the 1999 ammonia criteria document. Whenever possible, that document used the combined survival and growth endpoint, which it termed "biomass." Given the Agency's past preference for combining the important endpoints, it seems reasonable to request the Agency to combine the survival and deformity endpoints as its first choice for the brown trout study.

   Production of normal healthy aquatic organisms would seem to be a goal everyone can understand and support. The counting of surviving normal individuals as a function of selenium concentrations thus seems the most logical approach, whenever it is possible. We recognize that the available studies do not always allow such a calculation. Data in the brown trout study, as revised (AECOM, 2012), however, do allow those calculations, which EPA has provided but not used.

   If TRAP had been unable to provide a good fit to the combined survival/normal endpoint, then we might understand why EPA could opt not to use it. But as EPA has demonstrated in Figure 13(e) and (f) on page 110 of the Draft Selenium Criterion Document, TRAP had no difficulty fitting the combined endpoint.

Parallel with our assessment, above, for deformities, we have examined the performance at key exposures 17.7 and 20.5 mg/kg dw for the combined endpoint. From Table 2:

At 17.7 mg/kg dw:

Worst-case percent surviving and normal (60.5% normal): 1.15-fold better performance (*higher percent normal survivors*) than the average of field samples at lower concentrations (52.8% normal).

"Optimistic" case percent surviving and normal (74.2% normal): 1.30-fold better performance than the average of field samples at lower concentrations (57.0% normal).

At 20.5 mg/kg dw:

Worst-case percent surviving and normal (55.8% normal): 1.05-fold better performance than the average of field samples at lower concentrations (53.3% normal).

"Optimistic" case percent surviving and normal (61.8% normal): 1.07-fold better performance than the average of field samples at lower concentrations (58.0% normal).

We believe that close examination of the results at these key exposures demonstrates the reasonableness of the TRAP EC10s for the combined endpoint. That is, there is no evidence that selenium is having effects on the combined survival and deformities endpoint at concentrations below 20.5 mg/kg dw.

There is no logical explanation, other than an artifact of random noise, for why the effect on deformities or survival alone would yield lower EC10s than the combined endpoint. We do not believe it is scientifically defensible for EPA to choose the lowest value. Given the EPA past precedent for use of a combined endpoint, when available, we believe that use of an EC10 for combined survival and deformities, worst-case 20.65 mg/kg dw, "optimistic" case 21.16 mg/kg dw, would enhance the scientific defensibility of the proposed criterion.

4. *There is no basis for the assumption that organisms lost in the laboratory accident were dead, dying, or deformed.* Prior to the review by the U.S. Fish and Wildlife Service (USFWS) (2011), EPA had assumed that the health of the organisms lost in a laboratory accident was the same as the health of those not lost. Subsequently, EPA accepted the USFWS (2011) idea, sans supporting evidence, that the organisms lost in the accident were dead, dying, or deformed.

   In contrast, in its response to the USFWS (2011) comments, Simplot (2013) noted that the fish that overflowed from their aquaria were observed to be alive and swimming in the temperature-control water bath surrounding the aquaria, and they did not appear to be deformed, as judged visually. Furthermore, Chris Mebane (ERG, 2012) has indicated that brown trout behavior would support what EPA calls the optimistic assumption:

healthy swim-up brown trout fry are not strictly benthic and will move throughout the water column of an aquarium; moribund or dead brown trout fry do not float, but sink to the bottom.

Consequently, we believe that what EPA has called the "optimistic" case should be relabeled the "unbiased probable" case, and being both unbiased and probable, should by default be the favored assumption.

Given that EPA has no scientific evidence to support its assumption that individuals lost in overflows of the aquaria were unhealthy, a robust approach could sidestep the issue by using the effects endpoint that is not sensitive to the assumption, specifically the combined survival and deformities endpoint. For this endpoint, there is only a 2.5% difference between EC10s for the worst-case and the unbiased probable case (that is, what EPA calls the optimistic case): respectively, 20.65 mg/kg dw versus 21.16 mg/kg dw. This small difference cannot be environmentally significant relative to possible population-level effects. With the combined survival and deformities endpoint EC10s, we believe EPA could retain its worst-case assumption, thereby addressing USFWS (2011)'s concern. Since EPA cannot impose risk management assumptions having no scientific basis upon the states, use of the combined survival and deformities endpoint would have the added advantage of assuring greater consistency between states, since the criterion would be nearly the same under either assumption.

By contrast, if EPA were to replace the questionable 15.91 mg/kg dw deformities worst-case EC10 with the 16.79 mg/kg dw mortality worst-case EC10, the scientific defensibility of the criterion would be undermined by the observation that the overflowed fish were alive, thereby indicating that the mortality unbiased probable EC10 (EPA's mortality "optimistic case") of 20.40 mg/kg dw would be more appropriate. Thus, the combined mortality and deformity endpoint is not only the most comprehensive endpoint, but also the only one available that can finesse the entire overflowed fish issue.

3.6.2 Bluegill EC10 from the Hermanutz *et al.* (1992, 1996) Studies

EPA has obtained its bluegill reproductive GMCV as the geometric mean of EC10s from three studies: 20.05 mg/kg dw egg from Doroshov *et al*. (1992), 24.55 mg/kg dw egg from Coyle *et al*. (1993), and 12.68 mg/kg dw ovary from the combined Hermanutz *et al*. (1992, 1996) studies. The EC10 obtained from the Hermanutz *et al.* data is an outlier, and it is not scientifically defensible.

EPA has used its computer program TRAP for the calculation of this EC10. We do not believe EPA has followed the instructions provided by the TRAP program developer, Russell Erickson, however. That program (version 1.21) has a help screen titled "A Final Friendly Warning," which states:

> *In the end, to effectively use this (or any similar) program, the user should examine the fitted curve relative to the data and decide if*

> *the various parameter estimates and confidence limits appear reasonable. The value of this type of toxicity relationship analysis is to provide some quantitative objectivity and assessment of uncertainty to the estimation of parameters of interest that the user already can approximate by inspection of the data. The computed toxicity relationship should be close to what someone could get by just "eyeballing" the data; otherwise, some aspect of the data, model, or analysis might be causing problems. This kind of analysis demands some judgment from the user - if the results don't look good, they probably aren't and more evaluation is needed.*

We believe that the model fit that yields the 12.68 mg/kg dw EC10 is counterintuitive and cannot be claimed to be close to anything one would obtain by eyeballing the data. In contrast, if it were essential to obtain an EC10 from the information available from the study, we believe that EPA's alternate calculation of 18.40 mg/kg dw is closer to being appropriate.[10] Nevertheless, we are unsure that *any* EC10 can be confidently set forth for the Hermanutz *et al*. studies. Given that EC10s are available from the other two studies, Doroshov *et al*. (1992) and Coyle *et al.* (1993), neither of which is problematic in our view, we do not see how the use of the equivocal Hermanutz *et al*. (1992, 1996) data can reduce any remaining uncertainty about the sensitivity of bluegill. Consequently, we recommend that EPA dismiss the Hermanutz *et al*. (1992, 1996) results as inconclusive and unreliable when compared to the other two studies.

It is our understanding that EPA is now reexamining files of raw data that had been generated during the Hermanutz *et al.* studies. Given that the EC10 that EPA has published for these studies is uncertain, we believe it would be appropriate for EPA to release the newly compiled information as soon as it is available.

3.7 Inappropriate Expression of Tissue Criterion as "Never to Be Exceeded" on an Instantaneous Basis

EPA indicates that its draft tissue criteria are never to be exceeded.[11] Such a specification is unprecedented for EPA aquatic life criteria and EPA has presented no rationale for this recommendation. Nor has EPA explained how it should be interpreted or implemented. Because EPA's permit program incorporates assumptions of lognormal concentration distributions in its permit derivations, we are unable to understand how either the permit writer or the discharger can design for a criterion that is "never to be exceeded," since lognormal distributions have no concentration that is never exceeded.

Although EPA has provided no scientific support for its "never to be exceeded" recommendation, EPA does present its rationale for the instantaneous duration, labeling it an

---

[10] *Id.* at 111.

[11] *Id* at 4 and 97.

"Analysis Plan for Derivation of Duration."[12] We find that this section does not describe a plan for derivation from scientific data. Rather, it provides general statements for why EPA is recommending an instantaneous duration. The rationale hinges on two premises: (1) grab sample monitoring represents an instantaneous measurement; and, (2) duration is not important because tissue concentrations change so gradually. The rationale, based on these two premises, is not scientifically defensible.

We find the first premise, that monitoring is generally instantaneous, to be irrelevant. Although ambient monitoring of toxicant concentrations *in water* is likewise generally done with instantaneous grab samples, EPA has never cited this as being relevant to setting water criteria averaging periods. Take for example EPA's rationale for the 30-day averaging period for the ammonia chronic criterion. Rather than citing how ambient monitoring is usually done, it examines the toxicity tests from which the criterion concentration is derived. The results of those tests determine the averaging period. That is, the nature of toxic action of the pollutant determines the criterion, including its averaging period. The criterion, which includes its averaging period, describes the ambient condition to be attained, which then informs the interpretation of monitoring data. The selenium document's reversal of the process logic, basing its criterion derivation on common monitoring modes of operation, is not scientifically defensible.

The second premise, that duration is not important because practical (regulatory) outcomes would not change with different specifications of the averaging duration, is not supported by any information EPA provides in the Draft Selenium Criterion Document. EPA did not compare real-world outcomes stemming from different durations. Rather, EPA has asked the regulated community to accept EPA's worst-case specification of duration on the grounds that EPA is not concerned enough about the difference between worst case and reasonable case to provide an analysis of this difference. Because the sensitivity of real-world regulatory outcomes to EPA's worst-case assumption cannot be foreseen at this time (considering the diversity of situations to which the tissue criteria may apply), we cannot agree that EPA's approach is sound.

A particular concern about the instantaneous stipulation is that a single high outlier could completely subvert the weight of evidence from a much larger body of evidence. Although tissue concentrations vary gradually, when assessing monitoring data, there may be a substantial amount of random variability between samples at a site. To get an accurate determination of the risks at a site, these random variations need to be averaged out. Traditionally, EPA has cited concerns about lethal effects from brief concentration spikes to justify short averaging periods. However, because selenium tissue concentrations vary gradually, and because the effect of concern is the reduction in the average reproductive potential, not lethality to the standing crop of juvenile and adult individuals, the situation with selenium cannot be justified as requiring a short averaging period. Rather, because short-term spikes in tissue concentrations are not a concern, the tissue averaging period should be long, thereby allowing optimum consideration of the weight of evidence of all tissue concentrations that may be measured at a site.

---

[12] *Id.* at 34.

We point out that an EPA requirement to "never exceed a tissue criterion" has significant implications in implementation and engineering design of water treatment facilities. A "never to exceed" criterion means that water treatment facilities (industrial and municipal) have to over-design either in treatment capability or water holding capacity to avoid never having a tissue exceedance. This effectively lowers the selenium tissue criterion to a much lower level than proposed in the current document to ensure there is never an exceedance. We do not have the data at present to estimate the impact, but it would be significant. A "never to exceed" is unworkable.

More details on these concerns are provided below.

3.7.1 EPA's Own Guidelines Are Violated with Regard to Frequency

EPA's criteria derivation Guidelines (Stephan *et al*., 1985) are unequivocal about the inappropriateness of EPA's draft selenium recommendation:

> *A statement of a criterion as a number that is not to be exceeded any time or place is not acceptable because few, if any, people who use criteria would take it literally and few, if any, toxicologists would defend a literal interpretation.*

Absent any rationale for "never to be exceeded," this facet of the draft tissue criterion appears to be arbitrary and capricious.

We can find no precedent for the "never to be exceeded" stipulation. It is incompatible with EPA's Mercury Implementation Guidance (USEPA, 2010) and EPA's fish tissue monitoring guidance (USEPA, 2000), which addresses issues similar to those faced by the selenium tissue criterion. We also note that many human health criteria, which include bioaccumulative pollutants, are implemented as long-term arithmetic means (implying a harmonic mean flow for permit design conditions), per EPA's California Toxics Rule (40 C.F.R. Section 131.38), EPA's Permit Writers' Manual (USEPA, 1996), and EPA's Water Quality Standards Handbook (USEPA, 2014). Implementation as a long-term arithmetic mean implies a roughly 40% allowable exceedance frequency in a lognormal distribution of typical variability. We thus note that water quality criteria can allow substantial exceedance frequencies while achieving their protectiveness goal.

After examining the derivation of the proposed selenium water criteria from the fish tissue criteria, we believe that "never to be exceeded" is incompatible with that derivation. The water criterion derivation appears to treat the tissue criterion as a central tendency value for a site, not as an extreme upper limit. If the Appendix I criteria derivation procedures in the Draft Selenium Criterion Document are followed, we believe that the resulting water criterion would not achieve attainment of the tissue criterion *if* the tissue criterion were implemented as "never to be exceeded." Rather, the tissue criterion would need to be implemented as a central tendency value.

3.7.2 EPA's Own Guidelines Are Violated with Regard to Duration

As with frequency, it appears that EPA once again did not follow its own guidance when it incorporated an instantaneous (zero) duration into its tissue criterion. The 1985 Guidelines (Stephan *et al.*, 1985) state the following:

> *The Criterion Continuous Concentration (CCC) is intended to be a good estimate of this threshold of unacceptable effect. If maintained continuously, any concentration above the CCC is expected to cause an unacceptable effect. On the other hand, the concentration of a pollutant in a body of water can be above the CCC without causing an unacceptable effect if (a) the magnitudes and durations of the excursions above the CCC are appropriately limited and (b) there are compensating periods of time during which the concentration is below the CCC.*
>
> *[The approach taken by the Guidelines] is to require that the average concentration not exceed the CCC. The average concentration should probably be calculated as the arithmetic average rather than the geometric mean. If a suitable averaging period is selected, the magnitudes and durations of concentrations above the CCC will be appropriately limited.*
>
> *…it is the purpose of the averaging period to allow concentrations above the CCC only if the total exposure will not cause any more adverse effect than continuous exposure to the CCC would cause.*

We believe that the above material conclusively demonstrates that the instantaneous duration recommended in the Draft Selenium Criterion Document is inconsistent with EPA's own Guidelines. To accord with the Guidelines, it would be necessary to provide a genuine (*i.e.,* non-zero) averaging period.

In accord with its past approaches for deriving a pollutant-specific averaging period, we recommend that EPA consider the data used in the derivation of the tissue criterion concentration. They are of two types of data, laboratory studies and field studies. The Besser *et al.* (2012) study is an example of a comprehensive laboratory study. With eggs collected over a 60-day period, it is apparent that concentrations were measured over a substantial period of time, essentially a reproductive season. Coyle *et al.* (1993) observed bluegill reproduction over an 80-day period, again essentially equivalent to a reproductive season. Carolina Power and Light (1997) observed 56 successful largemouth bass spawns over a two-year period.

In contrast, some of the field studies collected gravid females over a short period of time, for example, the Formation Environmental (2011) brown trout study. This was by

necessity, because the reproductive season for brown trout is short. On the other hand, Holm (2002) and Holm *et al.* (2003, 2005) took observations over a multi-year period.

Considering the above, we find no evidence that "instantaneous" can in general be viewed as appropriate for the range of studies on which the criterion is based. On the other hand, we believe that "seasonal average" would likely be the most appropriate designation, while recognizing that, for some species, this is a short period of time, while for others it is substantially longer. We note that the criterion averaging period never establishes a duration over which samples *must* be taken. It designates the period over which averaging may be done when sufficient samples are available. That is, if tissue samples were collected on a single day during a year, those samples would constitute the seasonal average for assessment purposes. On the other hand, if samples were collected on multiple days, those samples should be averaged when they are part of the same reproductive season.

3.7.3 Rigorous Analysis of the Concept of Averaging

As implied above, it is necessary to consider both the variability of tissue sample concentrations of a species obtained during a single event and the variability of such concentrations over time. We do not believe that this is a significant complication, because the approach we advocate handles both types of variability in the same way.

Before we can be certain of our averaging recommendation, we need to address the question: is there some reason why averaging would prevent goals from being attained?

The goal we are seeking, as set forth in EPA's Guidelines, is to prevent *varied concentrations* having an arithmetic mean equal to the criterion from causing an effect greater than an *invariant concentration* at the criterion. To understand what is involved here, we have explored the implications of the type of probabilistic effects prediction EPA presented in its Campbell (2011) approval of Utah's Great Salt Lake (GSL) criterion. That analysis coupled (a) the concentration-response curve underlying the GSL criterion, with (b) a distribution of hypothetical tissue sample concentrations (a histogram). The aggregate effect caused by the distribution of concentrations was calculated as the summation of products of (a) the effect that each particular concentration in the histogram would have (based on the concentration-response curve), multiplied by (b) the probability that the particular concentration would occur.

Using that approach, Campbell (2011) found that use of the geometric mean ambient concentration would allow slightly greater than the 10% target effect level inherent in the EC10. While the amount of the effect was insufficiently above the target effect to form a basis for disapproval, the issue that it studied merits careful examination to determine the implications for the arithmetic mean we are recommending (as opposed to the geometric mean used by Utah). When comparing to the criterion, use of the ambient arithmetic mean is more stringent than use of the geometric mean because the ambient arithmetic mean is higher than the ambient geometric mean. Nevertheless, the question remains: will attainment as an arithmetic mean allow noticeably greater than 10% aggregate effect?

The inherent problem with averaging stems from the non-linearity in the concentration-response curve as it transitions from the flat region of no effect at low concentrations to the steeply sloping region where deleterious effects rapidly increase with selenium concentration. Consider a concentration-response curve having a zero-effect threshold at some concentration; curves based on rectangular or triangular distributions are such examples. Consider a concentration-response curve having ECzero=20 mg/kg dw and EC10=24 mg/kg dw. Now consider an ambient monitoring dataset consisting of three samples of 16, 20, and 24 mg/kg dw. When the three sample concentrations are averaged together, their average is 20 mg/kg dw, which corresponds to 0% effect on our example response curve. In contrast, if the individual samples are individually compared to the concentration-response curve, and their calculated effects averaged, we average three predicted levels of effect: 0%, 0%, and 10%, thus yielding an average effect of 3.33%, which is higher than the 0% effect predicted by the average concentration. (We are not implying that the example difference is necessarily significant, we are merely illustrating how the calculations work.)

The above example thus shows two different answers generated by two potentially reasonable approaches. The question is: which one is more accurate? If it is rigorously correct to average the sample concentrations before comparing to the concentration-response curve (and hence the criterion), then a comparison using individual sample concentrations will overstate the effect. On the other hand, if it is rigorously correct to use the individual sample concentrations with the concentration-response curve, then averaging the sample concentrations will understate the effect.

After careful consideration of the data used to construct the concentration-response curves underlying the criterion, we have concluded that a substantial amount of concentration averaging is inherent in creating the concentration-response curves. Consequently, unless the variability in the ambient concentrations being averaged exceeds the degree of averaging inherent in constructing the concentration-response curves, it is rigorously correct to average the samples before comparing to the criterion as explained below.

We have considered the data from which TRAP constructs a concentration-response curve. These data are of two types:

1. Data from laboratory studies usually (but not always) represent treatment averages (or sometimes replicate averages). A treatment essentially corresponds to a site, assuming that a site is defined as a location having similar water quality conditions within it. Consequently, when comparing tissue measurements for a species at a site to a concentration-response curve generated from treatment averages, the site concentrations should be averages, in order to correspond to how the concentration-response curve was derived.

2. Data from field studies, such as by Rudolph *et al.* (2008) (and one laboratory study, by Carolina Power & Light, 1997) involve measurements of individual adult female fish, either their eggs or ovaries. Initially this might suggest that individual sample measurements, not site averages, are appropriate for comparing to the concentration-response curve. It must be recognized that the concentration-

response curve is itself an averaging of observations, however. A prediction of effects from concentrations uses the central tendency curve that was fitted to the original data points. So the question becomes: how much noise variability existed in the data from which the concentration-response curve was derived? If the variability among samples at a site is equivalent to or less than the noise variability underlying the concentration-response curve, then it is appropriate to use the site average. If a site's sample variability is noticeably greater than the noise variability in the concentration-response curve, then the use of site averages may understate the effect.

We now ask the question: what are the key studies where concentration-response curves from *individual* fish concentrations strongly influenced the criterion derivation? These are Formation Environmental (2011), the combined Holm (2002) and Holm *et al*. (2003, 2005), Carolina Power & Light (1997), Nautilus Environmental (2011), and Rudolph *et al*. (2008), which are discussed in that order below:

- ***Formation Environmental (2011), brown trout.*** Given that the most appropriate concentration-response curve for this study remains to be determined, an analysis of the noise underlying its concentration-response curve must be deferred.

- ***Holm (2002) and Holm, et al. (2003, 2005), rainbow trout.*** The concentration-response curve is shown in Figure 5, which we use to explain the assessment approach.

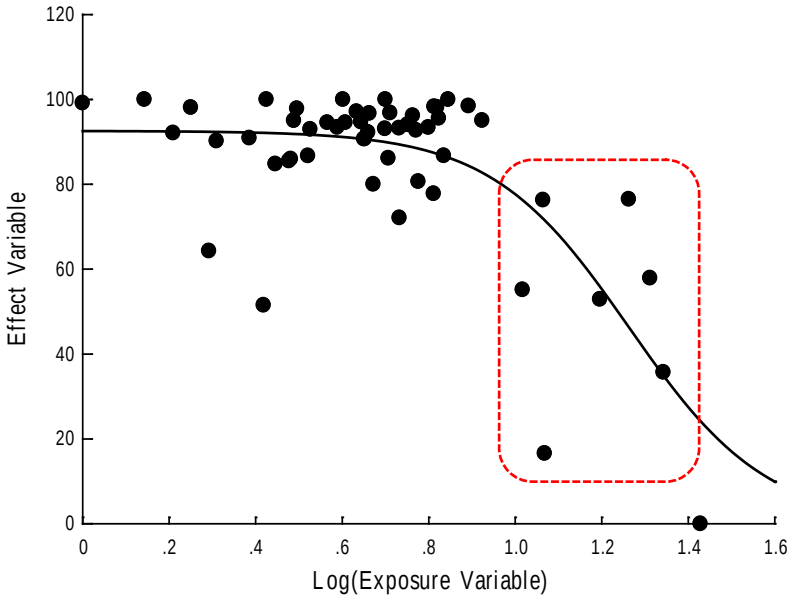


Figure 5. Holm (2002) and Holm, *et al.* (2003, 2005) concentration-response curve for rainbow trout skeletal deformities. Relevant data points for the current assessment are shown in the box.

Given the rainbow trout figure, we now ask the question: what is the range of individual measured concentrations that are associated with any particular level of effect? To answer this question, we examine the horizontal deviations from the central-tendency prediction line in the relevant portion of the graph. We calculate the central-tendency prediction associated with each relevant point by taking the logistic equation,[13] and solving it for x:

x(predicted) = x50+(ln(y0/y - 1))/(4*S)

For an observation having response y, the above equation shows the central-tendency concentration x(predicted) associated with that response; that is, it provides the exact value on the graphed line so that we do not have to work with the graph by eye. Values of parameters x50, y0, and S are given by the TRAP solution used by EPA.

The background response points in the upper left portion of the graph cannot be used because there is a wide range of concentrations that are predicted to be associated with the background response. The equation cannot be solved for y values very near, at, or greater than y=y0 (background response) because there is no unique concentration associated with such responses. The point at zero percent normal likewise cannot be used; there is no unique concentration associated with y=0 and the equation cannot be solved. We also reject points having depressed percent normal at low selenium concentrations, which merely reflect the noise inherent in y0, not the noise inherent in the downward sloping curve. The purpose of selecting data points for analysis is to assure that we do not overstate the noise underlying the curve, and thereby overstate the allowable amount of variability in samples that can be averaged when comparing to the criterion.

The approach is to treat the prediction line as the mean x associated with response y. Looking at the relevant points in Figure 5, we see the deviation from the line (measured horizontally) is small for some points and large for others. We measure the distance x – x(predicted), then square it. Summing the squares and dividing by N-1 provides the equivalent of a standard deviation.

- ***Carolina Power & Light (1997), largemouth bass.*** The concentration-response curve is shown in Figure 6. This dataset provides nine relevant points, which were evaluated as described above.

[13] *Id.* at 25.

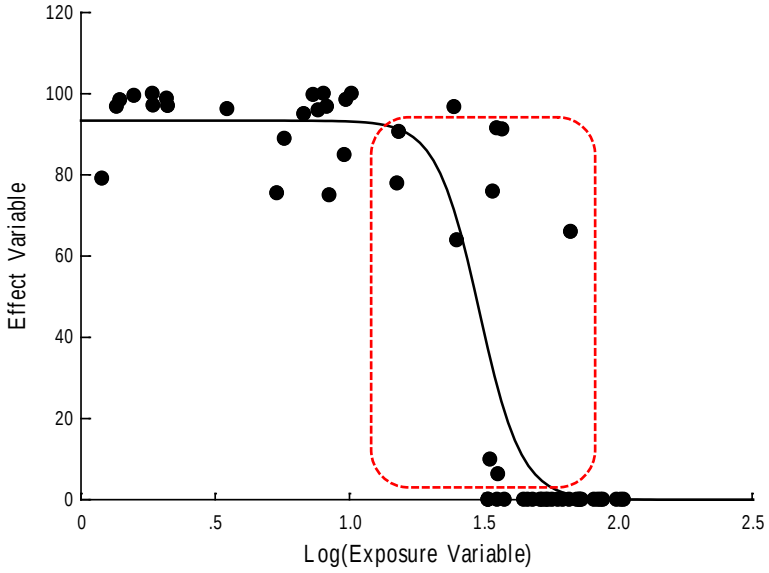


Figure 6. Carolina Power & Light (1997) concentration-response curve for largemouth bass, with relevant data points for this assessment shown within the box.

- ***Nautilus Environmental (2011), cutthroat trout.*** Inspection of the figure on page C-55 of the Draft Selenium Criterion Document indicates that this study provides only two relevant data points, too few to be meaningful. It was not further analyzed.

- ***Rudolph et al. (2008), cutthroat trout.*** The concentration-response curve is shown in Figure 7. This dataset provides four relevant points, which were evaluated as previously described.

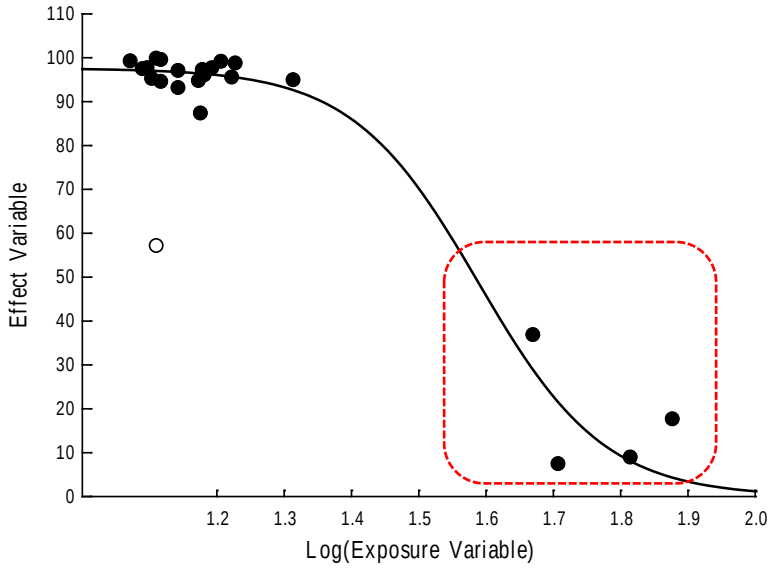


Figure 7. Rudolph *et al.* (2008) concentration-response curve for cutthroat trout, with relevant data points for this assessment shown within the box.

The combined results for the three studies represented by the above figures are shown in Table 5.

Table 5. Observations, TRAP prediction line concentration corresponding to the observed effect, horizontal error, and error squared for three studies.

	Observed		Calculated		
Study	x (base 10 logs)	y	x (predicted)	error, x - x(pred)	error sqr
Rainbow trout	1.0645	76.29	1.0169	0.0476	0.0023
	1.0170	55.15	1.2011	-0.1841	0.0339
	1.0682	16.59	1.5051	-0.4369	0.1909
	1.2625	76.5	1.0144	0.2481	0.0615
	1.3424	35.7	1.3370	0.0054	0.0000
	1.1959	52.9	1.2170	-0.0211	0.0004
	1.3118	57.9	1.1812	0.1306	0.0171
Large-mouth basss	1.1784	77.9	1.3545	-0.1761	0.0310
	1.1850	90.6	1.2065	-0.0215	0.0005
	1.4005	63.9	1.4213	-0.0208	0.0004
	1.5248	9.9	1.6520	-0.1272	0.0162
	1.5348	75.9	1.3662	0.1686	0.0284
	1.5495	91.5	1.1749	0.3746	0.1403
	1.5540	6.3	1.6913	-0.1373	0.0188
	1.5698	91.2	1.1868	0.3830	0.1467
	1.8248	66	1.4129	0.4119	0.1697
Cut-throat trout	1.6702	36.8	1.6356	0.0346	0.0012
	1.7076	7.4	1.8233	-0.1157	0.0134
	1.8149	8.9	1.8044	0.0105	0.0001
	1.8774	17.6	1.7306	0.1468	0.0215
Sum of squares (SS)					0.894
N					20
variance [base 10 logs] = SS/(N-1)					0.047
std dev [base 10 logs] = SQRT(variance)					0.217
std dev [base e logs] = std dev [base 10]*ln(10)					0.500
equivalent CV = SQRT(EXP(std dev[base e]^2)-1)					0.532

The horizontal deviations from the downward sloping prediction line are equivalent to a coefficient of variation (CV) of 0.53. This may now be compared to typical variation in tissue samples collected for a species within a site. We have examined data collected at various sites during annual tissue sampling for selenium in Sand Creek, Colorado, 1996-2011, and in the Arkansas River and its tributaries, 2004-2006 (GEI 2007, 2013 and supporting materials). In both water bodies, the annual mean for a species consisted of five samples. For Sand Creek, there were 134 species-site-years having five samples. The CV was calculated for each of the 134 species-site-years; the median CV was 0.18. For the Arkansas River, there were 39 species-site-years having five samples; their median CV was 0.24.

3.7.4 Conclusions Regarding Duration and Frequency

Regarding ***averaging***, both within sampling events and across sampling events (over time), we conclude:

- In studies where the concentration-response curve (and hence EC10) was derived from measurements of concentrations in individual fish, the individual concentration deviations from the central tendency curve generated by TRAP have a CV of 0.53. This CV is greater than typical variation within species tissue samples collected in a year.

- Consequently, irrespective of whether a concentration-response curve was generated from treatment average concentrations or from individual fish concentrations, the curve represents an average of variable observations.

- To be consistent with what the concentration-response curve represents, individual sample concentrations for a species should ordinarily be averaged before comparing to a concentration-response curve, and hence to the criterion.

  - If the aggregate effect is calculated using individual sample concentrations (rather than averages) using the approach shown in Campbell (2011), the effect will be overestimated.

  - If only one sample is available, the average used for assessment is the same as the individual sample concentration. The point here is not that multiple samples of a species are required, but rather that, when multiple samples are available, they should generally be averaged.

- For a species having sensitivity equal to the 5^{th} percentile hypothetical genus targeted by the criterion, if the average of the samples of that species does not exceed the tissue criterion, then the level of effect will not exceed 10% (the target level of protection for the criterion, representing insignificant effect), provided that the CV of the samples is equivalent to or less than approximately 0.53.

  - If the CV is somewhat greater than approximately 0.53, then the effect would be only slightly greater than 10%, as would be surmised from the type of calculations shown in Campbell (2011).

- Because the use of averages (comprised of individual samples having CV equivalent to or less than 0.53) will not allow the total exposure to "cause any more adverse effect than continuous exposure to the CCC would cause" (quote from EPA's 1985 Guidelines), it is fully consistent with the language and the intent of the Guidelines. That is, it will provide the level of protection intended by the Guidelines.

Regarding ***frequency***, we conclude:

- Because the "never to be exceeded" stipulation is without precedent or visible means of support, it is difficult for us to view it as other than arbitrary and capricious.

- Given the extreme nature of "never," we do not believe that EPA can develop a convincing rationale for why "never" is essential for attaining biological quality goals. Were EPA to press forward with an argument that it is necessary, we believe that a substantial body of evidence can be brought forth indicating that it is not necessary.

- The "never to be exceeded" provision is incompatible with the mathematics used for permit derivation, inconsistent with the selenium water criterion concentration derivation, and inconsistent with past EPA regulations and guidance (including the related Mercury Implementation Guidance).

- EPA needs to consider the Guidelines (Stephan *et al.*, 1985). They are unequivocal: "A statement of a criterion as a number that is not to be exceeded any time or place is not acceptable."

We recognize that a rigorous derivation of an allowable exceedance frequency is a difficult technical problem. Given that difficulty, we suggest that EPA apply its traditional "once-in-three-years" provision to an appropriate average selenium fish tissue concentration. When applied to a seasonal duration averaging period, as we are recommending here, we believe the once-in-three-year target is appropriately protective and scientifically defensible.

To be scientifically defensible, consistent with the 1985 Guidelines, internally consistent with the selenium water criterion derivation, compatible with other approaches EPA is using with bioaccumulative pollutants, and implementable without ambiguity by the state pollution control agencies, we recommend that EPA replace "instantaneous" with "seasonal average," and replace "never to be exceeded" with "not to be exceeded more than once in three years on average" applicable to the seasonal average concentration.

While we commend EPA for the level of effort it has put into deriving its draft recommended criterion *concentrations*, we believe that the level of rigor we have put into our rationale for duration and frequency far exceeds what EPA described as its basis for a duration and frequency, noting that EPA has supported its "instantaneous" provision with a footnote and one paragraph, and has provided no explanation for its "never to be exceeded" provision. More importantly, we believe that our recommendation provides a high degree of environmental protection and will assure attainment of biological quality goals. We are quite willing to discuss this further with the Agency and other parties.

3.8 Other Comments

Typically, selenium water concentrations are expressed as a single digit (*e.g.*, 4 µg/L, 5 µg/L, 6 µg/L). Thus, the reduction in the lotic water criterion from the current 5 µg/L to 4.8 µg/L is in practice a quantum jump -- reported values of 5 µg/L would now exceed rather than attain the water criterion.

It is noted within the text of the Draft Selenium Criterion Document that, for this review draft, EPA has conducted a new literature review and reanalyzed data considered in the 2004 and 2009 draft criteria documents. For what additional years new data were considered and reanalyzed is not stated, however. It would be useful to know the cut-off date for data consideration and reanalysis.

In Section 3.5 of the Draft Selenium Criterion Document, EPA discusses the interactions of mercury with selenium but these interactions, which can reduce selenium toxicity, were not fully evaluated. NAMC recommends that EPA fully evaluate the extensive evidence for modification of selenium toxicity by mercury relative to the potential for these interactions to comprise exposure and toxicity modifying factors that should be considered when applying the proposed criteria.

EPA should provide specific recommendations for assessing potential selenium effects and developing site-specific standards for water bodies where fish are not present due to water quantity, not water quality. Such water bodies include ephemeral or intermittent streams, and small headwater streams.

4.0 CONCLUSIONS

The primary conclusions from our review of the Draft Selenium Criterion Document are that EPA has made some good, technically defensible improvements compared to previous criteria, however, there are errors in the document and, in addition:

- Clarification is required regarding the definitions of lentic and lotic waters, and regarding tiered assessment beginning with water and proceeding to tissue selenium concentrations.

- We are very concerned that the lentic water criteria concentrations and the tissue criteria concentrations are not technically defensible and are demonstrably overly conservative. These too low criteria concentrations will result in a serious misallocation of resources, thereby reducing rather than enhancing the nation's ability to address environmental problems.

- We are also concerned, relative to addressing genuine environmental problems nationally, that the tissue criterion is inappropriately expressed as "never to be exceeded" on an instantaneous basis, particularly as this violates EPA's own Guidelines. We provide alternate recommendations that will provide a high

degree of environmental protection and will assure attainment of biological quality goals.

- Finally, we respectfully request that the planned second public comment time period be extended ahead of time to 60 days to allow adequate time for public review and comment.

5.0 REFERENCES

AECOM. 2012. Reproductive success study with brown trout (*Salmo trutta*). Data quality assurance report. Final.

Bennett PM, Janz DM. 2007a. Bioenergetics and growth of young-of-the-year northern pike (*Esox lucius*) and burbot (*Lota lota*) exposed to metal mining effluent. Ecotoxicol Environ Saf 68: 1-12.

Bennett PM, Janz DM. 2007b. Seasonal changes in morphometric and biochemical endpoints in northern pike (*Esox lucius*), burbot (*Lota lota*) and slimy sculpin (*Cottus cognatus*). Freshwater Biol 52: 2056-2072.

Besser JM, Brumbaugh WG, Papoulias DM, Ivey CD, Kunz JL, Annis M, Ingersoll CG. 2012. Bioaccumulation and toxicity of selenium during a life-cycle exposure with desert pupfish (*Cyprinodon macularius*). U.S. Geological Survey Scientific Investigations Report 2012–5033, 30 pp with appendices.

Biro PA, Post JR, Parkinson EA. 2003. Population consequences of a predator-induced habitat shift by trout in whole-lake experiments. Ecology 84: 691-700.

Brix KV, Toll JE, Tear LM, DeForest DK, Adams WJ. 2005. Setting site-specific water-quality standards by using tissue residue criteria and bioaccumulation data. Part 2. Calculating site-specific selenium water-quality standards for protecting fish and birds. Environ Toxicol Chem 24: 231-237.

Campbell .CI. 2011. Rationale for the EPA's Action on the Revisions to Utah Water Quality Standards. U.S. Environmental Protection Agency, Region 8, Denver, CO, USA.

Canton SP, VanDerveer WD. 1997. Selenium toxicity to aquatic life: An argument for sediment-based water quality criteria. Environ Toxicol Chem 16: 1255-1259.

Carolina Power & Light. 1997. Largemouth Bass Selenium Bioassay- Report. Environmental Services Section, New Hill, NC, USA.

Chapman PM, Adams WJ, Brooks ML, Delos CG, Luoma SN, Maher WA, Ohlendorf HM, Presser TS, Shaw DP (eds). 2010. Ecological Assessment of Selenium in the Aquatic Environment. SETAC Press, Pensacola, FL, USA.

Coyle JJ, Buckler DR, Ingersoll CG. 1993. Effect of dietary selenium on the reproductive success of bluegills (*Lepomis macrochirus*). Environ Toxicol Chem 12:551-565.

DeForest DK, Brix KV, Elphick JR, Rickwood CJ, DeBruyn AMH, Tear LM, Gilron G, Hughes SA, Adams WJ. 2014. Selenium partitioning between water and fish tissue in freshwater systems: Development of water-based selenium screening guidelines. Prepared for the North

American Metals Council - Selenium Working Group, Washington, D.C. http://www.namc.org/docs/Selenium%20Integrated%20Report%20-%20Final%20(2014-05-20).pdf

Doroshov S, Van Eenennaam J, Alexander C, Hallen E, Bailey H, Kroll K, Restrepo C. 1992. Development of Water Quality Criteria for Resident Aquatic Species of the San Joaquin River; Part II, Bioaccumulation of Dietary Selenium and its Effects on Growth and Reproduction in Bluegill (*Lepomis macrochirus*). Final Report to State Water Resources Control Board, State of California. Contract Number 7-197-250-0.

Driedger K, Weber LP, Rickwood CJ, Dubé, Janz DM. 2009. Overwinter alterations in energy stores and growth in juvenile fishes inhabiting areas receiving metal mining and municipal wastewater effluents. Environ Toxicol Chem 28: 296-304.

Eckmann R. 2004. Overwinter changes in mass and lipid content of *Perca fluviatilis* and *Gymnocephalus cernuus*. J Fish Biol 65: 1498-1511.

ERG (Eastern Research Group, Inc.) 2012. External Peer Review of the Interpretation of Results of a Study on the Effect of Selenium on the Health of Brown Trout Offspring. EPA Office of Science and Technology. Contract No. EP-C-12-021.

Formation Environmental. 2011. Brown Trout Laboratory Reproduction Studies Conducted in Support of Development of a Site-Specific Selenium Criterion. Prepared for J.R. Simplot Company by Formation Environmental. Revised October 2011.

GEI Consultants. 2007. Aquatic Biological Monitoring and Selenium Investigation of the Arkansas River, Fountain Creek, Wildhorse Creek, and the St. Charles River. Prepared for City of Pueblo.

GEI Consultants. 2012. Aquatic Biological Monitoring and Metals Bioaccumulation Analysis in Sand Creek and the South Platte River, Colorado 2008 – 2012. Submitted to: Suncor Energy.

Hermanutz RO, Allen KN, Roush TH, Hedtke SF. 1992. Effects of elevated selenium concentrations on bluegills *Lepomis macrochirus* in outdoor experimental streams. Environ Toxicol Chem 11(2): 217-224.

Hermanutz RO, Allen KN, Detenbeck NE, Stephan CE. 1996. Exposure of Bluegill (*Lepomis macrochirus*) to Selenium in Outdoor Experimental Streams. USEPA Report. Mid-Continent Ecology Division. Duluth, MN.

Holm J. 2002. Sublethal effects of selenium on rainbow trout (*Oncorhynchus mykiss*) and brook trout (*Salvelinus fontinalis*). Masters Thesis, Department of Zoology, University of Manitoba, Winnipeg, MB, Canada.

Holm J, Palace VP, Wautier K, Evans RE, Baron CL, Podemski C, Siwik P, Sterling G. 2003. An assessment of the development and survival of rainbow trout (*Oncorhynchus mykiss*) and

brook trout (*Salvelinus fontinalis*) exposed to elevated selenium in an area of active coal mining. Proceedings of the 26th Annual Larval Fish Conference 2003, Bergen, Norway. ISBN 82-7461-059-B.

Holm J, Palace VP, Siwik P, Sterling G, Evans R, Baron C, Werner J, Wautier K. 2005. Developmental effects of bioaccumulated selenium in eggs and larvae of two salmonid species. Environ Toxicol Chem 24: 2373-2381.

Kelly JM, Janz DM. 2008. Altered energetics and parasitism in juvenile northern pike (*Esox lucius*) inhabiting metal-mining contaminated lakes. Ecotoxicol Environ Saf 70: 357-369.

Lemly AD. 1985. Toxicology of selenium in a freshwater reservoir: Implications for environmental hazard evaluation and safety. Ecotoxicol Environ Saf 10:314-338.

Lemly AD. 1993. Metabolic stress during winter increases the toxicity of selenium to fish. Aquat Toxicol 27: 133-158.

Lemly AD. 1996. Winter stress syndrome: an important consideration for hazard assessment of aquatic pollutants. Ecotoxicol Environ Saf 34: 223-227.

McCollum AB, Bunnell DB, Stein RA. 2003. Cold, northern winters: the importance of temperature to overwinter mortality of age-0 white crappies. Trans. Am Fish Soc 132: 977-987.

McIntyre DO, Pacheco MA, Garton MW, Wallschläger D, Delos CG. 2008. Effect of Selenium on Juvenile Bluegill Sunfish at Reduced Temperature. Health and Ecological Criteria Division, Office of Water, U.S. Environmental Protection Agency, Washington, DC, USA. EPA-822-R-08-020.

Nautilus Environmental. 2011. Evaluation of the Effects of Selenium on Early Life Stage Development of Westslope Cutthroat Trout from the Elk Valley, BC. Report to Elk Valley Selenium Task Force.

Parrish DL, Hawes EJ, Whalen KG. 2004. Winter growth and survival of juvenile Atlantic salmon (*Salmo salar*) in experimental raceways. Can J Fish Aquat Sci 61: 350-357.

Rudolph B-L, Andreller I, Kennedy CJ. 2008. Reproductive success, early life stage development, and survival of Westslope cutthroat trout (*Oncorhynchus clarki lewisi*) exposed to elevated selenium in an area of active coal mining. Environ Sci Technol 42: 3109-3114.

Simplot. 2013. Responses to US Fish and Wildlife Service (USFWS) Comments on the Draft Interpretive Report, August 2010.

Sogard SM, Olla BL. 2000. Endurance of simulated winter conditions by age-O walleye Pollock: effects of body size, water temperature and energy stores. J Fish Biol 56: 1-21.

Stephan CE, Mount DI, Hansen DJ, Gentile GH, Chapman GA, Brungs WA. 1985. Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and their Uses. PB85-227049. National Technical Information Service, Springfield, VA, USA.

Toll JE, Tear LM, DeForest DK, Brix KV, Adams WJ. 2005. Setting site-specific water-quality standards by using tissue residue criteria and bioaccumulation data. Part 1. Methodology. Environ Toxicol Chem 24: 224-230.

USEPA (U.S. Environmental Protection Agency). 1995. Speed of Action of Metals Acute Toxicity to Aquatic Life. Office of Science and Technology, Office of Water. EPA-B22-R-95-002.

USEPA. 1996. U.S. EPA NPDES Permit Writers' Manual. Office of Water. EAP-866-B93-003.

USEPA. 2000. Guidance for Assessing Chemical Contaminant Data for Use in Fish Advisories, Volume 1, Fish Sampling and Analysis, Third Edition. Office of Water. EPA 823-B-00-007.

USEPA. 2010. Guidance for Implementing the January 2001 Methylmercury Water Quality Criterion. EPA 823-R-10-001.

USEPA. 2011. TRAP. Mid-Continent Ecology Division. [As referenced in EPA's criterion document.] http://www.epa.gov/med/Prods_Pubs/trap.htm

USEPA. 2013a. Letter from A Stanley Meiburg (EPA) to Leonard K Peters (State of Kentucky). November 15, 2013.

USEPA. 2013b. National Rivers and Streams Assessment 2008-2009: A Collaborative Survey (draft). EPA 841-D-13-001. Washington, DC: U.S. Environmental Protection Agency, Office of Water.

USEPA. 2014. Water Quality Standards Handbook - Chapter 5: General Policies (40 CFR 131.12). http://water.epa.gov/scitech/swguidance/standards/handbook/chapter05.cfm#section2. Accessed June 4, 2014.

USFWS (U.S. Fish and Wildlife Service). 2011. Technical Review: Smoky Canyon Mine Site-Specific Selenium Criterion Report. Division of Environmental Quality.

Attachment 1: De Forest *et al*. (2014)