DEPARTMENT OF BIOSTATISTICS
UNIVERSITY *of* WASHINGTON
School of Public Health

Casey Olives
University of Washington
Department of Biostatistics
1959 NE Pacific St
Seattle, WA 98195

Martha Hankins, Unit Supervisor
Policy and Technical Support Unit
Toxics Cleanup Program
Department of Ecology
PO Box 47600
Olympia, WA 98504-7600

September 25, 2012

Dear Martha Hankins:

Please find attached my review of the July 22, 2012 version of the report "Statistical Analysis of National and Washington State Fish Consumption Data" by Nayak L. Polissar, Moni Neradilek, Aleksandr Y. Aravkin, Patrick Danaher, and John Kalat. The primary objective of this review is to provide an independent assessment of the technical quality of the Polissar *et al* report.

Overall, my opinion is that the technical quality of the report is quite strong. The suggestions included herein are generally confined to additional and complementary analyses that would ultimately make both the methods used and the estimates reported clearer and more transparent.

I look forward to your feedback and would be happy to provide clarifications to my comments, should they be needed.

Sincerely,
Casey Olives, Ph.D.

Encl.

cc: Craig McCormack, Nayak Polissar, Verna Blackhurst, Tom Burbacher

On the application of the NCI Method to the NHANES data

1) *Definition of Fish Consumer:* Fish consumption rate (FCR) data in the National Health and Nutrition Examination Survey (NHANES) data are available both from the two-day recall period and a supplementary food frequency questionnaire (FFQ). The latter does not provide sufficient information to estimate the FCR, but is potentially useful for filtering out non-consumers of fish.

   The authors investigate the use of two possible definitions of fish consumers (the denominator in the FCR). The first defines anyone who reported eating any fish on at least one of the two days within the two-day recall period as a consumer. The second defines anyone who did not report themselves as a "Never-Consumer" of fish on the embedded FFQ as a consumer. The former definition is quite literal, whereas the latter benefits from capturing individuals who are fish consumers but happened to not eat fish in the previous two-days before the interview.

   The first definition would define 32% (n=7145*0.32 = 2286) of all adults in the NHANES survey as fish consumers (see Table 2 in the report). This very clearly appears to be an underestimate, as demonstrated in Table 2, where 66% of individuals who report themselves as "Ever-Consumers" of fish reported zero intake in the two-day recall period. Thus, use of the first definition seems problematic.

   The second definition would define 90% (n=6465) of adults in the NHANES survey as fish consumers. We note that the majority of additional 6465 – 2286 = 4179 individuals that are used to estimate FCR using the second definition reported zero fish consumption during the two-day recall period. As such, it is not surprising to see the rather stunning discrepancy between the USA Exposure Factors Handbook (EFH) and USA National Cancer Institute (NCI) analyses. The underlying data are quite different.

2) *Implications of Excluding Never-Consumers for the NCI method:* The authors discuss the implications of "sparse-fish" consumption items on the analysis. These are items such as Caesar salad dressing and fish sauce, which contain trace amounts of fish. On the bottom of page 13, the authors state:

   "These 'sparse-fish' consumption items and days have been retained in the analysis, even though it is likely that they are not from local harvest. Only a small percentage of fish-consuming respondents had consumption days with less than 1g/day"

   This is a reasonable assumption. However, it raises the question about whether or not it is consistent to exclude "Never Consumers" from the NCI method analysis? Indeed, it is likely that the 88% percent of "Never Consumers" that reported zero consumption during the two-day recall (Table 2) contains some proportion of consumers of sparse-fish items. Unfortunately, the NHANES data do not appear to be rich enough to estimate this proportion.

   Additionally, given the possibility for "Never-Consumers" to actually be "sparse-fish" item consumers, it would seem that excluding these individuals could violate the assumptions underlying the NCI method. More specifically, exclusion of these individuals could violate the assumption laid out by Tooze *et al* (Tooze 2006) which states

   "We assume that the 24-hour recall is an unbiased measure of the amount of food consumed on the consumption day."

Excluding, for example, the 12% of "Never-Consumers" that reported eating fish during the two-day recall implies that the recall measure is biased. The extent to which this will impact the analysis is likely to be negligible, given the small number of "Never-Consumers". But at the very least, including "Never-Consumers" will drive the estimated FCR down a bit. This would appear to be the more conservative and, moreover, defensible approach.

As such, for the purpose of consistency and in the interest of being conservative, I would suggest that the authors investigate the effect of including all "Never Consumers" on the estimates of FCR. This is equivalent to assuming that ALL adults in the NHANES are fish consumers. However, given the inclusion of consumers of "sparse-fish" items, this may not be an unreasonable assumption.

3) *Use of NCI Method on EFH Fish Consumer Definition Population*: An analysis that has not been included in this report is the analysis of fish consumers using the first definition with the NCI method. As is, the NCI and EFH results are not comparable as they rely on different data. However, it would be possible to apply the NCI method to the population of individuals that ate fish on at least one of the recall days.

   I do not expect that these numbers will be more reliable than the NCI results applied to the population of "Ever-Consumers" (or all adults if "Never-Consumers" are included). Nevertheless, it would be a useful analysis to better help understand the differences between the NCI and EFH (or direct) methods.

   Some effort is already made in the "Discussion" section to compare the methods by simulation. However, the authors admit that the non-NCI method "is not the approach used in the Exposure Factors Handbook, but the simulation is, nevertheless, useful as a comparison of the 'truth' to the two methods described." Likewise, Appendix 2 includes an investigation of the effect of using one or two days as the recall period for the EFH method. But neither of these attempts addresses the critical question of how these two methods perform on the *same* data. What I suggest would do just that and I think provide a nice complement to what is already included.

4) *Diagnostics for NCI Method:* The NCI method is well documented in a number of peer-reviewed articles (see Tooze 2006). However, there seems to be little mention in these references of regression diagnostics. Namely, how well does the model fit? At present, the FCR estimates produced using the NCI method appear to come from a "black-box" and there is little to lead the reader to believe the estimates are accurate.

   The NCI method is a two-part model. The first part models the probability of consuming fish using logistic regression with a person-specific random effect. The second part models the (transformed) consumption amount as a linear mixed model with a person-specific random effect. The random-effects in the two parts are linked through the specification of a bivariate normal. I.e., the random effects are correlated. Usual intake or the FCR is calculated by multiplying the estimated probability of consumption by the estimated amount consumed.

   Due to the complex nature of the model, it is not immediately clear what appropriate diagnostics might be. The approach I suggest draws from the Bayesian statistics literature (Gelman, Meng, and Stern 1996 and Gelman *et al* 2004), whereby simulated data is produced from the fitted model and compared to the observed data (to which the model was fit). A model that "fits well" will have the characteristic that it reliably simulates data that look like the observed data.

   To be more specific, assume we observe data Y and fit model M. Let [.] notate the "distribution of" a given variable. We wish to simulate data Y* from the estimated model given the observed data (i.e. simulate Y* from $[Y^* \mid Y] = \int [Y^* \mid M]\, [M \mid Y]\, dM$). Doing so repeatedly (say 1000 times) results in a set of simulated datasets, each of which maybe be summarized by, say, the mean and variance. Comparing the observed mean and variance to the distribution of simulated means and variances results in a type of p-value which can be used to determine whether or not the simulated data are consistent with the observed data.

Hopefully, applying this approach would show that the NCI method is reasonable for these data in the sense described above and provide some much-needed face validity to the resulting estimates.

5) *Use of NHANES data as representative of WA FCR:* In the “Discussion” section, the authors discuss the use of the US FCR to represent WA State. Due to the size and geographic sparsity of the NHANES data, it is not possible to test whether or not US FCRs are similar to WA FCRs. However, I would suggest that there is strong reason to believe that the US data are NOT representative of WA State. The report would benefit from emphasizing this point.

On the Analysis of the Native American Tribal and Asian and Pacific Islander Surveys

6) *Treatment of Outliers:* On page 19, the authors discuss the treatment of outliers in the Native American Tribal and Asian and Pacific Islander (API) surveys. In general, the authors’ strategy involves treating all data as plausible, rather than truncating according to the observed distribution of the data.

   I would tend to agree with the authors’ approach here. As an added justification, I would also note that in general, individual data is not available. Since most of the tribes only provide aggregated data, excluding extreme values is even more suspect, as each data point represents multiple survey participants. It is of course possible that aggregates are based on individual measurements that include erroneously coded or implausible values. However, this is better viewed as a limitation of using aggregated data, rather than justification for excluding that data point.

   Furthermore, the decision to include “spare-fish” items (potential outliers on the low end of the spectrum) is consistent with the decision to include plausibly large values.

7) *Tulalip Tribe*: I applaud the use of the Tulalip Tribe data to evaluate the impact of using individual versus aggregate data to arrive at FCRs (as described in Appendix 4). I suggest a handful of additional uses for the Tulalip data below.

8) *Squaxin Island Tribe*: It seems worth pointing out that Table A-3 suggests that use of group means systematically UNDERESTIMATES the FCR. Thus, reported rates for Squaxin Island tribe could be underestimated.

   The resulting table (Table A-3) also highlights the need for estimates of uncertainty. Even at the 95th percentile, where the apparent difference is large, it is quite possible that the two approaches result in statistically indistinguishable estimates. See comment “*Uncertainty in the quantile estimates*” below.

   It is unclear to me how the authors were able to apply their philosophy on outliers to the Squaxin Island data, which are reportedly extracted from the 1996 Toy *et al* publication and the 2006 Polissar *et a*l publication (Toy 1996, Polissar 2006). Is it the case that estimates both with and without outlier exclusion criteria applied are available in the Toy et al report? If not, the authors should be clear about this. My understanding is that actually, the Squaxin Island Tribe results are somewhat adjusted for outliers, given the use of the Toy *et al* data.

9) *Columbia River Tribes:* No comment.

10) *Suquamish Tribe*: The Suquamish Tribe rates are a mix of rates calculated from individual data and group means. As such, they strike me as somewhat incomparable, even within tribe. It is unclear to me why all rates were not calculated from individual data, particularly when that data were used in 2007 by the same authors?

The Suquamish data appear to be the source of some concern (see Lawrence McCone's response to the initial estimates on January 20, 2012). Given the extreme values in the data, it would seem that a review of the individual data might be useful here, if possible. If it is not possible to access the individual data, presentation of rates calculated using the group means approach for all sub-groups would likely result in more comparable rates. Rates calculated from the individual data might be included as a table in the Annex.

I also find it strange that sample size *ranges* are provided in rows 3, 6, and 8 of Table 8? If the assumed sample size was 90 for these rows, it would be better to simply put 90 in these cells with a footnote.

11) *Asian and Pacific Islanders*: Given the contrast between individual data and group means highlighted in all other areas of this report, it would be helpful to indicate whether results here reported are based on individual or group data.

12) *Limitations of Assumptions*: The authors point out two limitations in the use of the group means for calculation of FCR in the tribal data. The first assumption is that fish consumption increases linearly in the mean body weight (page 33, paragraph 4):

"The assumption implies that, on average in the population, a person who weighs 50% more than someone else would eat 50% more fish (by weight) than the other person."

The second assumption is that FCRs "are not dependent on the percentage of that consumption that is harvested (from Puget Sound, from the Columbia River, or just 'harvested')."

The authors go on to say that these "assumptions are untested for the populations for which we did not have access to individual-level data." This statement implies that they were tested when individual data were available (i.e. in the Tulalip Tribe). It would be useful to include in Appendix 4 a more explicit treatment of these assumptions in the Tulalip Tribe data. The findings might suggest that certain relationships are better approximated as log-linear, rather than linear, for example.

Additional Comments

13) *Uncertainty in the quantile estimates:* The authors' in their discussion rightly point out that all quantile estimates are informed by ALL of the data. The authors also point out (page 29, third paragraph) that

"… the sample sizes involved in the various studies provide some guidance as to which are more or less prone to random error."

I find this statement unsatisfying as clearly the uncertainty of each of the quantiles will be different with the same sample size.

As an example, consider the following plot. On the left, I have simulated data from a lognormal distribution with mean zero and standard deviation 0.5. On the right, I have used bootstrap replications from the simulated data to estimate the 50th-95th percentiles of the distribution and provided pointwise confidence intervals. I have let the sample size vary from 100 (which is similar to sample sizes observed in tribal data) to 3000 (which is around the sample size used in the USA EFH analysis).

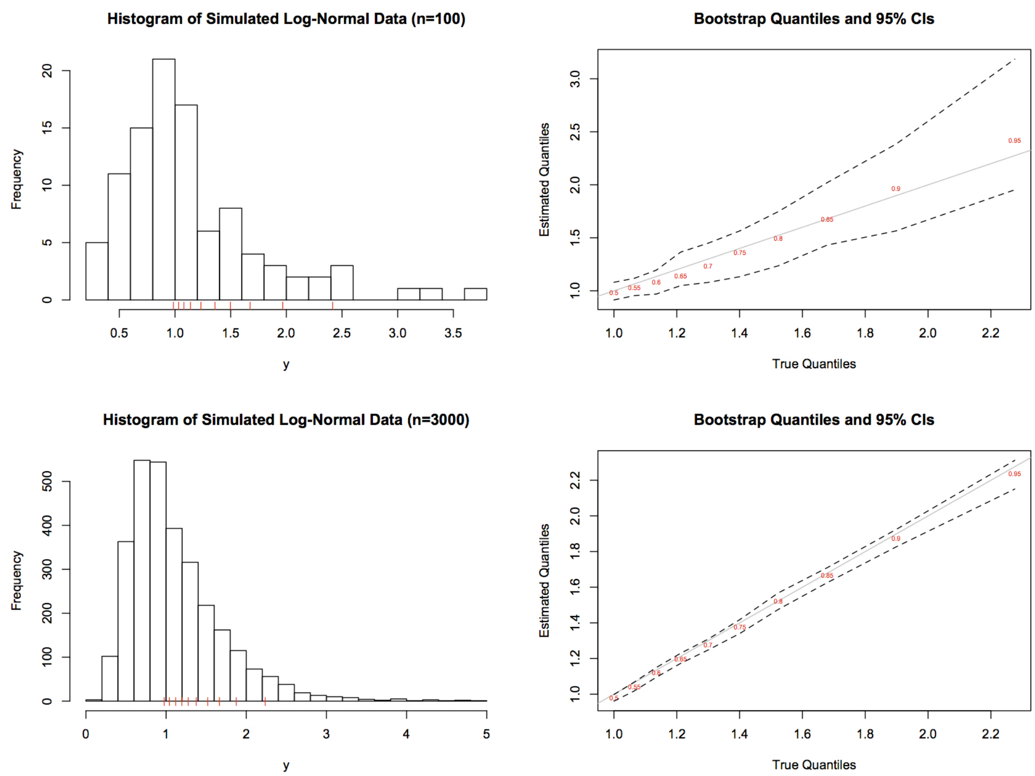

Figure 1: Simulation of lognormal data and resulting estimated quantiles with 95% bootstrapped confidence intervals. Grey line is the line of equivalence and the red numbers indicate the percentile.

With a sample of 3000, in this simulation there is little bias and little uncertainty at all quantile values. However, the same is not true for a sample size of 100. In fact, the uncertainty is quite large in the larger quantiles relative to the smaller ones (i.e. compare the 95th to the 50th percentiles).

Given the likelihood that 95th percentiles will be used for policy decisions, it would seem imperative that the estimates, particularly in the tribal data, be accompanied by measures of uncertainty. Indeed, in this simple example, when the sample is 100, there appears to be little statistical difference between the 85th and 95th percentiles. I emphasize that this is a toy example to illustrate a specific point and is not intended to approximate the level of uncertainty in the NHANES nor the tribal data.

Admittedly, this may be an impossible task. In most cases, published tribal data are not accompanied by estimates of uncertainty and individual-level data is in general not available. At the very least, a full treatment of uncertainty for the national data and for the Tulalip tribe data would provide some benchmarks which could help the reader understand the order of magnitude of uncertainty in the reported rates.

References

Gelman, A; Meng, XL; Stern HS. Posterior predictive assessment of model fitness via realized discrepancies. Statistica Sinica 6, 1996, pp. 733-807.

Gelman, A; Carlin, JB; Stern, HS; Rubin, DB. Bayesian Data Analysis. 2nd ed. Boca Raton: Chapman & Hall/CRC, 2004. 157-192. Print.

Polissar, NL; Stanford, D; Liao, S; Neradilek, B; Mittelstaedt, GD; Toy, KA. A fish Consumption Survey Of The Tulalip and Squaxin Island Tribes of the Puget Sound Region—Consumption Rates For Fish Consumers Only. Report by The Mountain-Whisper-Light Statistical Consulting, 2006. *This was designated as an EPA contractor report. A final report which is an extension of this work is expected to be released by EPA.*

Polissar, NL; Liao, S. 2007. Selected Suquamish Tribe Seafood Ingestion Rates, Consumers Only. The Mountain-Whisper-Light Statistical Consulting.

Tooze, JA; Midthune, D; Dodd, KW; Freedman, LS; Krebs-Smith, SM; Subar, AF; et al. A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. Journal of the American Dietetic Association 106:10, 2006, pp. 1575-1587.

Toy, KA, Polissar, NL, Liao, S, and Mittelstaedt, GD. A Fish Consumption Survey of the Tulalip and Squaxin Island Tribes of the Puget Sound Region. Tulalip Tribes, Department of Environment, 7615 Totem Beach Road, Marysville, Washington 98271. 1996.