Peer Review Report for

"Addressing MRIP Recommendations for the Puget Sound Recreational Fishery Monitoring Program"

Reviewed by

Dr. Kristen Olson, University of Nebraska-Lincoln
Dr. Kenneth Pollock, North Carolina State University
Dr. Lynne Stokes, Southern Methodist University

Introduction

This document combines the comments provided by three different peer reviewers of the MRIP Project Report entitled "Addressing MRIP Recommendations for the Puget Sound Recreational Fishery Monitoring Program." The document provides verbatim reviewer comments without identifying the source of each comment.

Reviewer 1

1. Introduction

In November 2010, several consultants met with the Washington Department of Fish and Wildlife (WDFW) staff to discuss their Puget Sound Sampling Program (PSSP). Their purpose was to review the program used to estimate fishing catch and effort in Puget Sound, and to make recommendations for improvement in the data collection and estimation program, if needed. The consultants identified several areas in need of improvement, mostly having to do with implementing a probability sample design, and developing appropriate estimators for the design. These suggestions were outlined in the consultants' report (Breidt, Opsomer, and Lesser (2010)).

The purpose of this review is to assess the proposed changes to WDFW's PSSP made in response to the consultants' report. To do this, I examined the description of these changes as they are described in the Final Report of their FY 2011 Project. This review contains my opinions about whether or not the consultants' concerns were effectively addressed and their suggestions implemented adequately.

This report has five additional sections, following this introduction. In Section 2, I review the four major suggestions made by the consultants for improving the PSSP. In Sections 3 and 4, I discuss the proposed changes to the sample design and estimation procedures of the PSSP that

were made in response to the consultant's suggestions. In Section 5, I discuss some issues with missing data. In Section 6, I include a listing of some minor errors and suggested edits in the report. I have labeled my major points with consecutive letters, throughout this review, so that they can be easily identified in discussions.

2. Review of Consultants' Suggestions

In Section 5 of the Consultants' report, four suggestions were made to improve the PSSP. They are:

> 1. Extend the probability sample design approach that is used for the intensive sampling program to the baseline sampling program. At the time of their report, the choices of sites for sampling in the baseline design were made mostly by supervisor judgment, making appropriate weighting in the estimation process difficult.
>
> 2. Modify the practice of sending access point samplers to fishing "hot spots," which was used to increase the number of angler encounters. The consultants suggested if it was still desired to collect data in this way, to keep out some sample that could be used opportunistically when evidence of these "hot spots" arose, but give them a weight of 1 in estimation.
>
> 3. Consider using a dual frame approach in the telephone frame to address undercoverage in the license frame for the telephone survey.
>
> 4. Coordinate data collection with other agencies responsible for Puget Sound Fisheries.

The report under review addressed only the first of these four suggestions at length. New methods were proposed to implement a probability (randomization-based) sample design for the access point data collection program. New estimators and their standard errors meant to be appropriate for that design were presented. Suggestion #2 in the list above was not directly addressed, but my understanding based on the report is that all sampled site-days would be chosen from the new randomized design. Thus these opportunistic access point samples would either not be collected at all or would not be used in estimation, so that the suggestion is implicitly implemented. Suggestion #3 dealt with the issue of undercoverage in the telephone survey, since the final report addressed changes only in the access point survey. Finally, suggestion #4 was not discussed, and appears to be outside the scope of the discussion of statistical methodology.

My comments will be focused on the proposed access point sample design and its estimators of catch and effort.

3. The proposed sample design changes

The PSSM access point data collection operation consists of two parts: a baseline survey and an intensive survey. The baseline survey was the one of concern to the consultants since it was not based on a formal randomized sample design, while the intensive portion of the sample was. The intensive portion selected two sites using a Murthy design (pps without replacement) on each of a stratified sample of days during a portion of the year in each sampling area. The modifications that are proposed in the final report are for the baseline design.

The proposed baseline design is described as a pps design, where the size measures are based on historical data on the number of anglers encountered at each site in the 2007-2011 time span. Specifically the probabilities are described (on p. 8) this way:

> "We calculated the probabilities of site selection as the proportion of the total unique interviews represented in the baseline database at a given site for each wave and MCA [Marine Catch area]. For cases in which a site was never sampled for a given MCA… yet the site was still listed as a potential sampling site, we assigned a sampling selection probability of zero….[F]or cases in which at least one interview was conducted but the calculated probability of selection was less than 0.01, we again adjusted the probability to 0.01..."

The report also indicates that when the baseline and intensive sampling operations overlap, the intensive sampling units are selected first using their own set of size measures, and the sampled units are then excluded before selecting additional units as part of the baseline sampling program, with readjusted probabilities that exclude these sampled units.

This design is a probability design, as suggested by the consultants. It is an advantage that the design is similar enough to the intensive design so that it should be transparent to the samplers, and so that training on the new procedure is expedited. The software that is described and illustrated in the report appears to be user-friendly and simple to implement. The merging of the two sets of sample data in the manner described can certainly facilitate appropriate estimation.

However, the description of the design as it is given is incomplete. Though the new design for the baseline access survey is described as pps and two-stage, it is unclear exactly how the sample selections are made. Specifically, here are some questions that I have after reading the report:

A. Is the selection of sites for the baseline sample a with-replacement or without-replacement design? If it is a without-replacement design, what kind of design is used? The Murthy design is used for the intensive sample in order that the weighting can be easily calculated without replacement, but it allows only a sample of size 2 psu's to be selected. The user interface for the new sampling software shows that any size sample is allowed for the baseline sample. This would necessitate a very complicated weight calculation scheme if without-replacement sampling is planned.

On the other hand, the fact that the estimator is referred to as a "Hansen-Hurwitz" estimator, as it is on p. 14, suggests that the design planned is a with-replacement design. (For example, see http://www.math.montana.edu/~parker/PattersonStats/PPS-Supplement.pdf for a nice discussion of the Hansen-Hurwitz estimator and when it is used.) If this is the case, some explanation of the practical implementation of a with-replacement design should be included. For example, what happens if there are replicated sites in the sample so that not all samplers are needed on a particular day?

In general, there should be a more exact description of exactly what kind of pps design is being proposed. (See Sections 6.2 through 6.5 of Lohr (2010) for a description of the variety of designs possible.)

Another more minor question is why would any site be given a selection probability of 0, as described in the excerpt above? I would suggest that all sites that are low pressure be included at some minimal probability of selection.

B. The sample data from the two surveys were put into a single database, and weights are provided for all sampled units. However, I do not find evidence from the report that the selection probabilities are calculated correctly for the purpose of weighting.

Table 3 on p. 19 provides an example of how the database looks. I am confused about the variable labeled *Site-weight*. Since it is less than 1, it appears to be a probability rather than a weight. Is it? The text reports that the reciprocal of the probabilities are used as weights, so if these are probabilities, then it appears that their reciprocals are used as weights. But these would be inappropriate weights for those site-days covered by both the intensive and baseline sampling. They may also be inappropriate if the sampling is without-replacement. The reason for these two assertions are explained below.

The selection probabilities for site-days that are covered by both access samples at the same time are not the same as the selection probabilities for each separately. Even adjusting the weights on the baseline sample multiplicatively is not sufficient to properly reflect the selection probability. To illustrate what I mean, consider the following (fabricated) example. Suppose a site is available for selection in either access survey, with probability 0.25 in the intensive survey. The probability that it is selected in one or the other is 0.25 + Pr[site is not selected in intensive sample]* Pr[it is selected in baseline sample | it was not selected in intensive sample]. This probability is not the same as even the adjusted probability of selection in the baseline survey. Since I don't know exactly how the baseline sample is selected (with or without replacement), I cannot produce this probability, but it should be explained in the report. (Note: The idea of this paragraph is relevant only if the two samples are combined and to compute a single estimator of catch. This will be discussed further in Section 4.)

The reciprocal of the inclusion probabilities may also not be the correct estimation weights if the baseline design is any kind of without replacement design, since the probabilities of selection at

each draw depends on what has been selected before. The use of a Hansen-Hurwitz estimator (and thus the simple reciprocal of selection probabilities) would be reasonable if the number of psu's in the frame is large enough that replicates are rare. but from the examples in the report, it appears that this is not the case, with some sites having selection probabilities near half.

4. The proposed estimator changes

The proposal to use weights in the estimator of CPUE to reflect the access sample design is a definite improvement in the data collection and estimation program. The SAS code on p. 12 will produce valid estimators of the CPUE by domain and species, as long as the weights have been calculated properly. However, I have some concerns about the weighting procedure, as discussed in Section 3. In addition, the displayed expressions for the estimators and their standard errors on pp. 14 – 16 are incorrect, even if the weights were correct.

C. The report is not clear about how the two samples will be used to produce estimates. It appears that the two samples are considered as though they are one sample design, and estimates calculated from the combined sample, but using their original weights. If this is the case, then the weights will not be correct, as noted in the previous section. Instead, the probabilities of selection must be recalculated to reflect the probability of being selected in either sample, for those domains or strata in which both baseline and intensive sampling operations are present.

On the other hand, it would be possible to combine the data in some other way. For example, one could calculated an average (or weighted average) of estimates of CPUE made from each sample design separately, for those domains/strata in which both surveys operate. If this is what is proposed, it should be described in the report.

Which of these two methods of combining the data is best from an efficiency point of view depends on properties of the two designs. A decision on how to combine the data in this instance is similar to a decision on the estimator from a dual frame design. For dual frame designs, one has a choice of a "single frame" estimator (which is analogous to an estimator that would be calculated from a combined sample with a single set of weights) or one that uses a weighted average of the overlap portion of the frame (as Hartley's original dual frame estimator did).

I recommend that how the data are being combined be more explicitly addressed in the report. If a "single frame" estimator is being produced, then the weights need to be corrected. In fact, I think that an investigation of this question (how to combine the data from the two samples) would be a more productive use of resources than one that tries to compare the newly weighted data with unweighted data. It is not clear how bias and variance of the old method could even be estimated, since the old estimators do not "match" the new sample design.

D. As mentioned in the introduction to this section, the SAS code on p. 12 will produce valid estimates of CPUE if the weights are correct. However, the expressions shown in Eq.'s 2, 5, and 9 will not be computed by the displayed SAS code. These estimators are not the correct

generalizations of the analogous unweighted estimators in Appendix A (which would be calculated by the SAS codes if there was no weight statement).

The SAS code on p. 12 computes the following estimator for CPUE:

$$CPUE = \sum_{d=1}^{g} \frac{C_d}{p_d} \Big/ \sum_{d=1}^{g} \frac{A_d}{p_d}, \qquad (*)$$

where g is the number of site-days sampled, C_d (A_d) is the total catch (anglers/trips) at site-day d, and p_d is the probability of selecting site-day d at each draw.

The estimator in Eq 2 would only be a Hansen-Hurwitz estimator if the total number of trips at every sampled site day (sum of the A's over the boats sampled) happened to be the same as the numerator of the size measure used to determine the probabilities of selection.

To demonstrate the calculations, I will use the illustrative data in the table below. It shows a sample of 3 site days whose catch, # of licensed anglers (trips), number of unlicensed anglers (trips), probability of selection for the site-day, weight, and site CPUE (from Eq. 1 on p. 14) are denoted by C, A_l, A_{ul}, p, W, and cpue:

C	A_l	A_{ul}	p	w	cpue
10	3	2	0.25	4	3.33
0	4	4	0.01	100	0
100	40	30	0.47	2.13	2.5

Then

CPUE from Eq. 2 = (3.33*4 + 0* 100 + 2.5*2.13)/(4+ 100 + 2.13) = 0.175

and

CPUE from (*) = (10*4 + 0* 100 + 100*2.13)/(3*4+ 4*100 + 40*2.13) = 0.508.

Your SAS code and its output are shown below:

```
data example;
input site_day domain catch n_angler n_licensed weight;
cards;
1 1 10 3 2 4
2 1 0 4 4 100
3 1 100 40 30 2.13
;
proc surveymeans;
cluster site_day;
domain domain;
var catch n_angler n_licensed;
ratio catch n_angler/n_angler n_licensed;
weight weight;
run;
```

Ratio Analysis

Numerator	Denominator	N	Ratio	Std Err	95% CL for Ratio	
catch	n_angler	3	0.508475	0.657613	-2.3210043	3.33795356
catch	n_licensed	3	0.535714	0.729958	-2.6050434	3.67647209
n_angler	n_licensed	3	1.053571	0.072996	0.7394957	1.36764721

Thus the SAS code produces the ratio estimate shown in (*).

E. On p. 15, it is suggested that R_{ij} and $CPUE_{ijk}$ can be combined into a single ratio estimator by multiplying them together, yielding the catch per licensed angler trip. Actually, if you multiply Eq. 2 and 5, you do not get Eq. 9. This is because these estimators are not defined correctly, as discussed in D. When the correct estimators are multiplied together, your statement will be true. In fact, there is no reason to arrive at the estimator of CAT by thinking of it as multiplying these four components together, but rather, just think of the total catch itself as being calculated using the number of licensed angler trips as an auxiliary variable to improve estimation. If we knew number of licensed angler trips (call it T_L), we could then use the simple ratio estimator for total catch as

$$\hat{CAT} = T_L * \hat{CPL},$$

where $\hat{CPL}$ is the catch per licensed angler in the sample. Since T_L is not known, it must be estimated as $\hat{T}_L = N_p * \hat{TR}$, and putting them together yields your Eq. 10. This seems to me to be a simpler justification for the proposed estimator of *CAT*. It also clarifies the fact that it is unnecessary to assume that the catch rate for licensed and unlicensed anglers is the same. Thinking of it in this way, it becomes clear that the number of licensed angler trips is simply an auxiliary variable for catch.

F. The estimated variance expression shown in Eq. 3 on p. 14 is not the variance estimate produced by SAS for two-stage designs. SAS produces a standard error for two stage designs that contains only the first stage variance component. From the section on Statistical Computation from PROC SURVEYMEANS documentation, this is described as follows:

> "For a multistage sample design, the Taylor series estimation depends only on the first stage of the sample design. Therefore, the required input includes only first-stage cluster (PSU) and first-stage stratum identification. You do not need to input design information about any additional stages of sampling. This variance estimation method assumes that the first-stage sampling fraction is small, or that the first-stage sample is drawn with replacement, as it often is in practice."

G. The report refers to the access sample as a two-stage design. However, the method is described as though all boats are intercepted on the sampled site-days, so there is no 2nd stage and thus no component of variance due to within site-day variation. I did notice that it was

mentioned that if the number of boats is too large, they can be subsampled randomly, but there is no indication that there is accommodation for this in the design and estimation, since no 2nd stage weight is described. If the sampling protocol does allow subsampling of boats (or anglers/trips) within the site-day, there must be an additional weight recorded in the file and used in estimation. This 2nd stage weight would require an additional field in the database, and additional training for the interviewers on how to calculate this weight would be needed. This could be difficult, since the interviewers may find it necessary to vary the skip interval depending on how variable the boat arrival rate varies over the day. It might be more feasible (and accurate) to treat the missed boats as missing data, rather than a planned subsample. This will be discussed further in the Section 5.

5. Imputation and non-response adjustment

Though the consultants did not make suggestions concerning the combining of the effort and catch surveys, the report addresses this issue. It contains a nice discussion of the investigation done on how to handle the situation when either trips or catch is missing in the sampled data for a particular stratum/domain. An investigation using regression showed that there was not a single imputation method that will be appropriate for all domains. The imputation approach suggested, which is to borrow data from other time periods, seems to be a reasonable one.

H. It does seem, however, that the assignment of 0 catch in those cases when one or the other of the surveys reports 0 (trips or catch), but the other reports a nonzero value, could be improved. The estimate of 0 in either rof these cases is definitely biased downward (assuming there is no measurement error in either sample's data). Thus it might make more sense to impute some value (besides 0) for the missing trip/catch.

I. There is no mention of nonresponse accommodation in the access surveys. Perhaps there is a high response rate. Nevertheless, there must be occasions when a boat within a site cannot be interviewed (e.g., refusal or language problem), or when a selected site-day is not included in the sample (e.g., sick interviewer). Another type of nonresponse could occur when there are so many boats returning to a site that they cannot all be interviewed. Rather than treating the missed boats as planned subsample, which would require calculating a 2nd stage weight, these cases could be treated as non-responses. In all these cases, some accommodation should be made for the nonresponse. The most common method for handling this type of unit nonresponse is by weighting class adjustment. That is, an additional weight can be applied to the site day observation to accommodate the missing data. This is easily done in SAS.

6. Other minor problems

There are several occasions where the language is not precise or the notation is not entirely correct. I have listed a few of them below.

a. p. 11, 1st paragraph in Data Preparation, the term "data parameters" is not a correct usage of the term parameter. Maybe you just mean "data".

b. p. A-38, Eq. 2. The notation s^2c_{ij} should have c_{ij} subscripted.

c. p. 13 In definition of w_d, index should be d, not s, in definition

d. p. 14&15, A's and L's in equations have a species subscript (k) while the notation definitions do not on p. 13. Wouldn't they be the same for all species?

References

Breidt, F. Jay, J. D. Opsomer, and V. Lesser (2010), "Consultant's Report: Preliminary Review of Washington's Puget Sound Sampling Program."

Lohr, S. (2010) *Sampling: Design and Analysis*, 2nd edition. Brooks-Cole, Boston.

Reviewer 2

1. Introduction

The Washington Department of Fish and Wildlife (WDFW) was charged with the Puget Sound Sampling Program (PSSP) changing the survey of anglers sampled in boats from a convenience sample to a probability sample by consultants Breidt, Opsomer and Lesser. The WDFW developed and implemented a new sampling procedure for the point intercept survey (the "Baseline" survey), following methods previously used only in special studies ("Intensive" survey).

The WDFW is to be commended for its development and implementation of a probability sampling approach for the Baseline survey. This change will help ensure the scientific validity and rigor of the catch and effort estimates from these data.

This report reviews the previous consultant recommendations and how these recommendations were implemented. It also identifies the strengths and areas for improvement or clarification in the WDFW report.

2. Previous Recommendations

Breidt, Opsomer and Lesser had a number of previous recommendations related to (1) sample design, (2) estimation, (3) coverage errors, and (4) coordination of estimates across agencies. I now review their recommendations and the corresponding work as developed for the PSSP, along with remaining issues that need to be clarified or reconsidered.

1. Develop a probability sampling method, extending the current two-per-stratum PPS design used for the Intensive survey. This should include stratification for space and time and indications of "fishing pressure." Low activity sites could be linked together to improve interviewing efficiencies.

1a. Develop formal protocols for sample design that can be implemented and repeated by multiple persons without subjectivity.

1b. If it is needed to have high pressure sites selected, incorporate them into the design by creating self-representing sites.

The WDFW developed a set of R programs to select sites for the Baseline design using a two-PSU-per-stratum PPS design. The site selection probabilities were developed from the total number of interviews conducted at each site from 2007 to 2011. That is, the measure of size (*Mos*) for each site (*i*) is $Mos_{i,xy} =$ # 2007-2011 interviews for site *x* at wave *y* in area *z* and the probability of selection at that site is

$$P(site\_selection) = \frac{Mos_{i,xy}}{\sum_{i=1}^{m_z} Mos_{i,xy}} = \frac{\text{\# 2007-2011 interviews for site } x \text{ at wave } y \text{ in area } z}{Total\text{\# 2007-2011 interviews for site } x \text{ at wave } y \text{ in area } z}.$$

The selection algorithms have been standardized and implemented through an R program with a GUI interface, and a training module for supervisors at four sites. This is consistent with the consultant recommendation to implement a procedure that could be replicated by persons who are not the current supervisors. The efforts were non-trivial and seem to have been successfully deployed across the four sites. This is where the bulk of the effort in the report is focused, and the WDFW has accomplished a successful new program.

There are a few outstanding issues with the sampling methods.

<u>New issue 1:</u> The report indicates that some sites were never sampled during the time period in question, with an assigned *Mos*=0. Given the discretion permitted in the previous design to supervisors and/or interviewers as to where interviews were conducted, it is not clear whether these sites were not sampled because there (accurately) was no fishing at the site, or whether there had been an erroneous exclusion of this site in the previous work. The report should clarify why this decision was made, and the potential for evaluating decisions about these zero *Mos* sites in the future. Alternatively, the WDFW could create linked clusters of low effort areas, per the Breidt, et al. recommendations, in which geographically proximate low effort areas are combined (linked) for purposes of selection. Linking procedures are described in Kish (1965), pp. 244-246.

New issue 2: The report describes changes in the probabilities of selection for the Baseline survey on days when the Intensive survey is conducted. The sites selected for the Intensive survey are excluded from those that could be selected for the Baseline survey. Their probability of selection for the Baseline survey is now zero, and the probabilities of selection for the remaining units is recalculated based on the previously selected sites. It is not clear how this set of conditional probabilities is accounted for in the calculations of the probabilities of selection (that is, both the probabilities of not being selected for the Intensive survey and the probability of being selected for the Baseline survey). It also makes the selection probabilities random, varying from sample to sample (and from day to day), because any single site remaining on the list now has a selection probability that depends on the other units on the list being selected. It is not clear, given that samples are being drawn daily, why the Intensive and Baseline samples need to be drawn separately. One could imagine a design in which the sites that are eligible for the Intensive sampling are in one strata and the sites for the Baseline are in a second strata. The Intensive selections are then independent from those of the Baseline. If there are only two sites that are possible for Intensive selections across the entire fishing season, then this stratum would have the two sites with probabilities equal to 1.0 (certainty selections or self-representing selections). If this is, in essence, what the program is currently doing, then the report should clarify this point.

New issue 3: It is unclear in the report (and in Appendix A of the report) whether the weights incorporate sampling for the interview stage. Although it is clear how the sites are selected and how the weights are calculated, the subsampling of anglers leaving the fishery is "instructed to be random and systematic" when anglers exceed the sampler's capacity. This introduces uncertainty into the weights and raises a number of questions. First, is this subsampling rate recorded? Is the decision made at the beginning of the day, and thus all cases at the site have the same subsampling rate or is it done at some later time during the day, at the sampler's discretion? For sites and anglers where subsampling occurred, is this reflected in the weights? This part of the data collection design, if not systematically determined prior to data collection, risks turning the probability sample into a nonprobability sample. As such, more discussion should be given to this part, as well as developing a formal protocol for subsampling.

2. Develop weights, and ensure that variance estimates reflect design.

The PSSP has developed a database with indicators for which survey (Intensive vs. Baseline) the site was selected and the probability of selection weights (inverse of probabilities of selection). The PSSP uses SAS's proc surveymeans to calculate weighted estimates for each wave, MCA and target species. This is an appropriate software procedure to use. The current report updates the previous notation and estimation procedures from the previous methodology report to reflect the weights.

The section on estimation requires editing and revision, as detailed below.

New issue 4: The method used to calculate the ratios of interest in SAS's proc surveymeans and the estimation formulas outlined on pages 14 and 15 do not match. To calculate a weighted ratio such as CPUE, SAS will use the formula $CPUE = \frac{\sum_{d=1}^{g_{ijk}} \sum_{l=1}^{f_{ijkdl}} w_d c_{ijkdl}}{\sum_{d=1}^{g_{ijk}} \sum_{l=1}^{f_{ijkdl}} w_d A_{ijkdl}}$. Equation 2 is a weighted mean of ratios, not a ratio of means, although the text immediately above Equation 1 states that a ratio of means estimator is better. Additionally, for unequal cluster sizes, SAS's proc surveymeans will use a Taylor Series Linearization (TSL) to estimate the variances (similar to that in the initial Lee, et al. methods report). The report states that the weights are considered constants rather than estimated quantities. The issue for a Taylor Series Linearization variance estimator is not that the weights are known, but that the value for the sum of the weights can conceptually vary from sample to sample (that is, it is random and the value depends on the sample that is drawn). The formula shown in Equation 3 does not adequately reflect the variation in the sum of the weights, and should be changed to the appropriate TSL formula (reflecting what the estimation appears to actually be doing via proc surveymeans).

New issue 5: The SAS proc surveymeans syntax does not include strata. Domain-specific estimates should only be done using the "by" statement in proc surveymeans when the "by" groups are also strata. The species variable is not listed in the report as a stratification variable. Thus, the species-level analyses should be conducted only using the domain statement, not the by statement.

New issue 6: At the top of page 15, the statement that an SRS is used is incorrect. The PSSP is using a stratified cluster design with unequal probabilities of selection.

New issue 7: The estimates for the trip rate from the angler survey states in Equation 7 that the variance is $V(TR_{ij}) = \frac{N_p^2(1-f)}{n_p} s_{T_{ij}}^2$, but the estimate shown in Equation 6 is only a mean, not a total. Thus, the variance estimate is actually $V(TR_{ij}) = \frac{(1-f)}{n_p} s_{T_{ij}}^2$. If trip rate is actually calculated using $TR_{ij} = \frac{N_p \sum_{m=1}^{n_{ij}} T_{ijm}}{n_p}$, then equation 7 is correct. If the trip rate is calculated as

described in Equation 6, then Equation 7 is incorrect and $V(TR_{ij}) = \frac{(1-f)}{n_p} s^2_{T_{ij}}$ should be used instead.

3. Evaluate coverage errors

3a. Investigate the possibility of a dual frame approach to account for potential undercoverage of the license frame for the licensed anglers telephone survey
3b. Identify whether coverage errors exist due to exclusion of shore fishing from estimates.

The report does not discuss efforts to evaluate or correct for coverage errors. The Lee, et al. paper in the appendix mentions efforts to evaluate these errors, but the Baltzell, et al. report does not discuss them.

4. Work with other agencies responsible for creating catch estimates

The report does not mention efforts to coordinate across agencies.

3. Additional Issues

The challenges and recommendations identified in the report are reasonable. Some (such as grouping low effort sites) are part of the initial recommendations from the Breidt, et al. report. Others, such as the development of a web interface, are logical extensions of the implementation of this work across multiple sites. Additional analyses of the new estimates, including their mean square error properties, are a necessary part of any new estimation scheme.

There are a few additional issues that arise from my reading of this report.

New issue 8: The original consultant report encouraged PSSP to consider potential coverage errors in both the intercept survey and the telephone survey. This is a good recommendation, and should be investigated. Similarly, the report does not mention response rates or efforts to evaluate nonresponse bias in estimates from either the intercept survey or the telephone survey. Are the anglers who participate in the telephone survey different from those who do not participate? Do the anglers who participate in the intercept survey have different catch/release information than those who do not participate? Unit nonresponse is an important source of errors in surveys that is not addressed in the new Baltzell, et al. report or in the original Lee, et al. report.

New issue 9: The report as currently written is difficult to follow, with selection of units separate from the data collection procedures, which are not fully incorporated into the estimation

methods. It is not clear whether this report is intended to replace the Lee, et al. report, or if it is an addendum to that report addressing only the consultant's comments.

New issue 10: Does the PSSP traditionally release estimates and their variances? It is more common to report estimates and standard errors (square roots of sampling variance) in tables.

4. Conclusion

The WDFW has developed and implemented a new sampling method for the Baseline survey of anglers, addressing the two largest concerns of consultants Breidt, Opsomer and Lesser. Two of the other concerns voiced by the previous consultants are not directly addressed in the new report. Other issues have arisen in the new report, as detailed above.

Reviewer 3

Executive Summary

The Overall complemented survey design used is a Telephone - Access design (Pollock et al. 1994). The Telephone survey is based on a saltwater angler license file. The access survey is carried out at public boat ramps. There is a baseline sampling design year round and also a more intensive sampling design during times of year where high sampling effort is expected. The Telephone survey is to estimate licensed boat angler effort and the access survey to estimate boat angler harvest rate and the ratio of total boat anglers to boat anglers with fishing licenses. Boat Effort is estimated as Licensed Boat Effort * Ratio of Total Anglers to Licensed Boat Anglers and requires both surveys for the estimation. Total Harvest can be estimated as Effort* Harvest Rate and requires both surveys for the estimation. Released Catch is also estimated in a similar manner to kept catch or harvest.

Breidt et al. (2011) recommended that sampling weights which are the inverse of the inclusion probabilities be formally incorporated into the design and estimation procedure for estimating catch rates. This is a version of probability proportional to size sampling. The probabilities of inclusion should be proportional to the relative uses of the sites. The authors have obtained these weights by using past baseline data on relative effort (see for example Table 1). The validity of the method of obtaining the weights is unclear to me as I do not know how the baseline sites were sampled in previous years? It seems there is a strong assumption that they were all sampled with equal frequency. I think that in Table 1 one would need to know the number of days each site was sampled and then use that information to adjust the numbers of interviews before one calculated the proportions. Another method that could be used is to use expert opinion which is gradually updated as information becomes available going forward. I am not advocating this but I do think the current method needs a much better justification and perhaps modification.

There are some very Low Use sites that have very small inclusion probabilities (0.01) and hence very large weights and this is of concern to me. You may need to look for outliers or unusual estimates because of sparse data estimating catch rates plus this issue of some catch rates being multiplied by very large weights. In Breidt et al (2011) they mention combining multiple low level sites into super sites that would be sampled as a group and in the report this has been noted as something to be explored in the future. This has not been done yet but I do think it is a very good idea. Perhaps some kind of modification of the bus route design (Jones and Robson 19, Pollock et al. 1974) could be used?

The results where estimates are presented in Table 4 and Appendix Dwould be much more useful if the units were given and if standard errors and then relative standard errors were presented rather than variances. I glanced through the Table and calculated standards errors by taking square roots of the variances and find that the precision of the estimates is sometimes quite poor. This should have been discussed and put in context of what the target precision levels were and for which species and areas and waves. Also I realize that it has been stated that future work will look at a comparison of this new method with the old one (P22) but I think some of that should have been done for this report.

Breidt et al. (2011) mention "flash fishing" where there is a problem with some sites getting a lot of fishing effort for a short time and in a way that is difficult to predict ahead of time I think the idea Breidt et al. (2011) suggest of giving some sites a weight of 1 within a wave so they are "self representing" would need some thought. For that to work information on the hot spot would be needed before the wave started and that might not always be available. I wondered if it might be possible to modify the adaptive sampling designs of Thompson and Seber (1996) but have not pursued that idea in depth.

The Telephone survey is based on a saltwater angler license file. There are some exemptions allowed which results in undercoverage and the current approach is to use an estimate of the ratio of total anglers to licensed anglers from the access point survey to adjust for the undercoverage. The Breidt et al (2011) review mentioned the benefits of a Dual Frame approach (see NRC 2006, Kott and Vogel 1996 or many sampling texts) to deal with undercoverage of the license file. I totally concur with this recommendation. I think it is likely a much stronger method than the current ratio adjustment method. Breidt et al (2011) also mentioned the importance of trying to minimize problems with the license file where phone numbers are incorrect or not provided. I totally agree with this recommendation. Response rates in the telephone survey will also have to be considered carefully going forward.

A concern about the current design is that private access boat anglers are not included directly in the on-site access survey and therefore the assumption is made that their catch rates are the same as those from the sampled public access points. As Private Access boat anglers are included in the effort calculation based on the telephone survey license frame there is not undercoverage in the usual sense but there is the potential for bias if catch rates are different.

Released catch is very important information for fisheries managers but it should be kept in mind that as it is based on angler supplied recall it likely contains a lot of measurement error. Therefore the released catch estimates will be biased in unknown ways.

More Detailed Review

1. Introduction

The Overall complemented survey design used is a Telephone - Access design (Pollock et al. 1994). The Telephone survey is to estimate licensed boat angler effort and the access survey to estimate boat angler harvest rate and the ratio of total boat anglers to boat anglers with fishing licenses. Boat Effort is estimated as Licensed Boat Effort * Ratio of Total Anglers to Licensed Boat Anglers and requires both surveys for the estimation. Total Harvest can be estimated as Effort* Harvest Rate and requires both surveys for the estimation. Released Catch is also estimated in a similar manner.

The Telephone survey is based on a saltwater angler license file. There are some exemptions allowed and there are license frame quality issues (like unavailable telephone numbers). The access survey is carried out at public boat ramps. There is a baseline sampling design year round and also a more intensive sampling design during times of year where high sampling effort is expected.

Private access boat anglers are not included directly in the on-site access survey and therefore the assumption is made that their catch rates are the same as those from the sampled public access points. As Private Access boat anglers are included in the effort calculation based on the telephone survey license frame there is not undercoverage in the usual sense but there is the potential for bias if catch rates are different. (Breidt et al. (2011) mention a roving boat survey as well but I see no mention of that in the document I am reviewing and don't consider it further.)

Shore angling is excluded from all aspects of the sampling and estimation.

The document I was asked to review primarily focuses on improving one aspect of the Access point survey design related to assigning access points formal weights and incorporating those weights in the estimation procedure. I focus mainly on that task and then broaden out my review to other aspects of the Access Point Survey and other on-site survey issues. I conclude with some brief thoughts on the telephone survey component as this is a complemented survey and the validity and precision of the final estimates depends on both components.

2. On-Site Survey Issues

PPS Sampling Design

Breidt et al. (2011) recommended that sampling weights which are the inverse of the inclusion probabilities be formally incorporated into the design and estimation procedure for estimating catch rates. This is a version of probability proportional to size sampling. The

probabilities of inclusion should be proportional to the relative uses of the sites. The authors have obtained these weights by using past baseline data on relative effort (see for example Table 1). The validity of the method of obtaining the weights is unclear to me as I do not know how the baseline sites were sampled in previous years? It seems there is a strong assumption that they were all sampled with equal frequency. I think that in Table 1 one would need to know the number of days each site was sampled and then use that information to adjust the numbers of interviews before one calculated the proportions. Another method that could be used is to use expert opinion which is gradually updated as information becomes available going forward. I am not advocating this but I do think the current method needs a much better justification and perhaps modification.

There are some very Low Use sites that have very small inclusion probabilities (0.01) and hence very large weights and this is of concern to me. You may need to look for outliers or unusual estimates because of sparse data estimating catch rates plus this issue of some catch rates being multiplied by very large weights. In Breidt et al (2011) they mention combining multiple low level sites into super sites that would be sampled as a group and in the report this has been noted as something to be explored in the future. This has not been done yet but I do think it is a very good idea. Perhaps some kind of modification of the bus route design (Jones and Robson 19, Pollock et al. 1974) could be used?

The estimation method seems quite standard and extends the work done earlier in appendix A. I did not check the variance equation by going back to Sarndal et al. (1992). I do worry about including very large weights (for sites with very small inclusion probabilities) as mentioned in the previous paragraph. In brief summary the Telephone survey is to estimate licensed boat angler effort and the access survey to estimate boat angler harvest rate and the ratio of total boat anglers to boat anglers with fishing licenses. Boat Effort is estimated as

Licensed Boat Effort * Ratio of Total Anglers to Licensed Boat Anglers

and requires both surveys for the estimation. I haven't had time to consider it in detail but I do wonder if there might be substantial bias in estimation of the ratio of total to licensed anglers from the access survey. Alternative approaches need to be considered and I briefly discuss that later in my review.

Total Harvest can be estimated as

Effort* Harvest Rate

and again requires both surveys for the estimation. As the estimation so crucially depends on both surveys they cannot be thought of in isolation and all sources of possible assumption violation need to be assessed and methods revised eventually

Released Catch is also estimated in a similar manner to kept catch or harvest. Released catch is very important information for fisheries managers but it should be kept in mind that as it is based on angler supplied recall it likely contains a lot of measurement error. Therefore the released catch estimates will be biased in unknown ways.

The results where estimates are presented in Table 4 would be much more useful if the units were given and if standard errors and then relative standard errors were presented rather than variances. I glanced through the Table and calculated standards errors by taking square roots of the variances and find that the precision of the estimates is sometimes quite poor. This should have been discussed and put in context of what the target precision levels were and for which species and areas and waves. Also I realize that it has been stated that future work will look at a comparison of the new method with the previous one (P22) but I think some of that should have been done for this report.

Flash Fishing

Breidt et al. (2011) mention there is a problem with some sites getting a lot of fishing effort for a short time and in a way that is difficult to predict ahead of time I think the idea Breidt et al. (2011) suggest of giving some sites a weight of 1 within a wave so they are "self representing" would need some thought. For that to work information on the hot spot would be needed before the wave started and that might not always be available. I wondered if it might be possible to modify the adaptive sampling designs of Thompson and Seber (1996) but have not pursued that idea. These designs have seen some use in wildlife sampling studies and there are many references in the literature but I haven't included them. The basic idea is that extra sampling is added in the field when a hot spot is detected. Inclusion probabilities are usually quite complicated and there can be practical issues (such as possibly needing more survey agents) can also be an issue.

Private Access Boat Fishing

A fairly serious defect of the current design is that it requires the assumption that catch rates are similar between private and public access points. Fortunately private access anglers are included in the telephone survey. One way to get at this question would be to get a subsample of these fishers involved in a longitudinal survey to obtain some catch rate information which could be compared to the catch rates at the public access sites. I would not recommend asking them cold recall questions on recent trips as the data quality would likely be poor. Another way this issue could be addressed is by collecting some incomplete trip catch information using a roving boat survey. Both of these approaches would involve additional cost but it seems to me that this is an important issue and should be addressed.

Shore Fishing

Shore Fishing is not included in the current survey and would require a new survey effort.

3. Off Site Survey Issues

Telephone Survey

The Telephone survey is based on a saltwater angler license file. There are some exemptions allowed which results in undercoverage and the current approach is to use an estimate of the ratio of total anglers to licensed anglers from the access point survey to adjust for the undercoverage. The Breidt et al (2011) review mentioned the benefits of a Dual Frame approach (see NRC 2006, Kott and Vogel 1996 or many sampling texts) to deal with undercoverage of the license file. I totally concur with this recommendation. I think it is likely a much stronger method than the current ratio adjustment method.

Breidt et al (2011) also mentioned the importance of trying to minimize problems with the license file where phone numbers are incorrect or not provided. I totally agree with this recommendation. Response rates in the telephone survey will also have to be considered carefully going forward.

4. References

Breidt, F. J, Opsomer, J. D., and Lesser, V. (2011). Consultant's Report: Preliminary Review of Washington's Puget Sound Sampling Program.

Robson, D., and C. M. Jones. 1989. The theoretical basis of an access site angler survey design. Biometrics 45:83–98.

Kott P. S., and *Vogel* F. A. 1995.Multiple-frame business surveys, in Cox, Binder, Chinnapa, Christianson, Colledge, *Kott* (Eds.), Business survey methods , Wiley,New York.

National Research Council, 2006. Review of recreational fisheries survey methods. National Research Council, National Academy of Sciences Press. Washington.

Pollock, K. H., C. M. Jones, and T. L. Brown. 1994. Angler survey methods and their applications in fisheries management. American Fisheries Society, Special Publication 25, Bethesda, Maryland.

Sarndal, C. , Swensonn, B., and Wretman, J. 1992. Model Assisted Survey Sampling. Springer-Verlag, New York.

Thompson, S. K, and Seber, G. A. F. 1996. Adaptive Sampling. Wiley, New York.