NAVFAC NW Comments on "Fish Consumption Rates Technical Support Document"
Public Review Draft, Version 2.0, August 27, 2012

Page	Comment
4	We appreciate that this document separates data and science from policy decisions.
7	We suggest that, in addition to values for the general population, Ecology consider estimate consumption rates for the identified subpopulations: Native American tribal nations, Asian and Pacific Islanders, and recreational fishers.
11	If "[t]ens of thousands of recreational sport clammers harvest razor clams" and 3,601,000 lbs are harvested annually, on average each clammer will obtain less than 400 lbs annually.
15	Ecology's estimates of fish consumption must be based on some assumptions, e.g., form of the distribution (normal or log-normal) and the characteristics of that distribution. A 90th percentile can be quite different depending on such assumptions, and they should be included in this document.
15 second bullet	This bullet discusses finfish and shellfish in the first sentence and "fish" in the second. It is not clear whether "fish" is meant to be inclusive or only refer to finfish. The glossary (Appendix D) suggests that "fish" does not include shellfish, in which case the method for determining the 90th percentile, as discussed in the first sentence is not complete: The source of information for the shellfish component is not given. Several times, the document states, "As noted, estimates of fish consumption that correspond to the 90$_{th}$ percentile of the distribution may vary depending on the statistical methods used to evaluate the national data." but does not provide the information to clarify what decisions were made. This issues is further obfuscated by the statement on page 17 that the evaluation is of consumption of "finfish and/or shellfish per day".
19	The last paragraph is unclear. Is the high consumption rate the amount that **may be consumed on any given day** or the amount that **high consumers eat every day?** Ecology should clarify this, as EPA's high consumption data are often interpreted as the amount that might be consumed on some, but not every, day. Similarly, this would apply to the comment in the last sentence on the page about high levels of consumption – as well as elsewhere in the document. This should be clarified.
22	Third bullet: Although we understand that the procedures have been reviewed by numerous government scientists, the issue of "technical defensibility" will always be unresolved as long as the methods used for the analyses are not clear and transparent, both by readily available document or within the document itself.
35	Respecting the desire of the tribal organizations to keep their raw data private, nevertheless just a little more data for the summary, in addition to the mean, median, and percentiles, would greatly enhance our ability to understand the data. Two additional data points, i.e., the highest data point and the lowest data point used in each analysis, would allow additional evaluations. Several techniques are available that require knowledge only of the range of the data, not the distributional form, that can enhance our understanding of the data. If Ecology or

	others are not familiar with these techniques, these data should be provided so that others can use those procedures if they wish.
40	Although these comments address the values in Table 17, they also apply to all of the similar tables in this document. • Given that the mean is larger than the median, the data are clearly skewed. Was particular distribution, e.g., a log-normal distribution, assumed? If so, please state both the distribution and the summary metric such as the geometric standard deviation. • Was the distribution truncated, and if so, how were the lower and upper bounds determined. The data look as though the distribution was not truncated. Not only might this affect the mean, but it would have a very significant effect on the 90th, 95th, and 99th percentile values of such a very skewed distribution – as can be seen from the large difference between the 95th and 99th percentile in this table.
41 & 42	Tables 18 & 19; Figures 1 &2: The significantly large differences in the two methods for analysis of the data are not well explained in the text. Thus, the reason for Ecology's choice of method is not clear and transparent to the reader.
45	The text states that the data were weighted by population size, but does not indicate how. At least two methods are possible, e.g., inversely by population size or inversely by the size of the uncertainty in the data. The process should be specified.
50	The exclusion of outliers that "seemed unreasonably high" is a judgment call. It would be useful to know how many outliers were excluded and their range (high and low) so an independent evaluation of the potential effect on the analyses could be performed (see comment for page 35). Furthermore, **this statement is not consistent with the statement on page 21 in Attachment C** [emphasis added], "First, even the *largest consumption rates* reported for these tribes and for other populations covered in the current report *are plausible*. They may be large, but there is no overriding reason to designate them as impossible. The second reason that the rates have been left intact (*with no adjustment for "outliers"*) is the potential for bias in any adjustment. ... If only the highest rates are adjusted downward, then the mean and the high-end percentiles calculated after such adjustments will be biased downward." This inconsistency should be addressed.
53, 59, & 60	• The tables and figures on this page each have one dataset for which the 75th percentile is lower than the mean. • In other datasets **and in other tables**, the mean and 75th percentiles are similar, and with the absence of confidence limits, may be statistically the same. While possible, these observations suggest that the tails of the distributions are very long and not very high. This is another indication that either (1) the distributions should be truncated, (2) the 90th and 95th percentiles may not be reasonable estimates, (3) the data may substantially benefit from analyses that do not assume a mathematically defined distribution (see comments for pages 35 and 40), and (4) the discarded outliers might provide information that would better substantiate these analyses.

65	While the 90th percentile is used as the RME, EPA Region 10 reanalysis uses the 95th. This is confusing, as the 90th percentile is available in the tables. Ecology should be consistent in its evaluation of the RME and include this definition in the glossary to provide clarity.
71	• No reference is provided either in the text nor in the reference list for the NCI "methodology"; please provide one. • Furthermore, the text does not discuss how the procedures differ between that used for the national data and that used for the regional data for the tier two subpopulations. Even if the data cannot be provided, a clear discussion of the differences in the procedures, as well as the implications on the resulting values, should be able to be provided by the experts who participated in the review of these statistical analyses.

77 and following pages	Chapter 5 discusses uncertainty and variability qualitatively. This analysis contains no quantitative indication of how any of these factors may affect results.
	• Even if an exact uncertainty analysis is not performed, the experts involved in these analyses should be able to provide a rough estimate of the potential effects, e.g., 2-fold, 10-fold, greater than 100-fold. This was done for some national data, as presented near the middle of page 4 in Attachment A. As the raw data are usually not available, such estimates from experts who evaluated the raw data are even more critical.
	• Such estimates are even more critical if the RME is assumed to be a daily consumption. If, as Harris and Harper (1997) state, that more than a pound of locally harvested fish and shellfish is regularly consumed daily, the analysis of variability and uncertainty should support such statements.
	• The number of people in each of the surveys of the discrete populations is quite small. Thus, the size of the uncertainty is likely to be proportionately larger, unless the populations are more uniform. A quantitative estimate of the uncertainty and variability in these data would allow readers to determine which of these two options is occurring in each of the datasets.
	• Notably absent from these analyses is any estimate of a modal value (see footnote, pg 92 of document).
	• This chapter does not examine the feasibility of the estimates of fish consumption. Several such analyses are possible.
	  o Does the population consuming the seafood multiplied by the amount consumed exceed the ability of the local source to provide that quantity?
	  o Does the amount consumed per person exceed the food mass or calories assumed for a diet for an RME?
	  o Is the amount consumed consistent with information on the height and weight of the population being evaluated?

98	The RAGS equations presented on this page are at best an approximation of reality, and the document should so state. • Risk is not exposure time toxicity, rather risk is a function of exposure and toxicity. If Ecology is improving the parameters for situation-specific risk evaluations, it should consider also improving its models that have

	become standard procedures since the RAGS models were developed in the 1980s. It is axiomatic among risk analysts that the best data can't rescue a faulty model. • If upper and lower bounds of a parameter are used, such as the RME parameters for fish consumption, inverting the risk equation (as done here) is inaccurate. This is easily demonstrated by the use of interval arithmetic. • The acceptable cancer risk for EPA is 1 in 10,000 to 1 in 1,000,000. While MTCA may use one end of the range, this document that discusses other regulatory uses of these data should indicate the range, rather than just the value used in one program which is only stated in the caption of the figure.
Appendix E	• The reference section indicates that a reasonably large number of sources used for this analysis are in draft or preliminary form. These may be altered, and the referenced information changed, in the final version. In particular, the 1981 "Preliminary Report" by Pierce et al. from 1981 should be in final form. If it is not, then the reader must assume that there was a problem with the initial report. • At least 2 of the references are labelled "personal communication". As regulations may be based on these data, it would be appropriate to include these communications in an appendix in this document so stakeholders can review the information. • Finally, several of the references that are available on line and free do not have their associated urls. These should be added for people not familiar with these documents.
	Attachment A
	The numbering of this attachment that contains multiple documents, each with its own numbering system makes commenting difficult.
9 & 10	Tables 2, 3, and 4 are calculated from primary data but are reported to three significant figures. The underlying data are unlikely to support this degree of accuracy. For example, the averaged fish consumption is listed to the tenth of a gram, and it is unlikely that anyone reported consumption rates that accurately. Thus, 19.9 g should be rounded to 20.
10	Typo: The "meat" in footnote "b" of Table 4 should be "fish".
11	Providing references is nice, but as many of these journals are not main-stream, a brief description of the major aspects of the process would be nicer, especially in a technical report in an attachment. Specifically, we recommend that the web sites in footnotes 14 and 15 of Attachment C be included here so that the reader is not required to read all of the attachments to find this information.
5	**Kissinger 2010 is a personal communication but is not labeled as such in the text.** Since it is from the EPA, it should be publically available and therefore should be included in this document.
7	Table 1. Fish and Shellfish Consumption Rates (g/day) for Marine Recreational Fishers in King County, WA: • If the original data only have 2 significant figures, the estimated data can only have 2 significant figures.

	• The standard deviations dwarf the data. The upper-bound estimates are more dependent on the assumption about the underlying distribution than the actual data.
10	Table 2: The standard deviations dwarf the data. The upper-bound estimates are more dependent on the assumption about the underlying distribution than the actual data.
11	Table 3 has the same problems as 1 and 2.
A-2 and following pages	Several of the supporting tables in Appendix A of Attachment A indicate that the tables are "modified" from the referenced source. As these tables are in support of the analyses in the appendix that supports the main document and as there is plenty of blank space on these pages, these footnotes should briefly explain how the data were modified.
B-3	Table B-1 • The column labeled "Count" is confusing. For example, under "Bass" the count for dioxin is 35 while the count for mercury is 403 and for PCB is 20. Were 403 bass caught and only 35 of those sampled for both dioxin and mercury? Alternatively, were 35 + 403 + 20 fish caught and each only sampled for one contaminant? Please clarify. • Several analyses report both dioxin-like TEQs and total PCBs. Were the dioxin-like PCBs included in the TEQ, and if so, doesn't this double count the PCB contamination?
	Attachment C

6	Table E-1:
	• The small number of people that comprise the samples of this information, as well as the large difference between the 50th and 90th percentile, suggest that the underlying distribution assumed by these analysts was not truncated. Fish consumption is not unlimited; for example, it is highly unlikely that even a subsistent fisher would consume much more than 4 pounds of fish per day. Truncating the data is expected to substantially affect the high-value estimates used for the RME.
	• Unlike the data provided for recreational fishers, the standard deviations or standard errors are not provided for these data.
	  o If these measures of uncertainty are as large as for recreational fishers, the high-end values would be significantly dependent on the distribution selected and uncertainty would be a major factor in these analyses.
	  o If, as expected, these measures of uncertainty are lower for the more homogeneous population, the uncertainty would be less of a factor in these analyses.

22	This appears to be the only place in the document that the assumption of a lognormal underlying distribution is proffered. • A text search of both "lognormal" and "log-normal" of the main text produced no findings of this term. As mentioned in many of the other comments on this document, choice of the distribution is critical, especially when uncertainty may be high due to low sample numbers.

	• Truncation of the lognormal distribution is critical, as these tend to have long tails, see comment for page 10.
22	The interpolation method is also untruncated and therefore may result in questionable values (see comment 52). The interpolation assumed "the set of percentiles from the standard normal distribution". **The "standard normal distribution", however, has a range from -∞ to +∞; fish consumption does not.** The percentile distribution should use a truncated normal with limits of some small value (not zero as these analyses are only of people who consume fish) to some reasonably high value of fish consumption (perhaps 4 pounds). The effect of this adjustment is unknowable until it is performed, but **as these data are likely to be used for numerous regulatory activities, these should be as accurate as possible; even small changes may be important for some decisions.**
23	Table 3 and other tables on fish consumption: It is unlikely that the data support the use of more than 3 significant figures. The estimates should be edited to reflect this.
52	The results of the validation study "from using the full Tulalip individual-level data vs. the summary statistics that result from using the "means" estimation method" in Table A-3 demonstrate many of the issues raised by previous comments. • The estimation method used to derive values from summary statistics consistently underestimates the value derived from raw data. Such results would be expected when the method does not use a truncated distribution when the actual distribution must be truncated, i.e., the method elongates the tails of the estimated distributions beyond plausible limits. • This lowering of the heights of the tail provides a logical explanation for the observation that the actual means reported from the raw data are occasionally below, and often near, the estimated 75th percentile. If truncating the distribution does not rectify the anomaly, the assumption that the data are lognormally distributed (like the national data) might need to be re-evaluated. • The methods developed for the heterogeneous national population are significantly inaccurate when applied to a more homogeneous population. **It is strongly recommended that the estimates from the tribal nation surveys be recalculated with adjustments.** The two anticipated to have the greatest effect on the results are described below. 1. Whenever an underlying distribution is assumed to estimate a parameter, the distribution is truncated at both ends. The lower bound should be some value above zero and the upper bound should be a plausible limit. From the data presented, an upper bound in the range of 4 pounds seems reasonable. 2. Absent data to the contrary, the variability within the more homogeneous populations would be expected to be tighter than the national population. As with Appendix 4 of Attachment C, perhaps the data from the Squaxin Island Tribe could be used to estimate this variability, as a better approximation than the variability of the national population.