

**Summary Report
Marine Recreational Information Program (MRIP)
Access Point Angler Interview Survey (APAIS)
Calibration Peer Review**

**Calibration Peer Review Workshop
March 20-22, 2018
Sheraton Hotel
Silver Spring, MD**

April 25, 2018

Panel Members

**Michael D. Murphy (Chair)
Carolyn Belcher
Mary Christman (CIE Reviewer)
James Chromy (CIE Reviewer)
Matthew Cieri
Paul Rago
John Whitehead (CIE Reviewer)**

Table of Contents

Executive Summary	3
1. Introduction	4
<i>1.1 Background</i>	4
<i>1.2 Review of Activities</i>	4
2. Review of MRIP APAIS Calibration Model	9
<i>2.1 Synopsis of Panel Review</i>	9
<i>2.2 Evaluation of Terms of Reference</i>	10
<i>2.2.1 Term of Reference 1</i>	10
<i>2.2.2 Term of Reference 2</i>	13
3. Bibliography	14
4. Appendices	16
Appendix 1. Terms of Reference.....	16
Appendix 2. Draft Review Meeting Agenda	17
Appendix 3. Individual Independent Peer Review Report Requirements (non-CIE).....	19
Appendix 4. Individual Independent Peer Review Report Requirements (CIE).....	20
Appendix 5: Calibration Workshop Attendees	28

Executive Summary

The MRIP team has addressed an extraordinarily difficult survey problem in an innovative way that has important implications for stock assessments of many of the nation's most valuable fishery resources. An essential ingredient for the proper management of fishery resources is an accurate estimate of the number of fish removals. Long time series of consistent estimates of recreational catch help establish the appropriate scale of the population and help characterize potential productivity. MRIP, and its predecessor MRFSS, have provided recreational catch information since 1981. Over time, a variety of technical and statistical problems were identified by various internal and external reviews. This process of improvement in the recreational fishery survey led to MRIP, a major improvement in the overall methodology. Such changes are welcome but inconsistencies with historical data have increased uncertainties in stock assessments and fisheries management. Hence there is a strong impetus to develop methods that can be used "calibrate" the more recent improved data with the less rigorous older estimates. "Calibration" herein is used in a colloquial sense to mean one or more adjustments to historical MRFSS data to improve their scientific credibility and consistency with the recent data collected under MRIP.

As part of the long-term objectives of MRIP to improve the precision and accuracy of recreational catch estimates, the access point angler intercept survey (APAIS) was revised (after wave 1 in 2013) to reflect a probability-based survey design that has been field tested to minimize logistic constraints that negatively affect productivity of angler interviewers but improve the probabilistic integrity of the survey. Because the earlier survey interviews were not conducted within this new, strict survey design there was a need to adjust the sample weights of the earlier observed catch per unit effort data to levels that would have been used if those earlier data had been collected under the current probabilistic design.

An innovative approach was used to adjust APAIS sampling weights using the relative observed total sampling efforts within domains (cell levels at the smallest units) and a well-known 'raking' algorithm to distribute these efforts across important domain dimensions. The proposed strategy also analyzed temporal trends in effort to determine the average over the appropriate time period to use as the basis for the weight adjustment ratios. This process was used to adjust the 2004-2012 'pseudo-weights' to new adjusted weights that theoretically would have been observed if the new APAIS statistical design had been employed during 2004-2012. The new 2004-2012 adjusted weights were then used as a basis for adjusting the weights for the 1993-2003 period and sequentially these new 1993-2003 adjusted weights were used to develop adjusted weights for the 1981-1992 period. Given that the proposed approach is conceptually superior to the ratio-based methods it can be concluded that the proposed method was preferred by the Panel.

The proposed methodology resulted in major changes to some of the sampling weights within domains. Such changes do not necessarily imply major changes in catch estimates but at present it is difficult to know if the revised weights exert undue influence. The Panel recommends further investigation of potential outliers and robust methods for revision. Outlier detection and truncation approaches should be objective, repeatable, and independent of the consequences of a single species.

1. Introduction

1.1 Background

The Review Panel for the MRIP-APAIS Calibration Peer Review Workshop met from March 20 to March 22 in Silver Springs, Maryland to evaluate an approach for converting historical estimates of mean angler catch rates to estimates that best represented what would have been produced had the new MRIP APAIS sampling design been in place prior to 2013. The review committee was composed of three scientists appointed by the Center for Independent Experts (CIE): Mary Christman, MCC Stats Consulting & Gulf of Mexico Fishery Management Council; James Chromy, RTI (retired); and John Whitehead, Appalachian State University; and representatives from the Mid Atlantic (Paul Rago, retired NMFS) and South Atlantic (Carolyn Belcher, Georgia Department of Natural Resources) Fishery Management Council's Scientific and Statistical Committees; and the Atlantic States Marine Fisheries Commission (ASMFC, Matthew Cieri, Maine Department of Marine Resources). The meeting was chaired by Michael Murphy, retired from the Florida Fish and Wildlife Conservation Commission.

The Panel reviewed supporting documents and presentations prepared by NOAA Fisheries' Office of Science and Technology (OST) staff, led by David Van Voorhees, and their contractors from the Department of Statistics at Colorado State University. Presentations were made by David Van Voorhees, Jason Didden (Mid Atlantic Fishery Management Council (MAFMC)), Katie Drew (ASMFC), Ryan Kitts-Jensen (OST), Tom Sminkey (OST), Jean Opsomer (Colorado State University), and John Foster (OST). NOAA Fisheries OST staff members Richard Cody, Ryan Kitts-Jensen, and Tom Sminkey acted as rapporteurs, providing valuable daily summaries for the Panel presentations and discussions. Other staff and contractors from the OST, notably Karen Pianka, assisted in the efficient handling of documents via a web-based Wiki site. Jason Didden (MAFMC) provided extensive technical and audio-visual support for the webinar. Approximately 20 people participated in the open sessions of the meeting and numerous members of the public followed the webinar on-line and contributed questions during the sessions. The meeting followed the agenda items in Appendix 2 with respect to the sequence but not necessarily the timing of the events as adjustments were made for differences in the duration of presentations and follow-up discussions.

1.2 Review of Activities

Panel members were given on-line access to a broad range of background documents including the two National Academy of Science reviews of the survey, descriptions and reviews of past calibration methods, the development, description and implementation of the current APAIS, and the supporting literature and description of the proposed method for developing revised weights to be applied to past trip data to best mirror the current statistical design. Prior to the meeting, an introductory conference call was made on 9 March where OST staff explained the peer review process, terms of reference, and the availability of documents on a Wiki site. At this meeting, the Panel decided to review the documents and terms of reference by Wednesday March 14 and then to develop a list of preliminary questions and remarks to be given to the analysts to consider in guiding the development of their presentations. These questions were forwarded to the NMFS analysts the following Friday and another conference call was held late that day to further discuss

the questions and the review process. The meeting was broadcast via webinar with the able assistance of Jason Didden (MAFMC).

David Van Voorhees started the meeting at 9AM on March 20, 2018 providing an introductory presentation highlighting the transition team's make-up and its approach to development of the MRIP statistical design change, general APAIS survey design, effort survey changes, and proposed methods for converting early catch estimates to the modern survey's 'currency'. This included a description of the program's evolution, external reviews, and past calibration attempts. The descriptive timeframes for the survey's collection of catch rate data were: 1) 1981-1990 using an unweighted, simple random sampling design, 2) 1990-2004, using unweighted, simple random sampling with quality assurance and control improvements, 3) 2004-2012 using 'pseudo-weights' developed later reflecting a true sample-weighted design, and 4) 2013-present, new probabilistic weighting under the recognized multistage cluster design. A clarification was requested about the 'calibration on calibrations' approach to the 2004-2012 conversion from unweighted MRFSS estimates to the new MRIP-design, 'pseudo-weighted' MRFSS estimates and this conversion was referred to as an 'incomplete calibration'. Michael Murphy officially opened the meeting after this introductory presentation and introduced the tasks at hand, the first day's planned activities, and introductions for those attending the workshop.

The early meeting presentations stressed the significance of producing an acceptable calibration for the older APAIS mean angler catch rates so that the estimated landings and biometrics would be comparable through time. Jason Didden emphasized the importance of consistency in the catch time series for management of in-season quotas, development of accountability measures, and allocation of catches to jurisdictions and stakeholders. Given the time constraints to the ability to produce new assessments, it is important to be able to convert in both directions between any new estimates and old estimates.

Katie Drew gave a presentation showing the significance of a stable, accurate catch series for input to stock assessments. Estimates of catch are typically the primary data that allow assessments to scale abundance trends to absolute levels, providing the basis for quotas and ACLs. Re-weighting of the data through a calibration can have significant effects on the estimation of catch-at-size, selectivity, mortality, and management benchmarks. A brief discussion was held on the complexities associated with a time-varying adjustment compared to a constant scalar adjustment associated with the calibration.

Finally, before the mid-morning break, the Panel reviewed the Terms of Reference and understood it as a task to investigate the suitability of conversions based on the three potential calibration methods: simple ratio, complex ratio, and proposed weight-adjustment method. The Panel understood 'suitability' to mean how likely the converted older survey results would reflect what would have been if the new APAIS design had been in place during the past.

Following the break, David Van Voorhees gave a presentation on the development of weights that reflected the true inclusion probabilities during the MRFSS, specifically during 2004-2012 when detailed sampling effort information was available. This required developing inclusion probabilities to calculate sampling weights that reflected the true multistage clustering design and the alternate site selection option employed by samplers during this period. Determining the

inclusion probabilities also required development of an innovative time-slice model that borrowed from fishing trip time-of-day data available from the Coastal Household Telephone Survey (CHTS). A small amount of historic sampling data was discarded because no inclusion probabilities could be logically determined for the infrequently used alternate-mode sampling option. Finally, after peer review of this new weighted estimation technique (Breidt *et al.* 2014), the procedure was accepted, and the results were certified as the best estimates available. The first calibration workshop utilized the ratio of the new ‘weighted MRFSS’ catch estimates and the unweighted MRFSS catch estimates during the 2004-2012 period as the basis for developing conversion ratios for earlier years (Boreman *et al.* 2012). The Panel had several questions at this point.

The Panel wanted to know why the MRFSS data prior to 2004 were not included in the re-weighting and learned that this earlier sampling was conducted under less quality control and often reflected quota sampling. This and the loss of important sample data meant that earlier years could not be confidently re-weighted. Dave Van Voorhees noted that an unavoidable bias existed because the time-slice model was used to scale up the sampling across the fishing day, when no actual observations of species caught during the unsampled portion of the day were available.

A presentation of the new APAIS using a fully formalized probability sampling design was given by Tom Sminkey. The new survey reflects the new survey design that was developed, reviewed, and accepted (Breidt *et al.* 2012) and included a series of revisions made following a side-by-side (MRFSS/MRIP) pilot study of the APAIS productivity and logistics. Important improvements included a new site-cluster sample frame, sampling across a selected time frame (from four, 6-hour time blocks and two overlapping time blocks needed to accommodate the typical charter fishing daily schedules), survivor-set sample draws that consider sampling effort limitation, mixed boat-mode sampling, and a new offshore stratum. All modifications were fully accounted for in the probability-proportional-to-size sample inclusions. Several questions and a general discussion followed. Further clarification indicated that the site-clusters were defined at the smallest level of sampling so specific cluster membership could change over the year reflecting changes in the distribution of fishing pressures. Sampling was not ended immediately at the end of the assigned 6-hour sampling window. The sample draw that accounted for sampler-availability limits was fully accounted for in the estimation design, so every sample-site had a chance of being included and the variance was properly estimable.

The final presentation before lunch on the first day was made by John Carmichael who reviewed the information and findings from the second calibration workshop (Carmichael and Van Voorhees 2012). The task for this workshop was to look at the changes made from the 1981-2012 MRFSS APAIS and the fully-implemented 2013 MRIP APAIS and evaluate three methods for converting the older estimates to reflect the new design. He briefly described the three methods: 1) simple peak-time catch rate ratio, 2) time-specific catch rate ratio method, and 3) regression model time-of-day classification model. The last method was never fully developed and was not considered further but the two ratio methods were more fully evaluated by the Panel later (see below).

The afternoon session began with a review of the two ratio methods proposed for ‘calibration’. Ryan Kitts-Jensen made a presentation describing the two ratio methods and provided examples of their use. Given that the methods are dependent on the time-specific breakdown in trip catches

under the new APAIS and contrasting those to catches made during the same sampled time periods in the past, changes in the sample times were examined for 2010 through 2013. Changes were seen over time, indicating the need to develop ratios within comparable time blocks. The first ratio method was developed to compare catch rates during a locally-defined ‘peak’ time-period. One complication was the changing definition of peak time-period and required sensitivity analysis. This method also excluded significant amounts of data collected outside this peak time window. Also, a significant number of rarer species would require some aggregation of data across strata (possibly including year) to get non-zero observed catch during the peak time period. Finally, these ratios would need to be developed for each species. The second ratio method was a complex version of the first ratio method that utilized the distribution of catches within several 3-hour sampling windows during the day, with distributions of catch prior to 2013 being raised to reflect the 2013 and later time distribution of catch. This method required higher proportion of aggregated data to be able to calculate defined time-specific ratio adjustments. The Panel was concerned about the inconsistencies (e.g., estimates by groups of species would not necessarily equal the sum of estimates for these individually), species-specific development of ratios, the high level of aggregation of data and loss of domain-specific adjustments, and how this would be extended when adjusting the estimated catch at size or weight data.

The proposed calibration method (also referred to proposed approach in this document) was described conceptually in a presentation made by Jean Opsomer. He provided an informative background of how various long-term surveys in other disciplines respond to changes in their design. Responses ranged from ignoring the survey change because the effects are thought to be insignificant, to adding a disclaimer so users were warned about the change, and least commonly, to producing calibrated results. The difficulty with the current ‘calibration’ is the lack of overlap between the old and new APAIS. For this reason, the proposed conversion is an attempt to re-weight all earlier observations at the angler-trip level to reflect the new survey design. If the re-weighting is successful then the older sampled data are now weighted as if they had been collected under a new, known sampling design.

The methodology used known-design total sample weights¹ within several important domains to determine sequentially (through a ‘raking’ algorithm; Deming and Stephan 1940) the distribution of past weights within these domains. These were then scaled to the mean weight and applied to angler-trip data for mean catch rate estimation. The sequence of developing the new weights was described in detail (see Opsomer *et al.* 2018). Due to changes in the survey and availability of methods for deriving angler-trip weights used at the time, the method first calibrated the MRFSS ‘pseudo-weight’ period of 2004-2013_(wave 1) to the MRIP 2013_(wave 2)-2016 period then used the newly derived 2004-2013_(wave 1) weights to calibrate the 1993-2003 weights (considered a different quality to earlier data), then finally used the newly derived 1993-2003 weights as the basis for calibrating the 1981-1992 angler-trips.

An important feature described later was the introduction of additional ‘raking’ domains as the process went back in time. These new domains were considered the most important in influencing how the choices for sample site selections were made during the earlier periods. Also, a sequence

¹ All the references to “weight” in the following paragraphs refer to sample weights, i.e., the inverse of the inclusion probability. References to revisions to the weight of individual fish in a sample or total weight of fish caught will be noted explicitly.

of checks for linear changes in sample effort within each time block was used to determine the lengths of the time-period that served as the basis for the calibration. The Panel questioned the choice of when the time-period boundaries occurred. A suggestion was made that when the 2017 data become available, the analysts might consider using only the 2014-2017 period as the initial calibration basis. This would allow for a sampler-adjustment period before the routine implementation of the new APAIS sampling scheme. The analysts reminded the Panel that the analyses that they presented during the workshop were still preliminary and were subject to some revisions. Others Panel members questioned the choice of using linear regression tests. This was defended for simplicity, with no real way to evaluate non-linear effects. A finer domain breakdown for the raking algorithm (e.g., area fished x sub-state region, e.g., in Florida and North Carolina) was suggested. Analysts believed that this was not feasible due to small-domain estimation issues. Several advantages of this proposed calibration method were pointed out: the maintenance of the public-use 'micro data' with just a weighting change, the better consistency of conversion at the angler-trip level rather than at the species level, and the continuity in using the same routines to generate landings and effort estimates from the survey. In this report 'micro-data' refers to the detailed APAIS sample data at a level of resolution needed to reconstruct estimates of mean catch per angler-trip using the appropriate algorithms.

The final presentation of the workshop, given by John Foster, began Tuesday afternoon and continued through mid-day Wednesday. This presentation provided the results of the proposed re-weightings and the distribution of preliminary changes seen across many different strata. This presentation was very detailed, partially in response to requests made by Panel members before the workshop after the initial review of the available documents. Though the effort was staggering, the Panel struggled with using this detail to review the efficacy of the proposed calibration method. The Panel was most concerned about the presence of extreme adjustments to sample weights and their impact on accumulated results, e.g., annual landings. However, it was unclear what the cause of some of the large changes was and how significant it would be to the final survey products. For instance, was the absolute adjustment relatively small but made to an extremely small initial weight (very small number in denominator) or is the adjustment applied to angler-trips with little to no catch or to catch of species that are not adequately sampled within the survey, e.g., cast-net caught herrings? For these reasons, the Panel turned to questions and a discussion about techniques that could be used to trim outliers and to whether this would or should be done.

The summary from the final presentation noted the comparisons between the results of conversions using the simple and complex ratio methods and the proposed weight-adjustment method were: 1) the simple ratio generally increased catch estimates and their variance, 2) the complex ratio resulted in the smallest change to catch estimates, and 3) the proposed method resulted in catch estimates that were between those for the two ratio estimators and showed the least change in their variances.

The final presentation and general discussion and questions ended at about 1PM on Wednesday, so the workshop broke for lunch. Upon return the Panel met in closed session from about 2:30PM-4:00PM to discuss the workshop, resolve questions and begin writing reviews. Discussions during the closed session included questions about details of the calculations for the proposed methods, the utility of judging change ratio distributions (numerator vs denominator values) and some of the specific meaning of the Terms of Reference. Questions about when the proposed weight-

adjustment method used input data from the Fishing Effort Survey (FES) or CHTS, the meaning of ‘consistently’ in the Terms of Reference, trimming options, and the significance of extreme ratio values required another open discussion session with the analysts and presenters. At about 4PM the workshop reconvened in open session to allow for these questions and to allow the public (both in attendance and online) to ask questions and comment on the proceedings. The meeting was adjourned at about 5:30PM. Over the course of the evening, panelists were requested to summarize their responses to the Terms of Reference and, if possible, forward rough drafts to the chair. Excerpts from the responses were included in the draft of the Workshop summary report and distributed to panelists prior to meeting on Thursday morning.

The final portion of the workshop was a closed session running from 9AM-12PM on Thursday where panelists continued discussions and began writing individual and summary reviews. An initial draft of the Panel report was distributed to the Panel for review and comment. The report was completed by correspondence among the panelists following the meeting. It should be noted that this report reflects the breadth of opinions of independent reviewers. No efforts were made to achieve consensus but for many topics there were no substantive disagreements among reviewers. As the sections of this report were assembled, in part, from individual submissions from the Panel, there is likely to be some redundancy with the individual reports. The individual reports should be consulted for additional details on some topics raised at the meeting and those that arose after the meeting as reviewers crafted their reports.

2. Review of MRIP APAIS Conversion Model

2.1 Synopsis of Panel Review

The Panel was impressed with the amount of work conducted toward the goal of developing a valid calibration method that extended the effective statistical design back through the entire survey using the raking algorithm and the defined domains. Concerns were raised about the presence of some extreme changes and impact of the new adjusted sample weights on final estimates (assumed to be for minor components, but this needs to be checked) and discussed weight trimming options. Other panelists wanted to point out that the calibrations might not be effective for all species and that assessment analysts needed to be aware of that moving forward. Further exchanges during the writing of this report also indicated that several panelists wanted to indicate that for some purposes, e.g., small area estimation of rare species, MRIP will never be adequate and that the proposed calibration could further exacerbate this issue.

Given these reservations, the Panel concluded that the proposed weight-adjustment raking algorithm method was an improvement over the two ratio-based conversion methods and should be implemented to create a consistent time series of data for developing estimates of catch, effort, and catch size composition. The calibration method had the advantage of preserving much of the public-use ‘micro’ data and was designed to adjust angler-trip level observations and maintain the current MRIP techniques for calculating catch. It also had the advantage of consistency that was lacking for the two ratio conversion methods. Although the Panel suggested modifications to the preliminary analysis (some that were already under consideration), the Panel acknowledged that the proposed method was a reasonable and defensible estimation approach.

It was agreed between the Panel and analysts that further investigation into the use of data trimming techniques and how this is conducted through the sequential sample weight adjustment process (raking) was warranted. Additionally the Panel recommends that analysts investigate the effect of using different time periods as the ‘base’ modern APAIS period (2013 or 2014-2017), determine whether the extreme changes between original and adjusted sample weights are for rare domains (currently assumed unimportant), and consider the potential to use disclaimers to advise users of the data about species whose estimated catch was highly sensitive to the adjusted sample weights, i.e., species not well suited for the MRIP survey.

2.2 *Evaluation of Terms of Reference*

2.2.1 *Term of Reference 1*

Evaluate the suitability of the proposed approach for converting historical estimates of mean angler catch rates obtained using the old MRFSS Access Point Angler Intercept Survey (APAIS) sampling design to estimates that best represent what would have been produced had the new MRIP APAIS sampling design been in place prior to 2013.

- The Panel concurs that this TOR and its subcomponents listed below (1a,1b, 1c, 1d) were met.
- a) *Does the proposed approach adequately account for consistent differences in estimates that would have been observed if the old MRFSS APAIS had been conducted side-by-side with the new MRIP APAIS in 2013-2017?*
- The proposed approach appeared to adequately reweight the old MRFSS APAIS angler trip data to reflect the new 2013-2017 APAIS sampling design.
 - The new raking ratio method attempts to identify those factors that might be expected to be distributed differently under the new design than what was reflected in the previously developed series of estimates. The method assumes that the distribution of angler fishing trips over the domains defined by these factors may be projected back from recent years. Adjustment factors are developed for each domain sequentially. The process is repeated until the required adjustments converge. The method is particularly suitable for the APAIS data since the estimated angler trip distributions under the new design are supported by statistical sampling theory.
 - Although the procedure appears to be robust because it is not species specific, there are concerns with how the new approach will affect estimates for rare species. It is not possible to review all species, state, wave, mode, etc. combinations for changes in catch rate. The sheer volume of data and estimates makes it likely that the full complexity of revisions will not be fully known for several years as the estimates are incorporated into stock assessments. As such, moving forward, analysts should be made aware that large changes in estimated removals are possible within species, waves, or geographic areas. Such changes should also be communicated to the entire user community when they occur. The web portals might include a section to advise on the nature of revisions and their scope.

- While stability in the time series is important, infrequent improvements necessitating calibration is encouraged. There is no perfect solution to this problem and the value of stability should not prevent the introduction of new improved methodology as they are developed in the future. Any proposed future changes should be considered considering potential costs and benefits to all stakeholders. Moreover, future revisions should be subject to rigorous external peer review.
- Given that the proposed approach is conceptually superior to the ratio-based methods it can be concluded that the proposed method was preferred by the Panel. Further, it need to we reiterate that this method will not ‘recover’ earlier data collected for poorly sampled, small domains.

b) *Is the proposed approach a suitable alternative to the calibration models that were originally developed in the 2014 MRIP Calibration Workshop and later evaluated by MRIP?*

- The proposed calibration approach is a significant improvement over the two ratio-based approaches originally developed during the 2014 MRIP Calibration workshop. The proposed approach could not be evaluated against the third method (regression model approach) which was described but never implemented.
- Not only did the simple (M1) and complex (M2) ratio methods require strong assumptions but they also failed to utilize the ancillary information collected under the original sampling design. The proposed methodology appears to include this information in a reasonable way and applied standard statistical approaches to adjust the surveys time intervals. Although such methods may be standard in other survey contexts, this would appear to be the first time such procedures have been applied in fisheries.
- The original adjustments were based on ratios using the relationship of peak time of day data to total day data and adjust prior years’ peak data to the level of full-day data. The approach is reasonable, but results in excluding some of the data from the off-peak periods. The complex ratio used some aspects of the design change and utilizes all the data. Both methods require the assumption that the available data for prior years can represent both peak and nonpeak periods.
- The two ratio estimation techniques were complicated by inconsistencies, e.g., ratio adjusted groups of species were not the summation of individually adjusted species in that group. The ratio methods are too simplistic because they ignore many of the factors in the original survey design.
- The proposed raking weight approach uses all available information and is an improvement over the simple ratio weights. One reason the raking procedure is preferred is that the simple ratios do not account for the differences between coastal and non-coastal anglers.
- As the proposed calibration approach is applied at the survey design level and is not a single applied correction value, it is considered an improvement over the ratio methodologies that were recommended at the previous calibration

workshop. The interim ratio method (Carmichael and Van Voorhees 2012), as developed in a previous review was appropriate in the absence of a more rigorous approach. Differences between the proposed calibration approach and the interim ratio method can only be evaluated with regard to which has the best chance of utilizing the data within a valid statistical survey design.

c) *Is it reasonable to conclude that revised 2004-2012 APAIS estimates based on the application of the proposed approach would be more comparable (statistically?) than the current ones to estimates produced since 2013 under the new APAIS design?*

- The proposed approach attempts to compensate for several aspects of design change and should provide more consistent differences in estimates though empirical results are somewhat mixed in terms of direction and magnitude of change for different species.
- This is a reasonable conclusion based on the goal that the proposed approach would replace the 2004-2012 sample ‘pseudo weights’ with sample weights that reflect the new probability-based experimental design for the 2013-2017 APAIS. The development of the ‘pseudo-weights’ for the MRFSS using the time-slice model was a good interim attempt to align the data with a fully probabilistic sampling design; however, the previous re-weighting is improved by utilizing the observed time-specific catch-rates on trips collected under the new 2013-2017 MRIP design. Close conformity in landings and release trends over time between estimates developed under the two approaches (the pseudo weights and raking algorithm) for several species could also indicate consistency in the direction of change. This could be used to conclude adequate ‘consistency’ however additional species, including rare occurrence species, will need to be investigated singularly during the assessment process.
- While this conclusion appears reasonable, this judgement is deduced from first principles rather than from species specific examples or simulation.

d) *Given the limitations of the available data, is it reasonable to apply the proposed approach to revise APAIS estimates prior to 2004 (back to 1981)?*

- Yes. The proposed methodology is rigorous and reproducible. It utilizes available survey information to adjust sampling weights. This preserves the fundamental aspects of the design and prevents the development of adjustment factors that may be based on a single realization of a catch rate series. Thus, the proposed method is appealing from intuitive and statistical perspectives.
- The alternative would be to suppress all estimates prior to 2004, which the Panel found untenable. The proposed approach requires applying some of the same assumptions prior to 2004 as adjusted estimates for 2004 to 2012 period. While reasonable, it should be recognized that any attempt to calibrate a time series for methodological improvements implemented late in the series will be difficult to implement and difficult to evaluate; though more practical for comparing changes over time for the end user.

- Developing weights back to 1981 differs from developing weights back to 2004 because there are no starting weights from 1981 to 2003. It should be expected that the proposed raking method will be less accurate for 1981-2004 than for 2004-2013. What is missing is a validity study that compares the various estimates to a benchmark true estimate. Without that sort of study, which is not feasible in this case, there is little to inform a determinant of accuracy of these estimates.

2.2.2 Term of Reference 2

Briefly describe the panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.

The Review of Activities section provides an overview of the presented talks and pertinent issues identified during those talks. The following bullets are comments provided after the meeting.

- The background on the MRIP development, importance of time series consistency to stock assessment and management, and design of the modern APAIS was succinctly presented and provided the panelists with a good appreciation for the importance of the proposed calibration.
- The technical description of the ratio calibration methods was important but was not followed by examples of its implementation, which would probably have been a good idea given their relative importance.
- The most significant meeting presentations, in terms of reviewing the adequacy of the proposed conversion methods, were the detailed mathematical description of the proposed method and the results from that method and its generated changes in the weightings for angler-trips within the various raking dimensions, and the effects on effort, landings, and releases. Although the extensive amount of results presented were very useful, there were some additional analyses that could have been presented which could have helped the Panel in its evaluation. For instance, the extreme ratio changes were thought to be derived from insignificant small trips catching rare species, but it was difficult to understand how these were then judged insignificant.
- The Panel appreciated the plans to retain the “micro data” during the transition. This will both facilitate the communication of the process and allow analysts to interrogate the data to understand the changes.
- The Panel wishes to thank the OST staff for their outstanding support during the course of this review by acknowledging their instructive presentations, valuable meeting notes, and responsiveness to the requests of the Panel.

3. Bibliography

Background and Working Papers

The following background documents may be found at:

<https://www.st.nmfs.noaa.gov/confluence/display/APAISCALIB/MRIP+APAIS+Calibration+Review>

Atlantic Coastal Cooperative Statistics Program. 2018. Field Procedures Manual: Access-Point Angler Intercept Survey. 104 pp.

Boreman, J. 2012. Consultant's Report: Summary of the MRFSS/MRIP Calibration Workshop. March 27-29, 2012, Raleigh, NC. Documents and presentations are at: <http://sedarweb.org/otw-01-mrfssmrip-calibration-workshop-2012>.

Breidt, F.J., H.L. Lai, J.D. Opsomer, and D.A. Van Voorhees. 2011. A Report of the MRIP Sampling and Estimation Project: Improved Estimation Methods for the Access Point Angler Intercept Survey Component of the Marine Recreational Fishery Statistics Survey. 83pp. NMFS/S&T/MRIP, Silver Spring, MD. Including comments from reviewers and responses to reviewer comments.

Breidt, F.J., et al. 2012. A Pilot Study of a New Sampling Design for the Access Point Angler Intercept Survey. Report to the Marine Recreational Information Program's (MRIP) Design and Analysis Workgroup. Including comments from reviewers and responses to reviewer comments.

Carmichael, J and D. Van Voorhees (eds.). 2012. MRIP Calibration Workshop II. Final report to of a workshop hosted by NOAA Fisheries, Office of Science and Technology, MRIP and SEDAR: Southeast Data, Assessment and Review. September 8 –10, 2014, North Charleston SC.

Deming, W.E. and F.F. Stephan. 1940. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, 11(4):427-444.

Foster, J., F.J. Breidt, and J.D. Opsomer. 2018. APAIS data calibration methodology report. Manuscript dated March 11, 2018.

Hernandez-Stumpfhauser, D., F.J. Breidt, and J.D. Opsomer. 2016. Hierarchical Bayesian small area estimation for circular data. *Canadian Journal of Statistics* 44(4):416–430.

MRIP staff. 2014. A Descriptive Analysis of the Access Point Angler Intercept Survey. 2013 Design Change. Internal Report. 48 pp., August 20, 2014, Silver Spring, MD.

National Academies of Sciences, Engineering, and Medicine. 2017. Review of the Marine Recreational Information Program. Washington, DC: The National Academies Press. doi: <https://doi.org/10.17226/24640>.

National Research Council. 2006. *Review of Recreational Fisheries Survey Methods*. Washington, DC: The National Academies Press.

Papacostas, K.J. and J. Foster. 2017. National Marine Fisheries Service's Marine Recreational Information Program Survey Design and Statistical Methods for Estimation of Recreational Fisheries Catch and Effort, Version 1.0. Prepared for: U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service.

Presentations and Webinar Recordings

The presentations and webinar broadcast recordings of the open sessions of the workshop can be found at:

<https://www.st.nmfs.noaa.gov/confluence/display/APAISCALIB/MRIP+APAIS+Calibration+Review>

Additional materials

Also available on the web page for the links above are: the Terms of Reference (revised 1-25-2018); the Transition Plan for the Effort Survey; a spreadsheet of effort ratio adjustments that were requested at the workshop; detailed workshop notes taken by rapporteurs; and the attendance list.

4. Appendices

Appendix 1. Terms of Reference for the MRIP APAIS Peer Review Workshop

The Review Panel shall assess whether or not the MRIP Working Group has reasonably and satisfactorily completed the following actions.

1. Evaluate the suitability of the proposed approach for converting historical estimates of mean angler catch rates obtained using the old MRFSS Access Point Angler Intercept Survey (APAIS) sampling design to estimates that best represent what would have been produced had the new MRIP APAIS sampling design been in place prior to 2013.
 1. Does the proposed approach adequately account for consistent differences in estimates that would have been observed if the old MRFSS APAIS had been conducted side-by-side with the new MRIP APAIS in 2013-2017?
 2. Is the proposed approach a suitable alternative to the calibration models that were originally developed in the 2014 MRIP Calibration Workshop and later evaluated by MRIP?
 3. Is it reasonable to conclude that revised 2004-2012 APAIS estimates based on the application of the proposed approach would be more comparable than the current ones to estimates produced since 2013 under the new APAIS design?
 4. Given the limitations of the available data, is it reasonable to apply the proposed approach to revise APAIS estimates prior to 2004 (back to 1981)?
2. Briefly describe the panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.

Appendix 2. Draft Review Meeting Agenda

MRIP APAIS Calibration Peer Review Workshop Silver Spring, MD March 20-22, 2018

Agenda

Tuesday, March 20:

OPEN SESSION

- 9:00am – Welcome, Introductions, and Overview of Workshop
Dave Van Voorhees, NOAA Fisheries Office of Science and Technology
Mike Murphy, Chair of Peer Review Panel
- 9:20am – Presentation: MRIP Transition Planning and the Access Point Angler Intercept Survey
Dave Van Voorhees, NOAA Fisheries Office of Science and Technology
- 9:40am – Presentation: Importance of Calibrated Catch for Fisheries Management
Jason Didden, Mid-Atlantic Fishery Management Council
- 10:00am – Presentation: Importance of Calibrated Catch for Fishery Stock Assessments
Katie Drew, Atlantic States Marine Fisheries Commission
- 10:20am – BREAK
- 10:30am – Presentation: Weighted Estimation for the APAIS (and Calibration Workshop I)
Dave Van Voorhees, NOAA Fisheries Office of Science and Technology
- 11:05am – Presentation: New Design of the APAIS
Tom Sminkey, NOAA Fisheries Office of Science and Technology
- 11:40am – Presentation: Calibration Workshop II
John Carmichael, South Atlantic Fishery Management Council & SEDAR
- 12:00pm – LUNCH BREAK
- 1:10pm – Presentation: Considered Ratio Calibration Methods
Ryan Kitts-Jensen, NOAA Fisheries Office of Science and Technology
- 1:50pm – Presentation: APAIS Calibration Methodology
Jean Opsomer, Westat and Colorado State University
- 2:40pm – Presentation: APAIS Calibration Results
John Foster, NOAA Fisheries Office of Science and Technology
- 3:30pm – BREAK
- 3:40pm – Follow-Up Questions for Presenters
- 4:20pm – Public Comment
- 4:50pm – Summary of Day 1
Mike Murphy, Chair of Peer Review Panel
- 5:00pm – ADJOURN

CLOSED SESSION

- 5:10pm – Review Panel Coordination and Writing
- 6:00pm – ADJOURN

Wednesday, March 21:

OPEN SESSION

- 9:00am – Overview of Day 1 and Preview of Day 2
Mike Murphy, Chair of Peer Review Panel

9:10am – Follow-Up Questions for Presenters
10:15am – BREAK
10:30am – Follow-Up Questions for Presenters
12:00pm – LUNCH BREAK

CLOSED SESSION

1:00pm – Review Panel Coordination and Writing

OPEN SESSION

2:30pm – Initial Summary Findings of Review Panel

3:30pm – BREAK

3:45pm – Public Comment

4:15pm – ADJOURN

CLOSED SESSION

4:30pm – Review Panel Coordination and Writing

6:00pm - ADJOURN

Thursday, March 22

CLOSED SESSION

9:00am – Review Panel Coordination and Writing

12:30pm – ADJOURN

Appendix 3. Individual Independent Peer Review Report Requirements

Appendix 4. CIE Statement of Work

Statement of Work

National Oceanic and Atmospheric Administration (NOAA)

National Marine Fisheries Service (NMFS)

Center for Independent Experts (CIE) Program

External Independent Peer Review

Calibration Model Accounting for a Recreational Fisheries Survey Design Change

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards.

(http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).

Further information on the CIE program may be obtained from www.ciereviews.org.

Scope

The Office of Science and Technology requests an independent peer review of a calibration model proposed for use in revising statistics produced by a survey of marine recreational fishing catch rates on the Atlantic coast and in the Gulf of Mexico. This calibration model is considered by the Marine Recreational Information Program (MRIP) to be very important to adjust historical time series of recreational catch estimates in order to account for biases in past sampling and estimation methods that have become apparent with the development of a new, more statistically sound method. The calibration model is intended to account for past biases in catch rate estimates for the shore, private/rental boat, and charter boat fishing modes that have resulted from the continued use of a legacy sampling design for the Access Point Angler Intercept Survey (APAIS). A more statistically sound sampling design for the APAIS was implemented in March of 2013

Calibration Model for the APAIS Design Change

In 2014, a Calibration Workshop was held to evaluate the potential consistent effects of implementing a new sampling design for the APAIS on the Atlantic and Gulf coasts in 2013. Workshop participants included three expert statistical consultants and representatives from NOAA Fisheries, the regional fishery management councils, the interstate marine fisheries commissions, and several state agencies. The participants determined that analyses conducted by the NOAA Fisheries Office of Science and Technology showed there was sufficient evidence that the more complete temporal coverage of the new design resulted in consistent changes in APAIS angler catch rate statistics for at least some species. They developed three different calibration models to evaluate for possible use in correcting the pre-2013 APAIS statistics. The statistical consultants concluded the simplest of the three proposed models was appropriate for use in the short term until a more complete evaluation of all three calibration models could be completed using three years of new APAIS data (2013-2015). The plan was to complete that evaluation by the end of 2016, so that one method could be selected as the best for use in 2017 to revise APAIS estimates prior to 2013.

Requirements

NMFS requires three (3) reviewers to conduct an impartial and independent peer review in accordance with the SoW, OMB Guidelines, and the Terms of Reference (ToRs) below. The CIE reviewers shall have working knowledge and recent experience in the design of sampling surveys and the evaluation of non-sampling errors (i.e., undercoverage, nonresponse, and response errors) associated with changes to survey designs over time. In addition, they should have experience with complex, multi-stage sampling designs, time series analyses, regression estimators, and small domain estimation methods. Some recent knowledge and experience in current surveys of marine recreational fishing is desirable but not required.

NMFS will provide a Chair who has experience with U.S. fisheries stock assessments and their application to fisheries management. The Chair would ensure that reviewers understand the importance of maintaining a comparable time series of marine recreational fisheries catch statistics for use in stock assessments and their application to fisheries management. The Chair will not be selected by the contractor and will be responsible for facilitating the meeting, developing and finalizing a summary report and working with the CIE reviewers to make sure that the ToRs are addressed in their independent reviews.

Tasks for Reviewers

Pre-review Background Documents

The following background materials and reports prior to the review meeting include:

APAIS Design Change Calibration Workshop Report:

http://www.st.nmfs.noaa.gov/Assets/recreational/pdf/MRIPCalibrationWorkshopII_FinalReport.pdf

NC APAIS Pilot Study Report: *A Pilot Study of a New Sampling Design for the Access Point Angler Intercept Survey.*

https://www.st.nmfs.noaa.gov/pims/main/public?method=DOWNLOAD_FR_PDF&record_id=772

Report on APAIS Calibration Model:

This report will be provided by the contractor (via electronic mail or make available at an FTP site) to the CIE reviewers.

Panel Review Meeting

Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The meeting will consist of presentations by NOAA and other scientists to facilitate the review, to

provide any additional information required by the reviewers, and to answer any questions from reviewers.

Contract Deliverables - Independent CIE Peer Review Reports

The CIE reviewers shall complete an independent peer review report in accordance with the requirements specified in this SoW and OMB guidelines. Each CIE reviewer shall complete the independent peer review according to required format and content as described in **Annex 1**. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in **Annex 2**.

Other Tasks – Contribution to Summary Report

The CIE reviewers may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. The CIE reviewers are not required to reach a consensus, and should provide a brief summary of each reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Foreign National Security Clearance

When reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for reviewers who are non-US citizens. For this reason, the reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: <http://deemedexports.noaa.gov/> and http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html. The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place of Performance

The place of performance shall be at the contractor's facilities, and at the NOAA Fisheries Service Headquarters in Silver Spring, Maryland.

Period of Performance

The period of performance shall be from the time of award through April 31, 2018. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Within two weeks of award	Contractor selects and confirms reviewers
Approximately 2 weeks later	Contractor provides the pre-review documents to the reviewers
March 2018	each reviewer participates and conducts an independent peer review during the panel review meeting
Within two weeks of panel review meeting	Contractor receives draft reports
Within two weeks of receiving draft reports	Contractor submits final reports to the Government

Applicable Performance Standards

The acceptance of the contract deliverables shall be based on three performance standards:

(1) The reports shall be completed in accordance with the required formatting and content (2) The reports shall address each ToR as specified (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (<http://www.gsa.gov/portal/content/104790>). International travel is authorized for this contract. Travel is not to exceed \$12,000.

Restricted or Limited Use of Data

The contractors may be required to sign and adhere to a non-disclosure agreement.

NOAA Fisheries Project Contact:

Dave Van Voorhees

NOAA Fisheries

1315 East West Highway

Silver Spring, MD 20910

dave.van.voorhees@noaa.gov

Annex I: Format and Contents of CIE Independent Peer Review Report

1. The report must be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether or not the science reviewed is the best scientific information available.

2. The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each ToR, in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the ToRs.
 - a. Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.

 - b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, but especially where there were divergent views.

 - c. Reviewers should elaborate on any points raised in the summary report that they believe might require further clarification.

 - d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

 - e. The report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The report shall represent the peer review of each ToR, and shall not simply repeat the contents of the summary report.

3. The report shall include the following appendices:

Appendix 1: Bibliography of materials provided for review

Appendix 2: A copy of this Statement of Work

Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Appendix 5: Calibration Model Review Workshop Participants

MRIP Calibration Model Peer Review Workshop
Sheraton Silver Spring Hotel
Silver Spring, MD
March 20-22, 2018
ATTENDANCE LIST

#	NAME	AFFILIATION
1	Carolyn Belcher	GA Dept of Nat. Resources
2	Richard Cody	NOAA support ECS
3	John Carmichael	SAFMC
4	James Chromy	retired (RTI)
5	Mary Christman	MCC Stat Consulting
6	Matt Cieri	ME Dept of Mar. Resources
7	Jason Didden	MAFMC
8	Laura Diederick	NOAA Fisheries support
9	Katie Drew	ASMFC
10	John Foster	NOAA Fisheries
11	Clifford Hutt	NOAA SF-1 (HMS)
12	Ryan Kitts-Jensen	NOAA Fisheries - ST1 support
13	Yong-Woo Lee	NMFS-ST1
14	Vivian Matter	NMFS - SEFSC
15	Michael Murphy	retired FL-FWCC
16	Jean Opsomer	CSU
17	Karen Pianka	NOAA Fisheries - ST1 support
18	Paul Rago	MAFMC SSC
19	Thomas Sminkey	NOAA Fisheries - ST1
20	David VanVoorhees	NOAA Fisheries
21	Chris Wright	NMFS - SF
22	John Whitehead	Appalachian State University

Center for Independent Experts (CIE) Independent
Peer Review Report on the Marine Recreational
Information Program (MRIP) Access Point Angler
Interview Survey (APAIS) Calibration Peer Review

Calibration Peer Review Workshop

March 20-22, 2018

Sheraton Hotel

Silver Spring, MD

May 2018

Mary C. Christman

MCC Statistical Consulting, Gainesville, FL 32605

marycchristman@gmail.com

Contents

Executive Summary..... 3

Background 6

Reviewers..... 9

Summary of the Proceedings of the Workshop..... 9

Summary of Findings for Each Terms of Reference (ToRs)..... 14

 Terms of Reference..... 14

Appendix 1 18

Appendix 2 20

 Statement of Work 20

Appendix 3 25

 Attendance List 25

 Proposed Agenda for the APAIS Calibration Review Workshop..... 26

Executive Summary

A workshop was held in March 2018 in Silver Spring, Maryland, to review a proposed method for calibrating estimates of population parameters such as CPUE for selected species or total effort (in angler-trips) from the 1981 to 2012 APAIS data collected under the MRFSS sampling design, to that which would have been obtained had the MRIP sampling design been in place. Such calibrations are critical to effective management of recreational fisheries since decisions rely on accurate and precise time series of the results of recreational activities. Changes to sampling designs that alter the sampling frame or population, such as sampling over all 24hrs in day versus only a subset of the 24hrs, can have a major effect on the estimates derived from the collected data and so calibration to account for such change is critical.

The current sampling design for APAIS is MRIP, a statistically valid sampling design that provided unbiased estimates of various parameters describing the recreational fishery along the east coast of the US and in the Gulf of Mexico. It was initiated in 2013 after a pilot study in NC in 2010, and after in-depth review by the National Academy of Sciences. It replaced the original survey design, known as the Marine Recreational Fishery Statistics Survey (MRFSS), which was an opportunistic sampling design that did not follow statistical sampling principles sufficiently closely to allow unbiased estimation of the desired parameters. In 2006, the National Research Council of the National Academies reviewed the methodology and concluded that the sampling strategies, collection methods and estimation approaches of MRFSS did not provide adequate information for management and policy decisions. As a result, NOAA modified the design of MRFSS to be more statistically sound and the changes were implemented in years 2004 to 2012 although the temporal coverage (time of day) was still limited.

At issue is how to use the entire time series back to 1981 to obtain indices that can be used in stock assessments given the changes that have occurred in the sampling design, temporal coverage, and estimation methods over time. Statisticians from Colorado State University with NOAA scientists developed an approach to calibrate the past data collected prior to the second wave of 2013 with the data collected under MRIP since then. This workshop was to review the recommended approach and determine its adequacy and statistical soundness for use in index development in future. The results are described herein.

There were three issues that the statisticians needed to address when determining if and how to combine the data collected under the different sampling approaches. First was the lack of valid sampling weights that could even be approximated for the early years (1981 – 2004). Second, the sampling frame changed with MRIP implementation, where the sampling frame was increased to include sampling around the clock, i.e. sampling over all 24 hrs. Third, there was no formal calibration experiment where both sampling designs were performed simultaneously for some period of time, such as was done for the conversion from CHTS to FES.

As a result, a classical calibration approach which compares two different sampling efforts on the same population could not be used. Instead, the statisticians recommended a method that has been used extensively in social science surveys: raking (Deming & Stephan 1940). Raking (or raking ratio estimation) is a method for adjusting the sampling weights of the sample data based on known population characteristics. By adjusting these weights, the survey sample is essentially forced to resemble (i.e. be more representative of) the population, thereby making unbiased inference to the entire population possible. This approach requires though that 1) there be initial sampling weights that can be adjusted, and 2) the important population characteristics used in the adjustments be correctly identified and quantified.

The approach that was recommended was innovative and addressed many of the issues associated with the lack of statistical rigor for the MRFSS data collection effort. The method adjusts the initial weights for each of several periods within the time series using information from the later periods. The adjusted weights are then used to calibrate data from the years prior to the implementation of MRIP. This is accomplished by using the adjusted sampling weights in design-based estimation of a multi-stage cluster sampling design (e.g. the Horvitz-Thompson estimator of mean CPUE).

This approach was applied to the MRFSS and MRIP data collected between 1981 and 2016 to obtain estimates of annual mean CPUE for specific combinations of state, wave, mode and possibly additional strata such as were used in the raking algorithm. In addition, during the workshop several diagnostics and fit metrics were reviewed to determine the adequacy of the approach.

Earlier efforts have been made to calibrate or adjust data collected in years prior to implementation of MRIP. These were ratio adjustment methods applied to the estimates of desired parameters, for example, adjusting the CPUE for a selected species for a given year, rather than adjustments to sampling weights that would not rely on the estimates. As a result, the proposed calibration method is an advance over the earlier approaches.

The proposed approach is a complex and sophisticated method for adjusting initial sampling weights so that estimates of desired parameters are believed to be relatively unbiased over the entire time series, and similar to that expected had MRIP been in place in earlier years. An advantage of the new approach is that the adjustments to the initial weights are species independent and allow for utilization of all the data collected in MRFSS, not just those during specific time periods. As a result, every record in the APAIS dataset will have a sampling weight that can be utilized in estimation of parameters over desired domains or within strata.

A disadvantage of this approach is that it is not a true calibration since the two sampling designs could not be simultaneously implemented, and so unbiasedness of the weighted estimators is difficult to assess. The approach also generates estimates of the desired parameters with large estimated variances, but this is to be expected given the multi-stage

cluster sampling design. Even with these caveats, the proposed method is a more thoughtful and appropriate calibration approach for the APAIS dataset.

Overall, this approach is recommended as the method for calibrating past data collected under MRFSS with MRIP. There are a few recommendations related to assessing sensitivity of the approach to some choices associated with the proposed method, namely the choice of initial weights for the years prior to 2004, and the effect of the choices of domains used in the raking algorithm.

Background

NOAA fisheries has been conducting access point intercept sampling (APAIS) of recreational fishermen/women since 1981 to estimate recreational effort and removals (both landed and discarded) of important fish species in the private boat, charter boat and shore fisheries. The original survey design, known as the Marine Recreational Fishery Statistics Survey (MRFSS), was an opportunistic sampling design that did not follow statistical sampling principles sufficiently closely. In 2006, the National Research Council of the National Academies reviewed the methodology and concluded that the sampling strategies, collection methods and estimation approaches of MRFSS did not provide adequate information for management and policy decisions. As a result, NOAA modified the design of MRFSS to be more statistically sound and the changes were implemented in years 2004 to 2012 although the temporal coverage (time of day) was still limited.

More recently, NOAA developed a new statistically valid sampling strategy with more complete time of day coverage known as the Marine Recreational Information Program (MRIP). MRIP was implemented in 2013 and later evaluated by the National Academy of Sciences (NAS) in 2017. The design and estimation methods were found to be statistically sound.

Both MRFSS and MRIP are based on a multi-stage cluster sampling strategy which requires the use of sampling weights to obtain an unbiased estimate of the desired parameters (e.g. mean CPUE of an angler-trip). Intuitively, the weights are the number of sampling units in the population that the observations (e.g. an individual angler-trip) represents, where those with a small selection probability represent more of the population and those with larger selection probabilities represent fewer units. In the years between 1981 and 2004, MRFSS was implemented such that the required sampling weights were not estimable. As a result, the estimates obtained from those data were calculated assuming observations were collected via simple random sampling (SRS).

In the later periods of the time frame, the changes made to MRFSS allowed derivation of approximate weights that could be used in estimation for data collected between 2004 and 2012. The temporal coverage, namely time of day, though was still limited.

The MRIP implementation has full temporal coverage and is statistically sound sampling. Hence, sampling weights can be calculated, and unbiased estimates derived.

At issue is how to use the entire time series back to 1981 to obtain indices that can be used in stock assessments given the changes that have occurred in the sampling design, temporal coverage, and estimation methods over time. Statisticians from Colorado State University with NOAA scientists developed an approach to calibrate the past data collected prior to the second wave of 2013 with the data collected under MRIP since then. The workshop convened in March 2018 was to review the recommended approach and determine its adequacy and statistical soundness for use in index development in future. The results are described herein.

There were three issues that the statisticians needed to address when determining if and how to combine the data collected under the different sampling approaches. First was the lack of valid sampling weights that could even be approximated for the early years (1981 – 2004). Second, the sampling frame changed with MRIP implementation, where the sampling frame was increased to include sampling around the clock, i.e. sampling over all 24 hrs. Third, there was no formal calibration experiment where both sampling designs were performed simultaneously for some period of time, such as was done for the conversion from CHTS to FES.

MRFSS was stopped in early 2013 and MRIP implemented subsequently. As a result, a classical calibration approach which compares two different sampling efforts on the same population could not be used. Instead, the statisticians recommended a method that has been used extensively in social science surveys: raking (Deming & Stephan 1940). Raking (or raking ratio estimation) is a method for adjusting the sampling weights of the sample data based on known population characteristics. By adjusting these weights, the survey sample is essentially forced to resemble (i.e. be more representative of) the population, thereby making unbiased inference to the entire population possible. This approach requires though that 1) there be initial sampling weights that can be adjusted, and 2) the important population characteristics used in the adjustments be correctly identified and quantified.

The raking method is also known as the raking ratio estimation method due to the fact that the initial sample weight for any given year is adjusted by multiplication with a ratio of the true total number of angler-trips within a domain (a defined subset of the population such as state-wave-mode combination coupled with “resident of a coastal county”) to the estimated number based on the sample collected that year (slide 12 in the workshop presentation “8OpsomerCalibration APAIS.pdf”). Since the true value of the total effort within a domain for a given year cannot be known, the method was modified so that, for any given year for which an adjustment is needed, the numerator, true total number of angler-trips within a domain, was replaced by an estimate derived from data collected during a specified range of subsequent years. The denominator, the estimated number for a given year, was replaced by an estimate based on several years. Specifically, the initial estimates of the components of the raking ratio for a given year were proposed to be calculated as follows:

- 1) for the years 2004 to 2012, the raking algorithm was based on a ratio with a numerator based on all the MRIP data collected in 2013 – 2016 and a denominator based on averaging over 2004 to 2012;
- 2) for the years 1993 to 2003, the raking algorithm was based on a ratio using a numerator based on MRFSS data collected in 2004 – 2012 and a denominator based on the years 1993 to 2003;
- 3) for the years 1981 to 1992, the raking algorithm was based on a ratio using a numerator based on MRFSS data collected in 1993 to 2003 and a denominator based on the years 1981 to 1992.

The set of years over which the numerators of the ratio for each time segment could be further modified, namely shortened to the most recent three years subsequent to a time segment, if there was evidence that there was a linear trend in the indices over time. For specific details of how that was determined see Foster, et al. (2018) or slide 19 of the workshop presentation “8OpsomerCalibration APAIS.pdf”.

For initial sampling weights, the statisticians and scientists proposed using:

- 1) the approximate weights previously developed and used for estimation of MRFSS data for the time period 2004 to 2013;
- 2) for the time period 1993 – 2003, the number of site-days sampled within each state-wave-mode-year combination standardized to the maximum value in the years 1993 to 2004 within the same state-wave-mode combination; and
- 3) for the time period 1981 – 1992, the number of site-days sampled within each state-wave-mode-year combination standardized to the maximum value in the years 1981 to 1992 within the same state-wave-mode combination.

The reasoning for the proposed initial weights in (2) and (3) is that these should capture the sampling variability among combinations of state-wave-mode and over years.

Once some initial weights were obtained for a particular state-wave-mode combination and the method for adjusting these weights using data from subsequent years identified, the population characteristics for which the sample weight adjustment should be performed needed to be identified. For these, the statisticians and scientists recommended the following:

- 1) for the time period 2004 to 2013, use the domains
 - a. AF (state, wave, mode, area fished),
 - b. HS (state, wave, mode, coastal/non-coastal household status),
 - c. FH (state, wave, mode, for-hire boat frame status), and
 - d. RE (state, wave, mode, sub-state region).
- 2) For the time periods 1981 to 1992 and 1993 to 2003, use the above domains plus
 - a. KOD (state, wave, mode and kind-of-day),
 - b. MG (state, wave, mode and month groups), and
 - c. AC (state, wave, mode and site activity classes).

Once these population characteristics/domains are identified, the raking algorithm is iterative. The first step is to adjust the initial weight using the estimated raking ratio for the first domain in the list. Once a modified weight is calculated within that domain, it is then used as the initial weight for the adjustment based on the second domain. The results of the adjustment for the second domain is now used as the initial weight to be multiplied by the raking ratio estimated for the third domain. This continues until all domains that identify the important population characteristics have been used to adjust the sampling weight. The procedure then repeats

starting with the first listed domain again, until there are no substantive changes to the sampling weight, i.e. once convergence of the sampling weight is reached.

These adjusted weights are now used to calibrate data from the years prior to the implementation of MRIP. This is accomplished by using the adjusted sampling weights in design-based estimation of a multi-stage cluster sampling design (e.g. the Horvitz-Thompson estimator of mean CPUE).

This approach was applied to the MRFSS and MRIP data collected between 1981 and 2016 to obtain estimates of annual mean CPUE for specific combinations of state, wave, mode and possibly additional strata such as were used in the raking algorithm. In addition, several diagnostics and fit metrics were reviewed to determine the adequacy of the approach.

A review of the approach and findings is given in the section on the Terms of Reference (ToRs).

Reviewers

Panel Members and their roles in () were:

Michael D. Murphy (Chair),
John Whitehead (CIE Reviewer),
Mary Christman (CIE Reviewer),
James Chromy (CIE Reviewer),
Carolyn Belcher (NOAA Reviewer representing SAFMC),
Matthew Cieri (NOAA Reviewer representing ASMFC), and
Paul Rago (NOAA Reviewer representing MAFMC SSC).

See Appendix 3 for their affiliations.

Summary of the Proceedings of the Workshop

A list of the attendees and the agenda for the workshop are provided in Appendix 3. During the three-day workshop (20 – 22 March 2018), Panel members were given on-line access to a broad range of background documents including the documents provided prior to the meeting (Appendix 1) and copies of the presentations made during the meetings. Before the in-person workshop, two conference calls were arranged for NOAA staff to explain the peer review process, the terms of reference, and the availability of documents on the Wiki site and to review any questions that the panelists had before the workshop. The in-person workshop was broadcast via webinar during the open sessions.

The workshop began at 9AM on March 20, 2018 with an introductory presentation by Dr. Van Voorhees. He introduced members of the transition team and gave an overview of the approach to development of the MRIP experimental design change, general APAIS survey design, effort survey changes, and proposed methods for converting early catch estimates to

the modern survey's 'currency'. Subsequent presentations by Jason Didden and Katie Drew provided overviews of the importance of the calibration effort to fisheries management and stock assessments. Jason Didden emphasized the need for consistency in the time series while Katie Drew noted that estimates of catch are the primary data for estimating absolute removals by the recreational fishery. She also noted that the choice of weights in the calibration can have significant effects on the estimated inputs to stock assessments including catch-at-age, selectivity, and mortality, and, so, affects the development of quotas and ACLs from these data. After Drew's presentation, a review of the Terms of Reference of the workshop occurred. Van Voorhees explained the meaning of the term "suitable" in Term 1(a) and the panel understood 'suitability' to mean how likely the converted older survey results would reflect what would have been the estimates had the new APAIS design had been in place at those times.

Following the break, Dave Van Voorhees gave a presentation on the development of pseudo-weights that approximated the inclusion probabilities for the MRFSS design for the time period 2004-2012 when detailed sampling effort information and changes to the MRFSS implementation was available (Breidt et al. 2011). The weights were based on a multi-stage clustering design of the selected sampling locations and removed the small amount of data associated with the alternate site selection option employed by samplers during this period. Determining the inclusion probabilities also required development of a Bayesian time-slice model (Hernandez-Stumpfhauer et al., 2016) that borrowed from fishing trip time-of-day data collected in the CHTS program. One difficulty of this approach is that no actual data were collected under MRFSS during certain times of the day, and so assumptions about catch rates were unavoidable and possibly is a source of bias in the new estimate procedure.

A peer review of this new weighted estimation technique was conducted; as a result, the procedure was accepted and certified as best available science (see slide 15 of "4Weighted Estimation Methods for the MRFSS APAIS - 03-19-2018.pptx"). The Panel learned that sampling prior to 2004 was not statistically valid and often based on quota sampling, hence it could not be applied to prior years.

Simultaneous with the development of pseudo-weights for the MRFSS 2004–2012 time period, a new design (MRIP) was being developed to address on-going concerns with the MRFSS implementation. A pilot study was conducted in 2010 in North Carolina to test the feasibility of implementing MRIP. The study "focused on developing a better understanding of how the changes to the new design would potentially affect sampling efficiency, statistical accuracy, and statistical precision going forward". A workshop was convened (Breidt et al., 2012) which provided specific recommendations for implementing MRIP coast-wide along the Atlantic and Gulf states.

A presentation of the new MRIP design, as implemented in 2013, was given by Tom Sminkey. Several questions and a general discussion followed in order to fully understand the distinctions between the various designs and their implementations.

Once MRIP had been implemented, it was necessary to develop a calibration effort to upweight data from the MRFSS years (1981 – 2012) in order to have a consistent time series for the entire dataset. As a result, three estimation approaches were considered before the most recent approach under review in this workshop. These were 1) a simple peak-time catch rate ratio, 2) a time-specific catch rate ratio method, and 3) a regression model time-of-day classification model. The last method was never developed and was not considered further. Ryan Kitts-Jensen gave a presentation describing the two ratio methods and provided examples of their use. One complication was that during MRFSS, the definition of peak time-period changed. This proposed ratio method also excluded data collected outside this time window. It also required post-stratification (aggregating across design strata) for obtaining sufficient data for rarer species. In addition, the ratios would need to be developed for each species of interest.

John Carmichael listed the findings from a calibration workshop (Carmichael and Van Voorhees 2012) that reviewed these three approaches. The tasks for the workshop were to look at the changes made from the 1981-2012 MRFSS and the fully-implemented 2013 MRIP APAIS, and evaluate these methods for converting the older estimates to reflect the new design. The recommendation from the second calibration workshop was to use the direct peak-time catch rate ratio method as it was simple to explain and it “preserve[s] the ability to calculate estimates consistent with “old” survey methods until calibration and adjustment methods are developed, peer reviewed and approved to address changes in estimates due to “new” survey methods” (pg. 9 of Carmichael & Van Voorhees, 2012).

Jean Opsomer then introduced the proposed calibration (or more accurately adjustment) method. A difficulty with the current ‘calibration’ is the lack of overlap between the old and new survey except for the small pilot study performed in North Carolina in 2010. For this reason, the adjustment re-weights observations at the angler-trip level to reflect the survey change and to use the expected sample weighting had those earlier data been collected under the new fully probabilistic design. See the background section of this report for more details on the proposed approach. To quickly recap, the method is first applied to the pseudo-weights of the MRFSS period of 2004-2012 using information from MRIP 2013-2016. Then, the newly derived 2004-2012 weights are used to calibrate the 1993-2003 weights. Finally, the newly derived 1993-2003 information is used as the basis for calibrating the 1981-1992 angler-trips. An important feature described later was the introduction of additional ‘raking’ domains as the process went back in time. These new domains were considered the most important in influencing how the choices for sample site selections were made during the earlier periods. Also, a sequence of checks for linear changes in sample effort within each time block was used to determine the lengths of the time-period that served as the basis for the calibration. A discussion of the design occurred, and members of the panel questioned the choice of when the time-period boundaries occurred. A suggestion was made that when the 2017 data become available, the analysts might consider using only the 2014-2017 period as the initial calibration basis. This presupposes a sampler-adjustment period in 2013 before implementation of the MRIP design became routine. The analysts reminded the Panel that the analyses that they

presented during the workshop were still preliminary and were subject to some revisions. Other Panel members questioned the assumptions that any trend would be linear when testing for trend. This was defended for simplicity, with no real way to evaluate non-linear effects. An additional point is that it is only applied to a sequence of 3 consecutive years, for which linearity is a reasonable assumption.

A discussion of whether a finer domain structure for the raking algorithm (e.g., area fished x sub-state region in Florida) was feasible, but it was considered not appropriate due to small domain estimation issues. Several advantages to this method were pointed out by Opsomer and others: the maintenance of the public-use 'micro data' with just a weighting change, re-weighting was based on the sampling unit of angler-trip and is not species dependent like the ratio adjustment methods, and, because it is not species dependent, the same adjusted weights can be used for any estimates desired from the survey.

The final presentation of the workshop, given by John Foster, began Tuesday afternoon and continued through mid-day Wednesday. This presentation provided some results of the proposed re-weightings and the distribution of preliminary changes seen across many different strata. This presentation was very detailed, partially in response to requests made by Panel members before the workshop after the initial review of the available documents. Though the effort was staggering, the Panel struggled with using this detail to review the efficacy of the proposed conversion method. The Panel was most concerned about the presence of extreme adjustments to weights and their impact on accumulated results, e.g., annual landings. However, it was unclear at the time what the cause of some of the large changes was and how significant it would be to the final survey products. For instance, was the absolute adjustment relatively small, but made to an extremely small initial weight (very small number in denominator) or is the adjustment applied to angler-trips with little to no catch or to catch of species that are not adequately sampled within the survey, e.g., cast-net caught herrings. For these reasons, the Panel turned to questions and a discussion about techniques that could be used to trim outliers and to whether this would or should be done.

The summary from the final presentation noted the comparisons between the results of conversions using the simple and complex ratio methods and the proposed weight-adjustment method were: 1) the simple ratio generally increased catch estimates and their variance, 2) the complex ratio resulted in the smallest change to catch estimates, and 3) the proposed method resulted in catch estimates that were between those for the two ratio estimators and that showed the least change in their variances.

The final presentation and general discussion and questions ended at about 1PM on Wednesday. Subsequently, the panel met in closed session for most of the remainder of the time Wednesday and throughout Thursday morning. The panel discussed the workshop presentations, resolved questions and began writing reviews. Discussions during the closed session included questions about details of the calculations for the proposed methods and the utility of judging change ratio distributions as provided by Foster. A short open session was

provided at which time questions about whether the proposed weight-adjustment method would use data from the FES or CHTS moving forward, the meaning of 'consistently' in the Terms of Reference, proposed trimming options, and the significance of the extreme values observed in the Forster presentation. During the closed session on Thursday, it was determined that the simulations and calculations presented by Foster included all species regardless of whether the MRIP program could provide reasonable data for use in estimation of mean CPUE for all species.

Summary of Findings for Each Terms of Reference (ToRs)

Terms of Reference

1. Evaluate the suitability of the proposed model for converting historical estimates of mean angler catch rates obtained using the old MRFSS APAIS sampling design to estimates that best represent what would have been produced had the new MRIP APAIS sampling design been in place prior to 2013.

b) Is the proposed approach a suitable alternative to the calibration models that were originally developed in the 2014 MRIP Calibration Workshop and later evaluated by MRIP?

As was described in the summary of the workshop proceedings, the two ratio methods (Method 1 using only peak time data and Method 2 using more of the daily time periods) have both flaws and advantages. These are described in detail in Appendix I of the report from the second calibration workshop (Carmichael & Van Voorhees, 2012). Method 1 is simple to apply but requires identification of peak time period for various domains and species, excludes possibly important information for other time periods, and assumes that the peak time period information is representative of the non-peak period times.

Method 2 is a more complex version of Method 1 that utilizes the distribution of catch within several 3-hour sampling windows during the day. Because of the use of finer time periods within a day, this method requires larger post-stratification in order obtain non-zero estimates for the less common species.

In both methods, the “tweaking” required to develop domains that are species-specific implies that the estimates of CPUE based on these ratios would not be additive across species. Further, when reviewing these methods, the panel found it difficult to see how the ratios could be extended to adjusting other important parameters such as catch at size or weight data.

The proposed model that was described at the March 2018 APAIS calibration workshop is an advancement on the two ratio methods. The approach is a complex and sophisticated method for adjusting initial sampling weights so that estimates of desired parameters are believed to be relatively unbiased over the entire time series, and similar to that expected had MRIP been in place in earlier years. An advantage of the new approach is that the adjustments to the initial weights are species independent and allow for utilization of all the data collected in MRFSS, not just those during specific time periods. As a result, every record in the APAIS dataset will have a sampling weight that can be utilized in estimation of parameters over desired domains or within strata.

A disadvantage of this approach is that it is not a true calibration, and so unbiasedness of the weighted estimators is difficult to assess. It also generates estimates of the desired parameters with large estimated variances, but this is to be expected given the multi-stage cluster sampling

design. Another source of error is in the estimation of the calibrated weights, but the method assumes that the weights calculated from the calibration methodology are fixed and known, and not based on estimates obtained from raking. As a result, it is not possible to know the effect of the estimation of weights on the true variability of the parameter estimates such as CPUE or effort. Even with these caveats, the proposed method is a more thoughtful and appropriate calibration approach for the APAIS dataset.

a) Does the proposed approach adequately account for consistent differences in estimates that would have been observed if the old MRFSS APAIS had been conducted side-by-side with the new MRIP APAIS in 2013-2017?

c) Is it reasonable to conclude that revised 2004-2012 APAIS estimates based on the application of the proposed approach would be more comparable than the current ones to estimates produced since 2013 under the new APAIS design?

d) Given the limitations of the available data, is it reasonable to apply the proposed approach to revise APAIS estimates prior to 2004 (back to 1981)?

All three of these questions are answered below since the topics overlap with respect to the proposed approach.

Overall, the answer to all three questions is a qualified yes, with more confidence for recent years than for older years. Had the 2 sampling designs been implemented simultaneously over the 2013-2017 period, the results most likely would have been very similar to that which was obtained from the proposed calibration approach. In addition, the proposed calibration approach appears to be appropriate for the 2004-2012 APAIS data given the use of the pseudo-weights as the initial weights for the raking algorithm. For earlier years, it is a bit more problematic. This is discussed below. First, though, is an overview of the concerns and discussions of the panel about the proposed method.

During discussion by the panel, several topics of concern were raised. First, was the decision regarding domain choices to be used in the raking algorithm. These varied by time period (the raking algorithm for the years 1981 to 2003 used 7 domains compared to 4 domains for later years). After much discussion it was decided that the choices of domains were reasonable given the information available to the panel. Related to this was the question of whether the order of the domain listed for the raking algorithm affected the re-weighting adjustments or convergence of the algorithm. The panel agreed that it is unlikely.

A second concern was the choice of splitting the 1991 to 2003 period into two 10-year periods, for purposes of applying the proposed calibration. Given that there are no clear alternatives, this was deemed a reasonable approach.

A third concern was the choice of initial sampling weights for the years prior to 2004, namely using the number of site-days sampled within each state-wave-mode-year combination standardized to the maximum value within each of the 2 time periods for the 1981 to 2003

years. The panel accepted this approach for obtaining initial weights. This reviewer (MCC) would recommend that the developers consider whether the calibration approach is robust to this decision. For example, annual initial weights based on treating the data as though it had been collected using a stratified random sample of primary sampling units (site-days) might be reasonable alternatives. In that case, the initial weights would be the proportion of total site-days sampled in a year that were sampled within each state-wave-mode combination. It is unlikely that the proposed calibration method would be sensitive to changes in initial weights in the earlier years, but it is important that the scientists demonstrate that the method is robust to choices of the initial conditions for the raking algorithm.

A final concern of the panel was that some of the initial results shown by Foster indicated some extreme outliers in several of the comparisons of the adjusted to pre-adjustment estimates of various parameters. In most of these instances, the results were for small domains within the APAIS sampling frame. For example, in the Foster presentation (“9FosterAP AIS.Calibration.Review.pptx”), slide 42 indicates that some estimates of effort (angler-trips) change by a factor of 5 or more after adjustment for a very defined set of conditions, namely in subregion 5, specific areas fished (inshore, offshore, etc.), set of years (e.g. 1981 to 1992) and modes of fishing. The discussion among the panel was the cause of these outlying values.

A discussion ensued as to whether the “extreme” values should be trimmed from the dataset and estimates based on the trimmed data. This reviewer (MCC) disagrees with this suggestion. Trimming or truncation of high or extreme weight values is usually done to reduce their impact on the variance of the estimates, especially for subgroup estimates. If some extreme weights are removed, the remaining weights need further adjustment in order to sum to the marginal population values. Hence, a repeat of the entire re-weighting process must be performed and could result in yet more “extreme” weights. This could continue for several iterations. The entire process of trimming is subjective – the decision of how much to trim and how many iterations of trimming depends on the analysts and is not reproducible by others. The cause(s) of the extreme weight values should be thoroughly reviewed before any decision for trimming is made. This has not been done. In fact, due to the lack of sensitivity analyses (except for the choice of which MRIP years to use in the first step of the calibration, namely adjusting the pseudo-weights for the years 2004 to 2012; see later) it is difficult to assess whether these extreme values are due to the choices made for calibration (which domains to use in the raking, the initial weights for the years 1981 to 2003, the choice of dividing the 20-yr time span into two 10-yr time spans, using a period of 3 years if a linear trend was found in the 10-yr periods, the decision to look for linear trend) or because of other reasons. The panel concluded that a possible reason is that for some years, species and domains (combinations of factors that may or may not have been strata identified in the sampling design), the number of observations was so small that the adjustment would need to be large. Hence, these are years, species or domains either where MRIP should not be used for estimation even under the new statistically valid design or where sufficient (or any) data were not available in the past. In other words,

since the raking adjustments depend to some extent on the initial weights and the parameter estimates provided for some domains may not be appropriate, other aspects of the proposed calibration method should be reviewed as possible sources of these outlying values before trimming is considered.

2. Briefly describe the panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.

The review proceedings were described in an earlier section and are not repeated here. Overall, the approach used (presentations followed by discussion and question and answer periods) was very effective at informing the panel of the proposed methodology and some results from initial application. In addition, it gave the panel the opportunity to engage the developers as needed for answering concerns and allowing for adequate time to review impressions and initial conclusions for the proposed approach.

Appendix 1

Atlantic Coastal Cooperative Statistics Program. 2018. Field Procedures Manual: Access-Point Angler Intercept Survey. 104 pp.

Boreman, J. 2012. Consultant's Report: Summary of the MRFSS/MRIP Calibration Workshop. March 27-29, 2012, Raleigh, NC. Documents and presentations are at: <http://sedarweb.org/otw-01-mrfssmrip-calibration-workshop-2012>.

Breidt, F.J., H.L. Lai, J.D. Opsomer, and D.A. Van Voorhees. 2011. A Report of the MRIP Sampling and Estimation Project: Improved Estimation Methods for the Access Point Angler Intercept Survey Component of the Marine Recreational Fishery Statistics Survey. 83pp. NMFS/S&T/MRIP, Silver Spring, MD. Including comments from reviewers and responses to reviewer comments.

Breidt, F.J., et al. 2012. A Pilot Study of a New Sampling Design for the Access Point Angler Intercept Survey. Report to the Marine Recreational Information Program's (MRIP) Design and Analysis Workgroup. Including comments from reviewers and responses to reviewer comments.

Carmichael, J and D. Van Voorhees (eds.). 2012. MRIP Calibration Workshop II. Final report to of a workshop hosted by NOAA Fisheries, Office of Science and Technology, MRIP and SEDAR: Southeast Data, Assessment and Review. September 8 –10, 2014, North Charleston SC.

Demming, W.E. and F.F. Stephan. 1940. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, 11(4):427-444.

Foster, J., F.J. Breidt, and J.D. Opsomer. 2018. APAIS data calibration methodology report. Manuscript dated March 11, 2018.

Hernandez-Stumpfhauser, D., F.J. Breidt, and J.D. Opsomer. 2016. Hierarchical Bayesian small area estimation for circular data. *Canadian Journal of Statistics* 44(4):416–430.

MRIP staff. 2014. A Descriptive Analysis of the Access Point Angler Intercept Survey. 2013 Design Change. Internal Report. 48 pp., August 20, 2014, Silver Spring, MD.

National Academies of Sciences, Engineering, and Medicine. 2017. Review of the Marine Recreational Information Program. Washington, DC: The National Academies Press. doi: <https://doi.org/10.17226/24640>.

National Research Council. 2006. Review of Recreational Fisheries Survey Methods. Washington, DC: The National Academies Press.

Papacostas, K.J. and J. Foster. 2017. National Marine Fisheries Service's Marine Recreational Information Program Survey Design and Statistical Methods for Estimation of Recreational Fisheries Catch and Effort, Version 1.0. Prepared for: U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service.

Additional material that was provided at the meeting included the power point files from each of the presenters (see final agenda in Appendix 3).

Appendix 2

Statement of Work

Statement of Work
National Oceanic and Atmospheric Administration (NOAA)
National Marine Fisheries Service (NMFS)
Center for Independent Experts (CIE) Program
External Independent Peer Review

Calibration Model Accounting for a Recreational Fisheries Survey Design Change

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards.

(http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).

Further information on the CIE program may be obtained from www.ciereviews.org.

Scope

The Office of Science and Technology requests an independent peer review of a calibration model proposed for use in revising statistics produced by a survey of marine recreational fishing catch rates on the Atlantic coast and in the Gulf of Mexico. This calibration model is considered by the Marine Recreational Information Program (MRIP) to be very important to adjust historical time series of recreational catch estimates in order to account for biases in past sampling and estimation methods that have become apparent with the development of a new, more statistically sound method. The calibration model is intended to account for past biases in catch rate estimates for the shore, private/rental boat, and charter boat fishing modes that have resulted from the continued use of a legacy sampling design for the Access Point Angler Intercept Survey (APAIS). A more statistically sound sampling design for the APAIS was

implemented in March of 2013

Calibration Model for the APAIS Design Change

In 2014, a Calibration Workshop was held to evaluate the potential consistent effects of implementing a new sampling design for the APAIS on the Atlantic and Gulf coasts in 2013. Workshop participants included three expert statistical consultants and representatives from NOAA Fisheries, the regional fishery management councils, the interstate marine fisheries commissions, and several state agencies. The participants determined that analyses conducted by the NOAA Fisheries Office of Science and Technology showed there was sufficient evidence that the more complete temporal coverage of the new design resulted in consistent changes in APAIS angler catch rate statistics for at least some species. They developed three different calibration models to evaluate for possible use in correcting the pre-2013 APAIS statistics. The statistical consultants concluded the simplest of the three proposed models was appropriate for use in the short term until a more complete evaluation of all three calibration models could be completed using three years of new APAIS data (2013-2015). The plan was to complete that evaluation by the end of 2016, so that one method could be selected as the best for use in 2017 to revise APAIS estimates prior to 2013.

Requirements

NMFS requires three (3) reviewers to conduct an impartial and independent peer review in accordance with the SoW, OMB Guidelines, and the Terms of Reference (ToRs) below. The CIE reviewers shall have working knowledge and recent experience in the design of sampling surveys and the evaluation of non-sampling errors (i.e., undercoverage, nonresponse, and response errors) associated with changes to survey designs over time. In addition, they should have experience with complex, multi-stage sampling designs, time series analyses, regression estimators, and small domain estimation methods. Some recent knowledge and experience in current surveys of marine recreational fishing is desirable but not required.

NMFS will provide a Chair who has experience with U.S. fisheries stock assessments and their application to fisheries management. The Chair would ensure that reviewers understand the importance of maintaining a comparable time series of marine recreational fisheries catch statistics for use in stock assessments and their application to fisheries management. The Chair will not be selected by the contractor and will be responsible for facilitating the meeting, developing and finalizing a summary report and working with the CIE reviewers to make sure that the ToRs are addressed in their independent reviews.

Tasks for Reviewers

Pre-review Background Documents

The following background materials and reports prior to the review meeting include:

APAIS Design Change Calibration Workshop Report:

http://www.st.nmfs.noaa.gov/Assets/recreational/pdf/MRIPC Calibration WorkshopII_FinalReport.pdf

NC APAIS Pilot Study Report: *A Pilot Study of a New Sampling Design for the Access Point Angler Intercept Survey.*

https://www.st.nmfs.noaa.gov/pims/main/public?method=DOWNLOAD_FR_PDF&record_id=772

Report on APAIS Calibration Model:

This report will be provided by the contractor (via electronic mail or make available at an FTP site) to the CIE reviewers.

Panel Review Meeting

Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The meeting will consist of presentations by NOAA and other scientists to facilitate the review, to provide any additional information required by the reviewers, and to answer any questions from reviewers.

Contract Deliverables - Independent CIE Peer Review Reports

The CIE reviewers shall complete an independent peer review report in accordance with the requirements specified in this SoW and OMB guidelines. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

Other Tasks – Contribution to Summary Report

The CIE reviewers may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. The CIE reviewers are not required to reach a consensus, and should provide a brief summary of each reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Foreign National Security Clearance

When reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for reviewers who are non-US citizens. For this reason, the reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: <http://deemedexports.noaa.gov/> and http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html. The contractor is required to use all appropriate methods to

safeguard Personally Identifiable Information (PII).

Place of Performance

The place of performance shall be at the contractor's facilities, and at the NOAA Fisheries Service Headquarters in Silver Spring, Maryland.

Period of Performance

The period of performance shall be from the time of award through April 31, 2018. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Within two weeks of award	Contractor selects and confirms reviewers
Approximately 2 weeks later	Contractor provides the pre-review documents to the reviewers
March 2018	each reviewer participates and conducts an independent peer review during the panel review meeting
Within two weeks of panel review meeting	Contractor receives draft reports
Within two weeks of receiving draft reports	Contractor submits final reports to the Government

Applicable Performance Standards

The acceptance of the contract deliverables shall be based on three performance standards: (1) The reports shall be completed in accordance with the required formatting and content (2) The reports shall address each ToR as specified (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (<http://www.gsa.gov/portal/content/104790>). International travel is authorized for this contract. Travel is not to exceed \$12,000.

Restricted or Limited Use of Data

The contractors may be required to sign and adhere to a non-disclosure agreement.

NOAA Fisheries Project Contact:

Dave Van Voorhees
NOAA Fisheries
1315 East West Highway
Silver Spring, MD 20910
dave.van.voorhees@noaa.gov

Appendix 3

Attendance List

NAME	AFFILIATION
Carolyn Belcher ¹	GA Dept of Nat. Resources
Jim Chromy ²	retired (RTI)
Mary Christman ²	MCC Stat Consulting
Matt Cieri ¹	ME Dept of Mar. Resources
Michael Murphy ¹	retired FL-FWCC
Paul Rago ¹	MAFMC SSC
John Whitehead ²	Appalachian State University
Richard Cody	NOAA support ECS
John Carmichael ³	SAFMC
Jason Didden ³	MAFMC
Laura Diederick	NOAA Fisheries support
Katie Drew ³	ASMFC
John Foster ³	NOAA Fisheries
Clifford Hutt	NOAA SF-1 (HMS)
Ryan Kitts-Jensen ³	NOAA Fisheries - ST1 support
Yong-Woo Lee	NMFS-ST1
Vivian Matter	NMFS - SEFSC
Jean Opsomer ³	CSU
Karen Pianka	NOAA Fisheries - ST1 support
Thomas Sminkey ³	NOAA Fisheries - ST1
Dave VanVoorhees ³	NOAA Fisheries
Chris Wright	NMFS - SF

¹ NOAA reviewer

² CIE reviewer

³ Presenter

⁴ Rapporteur

Proposed Agenda for the APAIS Calibration Review Workshop

Tuesday, March 20:

OPEN SESSION

- 9:00am – Welcome, Introductions, and Overview of Workshop
Dave Van Voorhees, NOAA Fisheries Office of Science and Technology
Mike Murphy, Chair of Peer Review Panel
- 9:20am – Presentation: MRIP Transition Planning and the Access Point Angler Intercept Survey
Dave Van Voorhees, NOAA Fisheries Office of Science and Technology
- 9:40am – Presentation: Importance of Calibrated Catch for Fisheries Management
Jason Didden, Mid-Atlantic Fishery Management Council
- 10:00am – Presentation: Importance of Calibrated Catch for Fishery Stock Assessments
Katie Drew, Atlantic States Marine Fisheries Commission
- 10:20am – BREAK
- 10:30am – Presentation: Weighted Estimation for the APAIS (and Calibration Workshop I)
Dave Van Voorhees, NOAA Fisheries Office of Science and Technology
- 11:05am – Presentation: New Design of the APAIS
Tom Sminkey, NOAA Fisheries Office of Science and Technology
- 11:40am – Presentation: Calibration Workshop II
John Carmichael, South Atlantic Fishery Management Council & SEDAR
- 12:00pm – LUNCH BREAK
- 1:10pm – Presentation: Considered Ratio Calibration Methods
Ryan Kitts-Jensen, NOAA Fisheries Office of Science and Technology
- 1:50pm – Presentation: APAIS Calibration Methodology
Jean Opsomer, Westat and Colorado State University
- 2:40pm – Presentation: APAIS Calibration Results
John Foster, NOAA Fisheries Office of Science and Technology
- 3:30pm – BREAK
- 3:40pm – Follow-Up Questions for Presenters
- 4:20pm – Public Comment
- 4:50pm – Summary of Day 1
Mike Murphy, Chair of Peer Review Panel
- 5:00pm –ADJOURN

CLOSED SESSION

- 5:10pm – Review Panel Coordination and Writing
- 6:00pm – ADJOURN

Wednesday, March 21:

OPEN SESSION

- 9:00am – Overview of Day 1 and Preview of Day 2
Mike Murphy, Chair of Peer Review Panel
- 9:10am – Follow-Up Questions for Presenters
- 10:15am – BREAK
- 10:30am – Follow-Up Questions for Presenters
- 12:00pm – LUNCH BREAK

CLOSED SESSION

1:00pm – Review Panel Coordination and Writing

OPEN SESSION

2:30pm – Initial Summary Findings of Review Panel

3:30pm – BREAK

3:45pm – Public Comment

4:15pm – ADJOURN

CLOSED SESSION

4:30pm – Review Panel Coordination and Writing

6:00pm - ADJOURN

Thursday, March 22

CLOSED SESSION

9:00am – Review Panel Coordination and Writing

12:30pm – ADJOURN

**Center for Independent Experts (CIE) Independent
Review Report
Marine Recreational Information Program (MRIP)
Access Point Angler Interview Survey (APAIS)
Calibration Peer Review**

**Calibration Peer Review Workshop
March 20-22, 2018
Sheraton Hotel
Silver Spring, MD**

May 2018

Individual Peer Review Report

James R. Chromy

Table of Contents

Executive Summary

1. Background
 2. Findings for Each Term of Reference
 3. Clarifications (if any) of Points in the Summary Report
 4. Critique of the NMFS Review Process
 5. Conclusions
 6. References
- Appendix 1. Bibliography of Materials Provided for Review
Appendix 2. Statement of Work
Appendix 3. Panel Membership

Executive Summary

The calibration process for adjusting time series APAIS estimates for the period 1981 through 2013 wave 1 was reviewed by a panel of invited experts during a workshop held in Silver Spring, MD March 21-22, 2018. The workshop consisted of open and closed sessions. During open sessions, MRIP staff and consultants presented background information and described three candidate methods for performing the calibration: two ratio adjustment methods and a weight adjustment method. During closed sessions, the invited experts were able to develop additional questions and plan the remaining evaluation process. Each outside expert prepared an independent report. The panel chair prepared a group report based on inputs from the individual panel members.

Calibration through weight development and raking ratio weight adjustment appears to be the preferred method of calibrating prior years' estimates. This was described as method 3. The method uses an iterative procedure to adjust the distribution of angler trips for early years (2012 and earlier) to match the same distribution in the data for data generated after the implementation of new sample design and survey procedures (2013 and later). This only affects the estimation of catch per angler trip. The number and distribution of angler trips is obtained from the effort surveys and a process for calibrating early year data has already been implemented and peer reviewed.

Methods 1 and 2 depend on ratios of post- and pre-estimates of total catch for specified domains. They have intuitive appeal, but may require specialized treatment for particular estimates. They incorporate correction for selected control totals at the cell level. Method 3 works on the distribution of angler trips and should not require specialized treatment for any particular catch estimate. Method 3 controls estimates in more dimensions by working on marginal rather than cell totals.

Small domains may still have difficulty with extreme weight adjustments because of sparse data in some adjustment dimensions. Highly unequal weights may also inflate the sampling error of estimates. Some reasonable methods of checking for extremes and trimming the extreme weights when necessary should be considered.

1. Background

NOAA Fisheries continuously monitors marine fish catch and removal. Regional fishery management councils, states, interstate fishery commissions, and other regulatory agencies use the data to effectively manage the fisheries. NOAA Fisheries develops the required data both from surveys and from biological studies. Consistently estimated time series are required to support the models for estimating stocks by species and area.

The Access Point Angler Survey (APAIS) collects data to estimate the average catch per unit effort (CPUE). The unit of effort is defined as an angler trip concluding on a specified day. The number of angler trips is estimated from separate surveys of recreational fisherman and fishing boat for hire operators.

NOAA Fisheries surveys have been in a period of transition. Following a National Research Council (NRC) review in 2006, they have been testing and implementing improved sampling, survey, and estimation procedures for estimating recreational fishing effort and catch. The improved methods are being implemented in the Gulf of Mexico and the Atlantic marine fisheries from Louisiana to Maine.

The sample design, data collection, and design-weighted estimation methodology for both the CPUE and the fishing effort surveys have been implemented starting with wave 2 of 2013. Ideally, simultaneous data from the old and new methodology would be compared for one or more years to establish correction factors for the older data; this was not done for the APAIS.

Quasi-design-based weights were developed for the period 2004 through 2013 wave 1 to better reflect the probability of selection under the previous design. Data available for early surveys 1981 to 2003 were unweighted. CPUE estimates for both early periods are not comparable with the recent year estimates. The primary problems identified in the National Research Council (2006) review related to coverage of the angler trip population by time of day and region. Data collectors had the discretion to finish assignments early if they met a specified quota and to collect data at alternate sites or from alternate modes of fishing.

Rather than starting a new time series and ignoring the estimates from earlier years, NOAA Fisheries has developed some candidate calibration methodologies to make the early year estimates more comparable to the recent ones. Based on a prior peer review (Carmichael and Van Voorhees, 2014), an interim calibration methodology based on the simplest alternative was adopted along with recommendations for further study as more data accumulated under the new design. The method used the ratio of estimated catch based on a full day to the estimated catch during peak periods covered in prior years. The resulting ratio was applied to estimates of catch generated from the peak period data in earlier years. Some data were lost from nonpeak periods in prior years.

Since the implementation of the interim calibration methodology, a multi-dimensional weight adjustment method using the distribution of data from recent years to calibrate weights from prior years was developed and tested. Preliminary results of this development and testing along with comparisons to the currently implemented adjustment methodology were presented at the

calibration peer review workshop held March 20-22, 2018 at the Sheraton Hotel, Silver Spring, MD. The general objectives of the peer review were specified by questions in the draft terms of reference.

**Draft Terms of Reference for Peer Review
of the Approach Proposed to Account for the Change
in Design of the Access Point Angler Intercept Survey**

1. Evaluate the suitability of the proposed approach for converting historical estimates of mean angler catch rates obtained using the old MRFSS Access Point Angler Intercept Survey (APAIS) sampling design to estimates that best represent what would have been produced had the new MRIP APAIS sampling design been in place prior to 2013.
 - a. Does the proposed approach adequately account for consistent differences in estimates that would have been observed if the old MRFSS APAIS had been conducted side-by-side with the new MRIP APAIS in 2013-2017?
 - b. Is the proposed approach a suitable alternative to the calibration models that were originally developed in the 2014 MRIP Calibration Workshop and later evaluated by MRIP?
 - c. Is it reasonable to conclude that revised 2004-2012 APAIS estimates based on the application of the proposed approach would be more comparable than the current ones to estimates produced since 2013 under the new APAIS design?
 - d. Given the limitations of the available data, is it reasonable to apply the proposed approach to revise APAIS estimates prior to 2004 (back to 1981)?
2. Briefly describe the panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.

The review panel met with marine fisheries staff and contractors for two and a half days. The first day and much of the second day were an open session primarily to set the stage and identify issues to be addressed by the reviewers. The remaining time was for closed session discussions among reviewers. Fisheries staff presentations were accompanied by power point presentations, which were made available to reviewers and are listed in Appendix 1. Mike Murphy chaired the meeting.

Presentation 1: Dave Van Voorhees introduced the Marine Recreational Information Program (MRIP) transition planning for the APAIS. The transition process, begun in 2008, reviewed the current design, developed improvements, and pilot tested the improvements. Experts inputs, stakeholder engagement, and peer review provided guidance. A National Academies review in 2017 recognized impressive progress and recognized that redesigned APAIS methods “reflect state of the art methods in survey sampling.” The MRIP revised the 2004-2012 series using

pseudo-design-based weights. It also reviewed alternative calibration strategies and implemented a ratio adjustment method considered simplest for interim estimates.

Presentation 2: Jason Didden discussed the importance of calibrated catch for fisheries management. Recreational fishing is economically important to the states with marine fisheries. Most analyses involve time series of (1) assessments, (2) allocations, and (3) management measures.

Presentation 3. Katie Drew described the methods used to assess stocks. Changes in relative abundance (determined from tows) and estimates of removal (catch and bycatch) can be used to model stock at the beginning and the end of a period. Age and size can also be incorporated. Models for estimating stocks are tolerant to imprecise estimates in a time series; bias is a problem.

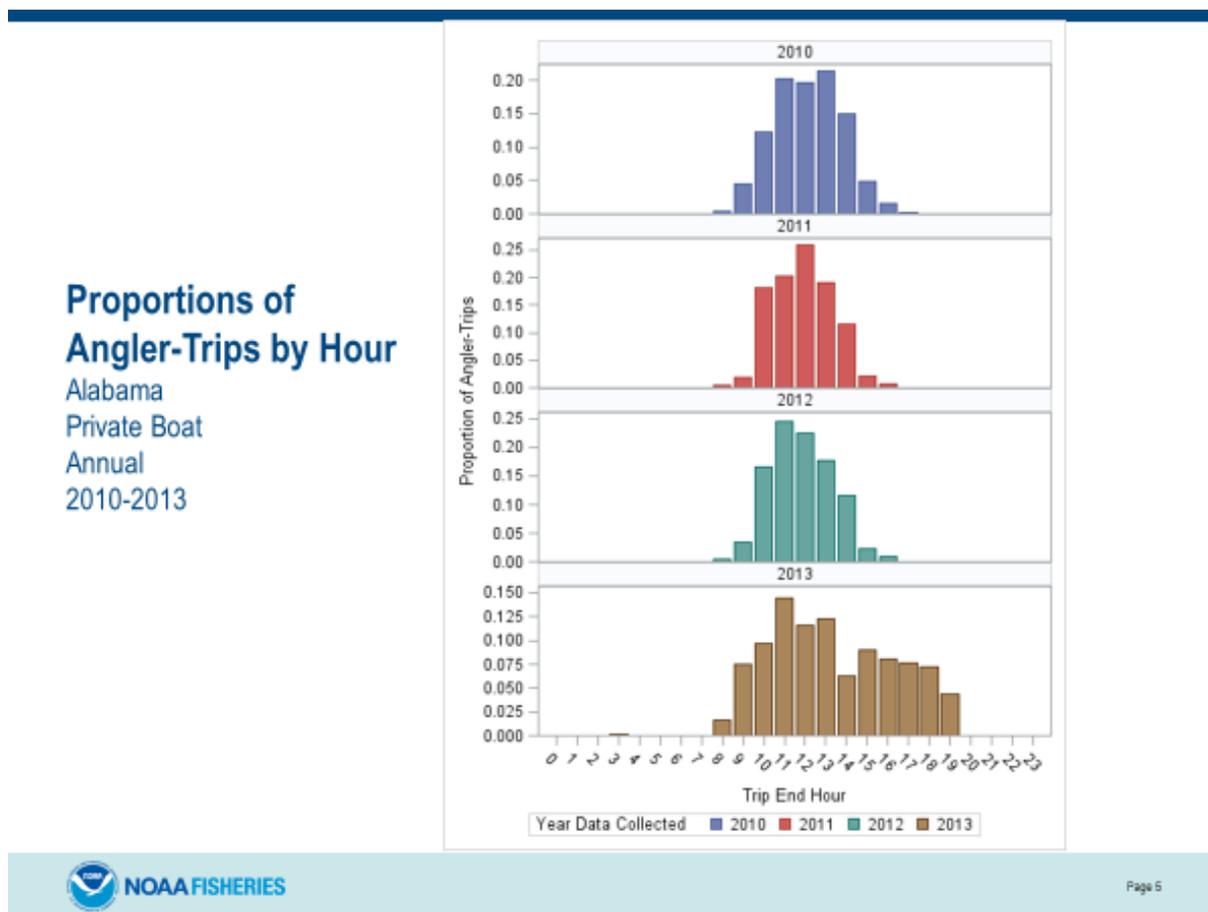
Presentation 4. Dave Van Voorhees discussed the methodology for developing pseudo-design-based weights for 2004-2012 APAIS estimates. These have been peer reviewed and accepted. The estimates developed using these weights may need to be adjusted further to make them comparable to estimates from the redesigned survey and weighting procedures in the recent period.

Presentation 5. Tom Sminkey discussed the new sample design for the APAIS. He summarized design changes that affect coverage and selection probabilities. Under the old design, single sites were sampled, but data collectors could use alternate sites at their own discretion. Under the new design, site clusters were selected. Each cluster contained one to three sites. Multi-site clusters were only formed when considered necessary to maintain productivity and procedures for randomly moving from site to site were specified to maintain a known selection probability for each site-time interval. Under the old design, data collectors were instructed to select peak times during the day to maximize yield. Under the new design, primary sampling units were a combination of site and six-hour time interval. Data collectors collected data over the entire six-hour interval. Under the old design, the fishing mode was specified for the site-day, but an alternate mode was allowed. Under the new design, initially a single mode was specified. Later, all available modes at the site were allowed. Under the old design, there were no tallies of eligible anglers. Under the new design, all eligible anglers were tallied. As a result, no overall selection probability could be computed under the old design. Under the new design, selection probabilities could be determined at each stage of sampling. The judgment of data collectors could determine the sample under the old design with no ability to compute a selection probability at most stages of sampling. Under the new design, the sampling process was controlled, and sufficient data was recorded to determine selection probabilities.

Presentation 6: John Carmichael summarized the process and recommendations of the calibration workshop held in North Charleston, SC in 2014. The workshop considered three calibration methods: a simple ratio of total catch to peak time catch to adjust estimates from years when only peak time data was collected, a more complex ratio method which incorporates effort distributions and trip sampling weights, and modeling approach, which uses regression modeling to classify trips into three general categories: morning, peak, and evening) based on

their characteristics. The workshop recommended use of some version of the simplest method during an initial transition period with further development and review of alternative methods.

Presentation 7. Ryan Kitts-Jensen presented an overview of the ratio methods considered for APAIS calibration. The graphs of trip return-hour distributions for private and charter boats showed a broader and flatter distribution in the 2013-2016 era when compared with 2010, 2011, and 2012 distributions. A graphic comparison of Alabama private boat data (copied below) showed a much greater difference with the 2013 distribution including a substantial portion of trips returning late in the day in 2013 and not covered in prior years. This may be an extreme case, but it illustrates the time of day coverage in the early years (before 2013).



This presentation described additional details about the three methods considered for calibrating 2004 to 2012 estimates.

Ratio Method 1, simple ratio adjustment. The following represents this reviewer's understanding of the methodology.

- Peak definitions are developed by year, sub region, state, and mode.
- Peak time intervals vary across years.
- Ratios are calculated by species and catch type (landed vs. released)

- The new survey design data used to compute adjustment factors could be combined across years.
- The following steps were computed for each year, sub region, state, mode, species, and catch type for the years to be adjusted (2012 back to 2004):
 - Identify the peak period for the target year: assume 2010 for illustration.
 - Estimate the peak period catch for the target year.
 - Estimate the peak period catch and total catch for 2013 or more years beginning with 2013. Compute the ratio of total catch to peak period catch under the new design.
 - Apply the ratio to the estimated peak period catch the target year (e.g., 2010).
 - Aggregate the catch estimates to higher levels by summing the catch estimates over sub region and mode.

This methodology is analogous to computing separate ratio estimates (e.g., Cochran 1977, pp. 164-165) over pseudo strata defined by year, sub region, state, mode, species, and catch type. As pointed out during the presentation, the method depends on developing estimates in some small cells based on the cross classification of the defining categories. Large adjustments can also arise when the catch in a target species occurs mostly outside the peak period. It was necessary to coarsen the classification in many cases to avoid problems with cell estimates of zero catch.

If 2013 is used to represent the post design change era, then an adjusted estimate in terms of total catch estimates for year y by sub region, state, and mode are produced as:

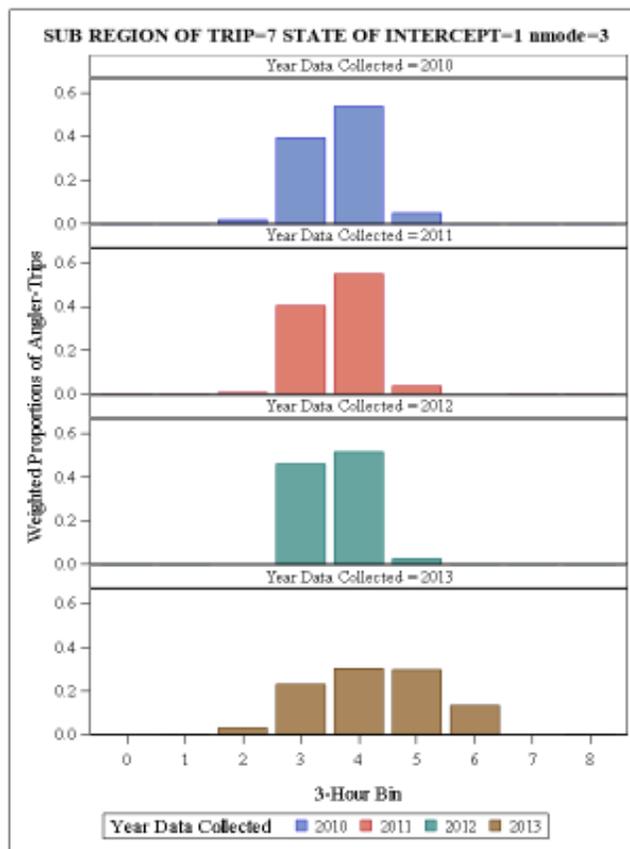
$$\hat{C}_{Ty}^* = \hat{R}_{2013|y} * \hat{C}_{Py}, y \in \{2004, 2005, \dots, 2012\}$$

where \hat{C}_{Py} is the estimated peak period catch estimate for year y , $\hat{C}_{P2013|y}$ is the estimated peak period catch estimate for 2013 using the year y peak period, $\hat{R}_{2013|y} = \frac{\hat{C}_{T2013}}{\hat{C}_{P2013|y}}$, and \hat{C}_{T2013} is the total catch estimate for 2013. All estimates are within sub region, state, and mode. This could also be written as $\hat{C}_{Ty}^* = \frac{\hat{C}_{Py}}{\hat{C}_{P2013|y}} * \hat{C}_{T2013}$ showing that trend is measured by the change in peak period catch applied to the total period catch for 2013. This would appear to cover those species caught only in nonpeak periods, but the trend ratio might be difficult to estimate if sample sizes are small. Also, if the species of interest is not reported in the peak period for the target year, the adjusted estimate will also be zero. Using more years in the reference period can reduce, but not necessarily eliminate, the problem with division by zero. For some species, peak period ratios may not be good indicators of trend in total catch which includes nonpeak hours. Some *ad hoc* rules are proposed to handle problems with ratio estimation.

Ratio Method 2: Complex ratio adjustment. Method 2 deals with time of day through time bins rather than just comparing peak time catch to total catch to adjust prior years. A copy of the graphic comparison of Alabama shore mode angler trip distributions before and after the design change illustrates the potential need for calibration. Method 2 reweights the 2013+ data to match the target year angler trip distribution, and then develops ratio adjustment factors from the 2013+ data to adjust prior year catch estimates.

Ratio Method 2 Intro

- **Weighted Proportions of Angler-Trips by 3-Hour Bins (Return Times)**
 - Alabama Shore Mode
 - Data from 2010-2016
 - 2013 represents combined APAIS years (2013-2016)
- Instead of only peak and non-peak designations, match a coarse APAIS (2013+) temporal distribution to a MRFSS year's coarse temporal distribution
- Example 3-Hour Time Bin Definitions
 - 3 – 9:00am-12:00pm
 - 4 – 12:00pm-3:00pm
 - 5 – 3:00pm-6:00pm



The following steps represents this reviewer's understanding of the methodology.

- Define time of day trip bins. As an example, three bins can be defined as 9 am to 12 pm, 12 pm to 3 pm, and 3 pm to 6 pm.
- Using available weights for the target year (2012 and earlier) estimate the number of angler trips within each time bin by sub region, state, and mode. For 2004 to 2012, pseudo design-based weights have been developed and are used at this step.
- Adjust the 2013+ weights through post-stratification to match the target year angler trip distribution by bin, sub region, state, and mode.
- Calculate adjusted 2013+ annual catch estimates by species, catch type, sub region, state, and mode using the adjusted weights.

- Use the ratio of the properly weighted 2013+ estimates to the adjusted 2013+ to adjust target year catch estimates by species, catch type, sub region, state and mode.
- Use this ratio to adjust the target year y peak data.

Method 2 forces the relative distribution of angler trips by time of day bin to exactly match the pre-adjustment distribution. Any change in annual catch estimates can only be based on changes in the catch per unit effort (CPUE) within the time bins.

The results of sensitivity analysis using a single year (2013) or multiple years (2013+) compute ratio adjustments were examined. Both methods were designed to correct for the distribution of angler trips by time of day.

Presentation 8. Jean Opsomer's presentation addressed APAIS Calibration Methodology. He noted the need for comparable time series data. Unlike the effort surveys which ran overlap surveys, new and old sampling and survey methodology had not been applied for the catch survey. He noted that introduction of new (improved) methods has been implemented for many government surveys. A variety of approaches had been followed to test, adjust, or ignore the effects on time series estimates. The following represents my understanding of the approach followed.

For the APAIS, pseudo weights were developed for 2004 to 2012 (wave 1) using available design information. No weights or useable design information were available for 1981-2003.

Method 3 operates on weight adjustments at the angler trip level. It requires known control totals for marginal distributions in more than one dimension. The process is iterative and is continued until it converges. Since the marginal control totals are not known, they are obtained as estimates of these domains from the APAIS surveys starting with 2013 wave 2 or some subset of those years with the new design and design-based weighting in place. The following steps are applied for 2004-2013 wave 1.

- Use the quasi design-based weights as a starting point.
- Compute marginal weight totals for four sets of domains:
 - AF defined by state, wave, mode, area fished.
 - HS defined by state wave, mode, coastal/non-coastal household status.
 - FH defined by state, wave, mode, and for-hire boat frame status.
 - RE defined by state, wave, mode, and sub region.
- Compute control totals from the new survey design.
- Apply the iterative raking procedure to each year's data.
- Use the revised weights for all estimates.
- Recognize that estimates for domains with a small number of sample points will have high sampling error.

Question: Do the domain estimates incorporate the effort estimates? At what levels are they applied? This was addressed to a limited extent in group discussions, but it would help to see more detail. For example, is this done within state, wave, and mode?

No initial weights are available. For 1993 to 2003. The following steps are applied:

- Develop an initial weight by counting the number of site days with intercepts by state, wave, and mode.
- Determine the maximum over years and weight all years up to that level within wave and mode.
- Add 3 additional sets of domains for the raking procedure:
 - KOD defined by state, wave, mode, and kind-of-day.
 - MG defined by state, wave, mode, and month group.
 - AC defined by state, wave, mode, and site activity classes.
- Use seven sets of control totals (AF, HS, FH, RE, KOD, MG, and AC) from 2004-2013 and apply the raking procedure from 1993 to 2003 data.
- Check for linear trend and break up the interval if significant trend is detected.
- Re-apply the procedure for 1981 to 1992 using 1993 to 2003 to develop control totals.

Presentation 9. John Foster discussed the implementation and results of APAIS calibration. Discussion of calibration scope indicates catch and effort are both calibrated, but calibration for new fishing effort survey was not included in the presentation. This requires further explanation. The results on the distribution of estimates were highly variable, perhaps because some very small domains were included in the distributions. The estimated number of angler trips was brought up to more comparable levels over periods studied. Were these adjustments just used to compute a revised catch per unit effort? In that case, only the relative distribution of angler trips is relevant. A key question is: Are the extreme ratios of estimates by species and sub region or other specific estimates just the result of small sample size and high variability for those estimates? Without looking at specific estimates at fine levels, it would appear prudent to implement some reasonable weight trimming approach to avoid extreme adjustments in general. The key advantage of the weight adjustment method is that it does not require separate adjustment for each statistic produced from the data. Some care must be exercised to avoid publishing official estimates which are highly imprecise because of small sample size.

The reviewers appreciated the high level of detail and the work completed to produce over 100 graphs to evaluate the outcome of the adjustment process.

2. Findings for Each Term of Reference

The responses are shown below. The “proposed approach” is assumed to be based on adjusting weights in several dimensions using the raking ratio methodology. This is method 3 in the discussion above.

Item 1a. Does the proposed approach adequately account for consistent differences in estimates that would have been observed if the old MRFSS APAIS had been conducted side-by-side with the new MRIP APAIS in 2013-2017?

Response: It is probably not possible to adequately account for all differences that would have been observed had the new procedures been introduced earlier. The new raking ratio method attempts to identify those factors distributed differently under the new design with associated design-based estimation. The method assumes that the distribution of angler fishing trips over the domains defined by these factors may be projected back to recent years. Adjustment factors to weights and consequently to the distribution of angler trips are developed for each domain sequentially. The process is repeated until the required adjustments converge. This allows several factors to be incorporated without requiring data at the cross classification of those factors. The method is particularly suitable for the APAIS data, since the estimated angler trip distributions under the new design are supported by the application of probability sampling theory.

Table 1 shows the control total definitions or the cells within which estimates or weights are adjusted using the three methods. The early designs were criticized for coverage problems including little or no coverage of fishing trips ending at night. Weight adjustment procedures depend on being able to redistribute the weights across observed domains. If a domain (*e.g.*, night fishing) is not included, then weights cannot be assigned to it. The attempts to adjust for night fishing with methods 1 and 2 also have computational problems. It might be possible to add some broad categories of time of day to the marginal total controls under method 3. Method 3 incorporates wave as a control total, which seems important for fisheries and sites with large seasonal variations.

The main problem with ratio methods 1 and 2 is that the adjustment factors cannot be computed when cell-level estimates are zero or when they are unstable for sparse domains. Both methods 1 and 2 work directly with cell estimates. Method 3 adjusts weights (ultimately angler trip counts) and is less susceptible to the computational problems associated with methods 1 and 2. Fewer *ad hoc* adjustment procedures should be required; they would generally involve some collapsing of one or more sets of marginal controls. Since method 3 works with angler visit counts rather than specific catch estimates, the number of data problems encountered during the process should be much lower and the process can be done once and apply to all domains, all species, and all other estimates. This does not mean that all small cells will have a nonzero estimate or an adequately precise estimate for publishing as an official statistic. Intense weight adjustment procedures can lead to large unequal weighting effects (UWEs). Calculation of UWEs and some reasonable procedures for trimming extreme weights should be considered. Official tabular presentation should require minimum sample size (both number of site days and number of angler trips) and not publish estimates with estimated relative sampling error estimates above reasonably specified levels. Micro data files will require some documentation to warn users about possibly unreliable estimates.

Kish (1992) defines the unequal weighting effect in terms of the squared coefficient of variation of sample weights, C_k^2 . The unequal weighting effect is estimated by:

$$1 + C_k^2.$$

He suggests checking the unequal weighting effect before and after trimming and checking how much it can be reduced by trimming. Like most government surveys, the APAIS is multipurpose and involves many estimates (*e.g.*, species) and many domains (*e.g.* state, mode, subregion, area fished, etc.) The reduction in variance should be judged in terms of the reduction in bias achieved with applying appropriate control totals. The raking procedure should be repeated after

each trimming step to maintain marginal weight totals. Computational difficulty in satisfying the control totals and achieving a reasonable trim may require some *ad hoc* adjustment, e.g., collapsing cells in the control totals or relaxing the trim limits.

Table 1. Control Total Definitions

Method 1. Simple Ratio	Method 2. Complex Ratio	Method 3. Raking Ratio
Year	Year	One or more years under new design: e.g., 2014 to 2017 Block of years under old design showing no detectable trend: e.g. 2003 to 2012
State	State	State
Sub region	Sub Region	Wave
Mode	Mode	Mode
Species	Species	Cross classification of above with marginal totals for: AF area fished HS household status (coastal/non-coastal) FH for hire boat frame status RE sub region
Catch type (landed vs. released)	Catch type (landed vs. released)	
Time of day Peak and nonpeak (new) Peak only (old)	Time of day in 3-hour bins based on year being adjusted. e.g.: 9 am to 12 pm 12 pm to 3 pm 3 pm to 6 pm	Additional marginal totals for early years (1981 to 2003): KOD kind of day MG month group AC site activity class
Adjustments made to pre-change by ultimate cell. Back adjusts estimates based on change in peak period estimates.	Adjustments made to post design change weights to match distribution angler trips to target year (a weight adjustment). Compute pre and post estimates from the comparable distribution of angler trips. Finally, back adjust prior years from the new survey total catch.	Adjustments made to weights to reflect align distribution of angler trip estimates with post design change distributions

Note that method 3 operates on angler domains and not on outcome variables such as average catch within domains. Use of method 3 would not prevent a fisheries scientist from applying ratio methods or modeling when direct weighted estimates are imprecise or estimated as zeroes. This would require access to micro data with time of day data fields.

It was noted that some form of bootstrap variance estimation might be required to account for the complex nonlinear adjustment process as well as sampling variability. The bootstrap weight replicates should attempt to incorporate the effects of sampling error and the application of the raking methodology.

Item 1b. Is the proposed approach a suitable alternative to the calibration models that were originally developed in the 2014 MRIP Calibration Workshop and later evaluated by MRIP?

Response: The original adjustments were based on ratios to reflect the relationship of peak time of day data to total day data and adjust prior year's peak data to the level of full day data. The approach is reasonable, but results in losing some of the data from the nonpeak periods. The two ratio methods work directly with estimated catch and must be applied at that level. Both methods assume that the relative change in peak catch can be applied to total catch estimates.

Proposed method 3 is particularly applicable to multi-purpose surveys such as the APAIS, since the procedure may be applied to the weights once and then the weights can be used to generate a large number of estimates. The raking ratio methodology has been in the literature since an early paper by Deming and Stephan (1941). Oh and Scheuren (1983) brought attention to it in their 1983 work. With advances in available computer power, the iterative raking procedure has gained applicability. Folsom and Singh (2000) present a generalized model for weight adjustment and trimming which includes methods similar to raking ratio methods; this methodology has been incorporated in the SUDAAN software package (Research Triangle Institute, 2012).

Item 1c. Is it reasonable to conclude that revised 2004-2012 APAIS estimates based on the application of the proposed approach would be more comparable than the current ones to estimates produced since 2013 under the new APAIS design?

Response: Yes, this should be true for most estimates. Since the new procedures attempts to correct for many aspects of the design changes rather than just time of day observed, it appears reasonable that the results will be more comparable to those that would be obtained had the current design been applied in prior years.

1d. Given the limitations of the available data, is it reasonable to apply the proposed approach to revise APAIS estimates prior to 2004 (back to 1981)?

Response: The alternative would seem to be to suppress all estimates prior to 2004. The proposed approach requires applying some of the same assumptions about prior year data given adjusted estimates for 2004 to 2012. The approach appears reasonable. It should be recognized that any attempt to correct a time series of estimates for methodological improvements implemented late in the series will be difficult to implement and difficult to evaluate. Corrections to the time series become more difficult further back in time. From a user perspective, an adjusted time series on a comparable scale should be much more useful for comparing changes over time.

General comment: There is no perfect solution to the problem of making a long-term timeseries of estimates comparable with recent estimates which utilize improved methods. This should not prevent the introduction of new improved methodology for current and future estimates. In the long term, the primary goal should be to improve the quality of data series going forward along with best efforts to compare with prior years' estimates.

Question: The APAIS is designed to produce estimates of catch per unit effort, but the adjustment process was discussed in terms adjusting the catch estimates in the time series. I would have like to see how the two sets of estimates are brought together. Is it done on the aggregate of within specific domains such as those used for adjusting the time series? The presentations on methods 1 and 2 indicated the adjustment of "catch" estimates. Method 3 is proposed to provide direct estimation of catch with additional blending with effort estimates that should be refined and reviewed.

3. Clarifications (if any) of Points in Summary Report

I reviewed a draft of the summary report and am satisfied that no clarifications are needed.

4. Critiques of the NMFS Review Process

This review is just one of several external reviews of the overall marine recreational fishing surveys and focused on calibrating data from earlier estimates to make them more comparable to estimates being generated by the revised sampling and survey designs. Effort estimation and calibration had already been reviewed; we were asked to address only the calibration of estimates obtained from the APAIS. I understand the product of effort estimates and catch per unit effort estimates is used to create the final estimate of total catch by various domains of interest. It was never clear to me whether this product was computed at some low level, e.g., year, state, wave, mode, sub region, or at some higher level. Incorporating the process into the weighting process was mentioned and seems logical, but perhaps this final step should also be elaborated and reviewed as a final step.

The process of providing information to the reviewers through a workshop held at a hotel near NOAA Fisheries facilities was excellent. Presenters were knowledgeable about each component of the APAIS process. Presentations were supported by accompanying power point or PDF projections. Notes on the proceedings were kept during the open meetings and were helpful in preparing this report. Presentation materials and notes were made available to reviewers on a web site. Some staff who were unable to get to the second day's workshop due to weather conditions were able to participate by telephone.

Terms of Reference were provided to reviewers to identify key issues and areas to be addressed in the review process. Reviewers were able to meet in closed sessions to discuss issues and begin to form opinions addressing the Terms of Reference. While I was able to draw some preliminary conclusions in discussions with other reviewers, I felt that it was necessary to re-examine selected presentation materials to more fully justify my opinions.

The request for independent and group reports is reasonable, but time consuming. The scope of work for external reviewers includes preparation of documentary sections that are likely to be repetitive if not identical. I tried to develop my opinions independently before referring to the group report.

5. Conclusions

Raking ratio adjustment of the angler trip weights is the preferred method of calibrating prior years' estimates. This was described as method 3. The method uses an iterative procedure to adjust the distribution of angler trips for early years (2012 and earlier) to match the same distribution in the data for data generated after the implementation of new sample design and survey procedures (2013 and later). This only affects the estimation of catch per angler trip. The number and distribution of angler trips is obtained from the effort surveys and the process for calibrating early year data has already been implemented and peer reviewed.

Methods 1 and 2 depend on ratios of post- and pre-estimates of total catch for specified domains. They have intuitive appeal, but may require specialized treatment for particular estimates. They incorporate correction for selected control totals at the cell level. Method 3 works on the distribution of angler trips and should not require specialized treatment for any particular catch estimate. Method 3 controls estimates in more dimensions by working on marginal rather than cell totals.

Small domains may still have difficulty with extreme weight adjustments because of sparse data in some adjustment dimensions. Highly unequal weights may also inflate the sampling error of estimates. Some reasonable methods of checking for extremes and trimming the extreme weights when necessary should be considered.

6. References

Carmichael, J. and D. Van Voorhees (2014). MRIP Calibration Workshop II, September 8-10, 2014, North Charleston, SC, NOAA Fisheries, Office of Science and Technology, MRIP.

Deming, W. E. and F. F. Stephan (1941). "On the Interpretation of Censuses as Samples." Journal of the American Statistical Association **36**(213): 45-49.

Folsom, R. E. and A. C. Singh (2000). "The General Exponential Model for Sampling Weight Calibration for Extreme Values, Nonresponse, and Poststratification." Proceedings of the Survey Research Methods Section, American Statistical Association: 598-603.

National Research Council (2006). Review of Recreational Fisheries Survey Methods. Washington, DC, The National Academies Press.

Oh, H. L. and F. S. Scheuren (1983). Weighting Adjustment for Sampling and Nonresponse Error Reduction. Incomplete Data in Sample Surveys. W. G. Madow, I. Olkin and D. B. Rubin. New York, Academic Press: 143-184.

Potter, F. J. (1988). "Survey of procedures to control extreme sampling weights." Proceedings of the Section on Survey Research Methods, American Statistical Association: 453-458.

Research Triangle Institute (2012). SUDAAN Language Manual, Volumes 1 and 2, Release 11. Research Triangle Park, NC, Research Triangle Institute.

Appendix 1. Bibliography of Materials Provided for Review

Workshop Presentations

1. Introduction to MRIP Transition Planning and the Access Point Angler Intercept Survey Presented by Dave Van Voorhees.
2. MRIP Calibration Workshop II held September 8-10, 2014, North Charleston, SC by John Carmichael and Dave Van Voorhees, editors and presented by John Carmichael.
3. Importance of Calibrated Catch for Fisheries Presented by Jason Gidden.
4. Importance of Calibrated Catch for Fishery Stock Assessments Presented by Katie Drew.
5. Weighted Estimation for the Access Point Angler Intercept Survey by Dave Van Voorhees, F. Jay Breidt, John Foster, Han-Lin Lai, and Jean Opsomer and presented by Dave Van Voorhees.
6. MRIP: A New Design of the Access Point Angler Intercept Survey by Tom Sminkey, Lauren Dolinger Few, John Foster, and Dave Van Voorhees and presented by Tom Sminkey.
7. Calibration Workshop II by John Carmichael.
8. APAIS Calibration Methods Considered presented by Ryan Kitts-Jensen.
9. APAIS Calibration Methodology presented by Jean Opsomer.
10. APAIS Intercept Survey Calibration: Results by John Foster and Jean Opsomer and presented by John Foster.

Other Background Material

Additional background material and reports were made available on a Wiki web site before the workshop and notes on the discussions were added after the workshop.

<https://www.st.nmfs.noaa.gov/confluence/display/APAISCALIB/MRIP+APAIS+Calibration+Review>

Appendix 2. Copy of Statement of Work

Statement of Work
National Oceanic and Atmospheric Administration (NOAA)
National Marine Fisheries Service (NMFS)
Center for Independent Experts (CIE) Program
External Independent Peer Review

Calibration Model Accounting for a Recreational Fisheries Survey Design Change

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards.

(http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).

Further information on the CIE program may be obtained from www.ciereviews.org.

Scope

The Office of Science and Technology requests an independent peer review of a calibration model proposed for use in revising statistics produced by a survey of marine recreational fishing catch rates on the Atlantic coast and in the Gulf of Mexico. This calibration model is considered by the Marine Recreational Information Program (MRIP) to be very important to adjust historical time series of recreational catch estimates in order to account for biases in past sampling and estimation methods that have become apparent with the development of a new, more statistically sound method. The calibration model is intended to account for past biases in catch rate estimates for the shore, private/rental boat, and charter boat fishing modes that have resulted from the continued use of a legacy sampling design for the Access Point Angler

Intercept Survey (APAIS). A more statistically sound sampling design for the APAIS was implemented in March of 2013

Calibration Model for the APAIS Design Change

In 2014, a Calibration Workshop was held to evaluate the potential consistent effects of implementing a new sampling design for the APAIS on the Atlantic and Gulf coasts in 2013. Workshop participants included three expert statistical consultants and representatives from NOAA Fisheries, the regional fishery management councils, the interstate marine fisheries commissions, and several state agencies. The participants determined that analyses conducted by the NOAA Fisheries Office of Science and Technology showed there was sufficient evidence that the more complete temporal coverage of the new design resulted in consistent changes in APAIS angler catch rate statistics for at least some species. They developed three different calibration models to evaluate for possible use in correcting the pre-2013 APAIS statistics. The statistical consultants concluded the simplest of the three proposed models was appropriate for use in the short term until a more complete evaluation of all three calibration models could be completed using three years of new APAIS data (2013-2015). The plan was to complete that evaluation by the end of 2016, so that one method could be selected as the best for use in 2017 to revise APAIS estimates prior to 2013.

Requirements

NMFS requires three (3) reviewers to conduct an impartial and independent peer review in accordance with the SoW, OMB Guidelines, and the Terms of Reference (ToRs) below. The CIE reviewers shall have working knowledge and recent experience in the design of sampling surveys and the evaluation of non-sampling errors (i.e., undercoverage, nonresponse, and response errors) associated with changes to survey designs over time. In addition, they should have experience with complex, multi-stage sampling designs, time series analyses, regression estimators, and small domain estimation methods. Some recent knowledge and experience in current surveys of marine recreational fishing is desirable but not required.

NMFS will provide a Chair who has experience with U.S. fisheries stock assessments and their application to fisheries management. The Chair would ensure that reviewers understand the importance of maintaining a comparable time series of marine recreational fisheries catch statistics for use in stock assessments and their application to fisheries management. The Chair will not be selected by the contractor and will be responsible for facilitating the meeting, developing and finalizing a summary report and working with the CIE reviewers to make sure that the ToRs are addressed in their independent reviews.

Tasks for Reviewers

Pre-review Background Documents

The following background materials and reports prior to the review meeting include:

APAIS Design Change Calibration Workshop Report:

http://www.st.nmfs.noaa.gov/Assets/recreational/pdf/MRIPCalibrationWorkshopII_FinalReport.pdf

NC APAIS Pilot Study Report: *A Pilot Study of a New Sampling Design for the Access Point Angler Intercept Survey.*

https://www.st.nmfs.noaa.gov/pims/main/public?method=DOWNLOAD_FR_PDF&record_id=772

Report on APAIS Calibration Model:

This report will be provided by the contractor (via electronic mail or make available at an FTP site) to the CIE reviewers.

Panel Review Meeting

Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The meeting will consist of presentations by NOAA and other scientists to facilitate the review, to provide any additional information required by the reviewers, and to answer any questions from reviewers.

Contract Deliverables - Independent CIE Peer Review Reports

The CIE reviewers shall complete an independent peer review report in accordance with the requirements specified in this SoW and OMB guidelines. Each CIE reviewer shall complete the independent peer review according to required format and content as described in **Annex 1**. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in **Annex 2**.

Other Tasks – Contribution to Summary Report

The CIE reviewers may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. The CIE reviewers are not required to reach a consensus, and should provide a brief summary of each reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Foreign National Security Clearance

When reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for reviewers who are non-US citizens. For this reason, the reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: <http://deemedexports.noaa.gov/> and http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-

[national-registration-system.html](#). The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place of Performance

The place of performance shall be at the contractor's facilities, and at the NOAA Fisheries Service Headquarters in Silver Spring, Maryland.

Period of Performance

The period of performance shall be from the time of award through April 31, 2018. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Within two weeks of award	Contractor selects and confirms reviewers
Approximately 2 weeks later	Contractor provides the pre-review documents to the reviewers
March 2018	each reviewer participates and conducts an independent peer review during the panel review meeting
Within two weeks of panel review meeting	Contractor receives draft reports
Within two weeks of receiving draft reports	Contractor submits final reports to the Government

Applicable Performance Standards

The acceptance of the contract deliverables shall be based on three performance standards: (1) The reports shall be completed in accordance with the required formatting and content (2) The reports shall address each ToR as specified (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (<http://www.gsa.gov/portal/content/104790>). International travel is authorized for this contract. Travel is not to exceed \$12,000.

Restricted or Limited Use of Data

The contractors may be required to sign and adhere to a non-disclosure agreement.

NOAA Fisheries Project Contact:

Dave Van Voorhees

NOAA Fisheries

1315 East West Highway

Silver Spring, MD 20910

dave.van.voorhees@noaa.gov

Annex I: Format and Contents of CIE Independent Peer Review Report

1. The report must be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether or not the science reviewed is the best scientific information available.
2. The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each ToR, in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the ToRs.
 - a. Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.
 - b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, but especially where there were divergent views.
 - c. Reviewers should elaborate on any points raised in the summary report that they believe might require further clarification.
 - d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.
 - e. The report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The report shall represent the peer review of each ToR, and shall not simply repeat the contents of the summary report.
3. The report shall include the following appendices:
 - Appendix 1: Bibliography of materials provided for review
 - Appendix 2: A copy of this Statement of Work
 - Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

Calibration Model Accounting for Changes in Recreational Fisheries Survey Methods

1. Evaluate the suitability of the proposed approach for converting historical estimates of mean angler catch rates obtained using the old Marine Recreational Fisheries Statistics Survey (MRFSS) Access Point Angler Intercept Survey (APAIS) sampling design to estimates that best represent what would have been produced had the new MRIP APAIS sampling design been in place prior to 2013.
 - a. Does the proposed approach adequately account for consistent differences in estimates that would have been observed if the old MRFSS APAIS had been conducted side-by-side with the new MRIP APAIS in 2013-2017?
 - b. Is the proposed approach a suitable alternative to the calibration models that were originally developed in the 2014 MRIP Calibration Workshop and later evaluated by MRIP?
 - c. Is it reasonable to conclude that revised 2004-2012 APAIS estimates based on the application of the proposed approach would be more comparable than the current ones to estimates produced since 2013 under the new APAIS design?
 - d. Given the limitations of the available data, is it reasonable to apply the proposed approach to revise APAIS estimates prior to 2004 (back to 1981)?
2. Briefly describe the panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.

Appendix 3. Panel Membership

The review panel consisted of seven members including four appointed by NOAA Fisheries and three appointed by the Center for Independent Experts (CIE).

Michael D. Murphy (Chair), Florida Fish and Wildlife Commission (retired)

Paul Rago, Mid Atlantic (retired NMFS)

Carolyn Belcher, South Atlantic. Georgia Department of Natural Resources, Fishery Management Council's Scientific and Statistical Committees

Mathew Cieri, Maine Department of Marine Resources and Atlantic States Marine Fisheries Commission

Three scientists appointed by the Center for Independent Experts (CIE):

Mary Christman, MCC Stats Consulting & South Atlantic Fishery Management Council

James Chromy, Statistician, RTI International (retired)

John Whitehead, Department of Economics, Appalachian State University

Marine Recreational Information Program (MRIP)

Peer Review of the Approach Proposed to Account for the Change in Design of Access Point Angler Intercept Survey

MRIP APAIS Review Meeting
March 20-22, 2018
Sheraton Hotel
Silver Spring, MD

Paul Rago
Mid-Atlantic Fishery Management Council
Science and Statistical Committee
Falmouth, MA
April 30, 2018

Executive Summary

The Access Point Angler Intercept Point Survey (APAIS) of the Marine Recreational Information Program (MRIP) is the primary survey for estimating angler catch rates for states along the Atlantic coast and Gulf of Mexico. The APAIS has evolved and improved over time in response to changing needs and the results of external peer reviews. In 2013 the MRIP implemented a rigorous design approach that addressed many of the previously identified limitations of earlier methods. In view of the overall economic and resource management aspects of recreational fishing, it is important to have comparable estimates of recreational catches over as long a period as possible. The purpose of this review was to examine a proposed methodology for calibrating historical catches with new and improved estimates.

The proposed methodology for adjusting catches from 1981 to 2013 (wave 1) relies on a statistical method known as raking ratio adjustment, described in a seminal paper by Deming and Stefan (1940). The method adjusts the weights associated with various domains defined by state/sub-state region, year and wave, mode, area fished, household type (coastal/non-coastal) and for-hire boat membership. The MRIP team has proposed a novel application of this method to the historic APAIS data as a way of deriving estimates of catch per angler trip that retain as much of the original survey design as possible. In contrast to methods that rely on ad hoc approaches, the raking procedure allows all sample estimates to be treated the same, irrespective of the species caught. The methodology accounts for potential non-stationarity in the estimates as a further refinement. Weights are adjusted by stanzas (1981-1992, 1993-2003, 2004-2012) to account for potential decadal changes. Moreover, adjustments within stanzas are examined for trend when sequentially adjusting the sample weights.

Overall, the approach represents a significant advance in the understanding of recreational catches. The methodology is innovative but the potential magnitude of changes in some domains could be large. The full range of adjustments and their implications for stock assessments will not be known for several years. Several methods for filtering the adjustments are proposed based on potential influence statistics and control chart methods. Consideration should be given to allowing users to subset catch estimates based on objective and repeatable algorithms. MRIP may also wish to consider implementation of more formal liaison positions within and outside NOAA Fisheries as a way of following up on questions from stock assessment scientists and managers.

1.Introduction

The MRIP team has addressed an extraordinarily difficult survey problem in an innovative way that has important implications for stock assessments of many of the nation's most valuable fishery resources. An essential ingredient for the proper management of fishery resources is an accurate estimate of the number of fish removals. Long time series of consistent estimates of recreational catch help establish the appropriate scale of the population and help characterize potential productivity. MRIP, and its predecessor MRFSS, have provided recreational catch information since 1981. Over time, a variety of technical and statistical problems were identified by various internal and external reviews. This process of improvement in the recreational fishery survey led to MRIP, a major improvement in the overall methodology. Such changes are welcome but inconsistencies with historical data have increased uncertainties in stock assessments and fisheries management. Hence there is a strong impetus to develop methods that can be used "calibrate" the more recent improved data with the less rigorous older estimates. "Calibration" herein is used in a colloquial sense to mean one or more adjustments to historical MRFSS data to improve their scientific credibility and consistency with the recent data collected under MRIP.

1.1 Background Material

Background documents for the meeting were made available prior to the meeting allowing Panelists to acquaint themselves with the history of MRIP and the extensive peer reviews and pilot studies. MRIP is notable in that it has relied extensively on an external peer review process and consultations with experts outside of the Agency to obtain necessary expertise, and to bring in external consultants as necessary. Moreover, pilot studies have been implemented at critical junctures. MRIP has extensive programs of outreach for partners and has had meaningful standing committees to provide advice.

The overall focus of the meeting was structured by the Terms of Reference (Appendix 1). Pre-meeting documentation on the proposed methodology was succinct and accurate. But it did not include additional information on the performance of the two ratio estimators. This information was provided during the meeting. Although it would have been desirable to have more of this information prior to the meeting, the sheer quantity of information to review makes it a difficult task to accomplish. It's not clear if the volume of information could have been reduced for the reviewers. As it was, the Panel spent considerable amount of time exploring the many dimensions of the revised data. The Panel recognized that it was viewing only a small fraction of the complexity of the MRIP database.

A background webinar presentation by Jean Opsomer and John Foster, given earlier to the Transition team, was made available to the Panel and found to be a helpful introduction to the material to be reviewed. When possible, such webinars should be part of standard practice for Office of Science and Technology reviews. They help orient the reviewers, identify potential concerns, and give the analysts and reviewers the opportunity to focus on the salient aspects of the methodology and its implementation.

The Panel met via webinar prior to the on-site meeting to identify important concerns. This meeting allowed MRIP staff and consultants to develop presentations and analyses better tailored to the interests of the Panel. The attention to these requests was much appreciated.

1.2 Review of Activities

The Panel Report provides a detailed summary of the individual presentations and ensuing discussions. Below is a more condensed summary relative to my review.

The review Panel met for two and half days to address the Terms of Reference. The first day and a half consisted of presentations by MRIP staff, their contractors, and representatives from the Atlantic States Marine Fisheries Commission and the Mid-Atlantic Fishery Management Council. All of the presentations (See Appendix 2) were well conceived and presented. Taken together they motivated the need for the proposed revisions, summarized the history of improvements to date, and provided progressively increasing details on the methodology and its implementation. Each presentation included interactions with the Panel both during and after each talk.

On the afternoon of the second day the Panel met *in camera* to summarize its major concerns about the proposed approach and review progress towards meeting the Terms of Reference. The Panel then met with MRIP staff and webinar participants to clarify additional issues.

Throughout the meeting, MRIP staff provide excellent support in terms of providing the Panel with necessary documentation and importantly, concise, comprehensive and thorough notes by rapporteurs. Technical aspects of meeting were extremely well done with support of Jason Didden of MAFMC who managed the webinar and audio visuals, accommodated the special needs of a snow day in Washington DC and also somehow managed to give a very informative talk. The Chair did an excellent job of running the meeting and allowing adequate time for in-depth discussions.

On the third morning of the meeting, the Panel met privately to discuss remaining issues and plans for finalizing the report. The discussions focused the preservation of historical data to allow future analysts the opportunity to interrogate the data and better understand the factors affecting the revised catch estimates. It was recognized that the calibration can enhance or reduce the importance of particular domains in the historical data. However it would not create fishing activity where there had not been any in the past. This stands in contrast to the ratio methods which implicitly extrapolate estimates into domains where no activities may have been observed. For example, a method that relies on the ratio of total day length to peak hours observed inflates the observed catches per angler to cover the entire day period. The raking method affects the weights applied to observed CPUE but preserves the marginal totals.

The panel also discussed the issue of trimming the weights generated by the raking procedure. A review of the NAS 2017 Review of MRIP indicated that trimming has been done for weights generated by the FES and CHTS. Similar trimming for the APAIS data has

not been done, but the Panel considered several alternatives that might be useful. It was agreed that the basis for any trimming approach would need to be carefully considered. A key concern is that the methodology would need to be applied across the entire database because it would be infeasible to examine the results for every possible estimate by species. Moreover, the metric of performance (i.e., are the estimates necessarily “better”?) is not immediately obvious.

2. Review of Proposed Methods

2.1 Evaluation of Terms of Reference

Term of Reference 1: Evaluate the suitability of the proposed approach for converting historical estimates of mean angler catch rates obtained using the old MRFSS Access Point Angler Intercept Survey (APAIS) sampling design to estimates that best represent what would have been produced had the new MRIP APAIS sampling design been in place prior to 2013.

Estimating the abundance of fish in the ocean is a difficult if not impossible proposition. Accurately estimating the number of fish removed from the ocean by humans is almost, if not equally difficult. The central challenge posed by TOR 1 is to develop a consistent measure of removals collected by methodologies that, to some extent, lacked consistency over time.

The MRIP team developed an elegant method to adjust sample weights for angler intercepts that preserves the design elements of the original MRFSS and MRIP surveys. Unlike simpler ratio methods, this adjustment approach does not obscure changes associated with the multiple design factors such as wave, mode, state, area fished, time of day etc. Instead it naturally addresses the original design without adding poorly supported alternative model that use less of the available information and hard to justify assumptions.

It builds upon the methodologies that were developed in 2012 to approximate sampling weights based on estimated inclusion probabilities for data from 2004 to 2012. These sampling weights were applied to the historic data collected under the MRFSS protocols. The method allowed for estimators that were consistent with intentions of the original design that recognized that harvest patterns would vary by state, by method of fishing, by area fished, by day of the week, and with season.

The method applies the “raking” method of Deming and Stephan (1940) to adjust the sample weights based on marginal totals for multiple factors. The result of this approach is to produce a consistent set of adjustments across all sampling stages. Moreover the scalar adjustments are examined for trend such that ‘raking’ in each period is adjusted for potential trends.

Overall, this term of reference was met. Details on each component are provided below.

- a. *Does the proposed approach adequately account for consistent differences in estimates (mean catch rates?) that would have been observed if the old MRFSS APAIS had been conducted side-by-side with the new MRIP APAIS in 2013-2017?*

In general, yes. Because the actual side-by-side comparison could not be done, it is impossible to actually know this. The proposed methodology however, does retain the salient aspects of the sampling design. Given the complexity of the survey design, the utility of ad hoc methods that do not incorporate the such information (e.g., a post hoc ratio estimator) should be considered inferior and potentially misleading.

The sheer volume of data and estimates makes it likely that the full complexity of revisions will not be fully known for several years as the estimates are incorporated into stock assessments. It was not possible to review all species, state, wave, mode, etc. combinations. See results from Foster's presentation on various factors and their influence.

- b. *Is the proposed approach a suitable alternative to the calibration models that were originally developed in the 2014 MRIP Calibration Workshop and later evaluated by MRIP?*

Yes. The proposed approach represents a considerable improvement over the originally proposed ratio methods that relied some rather heroic assumptions about the sampling conducted during so-called "peak" hours of fishing activity (actually the end times of fishing trips) and the expected trips that occur outside the peak period. Not only did the ratio methods M1 and M2 require strong assumptions they also failed to utilize the ancillary information collected under the original sampling design. The proposed methodology appears to include this information in a natural way and apply standard statistical approaches to adjust surveys. While such methods may be standard in other survey contexts, it would appear that this is the first time such procedures have been applied in fisheries.

Among the many advantages of the proposed raking method is the fact that calibrations for all species can be done simultaneously. In contrast the M1 and M2 methods must be done in a piecewise fashion, using ratio estimation methods whose properties are poorly known. Further it is anticipated that the result of such piecewise ratio estimation and adjustments would lead to inconsistent results. For example, the sum of piecewise M1 adjusted catches for several species might differ from the application of an M1 adjustment on the sum of catches for those same species.

- c. *Is it reasonable to conclude that revised 2004-2012 APAIS estimates based on the application of the proposed approach would be more comparable than the current ones to estimates produced since 2013 under the new APAIS design?*

Yes. But this is deduced from first principles rather than from an experiment that was not conducted (and perhaps could not have been conducted).

The first revision of the 2004-12 estimates are a function of pseudo weights based on ancillary data available in the APAIS documentation for these years. The proposed approach achieves consistency of the pseudo weights by revising them according to a multi-factor raking process. Such revisions are more sound statistically than the original weightings because they incorporate apparent changes that were revealed by the 2013 to 2016 implementation of APAIS.

There will always be concerns about post hoc adjustments to sample inclusion probabilities. Such concerns are likely to increase as one extrapolates further back in time. While the proposed sequential calibration approach accounts for non stationarity to some extent, it does not account for external sources of change that may in fact underlie some of the observed changes. This should not be construed as a defect in the proposed methodology. Instead it should be considered a cautionary tale for analysts and managers using the historical data.

d. Given the limitations of the available data, is it reasonable to apply the proposed approach to revise APAIS estimates prior to 2004 (back to 1981)?

Yes. The proposed methodology is rigorous and reproducible. It utilizes available survey information to adjust cell weights. This preserves the fundamental aspects of the design and prevents the development of adjustment factors that may be based on a single realization of a catch rate series. Thus the proposed method is appealing from intuitive and statistical perspectives.

One of the important limitations is likely to be the outliers that exist in some years. These arise for a variety of reasons, but most probably due to low sampling effort in a particular cell. Such cells will have low inclusion probabilities, even when adjusted by raking. Since the weights are the inverse of the inclusion probabilities, the resulting catch estimates may be abnormally high. To protect against this possibility, it may be important to develop some form of trimming to exclude outliers. Such a procedure should be algorithmic rather than to require inspection of each case. One option would be to exclude weights based on their departure from the mean or median rather than on a strictly percentile basis. For example, a rarefaction method that say excludes values in say the top or bottom 1% may exclude valid weights that lie within some normal distributional bounds.

The utility of potential trimming methods should be evaluated although it is not clear what criteria would necessarily be an appropriate metric of success. For example, the implications of a trimming scenario for variance estimation may be appropriate for some but not all instances. A bi-square weight function (used in Lowess method) in which truncation is a function of the median and IQR may be valuable to explore. Per the NAEP survey, a value is trimmed if it exceeds the 99th percentile and the value is 10X the median. This approach may be robust in the present application but it would need to be studied for its utility for the APAIS dataset.

It is difficult to determine a priori if the reweighting of a particular cell is important. A high weight or large change in weight may not have much effect on catch per trip.

In the recommendations section under TOR 2 I offer some suggestions about computing an influence statistic that could be used as a diagnostic tool.

Term of Reference 2: *Briefly describe the panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.*

Proceedings

The proceedings of the meeting were described above. Overall the MRIP staff did an outstanding job of ensuring that the meeting ran smoothly and that all staff were present to answer questions, to distribute reference information, to conduct analyses when necessary, and to serve as rapporteurs. Their dedication to duty was impressive given the inclement weather (snow day) that shut the Federal and local governments down in the Washington DC area during the review meeting. Dave Van Voorhees deserves credit for his leadership. The excellent support of Jason Didden of the MAFMC ensured that the meeting was available to the public, and recorded.

Utility of Presentations

All of the presentations were lucid and helpful. The schematic developed by John Foster was exceptionally helpful in describing the raking procedure and its application to the historic APAIS data.

Terms of Reference

TOR 1.c was not entirely clear and its ambiguity made it difficult for the Panel to answer. The wording of future TOR might include—Is the proposed method statistically sound? Are there other methods that might have been used?

One of the significant challenges is the sheer size of the underlying database and its inherent complexity. It does not easily lend itself to a simple working group environment since analyses may take a while to set up and validate and even longer periods to review and interpret results.

Documentation of Meeting

A primary document by Opsomer and Foster describing the proposed methodology was sent to the Panel about 10 days before the meeting. Supporting documents from the literature accompanied the report. In addition, a wiki site with a summary of previous reports, reviewer summaries, and background documents was prepared for the Panel by S&T staff. These were comprehensive and allowed the Panel to view the evolution of the science underlying the APAIS.

Ancillary Analyses

The MRIP staff conducted a number of analyses requested by the Panel prior to the meeting. These results proved valuable to the Panel in terms of showing key results and revealing the full complexity of the calibration task. Similarly, the MRIP staff prepared additional material for the Panel's consideration during the meeting itself. These efforts greatly improved the ability of the Panel to review the methodology and were much appreciated.

Improvements to Future Peer Review Process

The current Transition Team process that works so well for communications among all stakeholders might serve as a template for improving technical peer reviews. Agency staff and partners with strong statistical backgrounds could serve members of working groups that field test ideas and approaches, and identify potential problems before full scale implementation.

Data Availability

It is admirable that the MRIP intends to fully archive the historic "micro" data and thus allow future investigators the opportunity to interrogate the data and potentially devise alternative approaches. Training sessions, perhaps via webinar, would be a useful way of guiding initial data investigations. The feedback would be useful for clients who would avoid inappropriate analyses (i.e., analyses inconsistent with the underlying design) and for MRIP staff who could obtain insights into key concerns of users.

One way to guide users would be to show a completely worked example of the estimation for a particular species. This could show the various intermediate quantities for effort and CPUE and the effects of the various adjustments on both the raw data, means, and the variances. Perhaps a single year would suffice. Such a webinar/training document would give a dedicated user some key insights and allow for true understanding.

Overall it is highly desirable to finalize the process and resist the temptation to continuously revise the data. Stability in the estimates is desirable unless it can be conclusively shown that significant improvements in the catch estimates could be obtained by further analyses or adjustment.

Filtering of Estimates

It may be useful to allow user control on the specification of outliers resulting from the raking calibration. The following section on Influence Statistics provides some thoughts on how such filtering could be accomplished. As an analogy, users of the Northeast Fisheries Science Center bottom trawl survey can develop estimates of relative abundance conditioned on selection of one or more factors related to tow quality and survey design. Users are free to examine the performance of various estimates based on alternative subsets of the data. In the APAIS/MRIP context, one may wish to examine the potential effects of outliers on catch estimates. As noted in sections below analysts will need to be especially mindful of revised estimates that seem unreasonably high or low, and to be responsive to reviewers and stakeholders asking detailed questions.

This feature would give users a degree of control, it would identify explicitly what was done, and it would be repeatable by other investigators. It might also allow for explicit sensitivity analyses in stock assessments. Giving the user community an opportunity to filter the data according to some user-specified criteria would be valuable. For example, one might allow the user to restrict estimates to a subset of filtered weights, such as median \pm n IQR where n is user specified. This basic idea is examined further in the following section.

Recommendations on an Influence Statistic

As noted in TOR 1d, further investigation of the effects of reweighting may be useful. Here I describe potential Influence or Leverage statistics that may be useful for evaluating the effect of reweighting observations based on the revised weights resulting from the raking process. The influence of a particular weight can be considered in two ways. The first would be the relationship between weight and the sum of all weights for a particular domain. This approach has utility for filtering major changes in weighting independent of its consequences for CPUE.

Let w_i represent the original weight for cell i prior to adjustment by raking and w_i^* represent the adjusted weight. Let Θ define the set of factors that define a given domain (eg. Mode, Wave, State, etc). The influence of a particular cell i could be gauged by its contribution to the sum of all weights such that

$$L_W(i) = \frac{w_i}{\sum_{\Theta} w_i} \quad (1)$$

For the original weights. The influence statistic for the revised weights would be

$$L_W^*(i) = \frac{w_i^*}{\sum_{\Theta} w_i^*} \quad (2)$$

The original and revised leverage statistics could be considered as alternative measures of the same quantity and one would expect them to not vary greatly.

Let $X=L_W(i)$ and $Y=L_W^*(i)$. By general theory, if X and Y are random variables with equal variances, then the correlation between $X-Y$ and $(X+Y)/2$ is expected to be zero. A plot of $X-Y$ vs $(X+Y)/2$ should reveal departures from this assumption, including trends in the weighting factor difference or variability as a function of the mean. Bland and Altman (1986, 1995, 1999) proposed some useful methods for evaluating such relationships. Their work is widely used in the medical literature and may be a useful way of reviewing and screening the weights developed by raking.

The weights per se may not be of particular interest to users. Instead, the effects of the revised weights on estimated average catch per effort may be of greater importance to users. To examine this quantity one could express the influence of a particular weight on the overall mean as follows.

Using the basic notation of Papacostas and Foster (2017, page 20, section 2.3.2), the mean catch per angler trip for a given domain can be written as

$$\bar{y} = \frac{\sum_{\Theta} w_i y_i}{\sum_{\Theta} w_i} \quad (3)$$

The reweighted mean is expressed as

$$\bar{y}^* = \frac{w_i^* y_i}{\sum_{\Theta} w_i^*} \quad (4)$$

The leverage or influence of the original weight w_i can be expressed as

$$L_{\bar{y}}(i) = \frac{w_i y_i}{\sum_{\Theta} w_i} \left(\frac{1}{\bar{y}} \right) \quad (5)$$

The leverage or influence of the revised weight would be

$$L_{\bar{y}^*}^*(i) = \frac{w_i^* y_i}{\sum_{\Theta} w_i^*} \left(\frac{1}{\bar{y}^*} \right) \quad (6)$$

One could treat the influence statistics in Eq. 5 and 6 using the Bland-Altman method described above or one could simply examine

$$\delta_i = L_{\bar{y}}(i) - L_{\bar{y}^*}^*(i)$$

This statistic would be helpful for understanding the impacts of reweighting for a particular species or stock. It seems sensible to conclude that revised weights that alter the importance of an observation significantly would be a cause for concern or further investigation. Standard methods for control charts (eg Shewhart) might be useful for rapidly interrogating the data and devising composite metrics of the effects of reweighting on the catch per angler trip estimates.

All of the above methods should be considered exploratory in nature. Although I have some familiarity with the underlying data, I have ignored many of the details in this conceptual approach. However, I do think there will be a high demand for additional analyses once the catch data are revised. My experience in stock assessments is that when our perception of the state of nature changes dramatically, a lot of time will be spent answering questions. The above methods may be useful starting points.

Future Demands on MRIP Staff

As noted in a comment above, the full complexity of the adjustments resulting from APAIS changes will not be realized for several years. The complexity and potential inconsistencies with earlier perceptions of removals will not be revealed until the data are used in stock assessments. Historical trends in abundance, resource productivity and biological reference points are all likely to change. Across species, such changes will range from trivial to tectonic, but it should be anticipated that detailed questions will be addressed to MRIP staff. From my experience in stock assessments, it will not be sufficient to rely on multiple external reviews, the statistical validity of the design, or extensive quality assurance controls. The MRIP should expect a significant demand of staff time in the next few years.

One way of meeting this demand would be to assign one or more liaisons to various Councils, NMFS Regional Offices and Science Centers, and perhaps state agencies. The liaison would be responsible for doing detailed investigations of inquiries for various clients. Developing strong working relationships would go a long way towards creating acceptance for the new methodology and possibly uncovering errors in historical data.

The ideal liaison would strong statistical skills, a deep understanding of the MRIP database, and knowledge of the stock assessment process. The collaboration would be beneficial for both MRIP staff, the stock assessment community, managers, and the public at large. Survey sampling is often misunderstood, even by primary users. Users often have asymmetric interests; they focus more on the unexpected large estimates rather than the unexpected zero values. Formalizing a strong analytical partnership between MRIP and the primary analysts and managers could help shape future assessments and management plans towards more appropriate temporal and spatial scales consistent with the underlying data.

- There may be a temptation to view such an exercise as a diversion of resources from more critical issues or demands. Funding such positions poses obvious challenges as well. One solution might be to establish agreements with the various clients to jointly fund such a program for a limited number of years. Innovative ways of exchanging talent across the agency would be perceived as a way of addressing a critical future need.
- If the concept of on-site liaisons proved to be too costly or impractical, consideration should be given to development of a training course for the data analyses. This would have a minimal duration of one week and possibly be set up as one or more courses.

3. Bibliography

The OST staff provided a wealth of background material for review by the Panel via the Confluence Website. The papers and links to existing websites provide concise summaries of previous reviews and ongoing programs to implement and document changes in the MRIP. The links to this material is summarized in the Panel's Review Report for APAIS.

References cited in my review are provided below.

*Bland JM, Altman DG (1986). ["Statistical methods for assessing agreement between two methods of clinical measurement"](#) (PDF). *Lancet*. **327** (8476): 307–10.*

Bland JM and DG Altman. 1995. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet*. 346: 1085-1087

*Bland JM, Altman DG (1999). ["Measuring agreement in method comparison studies"](#). *Statistical Methods in Medical Research*. **8** (2): 135–60.*

Deming, W.E. and F.F. Stephan. 1940. On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known. *The Annals of Mathematical Statistics*. 11: 427-444

Papacostas, K. J. and J. Foster. 2017. Survey Design and Statistical Methods for estimation of Recreational Fisheries Catch and Effort. Version 1. Draft. National Marine Fisheries Service, Marine Recreational Information Program.

Smith, S.J. 1996. Bootstrap confidence limits for groundfish trawl survey estimates of mean abundance. *Can. J. Fish. Aquat. Sci.* 54: 616–630.

Smith, S. J. 1996. Analysis of data from bottom trawl surveys. *NAFO Sci Council Studies.* 28:25-53.

Appendix 1. Terms of Reference

Draft Terms of Reference for Peer Review of the Approach Proposed to Account for the Change in Design of the Access Point Angler Intercept Survey

Appendix 1. Terms of Reference for the MRIP APAIS Peer Review Workshop

The Review Panel shall assess whether or not the MRIP Working Group has reasonably and satisfactorily completed the following actions.

1. Evaluate the suitability of the proposed approach for converting historical estimates of mean angler catch rates obtained using the old MRFSS Access Point Angler Intercept Survey (APAIS) sampling design to estimates that best represent what would have been produced had the new MRIP APAIS sampling design been in place prior to 2013.
 - a. Does the proposed approach adequately account for consistent differences in estimates that would have been observed if the old MRFSS APAIS had been conducted side-by-side with the new MRIP APAIS in 2013-2017?
 - b. Is the proposed approach a suitable alternative to the calibration models that were originally developed in the 2014 MRIP Calibration Workshop and later evaluated by MRIP?
 - c. Is it reasonable to conclude that revised 2004-2012 APAIS estimates based on the application of the proposed approach would be more comparable than the current ones to estimates produced since 2013 under the new APAIS design?
 - d. Given the limitations of the available data, is it reasonable to apply the proposed approach to revise APAIS estimates prior to 2004 (back to 1981)?
2. Briefly describe the panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.

Appendix 2. Draft Review Meeting Agenda

MRIP APAIS Calibration Peer Review Workshop Silver Spring, MD March 20-22, 2018

Agenda

Tuesday, March 20:

OPEN SESSION

- 9:00am – Welcome, Introductions, and Overview of Workshop
Dave Van Voorhees, NOAA Fisheries Office of Science and Technology
Mike Murphy, Chair of Peer Review Panel
- 9:20am – Presentation: MRIP Transition Planning and the Access Point Angler Intercept Survey
Dave Van Voorhees, NOAA Fisheries Office of Science and Technology
- 9:40am – Presentation: Importance of Calibrated Catch for Fisheries Management
Jason Didden, Mid-Atlantic Fishery Management Council
- 10:00am – Presentation: Importance of Calibrated Catch for Fishery Stock Assessments
Katie Drew, Atlantic States Marine Fisheries Commission
- 10:20am – BREAK
- 10:30am – Presentation: Weighted Estimation for the APAIS (and Calibration Workshop I)
Dave Van Voorhees, NOAA Fisheries Office of Science and Technology
- 11:05am – Presentation: New Design of the APAIS
Tom Sminkey, NOAA Fisheries Office of Science and Technology
- 11:40am – Presentation: Calibration Workshop II
John Carmichael, South Atlantic Fishery Management Council & SEDAR
- 12:00pm – LUNCH BREAK
- 1:10pm – Presentation: Considered Ratio Calibration Methods
Ryan Kitts-Jensen, NOAA Fisheries Office of Science and Technology
- 1:50pm – Presentation: APAIS Calibration Methodology
Jean Opsomer, Westat and Colorado State University
- 2:40pm – Presentation: APAIS Calibration Results
John Foster, NOAA Fisheries Office of Science and Technology
- 3:30pm – BREAK
- 3:40pm – Follow-Up Questions for Presenters
- 4:20pm – Public Comment
- 4:50pm – Summary of Day 1
Mike Murphy, Chair of Peer Review Panel
- 5:00pm –ADJOURN

CLOSED SESSION

- 5:10pm – Review Panel Coordination and Writing
- 6:00pm – ADJOURN

Wednesday, March 21:

OPEN SESSION

- 9:00am – Overview of Day 1 and Preview of Day 2
Mike Murphy, Chair of Peer Review Panel

9:10am – Follow-Up Questions for Presenters
10:15am – BREAK
10:30am – Follow-Up Questions for Presenters
12:00pm – LUNCH BREAK

CLOSED SESSION

1:00pm – Review Panel Coordination and Writing

OPEN SESSION

2:30pm – Initial Summary Findings of Review Panel
3:30pm – BREAK
3:45pm – Public Comment
4:15pm – ADJOURN

CLOSED SESSION

4:30pm – Review Panel Coordination and Writing
6:00pm - ADJOURN

Thursday, March 22

CLOSED SESSION

9:00am – Review Panel Coordination and Writing
12:30pm – ADJOURN

Appendix 3. Contract

Statement of Work

National Oceanic and Atmospheric Administration (NOAA)

National Marine Fisheries Service (NMFS)

External Independent Peer Review 2/28/18

Calibration Model Accounting for a Recreational Fisheries Survey Design Change

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards.

(http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).

Further information on the CIE program may be obtained from www.ciereviews.org.

Scope

The Office of Science and Technology requests an independent peer review of a calibration model proposed for use in revising statistics produced by a survey of marine recreational fishing catch rates on the Atlantic coast and in the Gulf of Mexico. This calibration model is considered by the Marine Recreational Information Program (MRIP) to be very important to adjust historical time series of recreational catch estimates in order to account for biases in past sampling and estimation methods that have become apparent with the development of a new,

more statistically sound method. The calibration model is intended to account for past biases in catch rate estimates for the shore, private/rental boat, and charter boat fishing modes that have resulted from the continued use of a legacy sampling design for the Access Point Angler Intercept Survey (APAIS). A more statistically sound sampling design for the APAIS was implemented in March of 2013

Calibration Model for the APAIS Design Change

In 2014, a Calibration Workshop was held to evaluate the potential consistent effects of implementing a new sampling design for the APAIS on the Atlantic and Gulf coasts in 2013. Workshop participants included three expert statistical consultants and representatives from NOAA Fisheries, the regional fishery management councils, the interstate marine fisheries commissions, and several state agencies. The participants determined that analyses conducted by the NOAA Fisheries Office of Science and Technology showed there was sufficient evidence that the more complete temporal coverage of the new design resulted in consistent changes in APAIS angler catch rate statistics for at least some species. They developed three different calibration models to evaluate for possible use in correcting the pre-2013 APAIS statistics. The statistical consultants concluded the simplest of the three proposed models was appropriate for use in the short term until a more complete evaluation of all three calibration models could be completed using three years of new APAIS data (2013-2015). The plan was to complete that evaluation by the end of 2016, so that one method could be selected as the best for use in 2017 to revise APAIS estimates prior to 2013.

Requirements

NMFS requires three (3) reviewers to conduct an impartial and independent peer review in accordance with the SoW, OMB Guidelines, and the Terms of Reference (ToRs) below. The reviewers shall have working knowledge and recent experience in the design of sampling surveys and the evaluation of non-sampling errors (i.e., undercoverage, nonresponse, and response errors) associated with changes to survey designs over time. In addition, they should have experience with complex, multi-stage sampling designs, time series analyses, regression estimators, and small domain estimation methods. Some recent knowledge and experience in current surveys of marine recreational fishing is desirable but not required.

NMFS will provide a Chair who has experience with U.S. fisheries stock assessments and their application to fisheries management. The Chair would ensure that reviewers understand the importance of maintaining a comparable time series of marine recreational fisheries catch statistics for use in stock assessments and their application to fisheries management. The Chair will be responsible for facilitating the meeting, developing and finalizing a summary report and

working with the reviewers to make sure that the ToRs are addressed in their independent reviews.

Tasks for Reviewers

Pre-review Background Documents

The following background materials and reports prior to the review meeting include:

APAIS Design Change Calibration Workshop Report:

http://www.st.nmfs.noaa.gov/Assets/recreational/pdf/MRIPCalibrationWorkshopII_FinalReport.pdf

NC APAIS Pilot Study Report: *A Pilot Study of a New Sampling Design for the Access Point Angler Intercept Survey.*

https://www.st.nmfs.noaa.gov/pims/main/public?method=DOWNLOAD_FR_PDF&record_id=772

Report on APAIS Calibration Model:

This report will be provided by the contractor (via electronic mail or make available at an FTP site) to the CIE reviewers.

Panel Review Meeting

Each reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. Each reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The meeting will consist of presentations by NOAA and other scientists to facilitate the review, to provide any additional information required by the reviewers, and to answer any questions from reviewers.

Contract Deliverables - Independent CIE Peer Review Reports

The reviewers shall complete an independent peer review report in accordance with the requirements specified in this SoW and OMB guidelines. Each reviewer shall complete the independent peer review according to required format and content as described in **Annex 1**.

Each reviewer shall complete the independent peer review addressing each ToR as described in **Annex 2**.

Other Tasks – Contribution to Summary Report

The reviewers may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. The reviewers are not required to reach a consensus, and should provide a brief summary of each reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Place of Performance

The place of performance shall be at the contractor's facilities, and at the NOAA Fisheries Service Headquarters in Silver Spring, Maryland.

Period of Performance

The period of performance shall be from the time of award through April 30, 2018. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

Applicable Performance Standards

The acceptance of the contract deliverables shall be based on two performance standards:

- (1) The reports shall be completed in accordance with the required formatting and content
- (2) The reports shall address each ToR as specified

Travel

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (<http://www.gsa.gov/portal/content/104790>). Reviewers will be required to travel to Silver Spring, MD to attend and participate in a five business day peer review.

Restricted or Limited Use of Data

The contractors may be required to sign and adhere to a non-disclosure agreement.

NOAA Fisheries Project Contact:

Dave Van Voorhees

NOAA Fisheries

1315 East West Highway

Silver Spring, MD 20910

dave.van.voorhees@noaa.gov

MISC

Jean-Claude Deville, Carl-Erik Sarndal and Olivier Sautory. 1993. **Generalized Raking Procedures in Survey Sampling**. *Journal of the American Statistical Association* Vol. 88, No. 423, pp. 1013-1020

[https://bookdown.org/jespasareig/Book How to weight a survey/](https://bookdown.org/jespasareig/Book%20How%20to%20weight%20a%20survey/)

**Center for Independent Experts (CIE) Program
External Independent Peer Review**

*Independent Peer Review of the Calibration Model Accounting for a
Recreational Fisheries Survey Design Change*

Dr. John Whitehead

May 2018

Executive Summary

The proposed weighting approach (raking method) is conceptually superior to the two ratio approaches. It uses all available information in developing weights. The ratio methods used only a limited amount of information. The proposed weighting approach is most likely to be judged the best available science. Another benefit of the proposed weighting approach is that it would preserve the micro time-series data back to 1981.

A description of the individual reviewer's role:

As the lone economist on the review panel, my comparative advantage during the review was as a user of the MRFSS/MRIP data to conduct economic analysis. During the presentations, it became apparent to me that my statistical knowledge was far inferior to some of the other panel members. I focused almost all of my attention on the implications of the various weighting procedures on the two major areas where the MRIP data is used in economics: recreation demand modelling and economic impact analysis.

A brief summary of the findings, science, conclusions and recommendations:

The panel agreed that the proposed weighting approach (raking method) was conceptually superior to the two ratio approaches. The proposed weighting approach uses all available information in developing weights while the ratio methods used a limited amount of information. The panel raised concerns that estimates for some infrequently targeted species may be very sensitive to the weighting adjustment approach taken. The panel concluded that the proposed weighting approach is preferred over the ratio-based approaches. The panel felt that the proposed weighting approach is most likely to be judged the best available science. Another benefit is that the proposed weighting approach would preserve the micro data time-series back to 1981. The panel recommended that the proposed weighing approach should be implemented in an effort to develop a consistent data time-series back to 1981. Further, the panel recommended that an investigation be conducted to determine the extent of extreme changes in weighted and unweighted data.

Answers to the ToR questions:

1. Evaluate the suitability of the proposed approach for converting historical estimates of mean angler catch rates obtained using the old MRFSS Access Point Angler Intercept Survey (APAIS) sampling design to estimates that best represent what would have been produced had the new MRIP APAIS sampling design been in place prior to 2013.

a. Does the proposed approach adequately account for consistent differences in estimates that would have been observed if the old MRFSS APAIS had been conducted side-by-side with the new MRIP APAIS in 2013-2017?

Without a side-by-side comparison study, it is difficult to make a determination about the differences that would have been observed in an actual comparison study. The proposed weighting approach (the new raking method) makes a serious effort at adequately accounting

for differences. It employs a modelling approach using several important stratification variables used under the new MRIP sampling design. The method is developed so that the distribution of angler trips over state, year, and wave can be backcasted to 2004. Some evidence of the accuracy of the backcast, relative to estimates made with APAIS pseudo-weights, was presented. The panel was provided with a long series of graphical comparisons of weighted estimates of trips, landings and releases. A single case study was presented for Alabama private boat red snapper landings. The proposed raking approach generated an estimate of landings 29% larger than landings produced with the APAIS pseudo-weights. The confidence intervals for these two estimates overlap, indicating that there are no statistically significant differences between the approaches. The simple and complex ratio methods produced landings estimates 2% and 19% larger than the estimate with the APAIS pseudo-weights. Both of these point estimates appear to be within the confidence interval of the landings estimate from the proposed raking weight method. There do not appear to be systematic differences in the variations in estimated landings using the pseudo-weights (“current landings”) and landings estimated with the weights developed by the proposed raking method (“weight adjusted landings”).

b. Is the proposed approach a suitable alternative to the calibration models that were originally developed in the 2014 MRIP Calibration Workshop and later evaluated by MRIP?

The proposed raking weight approach uses most all available information and is an improvement over the simple ratio weights. One reason for the raking procedure to be preferred is that the simple ratios do not account for the differences between coastal and non-coastal anglers. The old MRFSS assumed that catch rates by non-coastal anglers were the same as catch rates from coastal anglers (Total catch = total trips by coastal county residents*mean catch per angler fishing trip*(1/proportion of trips by coastal county residents)). This is a questionable assumption since fishing households will make location decisions to minimize the costs of fishing. In other words, they will tend to move closer to their preferred fishing sites at the coast. Because they face lower costs, coastal anglers take more fishing trips and, as a result, become more expert at catching fish. It is likely that coastal anglers have greater catch per unit effort than non-coastal anglers. This characteristic does not seem to be captured in the ratio weighting approaches. In this way the raking procedure may address one component of avidity bias (that intercepted anglers are more expert), but the problem of avidity bias was not a topic covered during the review.

c. Is it reasonable to conclude that revised 2004-2012 APAIS estimates based on the application of the proposed approach would be more comparable than the current ones to estimates produced since 2013 under the new APAIS design?

The proposed approach has been applied to the 2004-2012 MRIP catch estimates. This has been compared to the 2004-2012 MRIP catch estimates adjusted by the pseudo-weights. This is a difficult question to answer since there is no data available to conduct a validity study. Accuracy of each time-series is left to the judgement of the reviewer. The proposed approach seems to be preferred due to its rigor and the vast amounts of information used in developing heterogeneous weights. There are differences in the point estimates of catch, but the

confidence intervals are wide for both estimates and overlapping. The review panel could recommend that users of the data and decision makers employ estimates from both methods as inputs in stock assessments, allocation and other analyses. To the extent that these analyses are insensitive to the two inputs, the question is moot. Where the analyses are sensitive to use of either the pseudo-weights or calibrated weights (the proposed methods) would provide useful information about which method is preferred for the 2004-2012 period.

- c. Given the limitations of the available data, is it reasonable to apply the proposed approach to revise AP AIS estimates prior to 2004 (back to 1981)?*

Developing weights back to 1981 differs from developing weights back to 2004 because there are no starting weights from 1981 to 2003. It should be expected that the proposed raking method will be less accurate for this time period than the 2004-2013 time period. That said, the Alabama private boat red snapper comparison showed that “adjusted landings” (i.e., adjusted by weights from the proposed raking method) are 29% greater than “current landings” but the confidence intervals are wide and overlap. While it is comforting that confidence intervals overlap, this does not mean that either estimate is correct. What is missing is a validity study that compares the various estimates to a benchmark true estimate. Without that sort of study, which is not feasible in this case, there is little to inform a determinant of accuracy of these estimates.

It was made clear that there is a need to develop weights back to 1981 for stock assessment purposes. It is not clear if the new raking method would produce weights that are superior to M1 and M2 for stock assessment. A study that investigates the sensitivity of a stock assessment to “current landings” and landings produced with M1, M2 and the proposed raking approach would be informative here. If the ratio methods and raking method produce similar stock assessments, then the stock assessment benefit of producing raking weights from the proposed approach back to 1981 is low. If the different methods produce different stock assessments, then the weighting decision is important and has economic implications. A study of this sort has not been conducted, so it is difficult to assess the benefit of the raking method in this context. A study of this sort would be informative if the decision is made to proceed with the proposed raking approach.

Economic analysis with the MRIP data is of two sorts. Second, the NMFS (and others) conducts an expenditure survey to estimate the economic importance of the recreational sector to the U.S. economy. Angler expenditure estimates are combined with effort estimates from the MRIP to estimate total expenditures. The second type of analysis is recreation demand, where the economic value of catch is estimated. Economists assess the effects of fishing costs and benefits (catch) on fishing behavior. These studies can be used for benefit-cost analysis of various management tools (e.g., bag limits), bioeconomic modelling and sector allocation (where the marginal value of commercial and recreational catch should be equalized in order to maximize the economic value of a sector allocation). Both types of studies (expenditure and demand) are typically ex-ante, i.e., forward looking, with a management purpose. Expenditure studies can be used to conduct an economic impact analysis to assess how a fishery management alternative may affect regional spending and jobs. Demand studies can be used to conduct a benefit-cost analysis, allocation or bioeconomic analysis of management

alternatives. Ex-post analysis is important, but this is rarely done in fisheries economics and we might assume that ex-post analysis has limited value (the USEPA has conducted several ex-post benefit-cost analyses of the Clean Air Act).

In the recreation economics literature, the MRFSS suffered from two types of bias: endogenous stratification (The National Academy review calls “coverage bias”, but recreation demand economists call it endogenous stratification since Shaw (1988)) and avidity bias. The MRFSS sampled fishing sites with stratification rules and interviewer effort was biased in favor of hot spots and times. While the MRFSS was a stratified sample with some convenience sample characteristics, it was treated as a random sample for estimation of catch. This problem was addressed by the NAS in 2006 leading to the APAIS survey. The APAIS survey samples sites and times of day focusing on accurate coverage instead of meeting quotas. The sampling of sites and times may not reflect differences in the demand for fishing. Assumptions that non-chosen sites and times are the same as chosen sites and times are wrong, so that sampling weights are needed to correct for the endogenous stratification.

Avidity bias results because anglers who are more likely to be at an interview site tend to fish more days increasing their chance of being interviewed. These anglers are also more likely to be better at catching fish. So, anglers who fish more and catch more fish are more likely to be represented in on-site surveys. The MRFSS intercepted anglers at their chosen site and time of day. The process of choosing an angler for the sample is related to the angler behavior behind that choice. Anglers tend to choose sites with lower travel costs and higher catch per unit effort. Thomson (1991) finds that avidity bias leads to inflated estimates of effort and catch in intercept (and other) surveys. More avid anglers are more likely to be interviewed and have higher catch. Corrections for avidity bias will decrease both. It is not clear how the APAIS addresses avidity bias. But, the new Fishing Effort Survey (FES) is designed to better estimate overall effort and should address avidity bias.

There is limited evidence that weighting of recreation demand functions to correct endogenous stratification with MRFSS/MRIP data is important for economic analysis. Hindsley, Landry and Gentner (2011) use the MRFSS data and develop their own weights from the intercept and coastal household telephone survey for 2003-04 private boat trips in the South Atlantic and Gulf of Mexico. They find that estimates of the economic value of recreational catch are biased upwards unless weighting for endogenous stratification and avidity bias. The new APAIS survey design was developed to address endogenous stratification. Lovell and Carter (2014) estimate recreation demand models with and without APAIS pseudo-weights for Gulf of Mexico trips in 2009. They find that there are differences in recreation demand parameter estimates, but the estimates of the value of snapper and grouper catch is not statistically different between weighted and unweighted models. It is not clear whether the results of Hindsley, Landry and Gentner (2011) is due to the weighting for endogenous stratification or avidity bias.

Finally, one concern with all the approaches is that the weights are being developed in one regulatory regime and applied to behavior in another. Bag limits, size limits and seasons affect fishing behavior. Other economic shifts also affect behavior. These issues will make the 1981 to 2013 weights less accurate than if this information was incorporated.

Backcasting weights would allow for the conduct of a number of historical studies and ex-post analysis of fishery management decisions. For example, weighted data back to 1981 would allow for a better understanding of fishing behavior under differing regulatory regimes, economic conditions, biological stock and climate change (if time-series of these variables are available). Note also that recreation demand models with the MRFSS/MRIP have relied on economic add-on surveys to collect information necessary to more accurately estimate the cost of a fishing trip (i.e., angler income is necessary to estimate the opportunity cost of travel time). The first add-on survey was conducted in 1994 so that recreation demand analysis with these older data is limited, but possible if inaccuracies in the cost of a fishing trip are acceptable, as in many historical studies. While these potential ex-post studies are interesting and valuable in an academic sense, there may be little value of these economic analyses to current fishery management issues. In that context, the weighting approach most appropriate for the 1981-2003 data is more of a stock assessment issue.

2. *Briefly describe the panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.*

The review panel heard a number of presentations. The morning session was focused on background material, setting the stage for presentation of the proposed weighting approach (Van Vorhees). First, there was background on MRIP transition and the APAIS. This presentation provided details of survey changes and the corresponding changes to the fishing effort survey. Next the panel heard presentations on the economic (Didden) and stock assessment (Drew) importance of a calibrated landings time-series for fisheries management. The panel then heard a presentation describing weighted estimation of the APAIS and its review in the first calibration workshop (Van Voorhees). This presentation included a description of problems with the old APAIS and the information available from old APAIS that is necessary to conduct the backcasting. The proposed weighting method was presented. An initial comparison of the MRIP and MRFSS estimates was made with the MRFSS estimates within the confidence interval of MRIP. Then, there was a presentation describing the new APAIS survey (Sminkey). This highlighted the improvements made in the APAIS from 2004 and 2013. The second calibration workshop was described with discussion of initial suggestion of the two ratio weighting methods and a third approach that was not implemented (Carmichael).

Following lunch, the panel heard presentations on the three weighting approaches. First, the ratio calibration weighting approach was presented (Kitts-Jensen). The ratio approaches are developed at the subregion, state, mode, wave and species. The simple ratio approach (M1) always results in a systematic increase in catch estimates because it is directed at accounting for temporal coverage bias (e.g., the MRFSS sampling did not intercept trips that ended late in the fishing day). The limitations are that the approach doesn't make use of all available MRIP data and it is difficult to develop weights for the microdata. The complex ratio approach (M2) leads to less of a difference between MRFSS and MRIP, and produces increases and decreases in catch estimates. The proposed APAIS calibration method (Opsomer) and results (Foster) was presented. While the ratio approaches can lead to constant differences across species, the proposed approach leads to heterogeneous differences across species.

The panel had ample opportunity for questions, but few were asked. The panel chose to have little discussion time until the presentations were completed at lunch on the second day. While the presentations lasted until lunch on the second day, a full morning past the agenda, the panel had plenty of time for discussion. The panel met in closed session in the afternoon of the second day and discussed various issues. Presenters returned for a question and answer session and the panel adjourned for the second day. The panel met again during the morning of the third day and reached a general consensus.

References

Hindsley, Paul, Craig E. Landry, and Brad Gentner. "Addressing onsite sampling in recreation site choice models." *Journal of Environmental Economics and Management* 62, no. 1 (2011): 95-110.

Lovell, Sabrina J., and David W. Carter. "The use of sampling weights in regression models of recreational fishing-site choices." *Fishery Bulletin* 112, no. 4 (2014): 243-253.

Shaw, Daigee. "On-site samples' regression: Problems of non-negative integers, truncation, and endogenous stratification." *Journal of Econometrics* 37, no. 2 (1988): 211-223.

Thomson, C.J. 1991. "Effects of the avidity bias survey estimates of fishing effort and economic value." In: *Creel and angler surveys in fisheries management*, D. Guthrie et al., eds. American Fisheries Society Symposium, 12: 356-366.

Elaboration on any points raised in the summary report that they believe might require further clarification:

When considering developing the weights back to 1981, the summary report mentions the absence and impossibility of a validity study. While this is true, it would be possible to conduct a comparative study of the different weighting approaches on the older data to determine the sensitivity of catch estimates to the different weighting approaches. Further, it would be interesting to conduct sensitivity analysis on the factors using in the ranking method, which is quite complex. Would similar weights be obtained with less researcher effort?

As mentioned above, it is not clear if there will be any demand amongst economists for MRFSS post-weighted micro data from 1981 to 2004.

A critique of the NMFS review process, including suggestions for improvements of both process and products:

In a pre-review conference call, the Panel asked for an illustration of results from the recommended weighting approach. After what appears to be a herculean effort, John Foster gave an hours long presentation presenting some of these results. On the reviewers' end, it was difficult to fully comprehend and digest these results without a briefing document.

The panel had an opportunity to read background material for all the other presentations. While I understand that developing a written document in such a short time period was not possible given time constraints, presenting a summary of the results along with the Foster, Breidt and Opsomer (March 11, 2018) methods paper would have been very helpful to the panel.

Appendix 1. Bibliography of Materials

F. Jay Breidt, Han-Lin Lai, Jean D. Opsomer, and David A. Van Voorhees, A Report of the MRIP Sampling and Estimation Project: Improved Estimation Methods for the Access Point Angler Intercept Survey Component of the Marine Recreational Fishery Statistics Survey (+3 review reports and response).

F. Jay Breidt et al., A pilot Study of a New Sampling Design for the Access Point Angle rIntercept Survey, Submitted by the MRIP Design and Analysis Workgroup: (+ peer reviews and response).

John Boreman, Consultant's Report: Summary of the MRFSS/MRIP Calibration Workshop, 27-29 March 2012, Raleigh, NC.

John Carmichael and Dave Van Vorhees, Editors, Final Report: MRIP Calibration Workshop II, September 8 –10, 2014, North Charleston SC.

MRIP Staff, Internal Report: A Descriptive Analysis of the Access Point Angler Intercept Survey 2013 Design Change.

Katherine J. Papacostas and John Foster, National Marine Fisheries Service's Marine Recreational Information Program Survey Design and Statistical Methods for Estimation of Recreational Fisheries Catch and Effort.

Atlantic Coastal Cooperative Statistics Program, Field Procedures Manual: Access-Point Angler Intercept Survey.

Deming, W. Edwards, and Frederick F. Stephan. "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known." *The Annals of Mathematical Statistics* 11, no. 4 (1940): 427-444.

Hernandez-Stumpfhauser, Daniel, F. Jay Breidt, and Jean D. Opsomer. "Hierarchical Bayesian small area estimation for circular data." *Canadian Journal of Statistics* 44, no. 4 (2016): 416-430.

John Foster, F. Jay Breidt, Jean D. Opsomer, APAIS data calibration methodology report.

Committee on the Review of the Marine Recreational Information Program, Review of the Marine Recreational Information Program.

Committee on the Review of Recreational Fisheries Survey Methods, National Research Council, Review of Recreational Fisheries Survey Methods.

Appendix 2. Statement of Work

National Oceanic and Atmospheric Administration (NOAA) National Marine Fisheries Service (NMFS) Center for Independent Experts (CIE) Program External Independent Peer Review

Calibration Model Accounting for a Recreational Fisheries Survey Design Change

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards.

(http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).

Further information on the CIE program may be obtained from www.ciereviews.org.

Scope

The Office of Science and Technology requests an independent peer review of a calibration model proposed for use in revising statistics produced by a survey of marine recreational fishing catch rates on the Atlantic coast and in the Gulf of Mexico. This calibration model is considered by the Marine Recreational Information Program (MRIP) to be very important to adjust historical time series of recreational catch estimates in order to account for biases in past sampling and estimation methods that have become apparent with the development of a new, more statistically sound method. The calibration model is intended to account for past biases in catch rate estimates for the shore, private/rental boat, and charter boat fishing modes that have resulted from the continued use of a legacy sampling design for the Access Point Angler

Intercept Survey (APAIS). A more statistically sound sampling design for the APAIS was implemented in March of 2013

Calibration Model for the APAIS Design Change

In 2014, a Calibration Workshop was held to evaluate the potential consistent effects of implementing a new sampling design for the APAIS on the Atlantic and Gulf coasts in 2013. Workshop participants included three expert statistical consultants and representatives from NOAA Fisheries, the regional fishery management councils, the interstate marine fisheries commissions, and several state agencies. The participants determined that analyses conducted by the NOAA Fisheries Office of Science and Technology showed there was sufficient evidence that the more complete temporal coverage of the new design resulted in consistent changes in APAIS angler catch rate statistics for at least some species. They developed three different calibration models to evaluate for possible use in correcting the pre-2013 APAIS statistics. The statistical consultants concluded the simplest of the three proposed models was appropriate for use in the short term until a more complete evaluation of all three calibration models could be completed using three years of new APAIS data (2013-2015). The plan was to complete that evaluation by the end of 2016, so that one method could be selected as the best for use in 2017 to revise APAIS estimates prior to 2013.

Requirements

NMFS requires three (3) reviewers to conduct an impartial and independent peer review in accordance with the SoW, OMB Guidelines, and the Terms of Reference (ToRs) below. The CIE reviewers shall have working knowledge and recent experience in the design of sampling surveys and the evaluation of non-sampling errors (i.e., undercoverage, nonresponse, and response errors) associated with changes to survey designs over time. In addition, they should have experience with complex, multi-stage sampling designs, time series analyses, regression estimators, and small domain estimation methods. Some recent knowledge and experience in current surveys of marine recreational fishing is desirable but not required.

NMFS will provide a Chair who has experience with U.S. fisheries stock assessments and their application to fisheries management. The Chair would ensure that reviewers understand the importance of maintaining a comparable time series of marine recreational fisheries catch statistics for use in stock assessments and their application to fisheries management. The Chair will not be selected by the contractor and will be responsible for facilitating the meeting, developing and finalizing a summary report and working with the CIE reviewers to make sure that the ToRs are addressed in their independent reviews.

Tasks for Reviewers

Pre-review Background Documents

The following background materials and reports prior to the review meeting include:

APAIS Design Change Calibration Workshop Report:

http://www.st.nmfs.noaa.gov/Assets/recreational/pdf/MRIPCalibrationWorkshopII_FinalReport.pdf

NC APAIS Pilot Study Report: *A Pilot Study of a New Sampling Design for the Access Point Angler Intercept Survey.*

https://www.st.nmfs.noaa.gov/pims/main/public?method=DOWNLOAD_FR_PDF&record_id=772

Report on APAIS Calibration Model:

This report will be provided by the contractor (via electronic mail or make available at an FTP site) to the CIE reviewers.

Panel Review Meeting

Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The meeting will consist of presentations by NOAA and other scientists to facilitate the review, to provide any additional information required by the reviewers, and to answer any questions from reviewers.

Contract Deliverables - Independent CIE Peer Review Reports

The CIE reviewers shall complete an independent peer review report in accordance with the requirements specified in this SoW and OMB guidelines. Each CIE reviewer shall complete the independent peer review according to required format and content as described in **Annex 1**. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in **Annex 2**.

Other Tasks – Contribution to Summary Report

The CIE reviewers may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. The CIE reviewers are not required to reach a consensus, and should provide a brief summary of each reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Foreign National Security Clearance

When reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for reviewers who are non-US citizens. For this reason, the reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: <http://deemedexports.noaa.gov/> and http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-

national-registration-system.html. The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place of Performance

The place of performance shall be at the contractor’s facilities, and at the NOAA Fisheries Service Headquarters in Silver Spring, Maryland.

Period of Performance

The period of performance shall be from the time of award through April 31, 2018. Each reviewer’s duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Within two weeks of award	Contractor selects and confirms reviewers
Approximately 2 weeks later	Contractor provides the pre-review documents to the reviewers
March 2018	each reviewer participates and conducts an independent peer review during the panel review meeting
Within two weeks of panel review meeting	Contractor receives draft reports
Within two weeks of receiving draft reports	Contractor submits final reports to the Government

Applicable Performance Standards

The acceptance of the contract deliverables shall be based on three performance standards: (1) The reports shall be completed in accordance with the required formatting and content (2) The reports shall address each ToR as specified (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (<http://www.gsa.gov/portal/content/104790>). International travel is authorized for this contract. Travel is not to exceed \$12,000.

Restricted or Limited Use of Data

The contractors may be required to sign and adhere to a non-disclosure agreement.

NOAA Fisheries Project Contact:

Dave Van Voorhees

NOAA Fisheries

1315 East West Highway

Silver Spring, MD 20910

dave.van.voorhees@noaa.gov

Annex I: Format and Contents of CIE Independent Peer Review Report

1. The report must be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether or not the science reviewed is the best scientific information available.
2. The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each ToR, in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the ToRs.
 - a. Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.
 - b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, but especially where there were divergent views.
 - c. Reviewers should elaborate on any points raised in the summary report that they believe might require further clarification.
 - d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.
 - e. The report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The report shall represent the peer review of each ToR, and shall not simply repeat the contents of the summary report.
3. The report shall include the following appendices:
 - Appendix 1: Bibliography of materials provided for review
 - Appendix 2: A copy of this Statement of Work
 - Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

Calibration Model Accounting for Changes in Recreational Fisheries Survey Methods

- 1.** Evaluate the suitability of the proposed approach for converting historical estimates of mean angler catch rates obtained using the old Marine Recreational Fisheries Statistics Survey (MRFSS) Access Point Angler Intercept Survey (APAIS) sampling design to estimates that best represent what would have been produced had the new MRIP APAIS sampling design been in place prior to 2013.
 1. Does the proposed approach adequately account for consistent differences in estimates that would have been observed if the old MRFSS APAIS had been conducted side-by-side with the new MRIP APAIS in 2013-2017?
 2. Is the proposed approach a suitable alternative to the calibration models that were originally developed in the 2014 MRIP Calibration Workshop and later evaluated by MRIP?
 3. Is it reasonable to conclude that revised 2004-2012 APAIS estimates based on the application of the proposed approach would be more comparable than the current ones to estimates produced since 2013 under the new APAIS design?
 4. Given the limitations of the available data, is it reasonable to apply the proposed approach to revise APAIS estimates prior to 2004 (back to 1981)?
- 2.** Briefly describe the panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.

Tentative Agenda

Calibration Model Accounting for a Recreational Fisheries Survey Design Change

TBD

NOAA Fisheries

Office of Science and Technology

1315 East-West Highway

Silver Spring, MD

March 2018

Point of contact: Front Desk

Appendix 3: Panel membership

Michael D. Murphy (Chair)

Carolyn Belcher

Mary Christman (CIE Reviewer)

James Chromy (CIE Reviewer)

Matthew Cieri

Paul Rago

John Whitehead (CIE Reviewer)