**APPENDIX 1: EPA's Guidelines for Deriving Numerical National Water Quality Criteria and Issues Common to All Criteria**

The following discussion and analysis examines the shortcomings of EPA's methodology for deriving the national criteria and is critical to understanding the relationship between the numeric criteria and the exposure-response analysis in this opinion. The discussion and analysis in this Section is separated into two main categories: (1) EPA's methodology for deriving the national aquatic life criteria, and (2) overview of the effects assessment methodology in EPA's BE for the Oregon criteria.

**Derivation of EPA Aquatic Life Criteria**

The foremost problem with EPA's national aquatic life criteria lies with the derivation methodology, which is set out in EPA's *Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses* (Stephan *et al.* 1985) (Guidelines). The extent of technical issues delineated in this section regarding the Guidelines produces far more uncertainty than predictability regarding the reliability of the criteria to protect aquatic life, and in particular, listed species. This analysis highlights the risks associated with use of the Guidelines and assesses how they are likely to influence the chemical and environmental stressors affecting the listed species evaluated in this opinion.

First, we look at EPA's general approach as described in the Guidelines. Second, we look at the risks or conservatisms associated with EPA's approach. Third, we provide a summary that qualitatively assesses the degree of uncertainty and likely influences on the effects associated with exposure-response risks to the listed species considered in this opinion.

The derivation methodology for EPA's water quality criteria, the basis of Oregon's proposed water quality criteria, is detailed in the Guidelines (Stephan *et al.* 1985). An overview of the Guidelines, as described in EPA's BE, is presented below.

> The first stage in deriving water quality criteria is to compile the available data on the chemical of interest regarding its toxicity to and bioaccumulation by aquatic animals and plants. These data then go through a review process to identify studies that should not be used to derive national criteria. Although there are a number of reasons why data are not included in the data sets used to develop national criteria, some of the more common ones are that one or more pieces of information regarding study methodology or calculation of results needed to assess the reliability of the study is missing; data quality of the study is less than acceptable (*e.g.* unacceptably high control mortality); the tested species does not have a reproducing population in North America; the test species was exposed to a chemical mixture or was previously exposed to the test chemical; the study reported effects on an endpoint other than survival, reproduction of growth; or the test duration was a non-standard test duration (*e.g.* fish toxicity test reporting a 24-hr $LC_{50}$ instead of the more standard 96-hr $LC_{50}$).

Once the available data have been reviewed and unacceptable or inappropriate study results have been removed from the data set, the data are reviewed to ensure that certain types of data are available. Specifically, for freshwater aquatic biota, the following eight types of toxicity data should be available:

- Data for a fish species in the family *Salmonidae* of the class *Osteichthys*
- Data for a fish species from a second family in the class *Osteichthys*
- Data for a third family in the phylum *Chordata* (may be a third fish species or an amphibian species)
- Data for a planktonic crustacean species
- Data for a benthic crustacean species
- Data for an aquatic insect species
- Data for a species in a phylum other than *Arthropoda* or *Chordata* (*e.g. Rotifera*, *Annelida*, *Mollusca*, etc.)

Data for a species in any family in any order of insect or any aquatic phylum not already represented.

Additionally, the following three other pieces of information are needed before a national water quality criterion can be developed for a given chemical (required to derive both freshwater and saltwater criteria). Unlike toxicity data, which must be from exposures of species to chemicals in freshwater in order to derive freshwater criteria, the following information can be either for freshwater data only or a specified mixture (Stephan *et al.* 1985) of freshwater and saltwater data.
Acute-chronic ratios (ACRs) for at least three different families of aquatic species. Toxicity data for at least one freshwater plant (can be either algal or a vascular plant)
At least one bioconcentration factor (BCF).

The eight taxa for which saltwater toxicity data are required prior to derivation of a saltwater criterion obviously differ from those for freshwater, and must be from the taxonomic groupings listed below:

- Data from two families in the phylum *Chordata*
- Data from a family in a phylum other than *Arthropoda* or *Chordata*
- Data from a species in either the *Mysidae* or *Penaeidae* family
- Data from three other families not in the phylum *Chordata* (may include data for a species from a phylum or family listed in taxa groups 1 - 3 above but which was not used)
- Data from any other saltwater family

Ideally, the above freshwater and marine species toxicity data have both $LC_{50}$ data of appropriate duration and chronic NOEC data available. In practice, most chemicals with water quality criteria have sufficient $LC_{50}$ data to permit derivation of an acute water quality criterion from measured $LC_{50}$ data, but do not have sufficient measured chronic NOEC to use the above procedure to directly calculate a chronic criterion. Instead, most chronic criterion are calculated by dividing the calculated acute criterion by the available ACR value.

If toxicity data are available from multiple studies (*e.g.* three $LC_{50}$ results are available for rainbow trout), a species mean acute value (SMAV) (or species mean chronic value if one is deriving a chronic criterion, although the rest of this discussion will assume that only measured acute toxicity data are available) is calculated as the geometric mean of the three available $LC_{50}$ values in this example. Similarly, if two or more $LC_{50}$ results are available for different species of the same genus (*e.g.* $LC_{50}$ data are available for rainbow trout and Chinook salmon, both members of the genus *Oncorhynchus*), a genus mean acute value (GMAV) is calculated from the geometric mean of all toxicity data for members of that genus. If only one $LC_{50}$ value is available for a species from a given genus, that single value becomes both the SMAV and GMAV for subsequent criteria calculations.

Geometric means are used to calculate central tendency species mean, genus mean, ACR and BCF values throughout the development of water quality criteria. This is because toxicity data and ratio data (ACRs and BCFs are ratios) tend to be lognormally distributed instead of normally distributed.

Acute water quality criteria are calculated by rank ordering the GMAV values from the lowest $LC_{50}$ to the highest $LC_{50}$, and using a formula given in Stephan *et al.* (1985) to estimate the $5^{th}$ percentile of the resulting species sensitivity distribution (SSD). This $5^{th}$ percentile of measured GMAVs is termed the final acute value (FAV) in the EPA criteria development documents. As a criterion based on a concentration causing mortality to 50 percent of a test species would not be a protective criterion, the FAV is divided by two to convert $LC_{50}$ values to concentrations expected to cause little or no mortality to test species. The FAV divided by two value becomes the EPA acute water quality criterion unless a commercially or recreationally important species, or an ESA listed species has a GMAV lower than the calculated water quality criterion. In these cases, the results of one or more individual species GMAVs is used to directly calculate an acute criterion.

If sufficient chronic NOEC data are available for the freshwater and/or saltwater taxa described earlier, the same approach described above is used with the measured NOEC data to calculate a final chronic value (FCV) from the $5^{th}$ percentile of the NOEC data. Final chronic values are not divided by two to obtain the chronic criterion, as unlike $LC_{50}$ data, NOEC values are already assumed to be concentrations that have no adverse effects on survival, reproduction and growth of the tested species. Much more common is the situation where the calculated acute criterion is divided by an acute-chronic ratio (ACR) to obtain the chronic criterion.

Additional details of the Guidelines to develop national water quality criteria and the assumptions that go into their derivation are provided in Stephan *et al.* (1985). Of all the assumptions that are made during the derivation of EPA water quality criteria, perhaps the most critical is that the species sensitivity distribution of

measured toxicity data used during the calculation of criteria values is representative of the range of toxicity of a chemical to all aquatic species. There are over 700 species of freshwater fish alone in North America, making it impractical to perform toxicity tests on all species with all chemicals for which criteria exist.

Water quality criteria calculated from the methodology described above have several levels of conservatism built into them, including:

- protection of 95 percent of all aquatic genera
- division of the 5$^{th}$ percentile of all genus mean acute values by two during the derivation of acute criteria
- use of no effect concentrations to derive chronic criteria
- short exposure durations at criteria concentrations relative to the lifespan of many aquatic species

However, water quality criteria are not designed to protect all aquatic species from exposure to chemical concentrations that may adversely affect some of the more sensitive species to a given chemical. Nor are criteria designed to protect all individuals of a given species, whether or not that species is a listed species. Despite these design aspects of the national water quality criteria, many of them are protective of more than 95 percent of aquatic genera from adverse effects, and are protective of all ESA listed species known to occur within many discrete geographical areas. ESA listed aquatic species as a group are generally not believed to be more sensitive to chemicals than aquatic species as a whole (Dwyer *et al.* 2005, Sappington *et al.* 2001, Dwyer *et al.* 1999).

The toxic criteria proposed by the State of Oregon for EPA approval are identical to the corresponding national toxic criteria developed by EPA as guidance for the states.

**<u>The following section provides NMFS' analysis on the Guidelines.</u>**

***Risks from Using Acute Criteria Based on $LC_{50}$ Concentrations and the EPA Acute Adjustment Factor.*** The acute criteria for aquatic life have been primarily based on compilations of toxicity study results reported in terms of the concentration resulting in 50 percent mortality over a fixed time period [usually 96 hours: *e.g.*, $LC_{50}$, effects concentration $(EC)_{50}$, EPA 1986a]. Although there are a number of reasons why data are not included in the data sets used to develop criteria, some of the more common ones are that one or more pieces of information regarding study methodology or calculation of results needed to assess the reliability of the study is missing; data quality of the study is less than acceptable (*e.g.* unacceptably high control mortality); the test species was exposed to a chemical mixture or was previously exposed to the test chemical; the study reported effects on an endpoint other than survival, reproduction or growth; or the test duration was a non-standard test duration (*e.g.,* fish toxicity test reporting a 24-hr $LC_{50}$ instead of the more standard 96-hr $LC_{50}$).

The acute criterion is based on acute toxicity tests, *i.e.*, 96-hour $LC_{50}$ toxicity tests, that indicate the concentration at which 50 percent of the test population was killed. However, what is often

not considered are exposure-related effects such as latent mortality, which can range between 15 and 35 percent greater than the $LC_{50}$ predictions compared to the control (Zhao and Newman 2004, Lee and Lee 2005). Furthermore, because 4- to 8-hour $LC_{50}$s are about the same as the 96-hour $LC_{50}$ for some compounds, *e.g.*, selenium, lead, arsenic (EPA 1991), criteria concentrations that do not take fast-acting compounds into consideration are likely to bias the magnitude of acute toxic effects. Theses factors create significant uncertainty regarding the reliability and predictability of the acute criterion to represent concentrations that are protective against acute toxic effects. Furthermore, these factors highlight the risks of toxicity data that are based on fixed durations instead of an exposure-response curve, and challenge the notion that $LC_{50}$ data that is above the acute criterion is protective against acute toxic effects based soley on a comparison of concentrations.

Acute water quality criteria are calculated by rank ordering the GMAV values from the lowest $LC_{50}$ to the highest $LC_{50}$, and using a formula given in Stephan *et al.* (1985) to estimate the $5^{th}$ percentile of the resulting SSD. This $5^{th}$ percentile of measured GMAVs is termed the FAV in the EPA criteria development documents. As a criterion based on a concentration causing mortality to 50 percent of a test species would not be a protective criterion, EPA divides the FAV by a safety factor of 2.27 (referred to as a factor of 2 in the below analysis) to convert $LC_{50}$ values into concentrations that EPA projects to be near or below lethality.

The database from which the safety factor was derived (actually the safety factor is 2.27) was published in the Federal Register in 1978. Table 10 from the Federal Register notice (43 FR 21506-21518) lumps data for freshwater and marine fish and invertebrates. The data are broken out by the chemicals tested. There are 219 data points, but a large proportion of them aren't for a specific chemical, but rather for whole effluents of various sources—115 of the 219 data points used to derive the acute adjustment factor are based on effluent studies where individual pollutants are not measured. Interestingly, effluent studies are one of EPA's "not pertinent" or "reject" categories identified in EPA (2005).

The assumption that dividing an $LC_{50}$ by 2 will result in effect concentrations near or below leathility rests on further assumptions of the steepness of the concentration-response slope. Several examples of tests with metals which had a range of response slopes are shown in Figure A1. These examples were selected from data sets that were relevant to salmonid species in Oregon and for which the necessary data to evaluate the range of responses could be located (Chapman 1975, 1978b, Marr *et al.* 1995, Marr *et al.* 1999, Mebane *et al.* 2010, Windward 2002). The citations given include both reports with detailed original data as well as the summarized, published forms of the same tests. The examples range from tests with some of the shallowest concentration-response slopes located to very steep response slopes. In the shallowest tests (panels A and E), an $LC_{50/2}$ concentration would still result in 15 to 20 percent mortality.

One challenge for deriving acute criteria for short-term exposures is that the great majority of available data is for mortality; that is, a concentration that kills 50 percent of a test population. A fundamental assumption of EPA's criteria derivation is that the FAV, which is the $LC_{50}$ for a hypothetical species with a sensitivity equal to the $5^{th}$ percentile of the SSD, may be divided by 2 in order to extrapolates from a concentration that would likely be extremely harmful to sensitive species in short-term exposures (*i.e.*, kill 50 percent of the population) to a concentration

expected to kill few, if any, individuals. This assumption must be met for acute criteria to be protective of sensitive species. It is difficult to evaluate from published literature if this assumption is met because so few studies report the data behind an $LC_{50}$ test statistic. While $LC_{50}$s are almost universally used in reporting short-term toxicity testing, they are not something that can be "measured," but are statistical model fits. An acute toxicity test is actually a series of 4 to 6 tests runs in parallel in order to test effects at these (usually) four to six different chemical concentrations. An $LC_{50}$ is estimated by some statistical distribution or regression model, which generates an $LC_{50}$ estimate, and some confidence interval, and then all other information is thrown away. Thus, while the original test data included valuable information on what were no, low and severe effects concentrations, that information is lost to reviewers unless the unpublished, raw, lab data are available. However, a more common pattern with the metals data was that an $LC_{50/2}$ concentration would probably result in about a 5 percent death rate (panels B and F), and in many instances, no deaths at all would be expected (panels C and D).

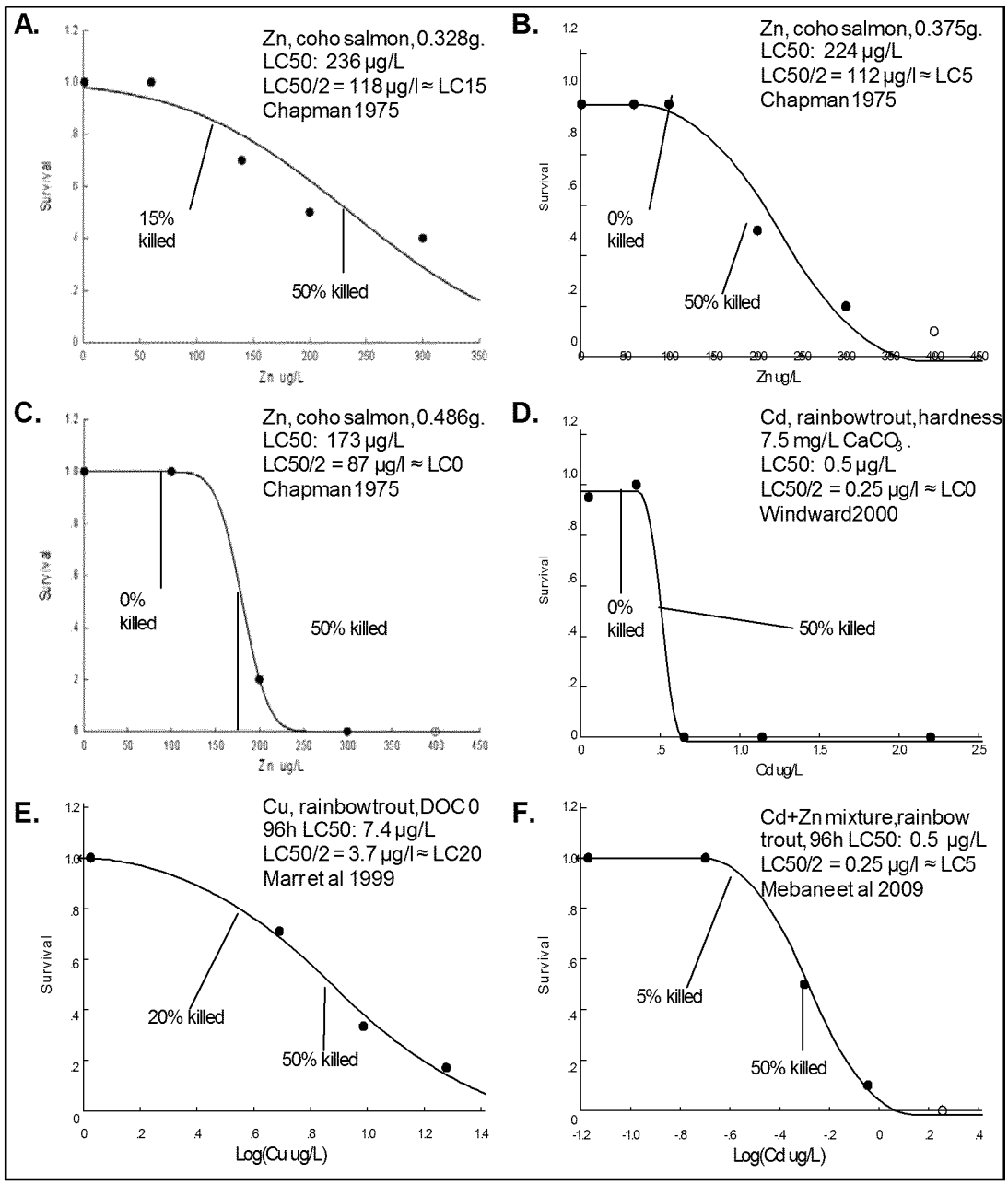


**Figure A1.** Examples of percentages of coho salmon or rainbow trout killed at one-half their $LC_{50}$ concentrations and at $LC_{50}$ concentrations with cadmium, copper, and zinc.

In one of the few additional published sources that gave relevant information, researchers happened to include effect-by-concentration information on the acute toxicity of chemical mixtures. Rainbow trout and the invertebrate zooplankton *Ceriodaphnia dubia* were exposed for 96 and 48 hours respectively to mixture of six metals, each at their presumptively "safe" acute CMC concentrations. In combination, the CMC concentrations killed 100% of rainbow trout and C. dubia, but 50% of the CMC concentrations killed none (Spehar and Fiandt 1986). This gives some support to the assumption that one-half the FAV divided by 2 is likely to kill a low

percentage of fish, although it raises questions about the overall protectiveness of criteria concentrations in mixtures.

Other relevant reviews include Dwyer *et al.* (2005b), who evaluated the $LC_{50/2}$ assumption with the results of the acute toxicity testing of 20 species with five chemicals representing a broad range of toxic modes of action. In those data, multiplying the $LC_{50}$ by a factor of 0.56 resulted in a low (10%) or no-acute effect concentration. Testing with cutthroat trout and Cd, Pb, and Zn singly and in mixtures, Dillon and Mebane (2002) found that the $LC_{50/2}$ concentration corresponded with death rates of 0 to 15 percent.

Summary: Based on this analysis, there are increased risks to listed species considered in this opinion from using acute criteria based on $LC_{50}$ concentrations and the acute adjustment factor, as acute criteria based on a hazard quotient—the acute adjustment factor, instead of acute toxicity tests that predict in $LC_{near\text{-}zero}$ concentrations, and are based on fixed duration toxicity tests instead of an exposure-response curve, are likely to underestimate the magnitude of effects for field-exposed fishes. Therefore, the risks identified in the above analysis are likely to result in mortality greater than the $LC_{50}$ test predictions and the presumed protection from the acute adjustment factor in deriving acute criteria.

***Risks from Using the Chronic Value Statistic in Setting Criteria***. An issue of concern with the derivation of the chronic criteria is the test statistic used to summarize chronic test data for species and genus sensitivity rankings. Literature on chronic effects of chemicals often contains a variety of measurement endpoints, different terms, and judgments by the authors of what constitutes an acceptable or negligible effect. While the Guidelines give a great deal of advice on considerations for evaluating chronic or sublethal data (Stephan *et al.* 1985, at p. 39), those considerations were not usually reflected in the individual national EPA-recommended ambient water quality criteria documents NMFS reviewed. In practice, for most of the criteria documents we reviewed, "chronic values" were simply calculated as the geometric mean of the lowest tested concentration that had a statistically significant adverse effect at the 95 percent confidence level (LOEC), and the next lower tested concentration (NOEC). The "chronic value" as used in individual criteria documents is effectively the same thing as the maximum acceptable toxicant concentration[20] (MATC) used in much environmental toxicology literature, even though the MATC term is never used in the Guidelines. This MATC approach has the potential to seriously underestimate effects because the statistical power in typical toxicity tests is fairly low. A bias in many ecotoxicology papers is to focus on avoiding "false accusations" of a chemical with 95 percent accuracy (*i.e.*, Type I error or false positive, the risk of declaring an effect was present when in fact there was no effect). Often no consideration whatsoever is given to the companion problem, known as Type II error, or false negatives (*i.e.*, declaring no adverse effects occurred when in fact they did occur, but because of the limited sample size or variability, they were not significant with 95 percent confidence).

The magnitude of effect that can go undetected with 95 percent confidence in a NOEC statistic can be large (greater than 30 percent on average for some endpoints), and much higher for individual tests (Crane and Newman 2000). This problem is compounded when the "chronic value" or MATC is calculated in its most common form as the geometric mean of a NOEC and

[20] The MATC is the range between the NOEC and LOEC.

LOEC. For instance, in one study, 100 percent of juvenile brook died after being exposed to 17 µg/L copper for 8 months; this was considered the LOEC for the test. The next lowest concentration tested (9.5 µg/L) had no reduced survival relative to controls. (McKim and Benoit 1971). Therefore, the only thing that can be said about the geometric mean of these two effect concentrations (*i.e.*, the chronic value of 12.8 µg/L that was used in the chronic copper criteria, EPA 1985) is that it represents a concentration that can be expected to kill somewhere between all and no brook trout in the test population. These factors create significant uncertainty regarding the reliability and predictability of the chronic criterion to represent concentrations that are protective against chronic toxic effects. Furthermore, these factors highlight the risks of toxicity data that are based on statistical hypothesis tests instead of an exposure-response curve (because the exposure-response curve describes the relationship between exposure and effect), and challenges the supposition that NOEC data that is above the chronic criterion is protective against chronic toxic effects based solely on a comparison of concentrations. Therefore, NOEC data that is above the chronic criterion does not necessarily ensure that there are no chronic toxic effects, but that the criterion concentration may result in chronic toxic effects to a subset of the test population, and therefore by inference, field-exposed individuals, relative to the criterion concentration in the range of 10 to 34 percent (Crane and Newman 2000). While the range of chronic effects predicted in these studies is likely to be less than or greater than the 10 to 34 percent range depending on compound and species, these studies highlight the inherent flaws associated with chronic toxicity tests, and provide evidence for long-term survival implications for field-exposed fishes.

Suter *et al.* (1987) evaluated published chronic tests with fish for a variety of chemicals and found that, on average, the MATC represented about a 20 percent death rate and a 40% reduction in fecundity. They noted that "although the MATC is often considered to be the threshold for effects on fish populations, it does not constitute a threshold or even a negligible level of effect in most of the published chronic tests. It corresponds to a highly variable level of effect that can only be said to fall between 0 and 90 percent." Barnthouse *et al.* (1989) further extrapolated MATC-level effects to population-level effects using fisheries sustainability models and found that the MATC systematically undervalued test responses such as fecundity, which are both highly sensitive and highly variable.

One implication of this issue is that because the MATC chronic values typically used in the EPA water quality criteria documents for aquatic life criteria may cause a substantial adverse effect for that test species, the criteria on the whole will be less protective than the Guidelines' intended goal of protecting 95 percent of the species. How much less protective is unclear and probably varies among the criteria datasets. One dataset from which a hypothetical NOEC-based chronic criterion could readily be recalculated and compared with the usual MATC criteria was a 2006 cadmium criteria update (Mebane 2006). In this comparison, Mebane determined that the MATC-based chronic criteria would protect about 92 percent of the aquatic species in the dataset at the NOEC level. Because the NOEC statistic also can reflect a fairly sizable effect (Crane and Newman 2000) it may be that at least with cadmium, the true level of protection is closer to about 90 percent than the 95 percent intended by the guidelines.

Summary: Based on this analysis, there are increased risks from using the chronic value statistic in setting criteria is high, as it is likely to result in sublethal effects, such as interference

in physiochemical processes, interruption of ecological interactions, changes in pathological stress, and toxicosis of listed species considered in this opinion.

***Risks from the CMC and CCC Duration and Frequency of Exposure.*** The CMC and the CCC are just two of six parts of an aquatic life criterion; the other four parts are the acute averaging period, the chronic averaging period, acute frequency of allowed exceedence, and chronic frequency of allowed exceedence (EPA 2006), refered to as the concentration-duration-frequency format (EPA 1991).

Concentration (magnitude) refers to how much of a pollutant, expressed as a concentration, is allowable. Duration refers to the period of time (averaging period) over which the instream concentration is averaged for comparison with criteria concentrations. This specification limits the duration of concentrations above the criteria. And, frequency refers to how often criteria can be exceeded (EPA 1991).

The 1-hour CMC averaging period means that the 1-hour average concentration of the compound does not exceed the CMC more than once every three years on the average. In other words, an organism should not be exposed to a pollutant concentration *greater* than the CMC for more than 1 hour, and an exceedence, *i.e.*, a concentration *greater* than the respective CMC, of the CMC 1-hour average concentration should not occur more than once every three years on the average. The 4-day CCC averaging period means the 4-day average concentration of the compound does not exceed the CCC more than once every three years on the average. In other words, an organism should not be exposed to a pollutant concentration *greater* than the CCC for more than 4 days, and an exceedence, *i.e.*, a concentration *greater* than the respective CCC, of the CCC 4-day average concentrations should not occur more than once every three years on the average.

This means that the averaging periods are average concentrations that are measured against the respective numeric parts of the criterion with the purpose being to minimize the duration of exposure above the CMC and CCC criteria concentrations. Figures A2 and A3 provide conceptual examples of the 1-hour and the 4-day chemical averaging periods for acute and chronic criteria, respectively. These figures show that excursions (short term concentrations above the CMC or CCC) can produce concentration "spikes" that, when compared to the available toxicity data, can result in exposure with lethal and sub-lethal responses in listed species, but that the average concentration is below the respective criterion and thus in compliance.

Figures A2 and A3 conceptually represent respective averaging concentrations for acute and chronic criteria. For example, the 1-hour averaging concentration must be evaluated for each hour of the day. That is, the average concentration in the acute example of 55.2 μg/L is a series of continuous (persistent) receiving water concentrations that occurs each hour on a continuum. The same holds true for the chronic average concentration, where the 4-day average concentration in the chronic example of 23.7 μg/L is a series of continuous (persistent) receiving water concentrations that occurs on a continuum. In these examples, the chronic criterion concentration will eventually determine the concentration outside the regulated mixing zone [defined as an area where an effluent discharge undergoes initial dilution and is…an allocated impact zone where water quality criteria can be exceeded as long as acutely toxic conditions are

prevented (EPA 1991)] boundary, and is a more accurate representation of ambient concentrations outside of regulated mixing zones. Inside regulated mixing zones, water quality criteria are permitted to be higher than criterion concentrations. While a particular toxic criterion must be met at the acute and chronic mixing zone boundaries, mixing zone boundaries vary with flow and discharge. For example, based on publically-available information from ODEQ analyzed by NMFS in this consultation, in the Willamette River mixing zone size varies greatly from a low of 1,089 square feet to a high of 1,000,000 square feet (n=19). So, meeting the aquatic life criteria at the edge of the mixing zone is a misleading protective assumption.

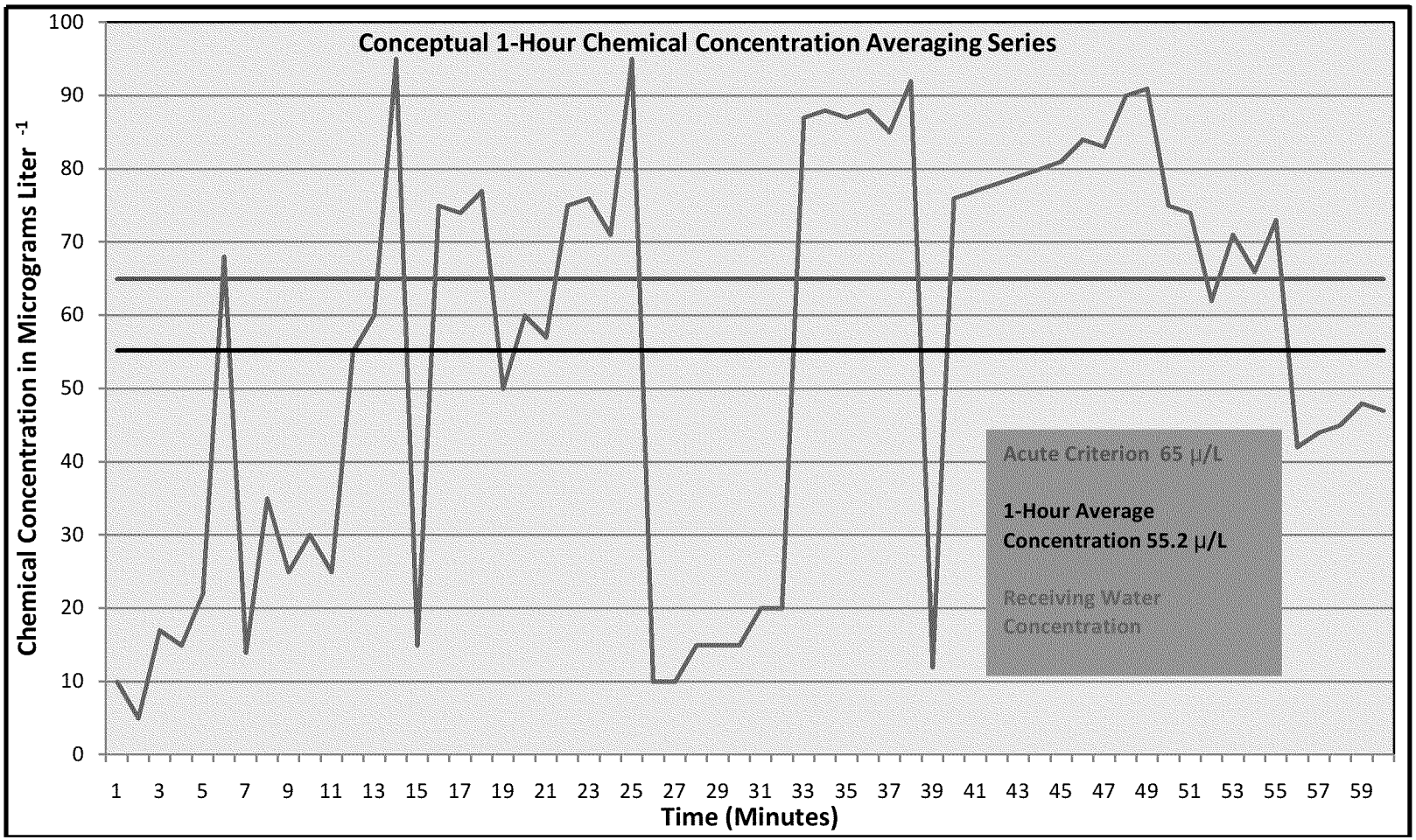


**Figure A2.** Conceptual concentration averaging series for acute criteria.

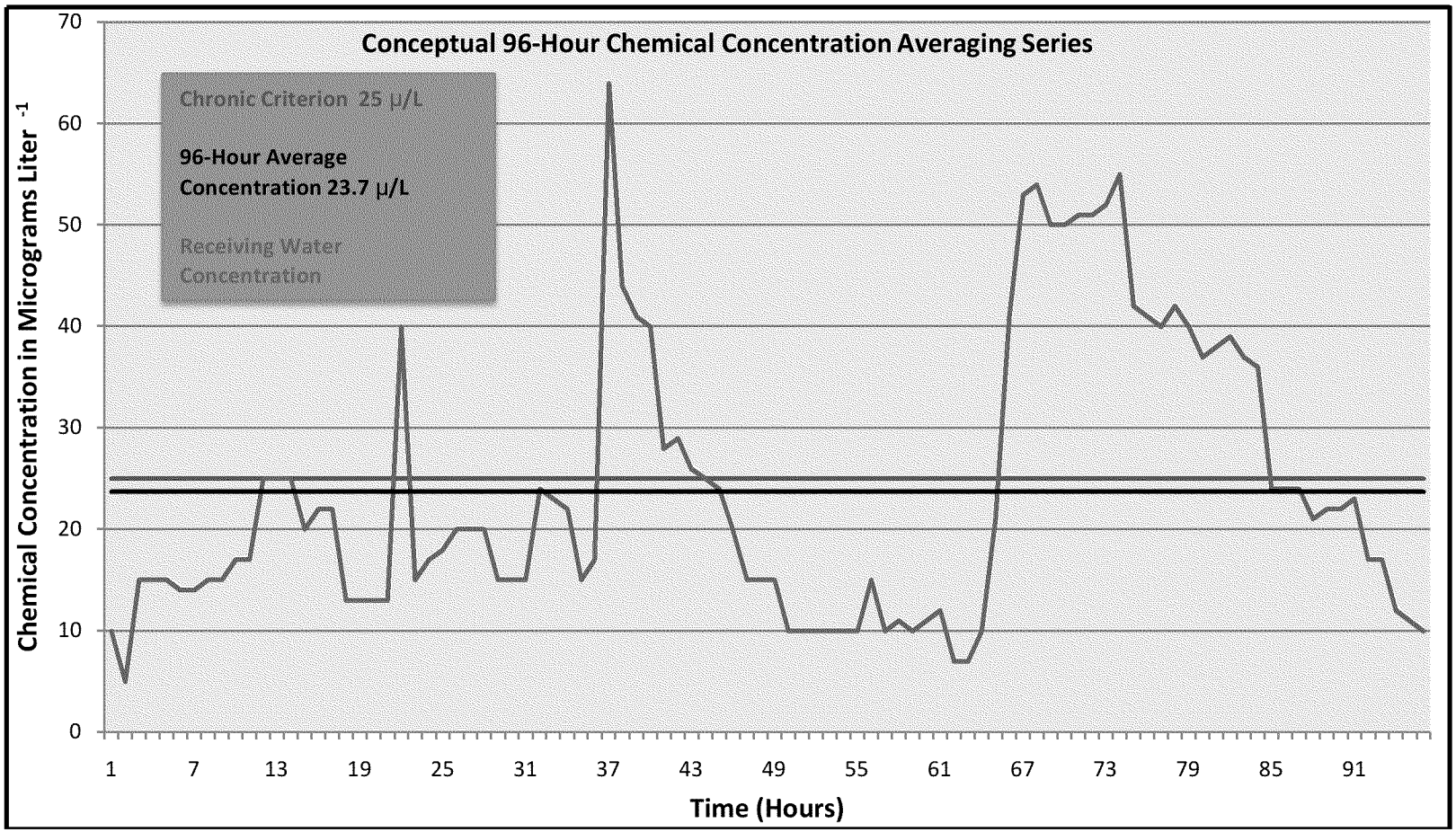


**Figure A3.** Conceptual concentration averaging series for chronic criteria.

Outside regulated mixing zones, chemical concentrations are theoretically lower than the proposed criteria, especially the acute criteria. However, waters that are 303(d)-listed for toxics do not meet water quality standards for toxics. So the assumption of lower concentrations at the edge of mixing zones is not met. That is, there is no assimulative capacity outside mixing zones.

The 1-hour and 4-day durations and averaging periods for criteria were based upon judgments by EPA authors that included considerations of the relative toxicity of chemicals in fluctuating or constant exposures. EPA's (1985) Guidelines considered an averaging period of one hour most appropriate to use with the criterion maximum concentration or (CMC or acute criterion) because high concentrations of some materials could cause death in one to three hours. Also, even when organisms do not die within the first few hours, few toxicity tests continue to monitor for delayed mortality after the exposure period is over. Thus it was not considered appropriate to allow concentrations *above* the CMC for more than one hour (Stephan *et al.* 1985). Recent criteria documents (*e.g.*, USEPA 2007) have used an averaging period of 24 hours for their CMC, although no explanation could be found for the deviation from the 1985 Guidelines.

A review of more recent information did not contradict these judgments. Some of the more relevant research relates the rapid accumulation of metals on the gill surfaces of fish to their later dying. When fish are exposed to metals such as cadmium, copper, or zinc, a relatively rapid increase occurs above background levels of metal bound to the gill. This rapid increase occurs on the order of <3 to 24 hours, and this brief exposure has been sufficient to predict toxicity at 120 hours (Di Toro *et al.* 2001, MacRae *et al.* 1999, Playle 1998, Playle *et al.* 1993). Acute exposures of 24-hours might not result in immediate toxicity, but deaths could result over the next few days. Simple examination of the time-to-death in 48 or 96 hour exposures would not detect latent toxicity from early in the exposures. Observations or predictions of appreciable mortality resulting from metals exposures on the order of only three to six hours supports the earlier recommendations by Stephan *et al.* (1985) that the appropriate averaging periods for the CMC is on the order of one hour.

The 4-day averaging period for chronic criteria was selected for use with the CCC for two reasons (Stephan *et al.* 1985): First, "chronic" responses with some substances and species may not really be due to long-term stress or accumulation, but rather the test was simply long enough that a briefly occurring sensitive stage of development was included in the exposure (Barata and Baird 2000, Chapman 1978a, De Schamphelaere and Janssen 2004, Grosell *et al.* 2006b, Mebane *et al.* 2008). Second, a much longer averaging period, such as 1 month would allow for substantial fluctuations above the CCC. Substantial fluctuations may result in increased adverse effects from those expected in constant exposures. A comparison of the effects of the same average concentrations of copper on developing steelhead, *Oncorhynchus mykiss*, that were exposed either through constant or fluctuating concentrations found that steelhead were about twice as resistant to the constant exposures as they were to the fluctuating exposures (Seim *et al.* 1984). The literature reviewed by NMFS either supports or at least does not contradict the Guidelines' recommendations on averaging periods.

In addition to the averaging periods, the Guidelines recommend for exceedence of the CMCs and the CCCs once every three years, on average. This recommendation was based on a review case studies of recovery times of aquatic populations and communities from locally severe

disturbances such as spills, fish eradication attempts, or habitat disturbances (Yount and Niemi 1990, Detenbeck *et al.* 1992). In most cases, once the cause of the disturbance was lifted, recovery of populations and communities occurred on a time frame of less than three years. The EPA has subsequently further evaluated the issue of allowable frequency of exceedences through extensive mathematical simulations of chemical exposures and population recovery. Unlike the case studies, these simulations addressed mostly less severe disturbances that were considered more likely to occur without violating criteria (Delos 2008). Unless the magnitude of disturbance was extreme or persistent, this three-year period seemed reasonably supported or at least was not contradicted by the information reviewed by NMFS.

A more difficult evaluation is the allowable exceedence magnitude, which is undefined and unlimited by the proposed criteria. Thus, theoretically, a once-per three year exceedence with no defined limits to its magnitude could be infinitely large, and have adverse effects on listed species. This is because environmental data such as chemical concentrations in water are not unpredictable, but can be described with statistical distributions and statements of exceedence probabilities. Commonly with water chemical data and other environmental data, the statistical distributions do not follow the common bellcurve or normal distribution, but have a skewed distribution with more low than high values. This pattern may be approximated with a log-normal statistical distribution (Blackwood 1992, Delos 2008, Helsel and Hirsch 2002, Limpert *et al.* 2001).

An important consideration that is often not addressed in water quality monitoring is the issue of sampling frequency. In order to accurately compare water quality samples with regulatory criteria, samples need to be collected at least at the same frequency as the criteria (*i.e.*, every hour for CMC and every four days for CCC). Otherwise, an exceedence could occur without detection. Samples, however, are not often taken at the specified frequency, and instead exceedence is detected indirectly through observed fish kills.

Summary: Based on this analysis, the duration and frequency parts of an aquatic life criterion seem like reasonable measures to keep the numeric criteria from exceeding criteria concentrations over long periods. However, the issue of excursions, exceedences with no defined limits on magnitude, and water quality monitoring and sampling sufficient to detect exceedences poses adverse risks likely to result in sublethal effects, such as interference in physiochemical processes, interruption of ecological interactions, changes in pathological stress, and toxicosis of listed species considered in this opinion.

***Metals Toxicity and Risks from Using Formula-based Metal Criteria.*** Pursuant to EPA policy, states may adopt criteria for metals measured as either the amount of metal dissolved in water or the total recoverable amount of metal. For dissolved criteria, water samples are filtered to remove any suspended solids before analysis, and a conversion factor (CF) is applied to add back a fraction of the suspended metal based on assumptions regarding bioavailability. Total recoverable metals criteria are a measurement of the suspended and dissolved amounts added together. In its National Toxics Rule (NTR) (58 FR 31177), EPA originally promulgated criteria for metals as total recoverable metals. Subsequently, EPA issued a new policy for setting water quality criteria for metals measured as dissolved metals and promulgated revised national metals criteria expressed in terms of dissolved metals (60 FR 22228, May 4, 1995). At the same time,

EPA promulgated recommended conversion factors for converting between dissolved and total recoverable criteria. The metals criteria in Oregon are expressed as dissolved metals, meaning that water samples are filtered to remove suspended solids before analysis.

Metals addressed in this consultation include: As, Cd, Cr(III), Cr(VI), Cu, Pb, Ni, Se, Ag, and Zn. The proposed ambient water quality criteria are formula-based, meaning that the criteria vary based on site-specific conditions, for the following metals: As, Cd, Cr(III), Cr(VI), Cu, Pb, Ni, Ag, and Zn. To determine criteria for these metals that are applicable to a given water body, site-specific hardness data must be obtained, input to a formula, and numeric criteria computed. There are three types of site-specific data that may be necessary to determine and/or modify the criterion for a metal at a site: water hardness, conversion factors (CF) and translators, and water effect ratios (WER). The following is a brief description of these types of data.

The general formula for a hardness-based acute (CMC) or chronic (CCC) criterion with respect to total metal concentration (dissolved and particulate) is:

$$\text{CMC or CCC (total recoverable)} = e^{(m[\ln(\text{hardness})]+b)}$$

Note that this is algebraically equivalent to the simpler expression:

$$\text{CMC or CCC (total recoverable)} = K\ (\text{hardness})^m$$

where $K = e^b$. When the m-exponent is close to 1.0, the relationship is approximately linear. Dissolved concentrations are evaluated using a total-to-dissolved CF that is based on the fraction of the metal that was in a dissolved form during the laboratory toxicity tests and that was used to develop the original total based criteria. The appropriate formula is:

$$\text{CMC or CCC (dissolved)} = CF \times e^{(m[\ln(\text{hardness})]+b)} = CF \times K \times (\text{hardness})^m$$

There is an added level of complexity in the computations of criteria for cadmium and lead because the CFs for these metals also vary with hardness.

If a total maximum daily load (TMDL) is needed to regulate discharges into an impaired water body, the dissolved criterion must be converted or translated back to a total value so that the TMDL calculations can be performed. The translator can simply be the CF (*i.e.*, divide the dissolved criterion by the CF to get back to the total criterion), or site-specific data on total and dissolved metal concentrations in the receiving water are collected and a dissolved-to-total ratio is used as the translator.

Formulae for all the metals listed above also include a WER, a number that acts as a multiplication factor. A WER is intended to account for the difference in toxicity of a metal in a site water relative to the toxicity of the same metal in reconstituted laboratory water. The reason is that natural waters commonly contain constituents which "synthetic" or "reconstituted" laboratory waters lack, such as dissolved organic compounds, that may act to bind metals and reduce their bioavailability. Where such constituents act to modify the toxicity of a metal in a site water compared to the toxicity of the same metal in laboratory water, a "water effect" is

observed. If no site-specific WER is determined, then the WER is presumed to be 1 and would not modify a formula result.

The EPA has provided specifications and guidance regarding procedures and requirements for determining "site-specific" WER values that include extensive comparative toxicity testing with several test organisms and statistical analysis of results. The example provided below only illustrates the basic principle in defining a WER value.

Example WER calculation:

Suppose the $LC_{50}$ of copper in site water is 30 µg/L
Suppose the $LC_{50}$ of copper in laboratory water is 20 µg/L
Assume a site hardness of 100 mg/L
The freshwater CF for copper = 0.96
Acute criteria (CMC) for total recoverable copper without the WER = 18 µg/L

$$\text{WER} = \frac{\text{Site } LC_{50}}{\text{Lab } LC_{50}} = \frac{30\ \mu g/L}{20\ \mu g/L} = 1.5$$

$$\begin{aligned} \text{Copper Site-Specific CMC} &= \text{WER} \times \text{CF} \times e^{(m[\ln(40)]+b)} \\ &= 1.5 \times 0.96 \times 18 \\ &= 24\ \mu g/L \end{aligned}$$

In the NTR, the EPA described and required minimum and maximum hardness values (25 mg/L and 400 mg/L as $CaCO_3$, respectively) to be used when calculating hardness-dependent freshwater metals criteria. Most of the data that the EPA used to develop the hardness formulae were in the hardness range of 25 to 400 mg/L. Therefore, the EPA stated that the formulae were most accurate in that range.

Formula-based metals criteria are discussed as a group here because the key issues of how dissolved metal criteria are derived and the implications of using the present formulae are similar for each of them. Issues include the influence of hardness, site-specific water quality characteristics, and the speciation of metal considered. The present formula-based metal method in the Guidelines does not consider the environmental fate, transport, and transformations of metals in natural environments (specifically for As, Cd, Cr (III), Cr (VI), Cu, Pb, Ni, Ag, and Zn), nor the influence of other water quality constituents on toxicity, and therefore affords incomplete protection for listed species.

A direct pathway for dissolved metals into aquatic organisms is through the gills. Dissolved forms of metals can adsorb to particulate matter in the water column and enter organisms through various routes. Metals adsorbed to particulates can also be transferred across the gill membranes (Lin and Randall 1990, Playle and Wood 1989, Sorensen 1991, Wright *et al.* 1986). Planktonic and benthic invertebrates can ingest particulate metals from the water column and sediments and then be eaten by other organisms. Thus, dietary exposure may be a significant source of metals to aquatic and aquatic dependent organisms.

Although metals bound to sediments are generally less bioavailable to organisms, they are still present, and changes in the environment (*e.g.*, dredging, storm events, temperature, lower water levels, biotic activity) can significantly alter the bioavailability of these metals. The feeding habits of fish can determine the amount of uptake of certain metals. Piscivorous fish are exposed to different levels of metals than omnivorous and herbivorous fish. For example, cadmium is more commonly found in omnivorous fish tissues than in carnivorous fish tissues from the same location (Enk and Mathis 1977).

Listed species are exposed to metals not only through the dissolved fraction in ambient waters, but they are also exposed to toxic effects of particulate metals through the mechanism of respiratory uptake in fish and by ingestion of contaminated particulate material. In addition, Finlayson *et al.* (2000) determined that metal-laden sediments in Keswick Reservoir, California were toxic to rainbow trout when re-suspended in moderately alkaline (pH 7.8) and soft (38 mg/L) water and elutriated. As fish respire, a nearly continuous flow of water passes across their gills (Moyle and Cech 1988) and particulate metals suspended in the water column may become entrapped. At the lowered pHs occurring near gill surfaces associated with gas exchange (Lin and Randall 1990, Playle and Wood 1989, Wright *et al.* 1986), entrapped particulate metals may release soluble metal ions, the form that is most bioavailable and efficiently taken up by aquatic organisms (EPA 1993a, 1997a). Although most research has been done on particulate exposures to gills of fish including salmonids, it is possible that other gill-breathing organisms (*e.g.*, aquatic macroinvertebrates) can be affected in the same way.

Current guidance for waste load allocation calculations (EPA 1996a) consists of simple dilution formulations using effluent metal loads, receiving water flows, and dissolved-to-total metals ratios in the receiving waters. Formula-based metal criteria are not protective of threatened or endangered aquatic species with respect to loading because the criteria development methods do not adequately consider the environmental fate, transport, and transformation of metals in natural environments. This concern is based in part on analyses conducted during the California Toxics Rule (CTR) consultation (USFWS and NMFS 2000), in which NMFS determined that substantial increases in total metals would be permitted in hypothetical discharges under the proposed criteria. The CTR analysis determined that as the fraction of particulate metal in the receiving water increases, the allowable discharge of particulate metals also increases rather than decreases. Such increases would be expected to occur through allowable TMDLs under the proposed ODEQ criteria because a TMDL is is based on the instream total metal concentration (EPA 1996a). Under Oregon's proposed water quality standards, total metal discharges may increase as long as the dissolved criteria are not exceeded.

Further, discharges from agricultural or urban non-point sources are largely uncontrolled through the discharge-permitting process. Metals criteria based only on dissolved concentrations provide little incentive for reducing non-point sources, which involve largely the particulate form. Thus, metals criteria based on dissolved concentrations in the absence of sediment criteria linked to total metals will not effectively prevent sediment contamination by metals and may lead to increased allowable loads of metals to sediments.

Formulae used to compute toxicity criteria for Cd, Cu, Cr(III), Pb, Ni, Ag, and Zn are presently functions of water hardness. By convention, hardness measurements are expressed in terms of

the equivalent concentration of $CaCO_3$ (expressed in mg/L) required to contribute that amount of calcium + magnesium hardness. Under the proposed criteria, hardness is determined for a site (expressed as mg/L of $CaCO_3$), and input to the criteria formulae for each metal. In natural waters considerable variation can occur in the calcium:magnesium ratio, contributing to site-specific water hardness. Studies show significant differences in toxicity for some metals depending on this ratio. In general, calcium provides greater reductions in toxicity. Site-specific hardness values with contributions from other multivalent cations (*e.g.*, iron, aluminum, manganese) that are evaluated using criteria based only on calcium + magnesium hardness result in site criteria that may not be protective. For example, in the case of cadmium, the presence of calcium is protective against toxicity whereas, magnesium, sodium, sulfate ions and the carbonate system appear to give little to no protection (Carroll *et al.* 1979). Welsh *et al.* (2000b) determined that calcium also afforded significantly greater protection against copper toxicity than magnesium.

The calcium:magnesium ratio in natural waters of Oregon varies substantially (Table A1).

**Table A1.** Total hardness for selected watersheds in Oregon in mg/L CaCO3. Data from USGS (1977).

| Watershed | Mean | Standard Deviation | Range |
|---|---|---|---|
| Snake River ID-OR Border | 141.3 | 33.7 | 97-190 |
| Rogue River (RM 25) | 37.5 | 5.1 | 30-45 |
| John Day River | 88.4 | 32.8 | 46-140 |
| Deschutes River | 41.5 | 2.7 | 37-45 |
| Columbia River (RM 140) | 69 | 11.8 | 45-94 |
| Tualatin River | 38.1 | 14.2 | 25-80 |
| Willamette River (RM 10) | 24 | 3.4 | 19-32 |
| Nehalem River | 18.9 | 6.5 | 12-32 |
| Umpqua River | 28.3 | 4.3 | 19-34 |

The majority of hardness data used to develop the EPA hardness-dependent criteria formulae were in the range of 25 mg/L to 400 mg/L (40 CFR Part 131). Consequently, EPA's regulations (40 CFR 131.36) specify that the minimum hardness that can be used in criteria equations is 25 mg/L. This requirement reflects that toxicity effects at hardness concentrations less than 25 mg/L are not known with a reasonable degree of certainty. Existing criteria formulae can result in toxic concentrations in water with hardness below the 25 mg/L lower threshold. There are some streams in Oregon where hardness concentrations average less than 25 mg/L, for which concentrations of contaminants with hardness ameliorated toxicity should be calculated on actual site conditions.

Comparable toxicity test data for hardness values greater than 400 mg/L appear to exist only for zinc, which precludes direct evaluation of the effects of extrapolating the criteria equations upwards. However, the ameliorating effect of increasing concentration of calcium ions means that the use of a default limiting value of 400 mg/L is protective for listed species in harder water in the case of metals for which toxicities are influenced by hardness.
The value of the site-specific hardness value will depend on where samples are collected. The calculated criteria may be less protective when samples are collected downstream of effluent

sources that may increase hardness locally (it is highly unlikely that discharges decrease downstream hardness). In otherwords, the use of hardness values measured downstream of the effluent source could lead to greater-than-intended site criteria. In some cases, certain effluents may alter ambient hardness, but not other important water quality constituents that influence metal toxicity (*e.g.*, pH, alkalinity, dissolved organic carbon, calcium, sodium, chloride, *etc.*). Alterations in receiving water chemistry by a discharge (*e.g.*, abrupt elevation of hardness, changes in pH, exhaustion of alkalinity, abrupt increases in organic matter *etc.*) could result, depending on the hardness value applied in the criteria formulae, in increased allowable discharges of toxic metals.

Water hardness and the hardness acclimation status of a fish will affect toxicity and toxic response. However the use of hardness alone as a universal surrogate for all water quality parameters that can modify metal toxicity will not always correlate well with the predicted toxic effect on listed species. The importance of water quality parameters other than hardness on metals toxicity has been understood for some time (Howarth and Sprague 1978). Numerous studies have been performed on the toxicity of metals in test waters of various compositions, and the results do not confer a singular role to hardness in ameliorating metals toxicity. Test water characteristics in most studies, including pH, calcium, alkalinity, dissolved organic carbon, chloride, sodium, suspended solids, and other chemical properties, are varied in a controlled manner while observing the responses of test organisms. It is likely that understanding metal toxicity in waters of various chemical makeups is not possible without the use of a geochemical model, and that a univariate regression formula will not suffice. It is also possible that simple toxicity tests (using mortality, growth, or reproductive endpoints) are not capable of discriminating the role of hardness relative to other water chemistry characteristics in modulating metals toxicity (Erickson *et al.* 1996).

Summary**:** Based on this analysis, using formula-based criteria for aquatic life criteria derived following the Guidelines are likely to be underprotective of listed species considered in this opinion. Formula-based metal criteria are discussed as a group here because the key issues of how dissolved metal criteria are derived and the implications of using the present formulae are similar for each of them. Issues include the influence of hardness, site-specific water quality characteristics, and the speciation of metal considered. The present formula-based metal method does not consider the environmental fate, transport, and transformations of metals in natural environments (specifically for arsenic, cadmium, chromium (III), chromium (VI), copper, lead, nickel, silver, and zinc), nor the influence of other water quality constituents on toxicity, and therefore affords incomplete protection for listed species and is likely to result in sublethal effects, such as central nervous system disruption, altered liver and kidney function, impaired reproduction, decreased olfactory response, delayed smoltification, impaired ability to avoid predation and capture prey, growth inhibition, growth stimulation, changes in prey species community composition (which will increase foraging budgets), and death of listed species considered in this opinion.

***Additive and Synergistic Toxicity.*** When two or more toxic pollutants are present, their combined effect may be either additive, synergistic (where the net effect exceeds the sum of effects), or antagonistic. The proposed water quality standards do not take these effects into account. Relatively few toxicity studies have addressed this issue, and some studies have

indicated conflicting results due to complex interactions that vary with the combination(s) and concentrations involved (Sorenson 1991). However, a number of studies have determined conclusively that adverse effects due to additive or synergistic toxicity mechanisms occur when several criteria are near or equal to acute criteria concentrations (*e.g.*, Alabaster and Lloyd 1982, Spehar and Fiandt 1986, EIFAC 1987, Enserink *et al.* 1991, Sorenson 1991). Spehar and Fiandt (1986) determined that rainbow trout embryo survival and growth were not reduced when exposed to combinations of arsenic, cadmium, chromium, copper, and lead at chronic concentrations, but production and growth of *Daphnid sp.* were reduced for the same test mixtures. Combinations of organic pollutants also have been shown to result in different toxic responses, as have combinations of organic and metals contaminants.

Alabaster and Lloyd (1982) observed from their data that the combined acutely lethal toxicity to fish and other aquatic organisms is approximately the simple addition of the proportional contribution from each toxicant. The median value of the effect on fish is 0.95 of that predicted; the collective value for sewage effluents, river waters and a few industrial wastes is 0.85. The range for effluents, river wastes, and industrial wastes is 0.4 to 2.8, which indicates that the combined effects of individual acutely toxic pollutants are from 0.4 to 2.8 times the effects predicted by adding the individual effects. The median combined effect is approximately additive (EPA 1991).

Summary: Based on this analysis, the aquatic life criteria derivied following the Guidelines do not take into account additive or synergistic effects, thus increasing the likelihood of acute toxic effects and sublethals effects, such as interference in physiochemical processes, interruption of ecological interactions, changes in pathological stress, and toxicosis of listed species considered in this opinion.

***Assumption that Effects in Laboratory Tests are Reasonable Predictors of Effects in Field Situations.*** The preceding discussion concerned whether compilations of laboratory test values were appropriate to treat as surrogates of the diversity of natural systems. A fundamental question in evaluating the Guidelines and the national criteria is whether tests of chemicals in laboratory aquaria with "domesticated" cultures of test animals are likely to produce similar effects as would exposure to the same substance on the same or closely related species in the wild. If the responses between animals in laboratory aquaria or the wild are different, is there a bias in the sensitivity of responses from either the lab or wild settings? That is, are the effects of chemical contamination likely more or less severe in the laboratory or wild settings? This question is important because water quality criteria are designed to apply to and protect ambient waters (that is, streams, rivers, and lakes), yet the data used to develop them are invariably compiled from laboratory testing under tightly controlled and thus quite artificial environments. There are myriad factors that may influence the effects of a chemical stressor on aquatic organisms, and this complexity makes the question of bias in sensitivity difficult or even impossible to answer with any certainty. The conclusion by Chapman (1983) regarding comparability of laboratory exposure-response effects and field exposure-response effects contributed to one the most fundamental assumptions in the Guidelines, that is, "the Guidelines have been developed on the theory that effects which occur on species in appropriate laboratory tests will generally occur on the same species in comparable field situations." A number of reasons why the effects of a criteria chemical could be more or less severe on listed species in laboratory or in wild settings are summarized in Table A2.

**Table A2.** Factors influencing the effects of a chemical stressor in a laboratory setting or in the wild.

| FACTOR | ARE EFFECTS LIKELY MORE SEVERE IN TYPICAL LAB SETTINGS OR IN THE WILD? |
|---|---|
| **Environmental Conditions** | |
| Nutritional state - acute test exposures | **In the wild:** In acute toxicity tests with fish fry, fish are selected for uniform size, and unusually underweight fish that might be weakened from being in poor nutritional state are culled from tests. For instance, if <90% of control fish survive the 4 days of starvation in an acute toxicity test, the test may be rejected from inclusion in the criteria dataset. In the wild, not all fish will be in optimal nutritional state. While perhaps counterintuitive, starvation can protect fish against waterborne copper exposure (Kunwar *et al.* 2009). Fish are routinely starved during acute laboratory tests of the type used in criteria development. |
| Nutritional state – chronic test exposures | **In the wild:** Fish in the wild must compete for prey, and if chemicals impair fish's ability to detect and capture prey because of subtle neurological impairment, this could cause feeding shifts and reduce their competitive fitness (Riddell *et al.* 2005). Fish in chronic lab tests with waterborne chemical exposures are often fed to satiation, and food pellets don't actively evade capture like live prey. Perhaps these factors dampen responses in lab settings. |
| Temperature | **In the wild:** In lab test protocols, nearly optimal test temperatures are recommended (*e.g.*, 12°C for rainbow trout, the most commonly tested salmonid). Fish may be most resistant to chemical insults when at optimal temperatures. At temperatures well above optimal ranges, increased toxicity from chemicals often results from increased metabolic rates (Sprague 1985); Under colder temperatures, fish have been shown to be more susceptible to at least Cu, Zn, Se and cyanide, although the mechanisms of toxicity are unclear (Dixon and Hilton 1985, Erickson *et al.* 1987, Hansen *et al.* 2002a, Hodson and Sprague 1975, Kovacs and Leduc 1982, Lemly 1993). |
| Flow | **In the wild:** Fish expend energy to hold their position in streams and to compete for and defend preferred positions that provide optimal feeding opportunity from the drift for the energy expended. Subordinate fish in the wild are forced to less profitable positions and become disadvantaged. Subordinate fish in lab settings still get adequate nutrition from feeding. Chemical exposure can reduce swimming stamina or speeds, as can exposure to soft water. (Adams 1975, De Boeck *et al.* 2006, Kovacs and Leduc 1982, McGeer *et al.* 2000). |
| Disease and parasites | **In the wild:** Disease and parasite burden are common in wild fish, but toxicity tests that used diseased fish likely were considered compromised and results likely were not used in criteria development. Chemical exposure may weaken immune responses and increase morbidity or deaths (Arkoosh *et al.* 1998, Stevens 1977). |
| Predation | **In the wild:** Fish use chemical cues to detect and evade predators; these can be compromised by some chemical exposures (Berejikian *et al.* 1999, Labenia *et al.* 2007, Phillips 2003, Scott *et al.* 2003) |
| **Exposure** | |
| Variable exposures | **In the lab:** Most toxicity tests used to develop criteria are conducted at nearly constant exposures. Criteria are expressed not just as a concentration but also with an allowed frequency and duration of allowed exceedences. In field settings, most point or non-point pollution scenarios that rarely if ever exceed the criteria concentration (*i.e.*, no more than for one 4-day interval per 3 yrs), will have an average concentration that is less than the criterion concentration. For some chemicals, such as copper, fish might detect and avoid harmful |

| FACTOR | ARE EFFECTS LIKELY MORE SEVERE IN TYPICAL LAB SETTINGS OR IN THE WILD? |
|---|---|
| | concentrations if clean-water refugia were readily available. |
| Metal form and bioavailability | **Uncertain:** Metals other than Hg and some organics are commonly more bioavailable in the lab because dissolved organic carbon, which reduces the bioavailability and toxicity of several metals, is low in laboratory tests that are eligible for use in criteria. The Guidelines call for <5 mg/L TOC (total organic carbon) in studies to be used in criteria (Stephan *et al.* 1985), but probably more often TOC is <2 mg/L in laboratory studies. |
| Chemical equilibrium | **Uncertain:** While results conflict, metals are usually considered less toxic when in equilibrium with other constituents in water, such as organic carbon, calcium, carbonates and other minerals. In the wild, daily pH cycles prevent full equilibria from being reached (Meyer *et al.* 2007). Likewise, in conventional laboratory flow-through tests, designs chemicals may not have long enough contact time to reach equilibrium. Static-renewal tests are probably nearly in chemical equilibria, although organic carbon accretion can lessen toxicity which may not reflect natural settings (Santore *et al.* 2001, Welsh *et al.* 2008). |
| Prior exposure | **Uncertain:** If fish are exposed to sublethal concentration of a chemical they could either become weakened or become more tolerant of future exposures. With some metals, normally sensitive life stages of fish may become acclimated and less sensitive during the course of a chronic test if the exposure was started during the resistant egg stage (Brinkman and Hansen 2007, Chapman 1983, 1985, Sprague 1985). |
| Life stages exposed | **In the wild:** Most lab studies are short term and realistically testing all life stages of anadromous fish is probably infeasible. Reproduction is often the most sensitive life stage with fish but most "chronic" studies are much shorter and just test early life stage survival and growth (Suter *et al.* 1987). At different life stages and sizes, salmonids can have very different susceptibility to some chemicals; even when limited to a narrow window of young-of-year fry, sensitivity can vary substantially. Unless the most sensitive life stages are tested, lab tests could provide misleadingly high toxicity values for listed species. |
| Chemical mixtures | **In the wild:** In field conditions, organisms never experience exposure to a single pollutant; rather, ambient waters typically have low concentrations of numerous chemicals. The toxic effects of chemicals in mixture can be less than those of the same chemicals singly, greater than, or have no appreciable difference. The best known case of one toxicant reducing the effects of another is probably Se and Hg (*e.g.*, Belzile *et al.* 2006). However, strongly antagonistic responses are probably uncommon, and much more common are situations where chemical mixtures have greater toxicity than each singly or little obvious interaction (*e.g.*, Borgert 2004, Laetz *et al.* 2009, Norwood *et al.* 2003, Playle 2004, Scholz *et al.* 2006). In general, it seems prudent to assume that if more than one toxicant were elevated, it is likely that lower concentrations of chemicals would be required to produce a given magnitude of effect than would be predicted from their actions separately. |
| Dietary exposures | **In the wild:** Toxicity test data used in criteria development have been mostly based solely on waterborne exposures, yet in the wild, organisms would be exposed to contaminants both through dietary and water exposures. With at least some organics and metals (*i.e.*, As, Se) dietary exposures are more important than water exposures. For some other metals (*i.e.*, Cd, Cu, Ni, Pb, Zn), at environmentally relevant concentrations that would be expected when waterborne concentrations are close to criteria, dietary exposures have not been shown to directly result in appreciable adverse effects on fish (Hansen *et al.* 2004, Schlekat *et al.* 2005). However, while dietary exposures of some metals have not yet been implicated in adverse effects on fish at or below criteria |

| FACTOR | ARE EFFECTS LIKELY MORE SEVERE IN TYPICAL LAB SETTINGS OR IN THE WILD? |
|---|---|
| | concentrations, they may in fact be both the primary route of exposure and an important source of toxicity for benthic invertebrates rather than fish (Buchwalter *et al.* 2008, Irving *et al.* 2003). For instance Besser *et al.* (2005a) found that the effects threshold for Pb to the benthic crustacean *Hyalella sp.* was well above the chronic criterion in water exposures, but when Pb was added to the diet, effects threshold dropped to near criteria concentrations. Ball *et al.* (2006) found that feeding Cd-contaminated green algae to the benthic crustacean *Hyalella* sp. caused a 50% growth reduction at about the NTR chronic criterion. |
| **Population Dynamics** | |
| Density effects | **In the lab:** Salmonid fishes are highly fecund (~500 to 17,000 eggs per spawning female). When abundant, overcrowding, and competition for food and shelter may result in relatively high death rates for some life stages, particularly young-of-year during their first winter. After many fish die in a density-dependent bottleneck, the survivors have greater resources and improved growth and survival. Conceptually, if an acute contamination episode killed off a significant portion of young-of-year fish prior to their entering a resource bottleneck, then assuming no residual contaminant effects, the losses to later life stages and to adult spawners could be buffered. |
| Meta-population dynamics | **In the lab:** If habitats are interconnected, as is the case in intact stream networks, and if pervasive contamination from discharges to a stream were to impair only some endpoints or life-stages, such as reproductive failure or YOY mortalities, immigration from source populations may make detection of population reductions in the affected sink population difficult (Ball *et al.* 2006, Palace *et al.* 2007). If an episodic contamination pulse were to kill a large proportion of fish in a stream, the proximity of refugia and donors from source populations affect recovery rates (Detenbeck *et al.* 1992). |
| | |

Summary: Based on this analysis, the assumption that effects in laboratory tests as reasonable predictors of effects to species in the wild is dependent upon the specific factor being considered. Overall NMFS finds that laboratory tests are likely to underpredict effects, as adverse effects are generally likely to be more severe in the wild than under laboratory conditions. Thus aquatic life criteria derivied following the Guidelines are likely to result in sublethals effects, such as interference in physiochemical processes, interruption of ecological interactions, changes in pathological stress, and toxicosis of listed species considered in this opinion.

***Risks of Using Flow-Through, Renewal, or Static Exposure Test Designs***. One area of controversy in evaluating toxicity test data or risk assessments, or criteria derived from them, has to do with potential bias in how test organisms are exposed to test solutions. Exposures of test organisms to test solutions are usually conducted using variations on three techniques. In "static" exposures, test solutions and organisms are placed in chambers and kept there for the duration of the test. The "renewal" technique is like the static technique except that test organisms are periodically exposed to fresh test solution of the same composition, usually once every 24 hours or 48 hours, by replacing nearly all the test solution. In the "flow-through" technique, test

solution flows through the test chamber on a once-through basis throughout the test, usually with at least five volume replacements/day (ASTM 1997).

The term "flow-through test" is commonly mistaken for a test with flowing water, *i.e.*, to mimic a lotic environment in an artificial stream channel or flume. This is not the case; rather the term refers to the once-through, continuous delivery of test solutions (or frequent delivery in designs using a metering system that cycles every few minutes). Flows on the order of about five volume replacements per 24 hours are insufficient to cause discernable flow velocities. In contrast, even very slow moving streams have velocities of around 0.04 ft/sec (an inch per second) or more. At that rate, a parcel of water would pass the length of a standard test aquarium (~2 ft) in about 48 seconds, resulting in about 9,000 volume replacements per day. A more typical stream velocity of about 0.5 ft/sec would produce over 100,000 volume replacements per day.

Historically, flow-through toxicity tests were thought to provide a better estimate of toxicity than static or renewal toxicity tests because they provide a greater control of toxicant concentrations, minimize changes in water quality, and reduce accumulation of waste products in test exposure waters (Rand *et al.* 1995). Flow-through exposures have been preferred in the development of standard testing protocols and water quality criteria. The Guidelines first advise that for some highly volatile, hydrolysable, or degradable materials, it is probably appropriate to use only results of flow-through tests. However, this advice is followed by specific instructions that if toxicity test results for a species were available from both flow-through and renewal or static methods, then results from renewal or static tests are to be discounted (Stephan *et al.* 1985). Thus, depending upon data availability, toxicity results in the criteria databases may be a mixture of data from flow-through, renewal or static tests, raising the question of whether this could result in bias. In the Guidelines, the rationale for the general preference for flow-through exposures was not detailed, but it was probably based upon assumptions that static exposures will result in $LC_{50}$s that are biased high (apparently less toxic) than comparable flow-through tests, or that flow-through tests have more stable exposure chemistries and will result in more precise $LC_{50}$ estimates.

With metals, renewal tests produce higher $EC_{50}$s (*i.e.*, metals were less toxic), probably because of accretion of dissolved organic carbon (DOC) (Erickson *et al.* 1996, Erickson *et al.* 1998, Welsh *et al.* 2008). However, in contrast to earlier EPA and ASTM recommendations favoring flow-through testing, Santore *et al.* (2001) suggested that flow-through tests were biased low because copper complexation with organic carbon, which reduces acute toxicity, is not instantaneous, and typical flow-through exposure systems allowed insufficient hydraulic residence time for complete copper-organic carbon complexation to occur. Davies and Brinkman (1994) similarly found that cadmium and carbonate complexation was incomplete in typical flow-through designs, although in their study incomplete complete complexation had the opposite effect of the copper studies, with cadmium in the aged, equilibrium waters being more toxic. A further complication is that it is not at all clear that natural flowing waters should be assumed to be in chemical equilibria because of tributary inputs, hyporheic exchanges and daily pH, inorganic carbon, and temperature cycles. Predicting or even evaluating risk of toxicity through these cycles is complex and seldom attempted (Meyer *et al.* 2007), in part because pulse exposures cause latent mortality (*i.e.*, fish die after exposure to the contaminant is removed), a phenomenon that is often overlooked or not even recognized in standard acute toxicity testing.

When comparing data across different tests, it appears that other factors such as testing the most sensitive sized organisms or organism loading may be much more important than if the test was conducted by flow through or renewal techniques. For instance, Pickering's and Gast's (1972) study with fathead minnows and cadmium produced flow-through $LC_{50}$s that were lower than comparable static $LC_{50}$s (~ 4,500 to 11,000 µg/L for flow-through tests vs. ~30,000 µg/L for static tests). The fish used in the static tests were described as "immature," weighing about 2 g (2000 mg). The size of the fish used in their flow-through acute tests were not given, but is assumed to have been similar. In contrast, 8 to 9 day old fathead minnow fry usually weigh about 1 mg or less (USEPA 2002b). Using newly hatched fry weighing about 1/1000$^{th}$ of the fish used by Pickering and Gast (1972) in the 1960s, and modern protocols, cadmium $LC_{50}$s for fathead minnows at similar hardnesses tend to be around 50 µg/L, with no obvious bias for test exposure. Similar results have been reported with brook trout. One each flow-through and static acute tests with brook trout were located, both conducted in waters of similar hardness (41 to 47 mg/L). The $LC_{50}$ of the static test which used fry was < 1.5 µg/L whereas the $LC_{50}$ of the flow-through test using yearlings was > 5,000 µg/L (Carroll *et al.* 1979, Holcombe *et al.* 1983).

Many studies on which the proposed criteria are based involve laboratory-based $LC_{50}$ bioassays using static exposure systems and nominal contaminant concentrations. Such studies often yield $LC_{50}$ values substantially higher than values obtained with flow-through tests or tests in which actual concentrations of contaminants in the system during the experiment are measured, with differences in some cases of an order of magnitude lower. For example, $LC_{50}$ values for static tests have been determined to be approximately 20 times higher than those from flow-through tests for DDT (Earnest and Benville 1971). Mercury toxicity testing of trout embryos has indicated that concentration-based endpoints (*e.g.*, $EC_{50}$) could be as much as one to two orders of magnitude lower in flow-through than static tests (Birge *et al.* 1979, 1981). Static assays were also found to underestimate the toxicity of endosulfan in comparisons with flow-through systems (Naqvi and Vaishnavi 1993). Several additional studies with a variety of compounds report increased toxicity in flow-through compared to static systems (*e.g.*, Erickson *et al.* 1998, Hedtke and Puglisi 1982, Vernberg *et al.* 1977, Randall *et al.* 1983, Burke and Ferguson 1969). Static conditions may underestimate the true exposure concentration because the fish will deplete the concentration in solution over time, causing a lack of steady-state exposure. There may also be important differences in energy expenditure and metabolism of test fish between static and flow-through tests, depending on the experimental setup. In the case of listed salmonids in Oregon, this may be an important source of variation because they typically live in flowing waters. Acute $LC_{50}$s for salmonids that are based on static tests could therefore underestimate toxicity, and water quality standards based on such tests may consequently not be sufficiently protective against conditions reasonably expected to occur in Oregon waters.

Summary: Based on this analysis, using flow-through, renewal, or static exposure test designs may result in greater than predicted effects.

***Effects of Acclimation on Susceptibility to Chemicals***. Exposure to sublethal concentrations of organic chemicals and other metals may result in pronounced increases in resistance to later exposures of the organisms. With metals the resistance may be on the order of two to four times greater for acute challenges, but for some organic contaminants may be much higher (Chapman 1985). However, the increased resistance can be temporary and can be lost in

as little as seven days after return to unpolluted waters (Bradley *et al.* 1985, Hollis *et al.* 1999, Sprague 1985, Stubblefield *et al.* 1999). For this reason, the Guidelines specify that test results from organisms that were pre-exposed to toxicants should not be used in criteria derivation (Stephan *et al.* 1985).

Effects from acclimation, however, are not precluded by the Guidelines and influence chronic values and thus chronic criteria. Several tests have shown that at least with fish and metals, if the toxicity tests were initiated during more resistant early life stages (ELS, *e.g.*, embryo stage), acclimation may occur, and later in the test when the more sensitive life stages become exposed (*e.g.*, fry stage), the usually sensitive life stages may be more resistant than the same life stages of fish which had no pre-exposure (Brinkman and Hansen 2004; 2007, Chapman 1978a; 1994, Spehar *et al.* 1978).

Chapman (1994) exposed different life stages of steelhead (*Oncorhynchus mykiss*) for the same duration (three months) to the same concentration of copper (13.4 µg/L at a hardness of 24 mg/L as $CaCO_3$). The survival of steelhead that were initially exposed as embryos was no different than that of the unexposed control fish, even though the embryos developed into the usually-sensitive swim-up fry stage during the exposure. In contrast, steelhead that were initially exposed as swim-up fry, without the opportunity for acclimation during the embryo state, suffered complete mortality. Brinkman and Hansen (2007) compared the responses of brown trout (*Salmo trutta*) to long-term cadmium exposures that were initiated either at the embryo stage (*i.e.*, early-life stage tests) or the swim-up fry stage (*i.e.*, chronic growth and survival tests). In three comparative tests, fish that were initially exposed at the swim-up fry stage were consistently two to three times less resistant than were the fish initially exposed at the embryo stage.

These studies support the counterintuitive conclusion that because of acclimation, longer-term tests or tests that expose fish over their full life cycle are not necessarily more sensitive than shorter-term tests that are initiated at the sensitive fry stage. Conceptually, whether this phenomenon is important depends on the assumed exposure scenario. If it were assumed that spawning habitats would be exposed, then the less-sensitive ELS tests would be relevant. However, for migratory fishes such as listed salmon and steelhead, life histories often involve spawning migrations to headwater reaches of streams, followed by downstream movements of fry shortly after emerging from the substrates, and followed by further seasonal movements to larger, downstream waters to overwinter (Baxter 2002, Quinn 2005, Willson 1997). These life history patterns often correspond to common human development and metals pollution patterns where headwater reaches likely have the lowest metals concentrations, and downstream increases occur due to point source discharges or urbanization.

From the discussion of the types of chronic data with fish that are acceptable for use in criteria development, it is clear that the intent was to capture information on the most sensitive life stage of a fish species. Unfortunately, the wording of the Guidelines could be interpreted to preclude the use of the more-sensitive chronic growth and survival tests that were initiated with salmonid fry stage, and specify the use of the less-sensitive ELS tests (Stephan *et al.* 1985, p. 44).

Summary: Based on this analysis, the risks of acclimation on susceptibility to chemicals are likely to result in sublethal effects, such as interference in physiochemical processes,

interruption of ecological interactions, changes in pathological stress, and toxicosis of listed species considered in this opinion.

***Toxic Responses of Different Species and Life Stages.*** The chemical concentrations causing toxic effects differ between taxa, with some species being more sensitive than others. The EPA's national water quality criteria, on which the proposed criteria are based, were developed from toxicity data compiled for a wide range of species and life stages and were determined on the basis of protecting roughly 95% of the species considered. However, because the criteria were not developed specifically to protect the most sensitive species or life stage present, it is possible that the proposed criteria may not be protective when that species and life stage is a listed species, *i.e.*, a species at risk of extinction. This is recognized in the Guidelines which indicate that it is possible to revise the criteria if it is determined that there is a more sensitive species and life stage present (EPA 1994a).

The EPA identified SMAVs in their criteria documents for most of the pollutants subject to this consultation that differ between species of salmon and trout. SMAV's for marine mammals, sea turtles, green sturgeon, and eulachon have not been developed. However, the SMAVs were in most cases based on limited toxicity testing data collected under varying conditions, and therefore may not be indicative of actual species differences. Moreover, SMAVs are not completely protective of listed species because they represent an average condition, where lower concentrations may be toxic to those species under certain test conditions. There is evidence that under similar testing conditions, some trout species have similar toxic responses (*e.g.*, rainbow and brown trout, Cohen *et al.* 1993). There is also evidence of differences in toxicity response between species when exposed to specific metals or organic compounds under similar conditions (*e.g.*, Chinook and coho salmon, Hamilton and Buhl 1990; Chinook salmon, Chapman 1978b; rainbow and brook trout, Holcombe and Andrew 1978; brown trout, Chinook and coho salmon, Macek and Allister 1970, Katz 1961; rainbow trout, and Chinook and coho salmon, Macek *et al.* 1969, Katz 1961), so species differences cannot be completely discounted. Overall, however, experimental evidence (including data presented in the various EPA water quality criteria documents) suggests that there is greater variation in toxic response between life stages than between species within the family Salmonidae.

Since a species can only be considered protected from acute toxicity if all life stages are protected, EPA's Guidelines recommend that if the available data indicate that some life stages are more resistant than other life stages by at least a factor of two, the data for the more resistant life stages should not be used to calculate species mean acute values (Stephan *et al.* 1985). Smaller, juvenile life stages of fish are commonly expected to be more vulnerable to metals toxicity than larger, older life stages of the same species. For instance, a standard guide for testing the acute toxicity of fish (ASTM 1997) recommends that tests should be conducted with juvenile fish (that is, post-larval or older and actively feeding), usually in the size range from 0.1 to 5.0 g in weight.

A review of several data sets in which salmonids of different sizes were similarly tested shows that even among juvenile fish in the 0.1 to 5.0-g size range, differences in sensitivity can approach a factor of 10. This emphasizes the importance of EPA's Guidance not to use the more resistant life stages. However, the data sets analyzed by NMFS indicated that in practice, there

were sometimes greater influences of life stage on the sensitivity of salmonids to some substances than was apparent to the authors of the individual criteria documents using the datasets available to them at the time. Some of the SMAVs and GMAVs which were used to rank species sensitivity and set criteria were considerably higher than $EC_{50}$s for salmonids that were tested at the most sensitive life stages (Figure A4).

For three Pacific salmonid species for which comparable test data were available for different life stages (coho salmon (*O. kisutch*), rainbow trout (*O. mykiss*) and cutthroat trout (*O. clarki*), the data suggest that swim-up fish weighing around 0.5 g to about 1 g may be the most sensitive life stage. None of the data sets or published studies NMFS examined in detail had sufficient resolution to truly define what weight fish was most sensitive to metals, but along with other data they suggest that larger fish are less sensitive than fish at 0.4 to 0.5 g. For instance, with zinc, rainbow trout in the size range of about 0.1 to about 1.5 g were consistently more sensitive to zinc in two studies with multiple tests in that size range. The paucity of data with salmonids in the size range of about 0.5 to 2 g prevents definitive identification of a most sensitive size across species or even tests. All data located for early swim-up stage *Oncorhynchus* in the 0.1 to 0.5 g range were consistent with increasing sensitivity with size. With Hansen *et al.* (1999b) rainbow trout studies, this relationship continued with fish up to about 1.5 g. However, with cutthroat trout, the few data available suggests that fish larger than about 0.5 g are less sensitive with increasing size.

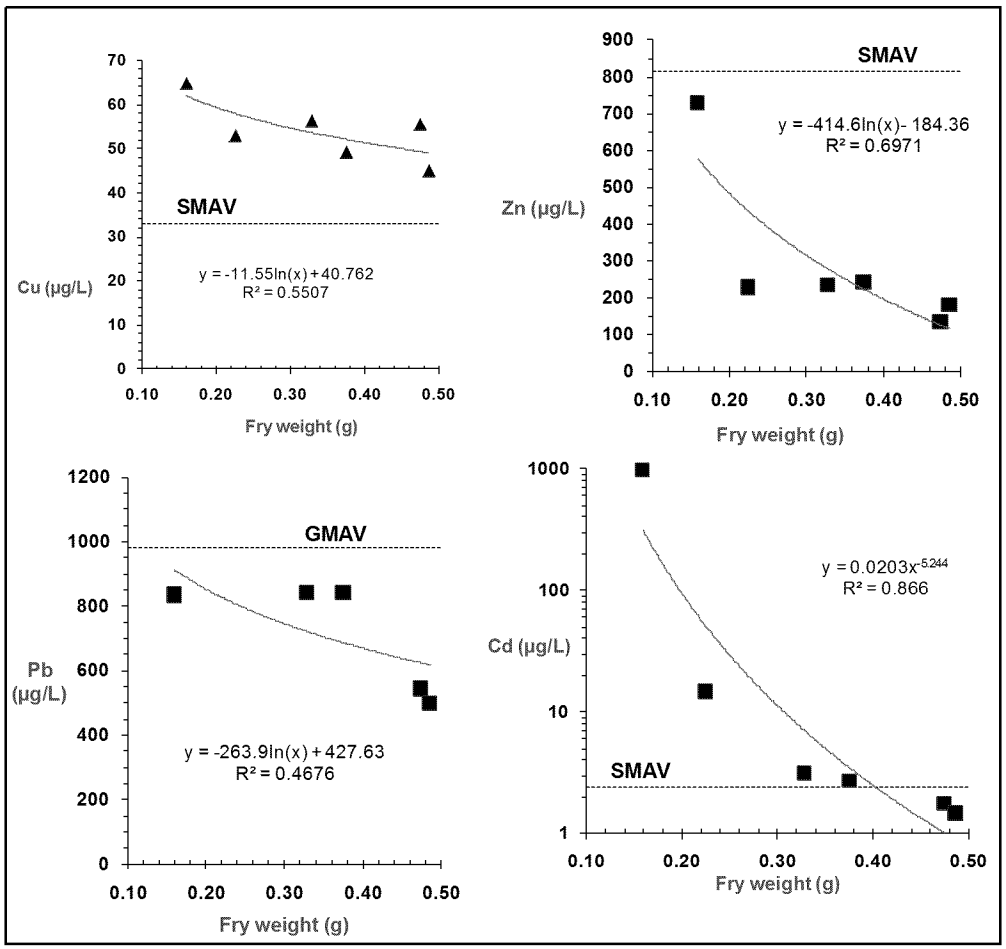


**Figure A4.** Size-developmental stage patterns SMAVs and GMAVs with coho salmon from 2 to 7 weeks posthatch, with data from Chapman (1975), and EPA (1984a, 1984b, 1985, 1987), adjusted to test water hardness. All tests used Willamette River water, TOC 3.4 mg/L, hardness 22 mg/L.

Some studies with older and larger rainbow trout have found that the fish became more resistant to zinc and copper (Chakoumakos *et al.* 1979, Chapman 1978b, Chapman and Stevens 1978, Howarth and Sprague 1978). Studies with copper all showed this trend, but the strength of size-sensitivity relations varied across studies. Chakoumakos *et al.* (1979) found that fish between about 1 and 25g in weight varied in their sensitivity to copper by about 8 times, but steelhead (*O. mykiss*) that were tested with copper at sizes of 0.2, 7, 70, and 2700 g showed little pattern of sensitivity with size (Chapman and Stevens 1978, Chapman 1978b). However, the large differences in sizes may have missed changes at intermediate sizes in the ranges compared (Figure A4). Similarly, with copper and rainbow trout, Anderson and Spear (1980) found that rainbow trout at sizes of 3.9, 29 to 176 g had similar sensitivities.

The NMFS reviewed several data sets indicated increasing susceptibility of salmonids to at least metals with increasing size and age as fish progressing from the resistant alevin stage. These

patterns indicate caution is needed when using SMAVs or GMAVs as a summary statistics for ranking species sensitivity or setting criteria.

Salmonids can have profound difference in susceptibility to chemicals at different life stages and in some instances SMAVs used in criteria may be skewed high because insensitive life stages were included. Across several good datasets, the most vulnerable life stage and size appeared to be swim-up fry weighing between 0.5 and 1.5g.

Summary: Based on this analysis, the risks from relying on toxicity data from species and life stages that are less sensitive than the most sensitive salmonid life stage is moderate to high, as aquatic life criteria derivied following the Guidelines is likely to result in sublethal effects, such as interference in physiochemical processes, interption of ecological interactions, changes in pathological stress, and toxicosis of listed species considered in this opinion.

***Bioconcentration and Bioaccumulation Factors Used in Determining and Evaluating Proposed Criteria Associated with High Variability and Uncertainty.*** An important problem with many of EPA's chronic criteria for organic pollutants is that the bioconcentration or bioaccumulation factors used in their determination may not be accurate. The BCFs determined in the laboratory based on water-borne exposure are typically much lower than field-derived values, and so may significantly underestimate uptake in the natural environment. Even among field-derived bioconcentration factors, estimates can vary by several orders of magnitude. Consequently, it is difficult to determine if BCF-based comparisons of water-borne and tissues concentrations are accurate when evaluating the chronic criteria proposed in this action.

The Guidelines include a component designed to assure that the water quality criterion for a substance is sufficiently low that residue accumulations will not impair the use of a waterbody by aquatic organisms, and specify that data from residue studies are to be considered alongside acute and chronic toxicity data in the criteria development process (EPA 1985a). However, metals criteria are presently based solely on results of aquatic toxicity tests (62 FR 42159), where metal exposures occur directly across gills or other respiratory surfaces.

Metals and organic contaminants can bioaccumulate, through either bioconcentration (an increase in concentration of a substance in relation to the concentration in ambient water) or biomagnification (a progressive increase in concentration from one trophic level to the next higher level in the aquatic food chain (Moore and Ramamoorthy 1984, Sorensen 1991).

All of the organic pollutants of concern in this action bioaccumulate. All biomagnify to some extent in the food chain, although this is more of a serious concern for some contaminants than others. The Guidelines include a component designed to address the risks of elevated fish tissue residues of organic compounds to humans and avian and mammalian predators, but not the risk of that residue to fish (EPA 1985a). In fact, this process drives nearly all of the numeric criteria established for organic contaminants. What is not considered in these evaluations, however, is whether these tissue residues would directly affect the health of the aquatic organisms. Similar to metals, the consumption of aquatic invertebrates by fish is never formally considered in the development of the criteria for organic compounds. It is well established that invertebrates may accumulate organic contaminants in aquatic systems, and that these contaminants are passed on

to fish through the diet (*e.g.*, Streit 1998). Consequently, if the water quality criteria do not protect invertebrate prey species from organic residue accumulations, they may not protect listed species from adverse effects associated with dietary exposure.

In particular, measuring compliance with the criteria through ambient water concentrations alone leaves exposure pathways to several organic pollutants un-regulated. For example, dieldrin, lindane, and heptachlor epoxide are not highly water soluble, and are persistent in both food and sediments. A number of the organic compounds reviewed here (*e.g.*, dieldrin, lindane, heptachlor epoxide), have considerable potential to biomagnify in aquatic systems (Suedal *et al.* 1994). The Guidelines for such compounds do not consider food web transfer and bioaccumulation with respect to the target species. Consequently, they may greatly underestimate the toxicity of these chemicals in the environment. This is particularly important for the juvenile life stage of anadromous salmonids while they reside in rearing habitat, if such exposure later influences their downstream migration and subsequent ability to osmoregulate as they enter saltwater. This is an especially significant concern for organic contaminants such as organochlorine pesticides (*e.g.*, dieldrin, lindane, heptachlor epoxide), for which exposure is primarily via sediments and tissues of prey organisms.

A biologically significant pathway for exposures of aquatic organisms to contaminants is through consumption of contaminated aquatic detritus, plants, invertebrates, and other food items (bioaccumulation). Invertebrates that can accumulate metals in aquatic systems are often prey consumed by salmonids and other fish species (*e.g.*, Moore *et al.* 1991, Luoma and Carter 1991, Cain *et al.* 1992, Kiffney and Clements 1993, Rainbow and Dallinger 1993, Timmermans 1993, Ingersoll *et al.* 1994, Dallinger 1994, Cain *et al.* 1995, Gerhardt and Westermann 1995).

In an experiment that shows how readily contaminated food items lead to elevated fish tissue concentrations, Woodward *et al.* (1994) held paired groups of age 0 rainbow trout in clean and contaminated over a range of metal-concentrations. They fed one group a diet of reconstituted, metals contaminated invertebrates, and the other group a comparable diet based on uncontaminated invertebrates. After 91 days, they observed that only fish fed the contaminated diet exhibited reduced survival and growth. These results demonstrate that exposure to a dissolved metal can be a secondary hazard pathway in cases where food is contaminated and fish can bioaccumulate the substance of concern. In cases where fish can bioaccumulate a metal, these results and similar results from other studies of diet-borne metal exposures to salmonids collectively indicate that toxic effects can occur through dietary pathways (*e.g.*, Dallinger and Kautzky 1985, Dallinger *et al.* 1987, Spry *et al.* 1988, Giles 1988, Harrison and Klaverkamp 1989, Harrison and Curtis 1992, Miller *et al.* 1993, Mount *et al.* 1994, Farag *et al.* 1994).

In general, the metals considered in this opinion do not appear to biomagnify in the food chain, with the exception of selenium. The Guidelines include a component designed to assure that the water quality criterion for a substance is sufficiently low that residue accumulations will not impair the use of a waterbody by aquatic organisms, and that data from residue studies are to be considered alongside acute and chronic toxicity data in the criteria development process (EPA 1985a). However, metals criteria are presently based solely on results of aquatic toxicity tests (62 FR 42159), where metal exposures occur directly across gills or other respiratory surfaces.

Risk management via water concentration-based water quality criteria is not protective of listed salmonids for toxic pollutants that strongly bioaccumulate (*e.g.*, selenium, and organic pollutants: Pease *et al.* 1992; Taylor *et al.* 1992, 1993; Canton 1997; EPA 2001). This is because the true potential for toxic hazards to fish and wildlife through bioaccumulation is determined not only by an immediate water-borne exposure and direct toxicity effects, but also by the rate of mass loading into an aquatic ecosystem, the corresponding environmental partitioning of mass loads between the water column, sediments, and biota (food chain), and how the toxic pollutant is assimilated and acts on the organism. A water column concentration of a toxic pollutant may not reflect mass loading or be reflected in food chain bioaccumulation. Therefore, water quality criteria are useful guides for risk management only to the extent that they protect aquatic food chains from bioaccumulation.

This is an especially significant concern for organic contaminants such as organochlorine pesticides, for which exposure is primarily via sediments and tissues of prey organisms. Indeed, environmental agencies in some other countries, including Canada, no longer recommend water quality guidelines for these substances, but regulate them through other media such as sediment, soil, or tissue (CCREM 2001a).

Because hydrophobic compounds are expected to show a similar or proportional affinity for the lipid of an organism as that for octanol (which is used to calculate the partition coefficient[21]), the degree of partitioning exhibited between water and octanol, as characterized by the partition coefficient $K_{ow}$, can be a useful means for evaluating and predicting bioaccumulation (Mackay 1982, Di Toro *et al.* 1991). For organic compounds that are not metabolized, the relationship between the bioconcentration factor (BCF) and $K_{ow}$ is strong (Mackay 1982). The expected wet-weight BCF for a non-metabolized hydrophobic compound is a function of the lipid content of an organism and the value of $K_{ow}$ for the compound. The standard equation for determining the expected BCF is:

$$BCF = 0.046 \times K_{ow}$$

which is derived from fish studies and is based on an average lipid content of 4.6% wet weight (McCarty 1986). This relationship is used in this opinion for evaluating effects related to exposure and bioconcentration of the toxic organic pollutants addressed by the ODEQ.

Sediment concentrations that would result in organic toxic pollutant concentrations in the water column can be calculated using the equation (Di Toro *et al.* 1991):

$$SQC_{oc} = K_{oc} \times F_{CV}$$

where:

$SQC_{oc}$ = sediment contaminant concentration in mg/kg organic carbon
$K_{oc}$ = partitioning coefficient for sediment organic carbon

[21] A coefficient representing the ratio of the solubility of a compound in octanol (a non-polar solvent) to its solubility in water (a polar solvent). The higher to $K_{OW}$, the more non-polar the compound. Log $K_{OW}$ is generally used as a relative indicator of the tendency of an organic compound to adsorb to soil. Log $K_{OW}$ values are generally inversely related to aqueous solubility and directly proportional to molecular weight.

$F_{cv}$ = the chronic water quality criterion in µg/L

$K_{oc}$ can be calculated from the octanol/water partitioning coefficient, Kow, using the formula:

$Log_{10}(K_{oc}) = 0.00028 + 0.983 \times Log_{10}(K_{ow})$

This equation is used in the analysis of effects later in this opinion, provided that the data necessary to conduct the analysis were available, to evaluate the potential for water-borne exposure concentrations of organic pollutants at or below criteria concentrations.

Summary: Based on this analysis, the risks of bioconcentration and biooaccumulation factors are likely to result in sublethal effects, such as interference in physiochemical processes, interruption of ecological interactions, changes in pathological stress, and toxicosis of listed species considered in this opinion.

***Insufficient Information on Behavioral and Other Sublethal Endpoints.*** In the case of chronic criteria, data are available for a range of sublethal effects such as growth and fecundity or sperm production. However, some important effects reported in mammals, such as immunosuppression and endocrine disruption, are inadequately studied in salmonids therefore were not considered in the development of the national criteria. These sublethal effects cannot be considered trivial, because they are associated with the potential for increased mortality (Arkoosh *et al.* 1998). Sublethal effects involving alterations in behavior can occur during relatively low concentration, short-term exposure, and can have profound biological implications (*e.g.*, chemical migration barrier, interference with spawning behavior). The NMFS recognizes that relevant data may not be available for all toxic substances, and that determination of a repeatable, detectable endpoint may involve a degree of subjectivity. Relatively little data are available to help elucidate these concerns; however, the research that does exist indicates that sublethal effects can be very serious for at least some toxicants.

Summary: Based on this analysis, the risks of sublethal effects will exacerbate adverse effects, and are likely to result in sublethal effects, such as interference in physiochemical processes, interruption of ecological interactions, changes in pathological stress, and toxicosis of listed species considered in this opinion.

***Influence of Temperature, pH, and other Water Quality Stressors on Fish Response to Toxicity.*** In addition to direct influences on toxic pollutant speciation and chemical toxicity mechanisms, several water quality parameters influence general fish health, and susceptibility and ability to acclimate to and depurate after short-term increases in toxic parameter concentrations. This is generally addressed indirectly (with respect to toxicity) through conventional water quality criteria (*e.g.*, water temperature, pH, dissolved oxygen, dissolved gases, ammonia, *etc.*). However, it is possible for fish to be stressed or become stressed more rapidly when conventional water quality parameters are near or exceed criteria limits. This effect pathway is not addressed by most existing toxic pollutant criteria, and represents a shortcoming of the proposed criteria.

Summary: Based on this analysis, the risk that temperature, pH, and other water quality stressors will exacerbate the effects of the proposed criteria is high, as aquatic life criteria derived following the Guidelines do not take these additional stressors into account and are therefore likely to result in sublethals effects, such as interference in physiochemical processes, interption of ecological interactions, changes in pathological stress, and toxicosis of listed species considered in this opinion.

***Toxicity of Total Recoverable vs. Dissolved Metal Concentrations and the Use of Conversion Factors and Translators.*** Acute and chronic criteria for metals may be interpreted using either total recoverable or dissolved metal concentrations, depending on the objective of the study. The term "total recoverable" metal refers specifically to metal concentrations determined in unfiltered samples that have been acidified (pH < 2) before analysis. The term "dissolved" metal refers specifically to metal concentrations determined in samples that have been filtered (generally a 0.45 micron pore size) prior to acidification and analysis. Total recoverable metal concentration includes both the dissolved form and the portion either attached to particles in the water or present in suspended insoluble form. Particulate metals can be single atoms or metal complexes adsorbed to or incorporated into silt, clay, algae, detritus, plankton, *etc.*, which can be removed from the test water by filtration through a 0.45 micron filter.

Only dissolved metals are immediately bioavailable and thus immediately toxic to freshwater organisms (however, the particulate form may still affect listed species, as discussed below). The non-dissolved form is generally not directly hazardous to listed salmonids except under certain circumstances were (1) changes in water chemistry conditions lead to increased solubility from particulate forms within the water column, or (2) metal contaminated particulates are ingested or encounter gill surfaces. Factors in addition to hardness that influence solubility, and thus bioavailability and toxicity, include suspended sediment concentration, pH, organic carbon content, and chemical speciation of the metal. Further, some metal compounds are less soluble than others for a given set of water quality conditions.

Studies indicate that particulate metals contribute to organism exposure to metals. Particulates may act as a sink for metals, but they may also act as a source. Through chemical, physical, and biological activity, particulate metals can become bioavailable (Moore and Ramamoorthy 1984). Particulate and dissolved metals that end up in sediments are not rendered entirely nontoxic nor completely immobile, and may still contribute to the toxicity of the metal in natural waters. Of special concern are situations where waters contain both high particulate metal concentrations and dissolved concentrations near the proposed criteria. Additionally, those metals that can bioaccumulate through food-chain organisms and can cause indirect effects through particulate metal contamination.

Particulate metals are removed from the proposed regulatory "equation" through at least two methods: the use of CFs to determine the dissolved metal criteria from total recoverable criteria, and the use of a translator to convert back to a total metal concentration for use in waste load limit calculations. When waste discharge limits are to be developed and TMDLs are determined for a receiving waterbody, the dissolved criterion must be "translated" back to a total concentration because TMDLs are based on total metals.

EPA originally used total metal concentrations to establish national criteria, as provided in the National Toxics Rule published in 1992. The EPA subsequently changed to use of dissolved metal criteria, as explained in a 1993 policy statement:

> [I]t is now the policy of the Office of Water that the use of dissolved metal to is now the policy of the Office of Water that the use of dissolved metal to set and measure compliance with water quality standards is the recommended approach, because dissolved metal more closely approximates the bioavailable fraction of metal in the water column than does total recoverable metal. This conclusion regarding metals bioavailability is supported by a majority of the scientific community within and outside the Agency. One reason is that a primary mechanism for water column toxicity is adsorption at the gill surface which requires metals to be in the dissolved form (Prothro 1993).

Because no supporting references were given in support of the policy, it is hard to evaluate. There is theoretical support for the assumption that metals need to be in dissolved form to adsorb to the gill surface (Wood *et al.* 1997), and it does seem logical to assume that metals bound to particulates would be less toxic. However, two studies that examined the toxicity of particulate metals in controlled experimental studies (Brown *et al.* 1974, Erickson *et al.* 1996) found toxicity associated with particulate bound copper.

Erickson *et al.* (1996) estimated that the adsorbed copper has a relative toxicity of almost half that of dissolved copper, and noted that the assumption that toxicity can be simply related to dissolved copper was questionable, and a contribution of adsorbed copper to toxicity cannot be generally dismissed (Erickson *et al.* 1996). One possible reason for the observed toxicity from particulate-bound copper is that the pH of water changes as it crosses the gills of fish, and at pH of 6 or greater in the water where a fish is living, the pH of water will be lowered as it crosses the gill (Playle and Wood 1989).

Attempting to define, evaluate and manage risks associated with dietary exposures of metals or contaminated sediments by basing criteria on total recoverable metals would likely be so indirect as to be ineffective. However, in the absence of such efforts, the stance that metals sorbed to particles are in effect biologically inert and can safely be ignored is questionable. The effect of this stance is to give up some conservatism in aquatic life criteria for metals.

***Conversion Factors.*** The EPA derived ambient dissolved metals criteria from aquatic toxicity tests that produced dose-response relationships in test organisms under controlled (laboratory) conditions. In most of these studies, organism responses were plotted against nominal test concentrations of metals or concentrations determined by analyzing unfiltered samples to which soluble metal compounds had been added. Thus, until recently, metals criteria have been expressed in terms of total metal concentrations. Current EPA metals policy (EPA 1993a) and the ODEQ stipulate that criteria be expressed on a dissolved basis. The CF used in the EPA formulae for computing criteria represents a corresponding adjustment so that criteria based on total metal concentrations used in laboratory testing can be "converted" to a dissolved basis actually present in the toxicity test solutions. Metals for which a CF has been applied include arsenic, cadmium, chromium, copper, lead, nickel, silver, and zinc.

CF values for the proposed metals criteria are near 1.0 for most metals, because they were determined using laboratory toxicity-test solutions prepared with purified, soluble metal compounds, rather than using natural waters where relative contributions of water-borne particulate metals are much greater. To develop the coversion factors, EPA reviewed test data that reported both total and dissolved concentrations in their test waters and also conducted simulations of earlier experiments to determine the dissolved to total ratios (60 FR 1536, 62 FR 42159). In this way, the historical toxicity database could be utilized and a large number of new toxicity tests would not have to be performed. However, the CFs in many cases (*e.g.*, As, Ni, Cr, Pb) developed based upon a small number of studies and samples compared to the historical database of toxicity tests. Although additional confirmatory studies were performed to develop the CFs, the database available appears to be limited and calls into question the protectiveness of the CFs determined for these metals in cases when site-specific water quality approaches toxic conditions.

***Translators.*** The EPA provides three methods to translate criteria based on dissolved metals to permit-specific criteria based on total recoverable metals. These three methods may result in greatly different outcomes relative to particulate metal loading. These methods are::

1. Determination of a site-specific translator by measuring site specific ratios of dissolved metal to total metal and then dividing the dissolved criterion by this translator. As an example, a site specific ratio of 0.4 (40 percent of the metal in the site water is dissolved) would result in a 2.5-fold allowable increase in the discharge of total metals. The higher the fraction of particulate metal in the site water the greater the allowable discharge of total metal. This is EPA's preferred method.

2. Theoretical partitioning relationship. This method is based on a partitioning coefficient determined empirically for each metal, and (when available), the concentration of total suspended solids in the site-specific receiving water.

3. The translator for a metal is assumed to be equivalent to the Guidance conversion factor for that metal (*i.e.,* use the same value to convert from total to dissolved and back again).

Since translators are needed to calculate discharge limits they become important in determining the total metals allowed to be discharged. In California, economic analyses performed by the EPA and evaluated by the State Water Resources Control Board (SWRCB 1997) indicated that translators based on site-specific data would decrease dischargers costs of implementing the new CTR criteria by an estimated 50%. This cost savings is "directly related to the less stringent effluent limitations that result from the use of site-specific translators," and implies a strong economic incentive for dischargers to reduce costs by developing site-specific translators and ultimately being allowed to discharge more total metals. This conclusion regarding the impact of site specific translators is supported by documents received by the NMFS in the CTR consultation from EPA (*i.e.*, EPA 1997c).

The EPA performed a sensitivity analysis on the effect of the site specific translator, which relies on determining the ratio of metal in water after filtration to metal in water before filtration in downstream waters. The EPA's analysis indicated that use of a site-specific translators to

calculate criteria would result in greater releases of toxic-weighted metals loads above the option where the CFs are used as the translators. The potential difference was estimated to be between 0.4 million and 2.24 million "toxic weighted" pounds of metals discharged to California waterways (USFWS and NMFS 2000). Lastly, the current use of conversion factors and site specific translators in formula-based metal criteria is not sufficiently protective of threatened and endangered aquatic species because:

- Particulate metals are not regulated, yet chemical, physical, and biological activity can subsequently cause these particulate metals to become bioavailable and cause adverse effects.
- Particulate metal concentrations are not always negligible in critical habitat in Oregon.
- The national criteria were developed using toxicity tests that expose test organisms to metal concentrations with very low contributions from particulate metals.
- Toxicity tests do not assess whether the toxic contributions of particulate metals are negligible when particulate concentrations are great and dissolved concentrations are at or near criteria levels.
- This method has the potential to allow point sources to significantly increase the discharge of total metal loads into the environment, even though dissolved metal criteria are being met by a discharger.
- Metal loading occurs from the water column to streambed sediments.

Summary: Based on this analysis, the risks of using conversion factors and translators is likely to result in sublethal effects, such as interference in physiochemical processes, interruption of ecological interactions, changes in pathological stress, and toxicosis of listed species considered in this opinion.

***The Water-Effect Ratio Provision***. The water-quality criteria for metals all include a WER in their formulas. The WER is the ratio of the test $LC_{50}$ in site water divided by the $LC_{50}$ in laboratory water; the ratio is then multiplied by the aquatic life criteria to obtain a WER-adjusted site-specific criteria. The approach has probably been most used with copper because of the profound effect of organic carbon (DOC) to ameliorate toxicity, which is not correlated with hardness. The purpose of WERs is to empirically account for characteristics other than hardness that might affect the bioavailability and thus toxicity of metals on a site-specific basis. Because the WERs are directly incorporated into the criteria equations, no separate action is needed to change the criteria values using a WER. The default WER value is 1.0 unless DEQ determines that a different value should apply.

The concept of adjusting metals criteria to account for differences in their bioavailability in site waters has long been a precept of water quality criteria (Bergman and Dorward-King 1997, Carlson *et al.* 1984, USEPA 1994). The WER approach uses one or more standard-test species (usually *Ceriodaphnia* and/or fathead minnows), which are tested in tandem in dilution waters collected from the site of interest and in standard reconstituted laboratory water. The results in the laboratory water are presumed to represent the types of waters used in tests relied on by EPA in criteria documents.

The main problem with this concept and approach is trying to define a single "typical" laboratory dilution water that reflects that used in criteria documents. Testing laboratories may generate valid results using all sorts of different dilution waters including dechlorinated tap water, natural groundwater (well water), natural surface water such as Lake Superior or Lake Erie, and reconstituted waters made from deionized water with added salts. The widely used "Interim Guidance on Determination and Use of Water-Effect Ratios for Metals" (Stephan *et al.* 1994) specified using recipes from EPA or American Society for Testing and Materials (ASTM) for making standardized test water that results in a water hardness with unusually low calcium relative to magnesium concentrations compared to that of most natural waters. This has the effect of making metals in the reconstituted laboratory water made by standard recipe more toxic than would be expected in water with more natural proportions of Ca and Mg. This is because, at least for fish and some invertebrates and copper, Ca reduces toxicity but Mg affords little or no protection (Borgmann *et al.* 2005, Naddy *et al.* 2002, Welsh *et al.* 2000). Lastly, the water-effect ratio seems to have always been recognized by EPA as an interim, operational substitute to establishing criteria on a more mechanistic basis that could directly account for a lot of the factors that affect toxicity. A major development toward this is the biotic ligand model (BLM) which is supposed to capture the major interactions between metals concentrations, competition, and complexation, which control bioavailability and thus toxicity (Di Toro *et al.* 2001, Niyogi and Wood 2004). For copper, the BLM was used as the basis of EPA's (2007) updated aquatic life criterion, which for copper at least, should negate much of the need for empirical WER testing.

Summary: Based on this analysis, the risks of using water-effect ratios is likely to result in sublethal effects, such as interference in physiochemical processes, interruption of ecological interactions, changes in pathological stress, and toxicosis of listed species considered in this opinion.

***Summary of the Derivation of the EPA Aquatic Life Criteria.*** Based on the analysis on the derivation of the EPA aquatic life criteria, NMFS concludes that predicted effects associated with the aquatic life criteria are likely to be significantly greater than asserted and are likely to have significant consequences for field-exposed species.