

**Response to Reviewers Comments**  
**Report on a Pilot Study of a New Sampling Design**  
**for the NOAA Fisheries Access Point Angler Intercept Survey**

We appreciate the careful review and insightful comments provided by the two reviewers. We have added clarifications to the report in some sections following reviewers' advice. Listed below are the reviewer's comments and our responses.

**1. *Time of day stratification:***

Time of day stratification versus sampling at peak intervals: This suggestion is quite important to eliminate interviewer judgment about when to interview. Without control in time of day stratification, there could be substantial bias in estimating mean catch and number of fishing trips. Is interviewing assigned at night? It would seem prudent to not sample in the middle of the night. Or to sample with a low frequency in the middle of the night. Perhaps the four time frames are not enough. What if you excluded 12-4am and divided the remaining 20 hours into 5 slots each of 4 hours?

**Response.** Data from the pilot study showed that for certain modes (e.g. shore mode) fishing activities targeting specific species do occur at night in some states and seasons. Therefore the inclusion of the night fishing corrects the temporal and spatial undercoverage of night fishing in the current MRFSS. We recognize the importance of having an appropriate sample allocation across all time intervals to accurately represent the expected effort at each time interval and anticipate a low sampling intensity in intervals A (2:00 am - 8:00 am) and D (8:00 pm - 2:00 am).

As explained in the report we selected the six-hour sampling interval because it would allow for a standard eight-hour workday when travel time to and from the sampling cluster is included.

**2. *Geographic stratification:***

Geographic stratification versus sampling across entire state: This suggestion is important for ensuring coverage of the whole state and allocating interviewers to areas.

Effective and creative selection of strata could lead to efficiency and adequate information for estimation in some sub-state areas. Stratification should be implemented separately in each state. It could reflect NOAA Fisheries Service divisional units in addition to major state-level geographic features. Heavy stratification into small strata could lead to gains in efficiency.

**Response:** We have considered geographic stratification as a means to ensure adequate spatial distribution of the study and to improve precision of the estimates at different levels. We agree with the reviewer that stratification could be implemented at different levels (state and regional) so data collected can be used to inform decision makers and stakeholders. As we move on with the implementation of the sampling design, we will be considering stratification to improve precision estimates.

### **3. Clustering sites:**

In the old design sites were not clustered together, and interviewers were allowed to visit other sites. Allowing interviewers to choose sites invalidates the assumptions of probability sampling and would be likely to lead to bias in estimation. Not clustering sites together could lead to high variability in probabilities of selection if selection is based on activity level. Although not related to bias, high variability in probabilities can be disadvantageous due to associated high uncertainty in estimators. Grouping low activity sites together and visiting all grouped sites should stabilize probabilities of selection, but still allow high rates of sampling among high activity sites. This seems like a good compromise. Control of routes and time spent by interviewers is a critical addition to the design.

Some high activity sites might be included with certainty every survey year due to their importance to state-level fishing. Other sites might appear every few years in surveys. It would be a good idea to review information used to create probabilities of selection every year. Given the availability of GIS and other technologies for recording time and location, it probably would be worthwhile to design measures of performance of interviewing assignments. That is, if you can collect some process information about travel and ease of collecting information by site, then it might be possible to improve resource planning in the future.

**Response.** When clustering the sites, we used a GIS algorithm that clusters sites in the sampling frame by minimizing the driving distance between sites while maintaining similar size measures (effort) across clusters. This algorithm allows for high activity sites to be placed in a single site cluster and to combine nearby sites so the total effort across clusters are similar. This clustering of the sites in the sampling frame will be performed for each sampling wave (two month period) and year, therefore there is no certainty that two sites will be in the same cluster during the same period. Also, a site that has high activity in a given mode in one wave not necessarily will have the same activity level the next wave. Furthermore, the sampling frame size measures are updated for sampling purposes before the beginning of each wave to reflect the changing trends in fishing practices across time and space. We will revise the clustering method section (section 3.1.1.3) in the report to better explain the clustering algorithm. Also, we will revise current examples and will add few more to cover the comments provided by the reviewer and other relevant instances that can clarify the clustering process to the reader.

Currently field interviewers collect information on time and length of interview, site location, and site visit arrival and departure times. This information is used to evaluate productivity and performance measures and to inform the allocation of resources for data collection purposes. We also perform a quality check on interviewer performance by selecting a random sample of their intercepts and verifying among other topics interviewer time and location as well as interviewing technique.

### **4. Sampling frame and probabilistic sampling; elimination of opportunistic sampling; angler trip counts:**

Specifying a formal sampling frame and implementing a probability sampling scheme is a major advance over the former method. Statistical models can be useful for improving efficiency and

removing some amount of bias, but it is best to start with a probability design. Use of models to improve efficiency can still be considered, but one is relying on models to a much smaller degree if the starting point is a probability sampling design.

Eliminating opportunistic switching of fishing modes removes a potential source of bias. Getting accurate numbers of angler trips ending in a given interval is important as stated in the report for assessing secondary probabilities of selection. It will be important to assess whether adequate resources are available for both counting anglers finishing trips as well as interviewing in sites. It might be quite difficult for a single individual to interview effectively and count over a broad area at the same time. According to Appendix A, page 28 (page 33 of PDF)

“At sites with low activity you should be able to both count and conduct interviews at the same time. At sites with moderate-to-high activity you will alternate between counting and conducting interviews by the hour. Be sure to record the start and stop time for the time spent counting and the time spent interviewing as two separate sampling periods, even if you do not switch sites.”

It is recommended that an experiment be conducted to compare the current new plan (simultaneous counting and interviewing) with dedicated counting plus interviewing. The latter likely will require a second person.

**Response:** We thank the reviewer for the review and positive comments about the sampling design. Obtaining accurate number of angler trips is very important since these are the base for the selection probabilities and estimation processes. During the pilot study we did not carry out a formal experiment to compare the efficiency of ‘simultaneous counting and interviewing’ (a single interviewer) versus ‘dedicated counting plus interviewing’ (two interviewers), however, both modes of counting and interviewing were implemented at some time during the pilot study. Feedback from the field interviewers suggested that simultaneous counting and interviewing was feasible and didn’t affect their productivity. Sites with high activity will be in a single site cluster and the field interviewer will count and interview in that site for 6 hours.

##### ***5. Issuing and completing assignments:***

The new design requires issued assignments to be completed without rescheduling. The former design allowed interviewers to reschedule. As stated in the report, “Eliminating assignment rescheduling greatly reduces the possibility of a nonresponse bias that could result from a failure to obtain observations from some of the selected assignments. It also eliminates possible temporal undercoverage biases that could result from the rescheduling of assignments.” It is the opinion of the reviewer that these statements most certainly are true.

**Response:** We appreciate the positive and encouraging comments from the reviewer regarding issuing and completing assignments.

## **6. Interviewing limits:**

The previous design had a cap on interviews. Instead, the pilot design has interviewers utilize the full time frame for sampling. There are a couple of possibilities motivating this change. First, for a given number of clusters, a larger sample size within each cluster is better. Second, interviews of a higher quality provide better information. The change in policy obviously should increase the number of interviews in some clusters. It might also positively impact interview quality as long as the interview period is not too long and tiring. If one knows that one can stop interviewing after a given number of interviews, then one might do the required number of interviews quickly in order to be done with the assignment. Requiring interviewing to continue for a given period of time eliminates the incentive to finish quickly.

It is recommended that available technology (GIS, computer time/date stamps, etc.) be used to monitor interviewers if there is any question about the legitimacy of data collection reporting.

**Response:** We appreciate the positive and encouraging comments from the reviewer regarding interviewing limits. See bullet (c.) for response to monitoring interviewers' issues.

## **7. Eligibility for interviews:**

As I understand the report, the new design allows interviewing children under age 5 and individuals who are returning from a contest. These considerations are beyond my expertise to evaluate. Of course, interviewing children usually requires parent/guardian permission and attention to what the children might or might not be able to reasonably answer. I could see one defining the ultimate sampling unit either as an individual angler or as a group of anglers together (in a boat, in a group such as a family with small children). There could be advantages and disadvantages to both arrangements in terms of ease of interviewing, clarity of definition of unit, and value of information in estimation.

The inclusion of tournaments is specified on page 30 (PDF page 35) of Appendix A, but I see no mention of ages of interviewees. Given that children under age 5 were not included before, it is likely important to add a section stating that they should be interviewed. Procedures for interviewing children (e.g., parental/guardian permission, presence of parent/guardian) should be included.

**Response:** The ultimate sampling unit is the angler. The current procedures manual instructs the interviewer on obtaining parental permission, or permission from a legal guardian to interview anglers of age 5 to 18 years. The interviewer is instructed to interview the young angler or child only after the permission is granted by the parents or legal guardian. We plan to include more detail instructions in the procedure manual for carrying out proxy interviews (an adult answers the questions for the children) when the child is very young or unable to provide the answers. These instructions will be valid for both the tournament intercepts and regular intercepts.

## **8. Complete vs. incomplete beach/bank interviews:**

The inclusion of incomplete angling trip under the old design probably was problematic for estimation. Even if anglers were asked what percent of their trip was completed it would have had some degree of speculation. It is likely as stated in the report that those fishing longer would have a higher probability of being selected for an incomplete trip and also have a higher number of fish, thereby producing bias. Although it makes finding someone to interview harder (they have to be done fishing), this change likely removes a source of bias.

**Response:** We appreciate the positive and encouraging comments from the reviewer regarding complete vs incomplete beach bank interviews.

## **9. Questionnaires and data forms:**

These seem to have been adapted appropriately for the new design. Are any thoughts being given to electronic data capture on, for example, a smart phone, GIS position recording, or time/date stamp to automate the recording of some information?

**Response:** We appreciate the positive and encouraging comments from the reviewer. We are considering implementing electronic data collection procedures in the future. We agree that this will be the next step to improve data collection, and that such will not only facilitate the recording and secure transmission of all information but also will improve data quality and facilitate data management as well.

## **10. Estimation methodology:**

A prime advantage of a probability sampling design is that design-based estimation should produce unbiased (or nearly unbiased in the case of ratio estimation) estimates of desired quantities. It is appropriate to switch to a design-based estimation method.

One could consider the use of statistical models in order to utilize auxiliary information available at all sites. It is recommended that one study correlations among survey outcome variables and information available at each site. Some of this information could be quantitative and other information categorical or dichotomous. Environmental: area of lakes, length of fishing bank, etc. Population: population size within 10, 25, and 50 miles of site, percent of population nearby below poverty and below 200% of poverty, etc. Administrative: is it a state park? Is a state permit required? Is hunting allowed at the site? Is there a swimming beach? It is uncertain what will be predictive, but thinking broadly about potential relationships could guide useful data collection. Even if not used in a statistical model to improve estimation, it still might be interesting to examine associations of auxiliary variables with outcomes.

**Response:** We appreciate the positive and encouraging comments from the reviewer. We recently developed new design-based estimation methods to produce catch and effort estimates by region, state, mode, species, etc. to inform policy makers for managing the east coast fish resources. As we move forward with the survey we will consider model based approaches to improve estimates or to compensate for non-response (e.g. imputation models).

## **11. Comments on some survey questions:**

The introduction to the potential respondent could be rephrased. Instead of, Hi, I'm representing NCDMF, it would be better to not use an acronym in the opening contact. See page 33 (PDF page 38) of the Appendix A (NOAA 2011).

The screening question (recreation) is as follows (page 33, PDF page 38 of the Appendix A, NOAA 2011):

“Was the primary purpose of your trip today for recreation, that is, for fun and relaxation, or was it to provide income either from the sale of fish or from the sale of the fishing opportunity?”

The screening criterion might be better stated, or additional instructions could be provided to interviewers. How would someone be classified who is fishing to have something to eat? Presumably some people fish regularly in order to supplement their diet. I would assume as recreation. What if they trade some fish to someone for vegetables or wild game? Does this qualify as recreation or income?

**Response:** We thank the reviewer for the feedback of the survey forms. Our current procedure manual instructs the interviewer to use the name of the institution and not the acronym. In addition, it instructs the interviewer to include non-commercial (not to provide income) “subsistence fishing” trips as “recreational” trips that are eligible for an interview.

## **12. Comments on results and recommendations:**

In the pilot study, the original design resulted in more interviews per assignment and per hour. Estimates of mean catch rates were not very different under the two methods. If measured simply by estimated sampling variance, the results using the new method are not superior. As Breidt et al. (2012; page 9) notes,

“The estimates generated from the MRFSS sampling design were more precise than the estimates generated from the Pilot design largely because a greater number of sampling assignments were completed under the MRFSS design.”

One can speculate, however, that the new design actually could be better in terms of selection bias and other problems. As in most surveys, it is not really possible to measure the amount of bias, because it could only be compared to a much more rigorous and large effort.

“It should be noted, however, that the potential for non-sampling errors was much larger under the MRFSS than under the new design, which is not reflected in these precision comparisons.” (Breidt et al 2012; page 9)

One also can speculate that the performance of the new design can be improved. Although sampling under the new design in this study yielded a much larger percentage of completed assignments with no angler trip interviews and a much smaller number of interviews

per positive assignment, changes in the allocation of sampling across sampling strata could greatly reduce these differences.

Importantly, one could try to analyze the existing pilot survey in order to better inform choices about the new sample design.

**Response:** We thank the reviewer for the insightful comments. This is indeed the intent expressed in the report.

### **13. Allocation of sample:**

The report gives suggestions on how to improve. The 2<sup>nd</sup> and 3<sup>rd</sup> recommendations are critical. One should study when, where, and for which mode the MRFSS survey gained such large amounts of sampled units. Time stratification and whether to exclude the middle of the night are topics that need consideration.

One should also identify resources before allocating sample. It sounds like resources were woefully inadequate in some areas. One can implement legitimate probability sampling schemes with controls connected to geography and resources, and more effort apparently is needed in this direction.

**Response:** We thank the reviewer for the insightful comments. As discussed before (bullet a.) we are working on obtaining accurate size measures (effort) for the sites for all modes and time intervals. These size measures will inform the sampling design and the sample allocation across kind of day and time interval. We anticipate that the observed fishing activity at night for some modes will result in few sites selected at intervals A and D.

We agree that identification of resources is essential in the success of a sampling design as the one proposed here. For example, a small number of interviewers would be able to successfully cover a large area and a diverse schedule that combine day and night sampling. We anticipate that some of the limitations encountered in the pilot study will not be an issue when implementing the new sampling design.

### **14. Splitting high intensity sites:**

Among issues to study in the future, splitting high intensity sites so that size is not so large among the largest probably is a good idea. Handling a mix of very small and very large sites is difficult when a limited amount of time is available for each and there are substantial travel times to the next site. It is recommended that stratification become finer and clusters be formed to be more uniform in size. Recommendations for future research #9, 12, and 13 likely should be moved up in priority. One aspect of this is time stratification. Surely one can omit 12-4am and do a better job. One could consider splitting the remaining twenty hours into five 4-hour blocks. One then could still visit two low intensity sites (up to 2 hours each) in a single cluster if they are geographically close enough.

**Response:** As discussed in bullet (c.) above the current clustering algorithm produces clusters by minimizing the distance between them while maintaining an almost homogeneous size measure across clusters. The NOAA staff in charge of running this algorithm works with the states to split large sites into smaller and manageable sizes sites that can be surveyed in 6 hour intervals.

### ***15. Adjusting allocation to mode, or allowing mode switching:***

In Breidt et al. (2012), it is noted that

“For the Pilot, assignments were allocated evenly across the four modes in each state subregion: Man-made (MM), Beach Bank (BB), Private/Rental (PR), and Charter (CH). Allocation of mode-specific assignments within each state subregion and day type (i.e. kind of day) was determined monthly.”

It seems to this reviewer that the actual modes used would not be close to even in every site. Being able to switch modes would have given the original survey a big advantage. Perhaps one can learn from the actual data collected (pilot versus original design) how mode restrictions impacted data collection.

**Response:** For the implementation of the proposed sampling design we are not proposing equal allocation across modes. In the pilot study we used equal allocation across modes to ensure enough sample to produce some estimates at different levels and to be able to make some comparisons between modes and across surveys. We are proposing to allocate the sample based on current and historical data as it is currently done in the current MRFSS.

### ***16. Performance of the proposed new design:***

The proposed new design and implementation rules undoubtedly produce estimates whose properties can be evaluated with much more precision than those of the MRFSS. I also believe that with adjustments and improvement, the new design will produce estimates that are, on average, better than those of the MRFFS. However, that message does not come through strongly enough in this report, in my opinion. The productivity measures (interviews per hour, etc.) and the standard errors reported (Figures 4,5,9 and 10) always show the pilot at a disadvantage. There are mentions of possible biases in estimates from the MRFFS, but then in other places the new estimation method developed by Breidt et. al. are described (p. 13) as approximately design unbiased.

The only statements about standard error are vague but suggest that even if sampling hours were the same, the Pilot would have been better but would still compare unfavorably to MRFSS (p. 60):

“If the number of PSU’s observed in the Pilot design had been increased to match the number of assignments in the MRFFS design, the variances of the mean catch rates would have been lower.”

but then:

“If both designs had completed the same number of assignments, the MRFSS design would still likely have provided estimates with greater precision”

But how much better? The report doesn't make it easy for the reader to make a fair comparison. It would be useful for evaluating to have something like design effects for the two methods, since there are competing factors make the Pilot both less efficient (lower interviews per assigned hour) and more efficient (less variable weights, as mentioned on p. 54). It isn't clear (to me) how to do this precisely from the data, but even some rough measure of “variance per sampler hour” would be helpful.

Another analysis that would be useful to see would be a prediction of how much the design effect could be improved after implementation of some of the suggested improvements in allocation to the various strata. It seems that this could be calculable for at least a few major species or total catch, to give an idea of the potential value of the new method.

And of course this doesn't even consider the possible residual biases in the MRFSS data. The similarity of both sets of estimates would seem to suggest that there is little bias left in the newly weighted MRFSS, but then we get statements like this one on p. 8 :“... the potential for non-sampling errors was much larger under the MRFSS...” So does the statement on p. 8 refer to the “old MRFSS” or does it mean that even the “new MRFSS” is believed to still have substantial non-sampling errors. If the latter, it would be helpful to briefly describe what kinds of non-sampling errors the authors think still remain. Presumably, the authors feel there are smaller non-sampling errors than with just the original MRFSS estimators. Is that the interpretation of the authors? I find that surprising, although it doesn't mean that even if it was true in NC in this year it will always be true. Does this mean then that the average catch per trip really does not vary much by time of day (night vs. day), or at alternative sites in the same vicinity (chosen by MRFSS samplers)? Or is it that the precision was so low we couldn't tell there was a difference? More discussion of what evidence there is about this would be helpful to explain the advantage of the new design to those who might not be so convinced of the superiority of a strictly probability sampling design on theoretical grounds alone.

**Response:** We appreciate the careful review and insightful comments provided by the reviewer. We agree with the reviewer that plots and figures need better interpretation to convey the message we tried to communicate to the reader. We will revise some of these paragraphs as suggested by the reviewer. We will also revise the reference to the MRFSS and “old MRFS” to be consistent through the report.

The MRFSS sampling design suffers from an inherent space and time undercoverage resulting from the discretion the interviewers have in the MRFSS to select the interviewing time and alternate sites. This undercoverage was removed in the new sampling design by incorporating stratification by mode, kind of day and time interval and clustering of sites.

The goal of the comparisons discussed in the report was to learn specific aspects to be improved in a future implementation of the new sampling design. For example, we need adequate sample allocation across modes and time intervals. Given the difference in number of assignments

(sample size) between the MRFSS in the pilot study, we expected the precision of the estimates of the MRFSS to be sometimes better than those obtained from the pilot. An implementation of the new sampling design will result in larger number of assignments for each state which will likely result in more precise estimates than those obtained in the pilot study. We will add some paragraphs to clarify this goal in the report.

### **17. Interpretation of results:**

I would like to see more interpretation of results, or explanation of most likely causes, when there are differences between MRFSS and Pilot results. Here are some places where that would be informative:

#### **a. Effort ratio estimator (p. 8):**

The summary states that the proportion of anglers reachable by the CHTS is comparable for all but beach/bank mode. Is there any intuition or speculation about why that would be true? Also, I believe it is true that an angler is not in the CHTS frame if he or she does not have a landline phone. If that is correct, it should be noted as part of the description of “the effort ratio estimator.”

**Response:** There is not enough information to fully explain the observed differences in the beach/bank mode discussed in page 8. Possible causes are the removal of the incomplete trip interviews and the inclusion of nighttime sampling. Local residents may fish for longer durations and may be more likely to fish at night than the out of state residents or non-coastal state residents. We will add text pointing out these possible explanations and suggest that further study is warranted.

#### **b. Average number of interviews obtained per two-hour interval (p. 37):**

Do you have any speculation on why the jagged curve for the pilot? Is it just because of the small sample size, or could it be because of the rules for clusters that have people driving from one place to another at certain hours?

**Response:** We thank the reviewer for bringing to our attention that we overlook explaining the jagged curve depicting the frequency of intercept per hour (Figure 3). We agree with the reviewer that the up and down pattern of the curve for the pilot might correspond to the driving between sites. We will add a sentence to the corresponding paragraph commenting on this.

#### **c. Average numbers of fish reported and observed (p. 38):**

It is hard to know what to make of the differences seen in Table 6. What is the difference between reported and observed? Were these fish chosen because they are ones that are particularly common? How is the average # of fish computed? Is it average per angler overall, or average per angler that targeted or has any of that fish? Do these differences seem reasonable or explainable (e.g., are some fish caught more at night, for example, and so may

be more likely to be seen in the Pilot?)

**Response:** We agree with the reviewer that adding a sentence explaining the definition of the ‘% Difference’ column shown in Table 6 (page 38) will help with the interpretation of the table results. We will add few sentences in the paragraph before the table to clarify that 1) the species showing in the table were selected because these are the most common species in North Carolina, and 2) the average was calculated as the average number of fish among anglers who caught that species.

*d. Differences in catch estimates (p. 47):*

The fact that the MRFFS estimates are higher when there is a difference is described as “interesting.” Is there a reason to believe this is a real difference and indicates a bias on the part of MRFFS data, due to either the noncoverage of night fishing or some other reason?

**Response:** We will add a sentence in the report to discuss some of the reasons that might have resulted in the observed differences.

*e. Percent of beach/bank trips by coastal residents (Table 7):*

Does it make sense that these estimates would differ due to noncoverage of night fishing in MRFFS, or is there some other reason? Maybe only locals fish from the beach at night?

**Response:** We appreciate the comment from the reviewer. Although non coverage of night fishing in the MRFSS is a possible cause for the observed difference, the data we obtained in the study did not show a significant difference in the proportion of coastal residents for sampled nighttime and daytime beach/bank trips.

*f. Estimation of total catch:*

On p. 7 and again on p. 38 are mentions that the telephone survey (CHTS) is used along with the intercept data to come up with total catch estimates. I believe these descriptions are too cryptic for a reader who is not already familiar with the estimation method to understand. Even a brief explanation that total catch is estimated as (total trips by coast county anglers) \*(catch per trip for all anglers) \*(# of all anglers/# of coastal county anglers) would be helpful, and would help elucidate some of the comments about why bias can result if certain assumptions (e.g., avg. catch the same for coastal and noncoastal) do not hold.

**Response:** We thank the reviewer for this comment and will add some sentences on page 38 (and will make a reference to it in page 7) to explain the calculation of total catch.

**18. Clarifications Requested:**

*a. Which version of MRFSS? (p. 4):*

When the report says that the Pilot compared the performance of the new sampling design and estimates of catch per trip, etc. with the “traditional MRFSS sampling design” (p. 4), does this mean that the comparisons are with the old design but the new estimators developed by Breidt et al.? The remainder of the report seems to suggest the comparisons are always to the new analysis method for MRFSS, but the use of the word traditional made me wonder as I was reading the executive summary.

**Response:** We agree with the reviewer that we need to be consistent through the report when mentioning the “traditional MRFSS sampling design” which refers to the current MRFSS. The estimates for the traditional MRFSS used in the comparisons were obtained by applying the new estimation procedures to the data collected using the old design. We will revise the report accordingly.

***b. Number of sites per assignment (p. 8):***

How does the new design obtain nearly 3 sites per assignment (2.46, from p. 8)? Does that mean that most assignments are 3-site clusters? It might be good to have a summary somewhere of how many units in the frame are of each size.

**Response:** We will add few sentences describing the frequency of clusters with 1, 2 and 3 sites visited during the pilot study.

***c. What is a “positive” assignment? (p.10 and p.61):***

What does the word positive mean in this statement on p. 10 “provide higher average number of interviews per positive assignment.” Also on p. 61.

**Response:** The term positive assignment refers to an interviewing assignment that obtained at least one interview. We will revise the corresponding sentences in the report.

***d. Recommendation on site clustering rules (p. 11, recommendation #6):***

It is not clear in what way cluster formation would be modified. More two-site clusters? Why would that happen?

**Response:** When clustering the sites, we used a GIS algorithm that clusters sites in the sampling frame by minimizing the driving distance between sites while maintaining similar size measures (effort) across clusters. This algorithm allows for high activity sites to be placed in a single site cluster and to combine nearby sites so the total effort across clusters are similar. In the pilot we didn’t enforce the constraint of achieving a total pressure across all sites, so sites were clustered based on proximity and having a pressure lower than a given threshold. If a threshold level of pressure was used to determine whether or not a given site could be clustered with one or two additional sites, more two-site clusters would be allowed. This could reduce the average driving time and increase average interviewing time per assignment. We will revise the explanation of this recommendation to clarify.

*e. Site list maintenance (p. 11, recommendation # 10):*

Apparently there was some indication that the site list had some flaws. Could these be clarified?

**Response:** The maintenance of the site registry is an ongoing task that needs to capture trends in space and time. Size measures for some sites might not reflect the actual fishing activity for given modes or site status (e.g. inactive and active) may not be updated. We are currently working on obtaining accurate size measures (effort) for the sites for all modes and time intervals; including new sites (e.g. new piers, private piers, etc) and making sites inactive based on the accessibility or closings due to maintenance or for ecological reasons (e.g. bird nestings). We will add few sentences explaining the nature of these maintenance tasks.

*f. Explicit vs. implicit stratification (p. 12 Recommendation #12):*

Clarify what is meant by “explicit” and “implicit”

**Response:** We will add few sentences and examples to illustrate the concept of explicit and implicit strata.

*g. One psu per month? (p. 23):*

[The text reads] “at least one psu was selected from each interval...” Clarify that this means “per month”

**Response:** We thank the reviewer for detecting the missing “per month” in page 23. We will add these two words to the corresponding sentence.

*h. X-axis in Figure 2 (p. 36):*

The x-axis scales are different on these plots. I don’t think they should be. In fact, why should n’t they all be from 0 to 24? Redraw graphs using 0-24 x-axis.

**Response:** We will redraw the plots and use the same 0-24 range in the x-axis.

*i. Meaning of “fishing area”(p. 39):*

I don’t understand the difference in the last two sentences of the paragraph before Table 7. They seem contradictory, but then I don’t know what “fishing area” means.

**Response:** Fishing area refers to location of the fishing site (ocean within 3 miles, ocean outside of 3 miles, and inland) as described in the last sentence of the first paragraph on page 39. We will add few sentences to clarify this definition and will review the two sentences referred by the reviewer to better convey the message to the readers.

*j. Figure labeling (Figures 4 and 5):*

Figures 4 and 5 should have a label of Pilot and MRFSS to be consistent with the other descriptions, rather than NC and MRFSS.

**Response:** We thank the reviewer for the detailed review of the report. We will change the word “NC” for “Pilot” in both figures to be consistent across the report.

**k. *Future recommendation #1 (p. 10):***

I do not understand the meaning of #1 under future recommendations.

**Response:** We will add few sentences in the report to clarify recommendation #1 to consider obtaining total counts of boat trips in addition to total counts of angler trips.

**l. *“Inappropriate” sampler choices (p. 53):***

The word *inappropriate* seems wrong. It is not that samplers are making bad choices, but that they are making choices at all that is wrong with MRFSS.

**Response:** We agree with the reviewer that the word “inappropriate” seems wrong in the sentence and paragraph context. We will rephrase the sentence to make emphasis that is the discretion the interviewers have to select alternate sites in the MRFSS is the main issue affecting the MRFSS.

**m. *Frame maintenance assignments (p. 69):***

The need for frame maintenance assignments is mentioned a couple of times, but no information on the reason for this suggestion is mentioned. Was there some problem with the list of sites that was discovered during the pilot?

**Response:** The site registry (frame of the MRFSS and new sampling design) must be updated to reflect the space and temporal trends on fishing activity across the different modes. New sites must be added, size measures (effort) must be updated, contact information, site closures and security issues must be updated to better inform the sample selection and the site clustering exercise. We will provide information on the rationale for this recommendation. We did not discover any specific problems with the list of sites used in the pilot study. We just think that the completeness of the list and the accuracy of the site pressure estimates in it would be improved by allocating more resources to frame maintenance tasks.

**n. *“Probabilistic” sampling? (p. 5):***

Probability sampling is referred to as probabilistic sampling on p. 5. Is this meant to imply something different than what we usually mean by probability sampling?

**Response:** No. we were referring to probability sampling. We will make the changes in the report.

*o. Quotes? (p. 15):*

Why are there quotes around the statement under item #3? I presume because it is quoting from the previous report, but it should say so.

**Response:** We will add the corresponding reference to item #3.

*p. Night sampling and the percentage of coastal county anglers (pp. 52, 53,55 and several other places):*

There are comments about the fact that one advantage of the Pilot covering all times is that it eliminates potential bias due to different catch rates at night. But it also eliminates potential bias caused by different coastal county anglers' percentage at night. I believe this should be emphasized.

**Response:** The reviewer is correct that we need to emphasize that with the new sampling design we will also eliminate potential bias by different coastal county anglers' percentage at night. We will add a sentence in the corresponding paragraphs.

*q. Overtime for samplers? (p. 52):*

"It was not necessary to require samplers to regularly work overtime." Was this a problem with MRFSS? If so, then state this advantage.

**Response:** This was not a problem with MRFSS sampling. We were simply pointing out that it was not necessary in the pilot study to require any samplers to work more than a 40-hour week to complete the assignments drawn under the new sampling design.

**19. Other comments:**

*a. Fishing trips ending on private property:*

Executive summary and p. 13 mentions that the MRFSS survey was "...not providing coverage of fishing trips ending on private property or at night." Much discussion was devoted to the issue of improving night fishing, but no other discussion of private property was included in the report. It seemed a bit like a "bait-and-switch" as I kept looking for discussion of innovations for the private property issue. Maybe this should be removed from the overview.

**Response:** We thank the reviewer for the careful review. The site registry which is the sampling frame for the MRFSS and the new sampling design is affected by possible under coverage of private landing sites (e.g. piers, marinas, etc). We will add a statement in both the summary and page 13 clarifying this issue.

***b. Recommendation #11 for future consideration (p. 12):***

The idea of basing pressure measures on # of fish landed seems a good one, especially if this allocation is very different than the one being used. It would be even better if the variability in fish landings could be used, though maybe mean and variance of fish landings are positively correlated.

**Response:** We appreciate the reviewer's comment and suggestion on our intent to explore new ways to obtain better size measures. As we move on with the new sampling design we will explore the impact of different size measures (e.g. # of fish landed) or using the variability in fish landing (as suggested by the reviewer) on sample size selection and sample allocation.

***c. Recommendation #8 for future consideration (p. 68):***

I like this idea a great deal. In fact, the potential for cost savings if the entire effort estimation were to be moved to the intercept survey is large. With the new design, my guess would be that number of trips could be more accurately estimated from the intercept survey than from telephone or mail, due to the large measurement error and nonresponse of the effort survey. The only problem I see is the private access sites. How much of the total effort is on private access sites? Maybe a telephone survey would still be necessary but only to estimate this ratio rather than the current coastal county ratio estimated from the intercept. Why not just compare total public access estimates from telephone and the intercept survey rather than needing to estimate by site (thought this is perhaps not a discussion for this report)? If the intercept survey were to be used for both effort and catch, one might consider counting both arriving and departing anglers, since the spatial-temporal time frame would be appropriate for either. The two estimators would be correlated but not identical, so should provide a little extra information basically for free.

**Response:** We agree with the reviewer that in order to get information from the private access sites, other means of data collection (such as a phone or mail survey) in addition to the intercept survey is needed. As we move on with the sampling design we will explore the reviewer's suggestions and other approaches to obtain better estimates of effort.