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

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.

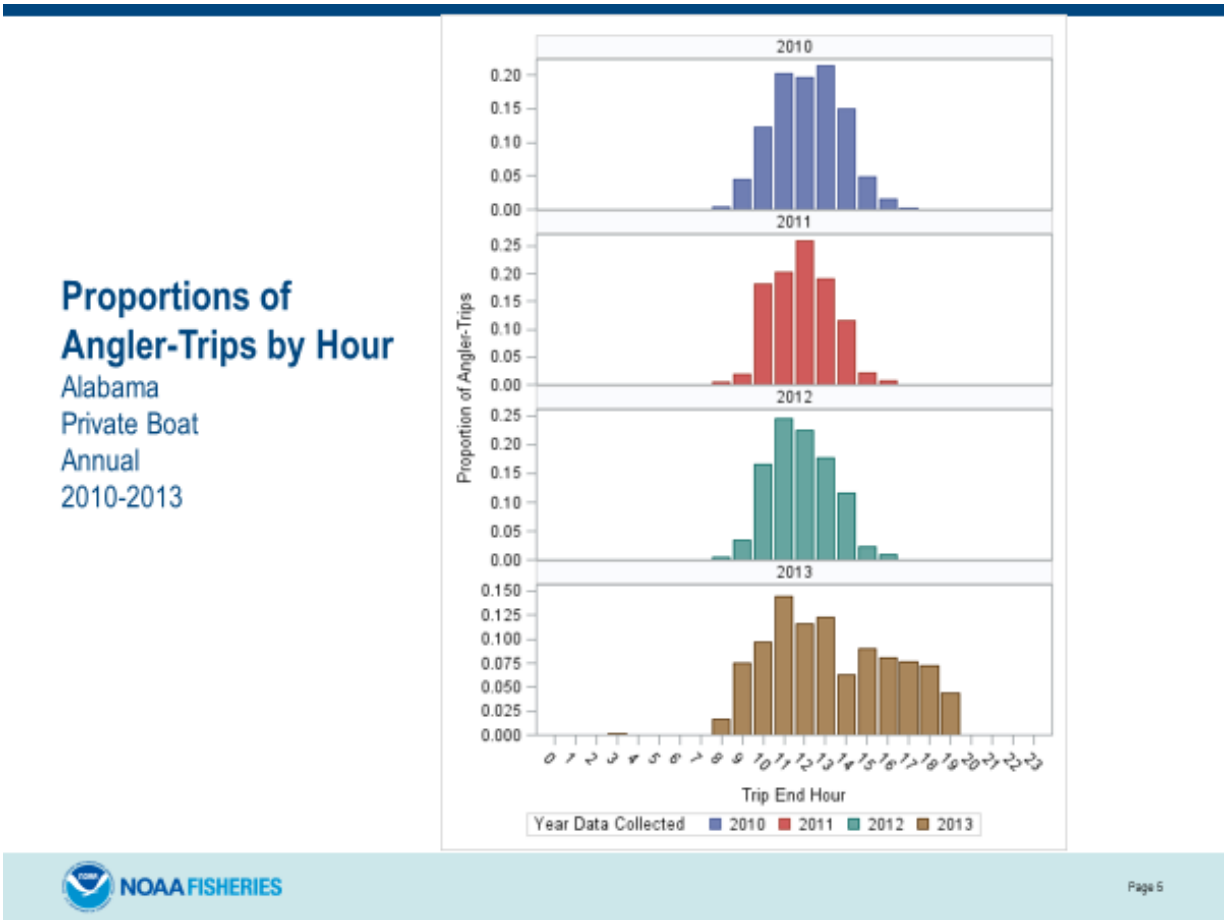
Presentation 4. Dave Van Voorhees discussed the methodology for developing pseudo-design-based weights for 2004-2012 APAIS estimates. These have been peer reviewed and accepted. The estimates developed using these weights may need to be adjusted further to make them comparable to estimates from the redesigned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

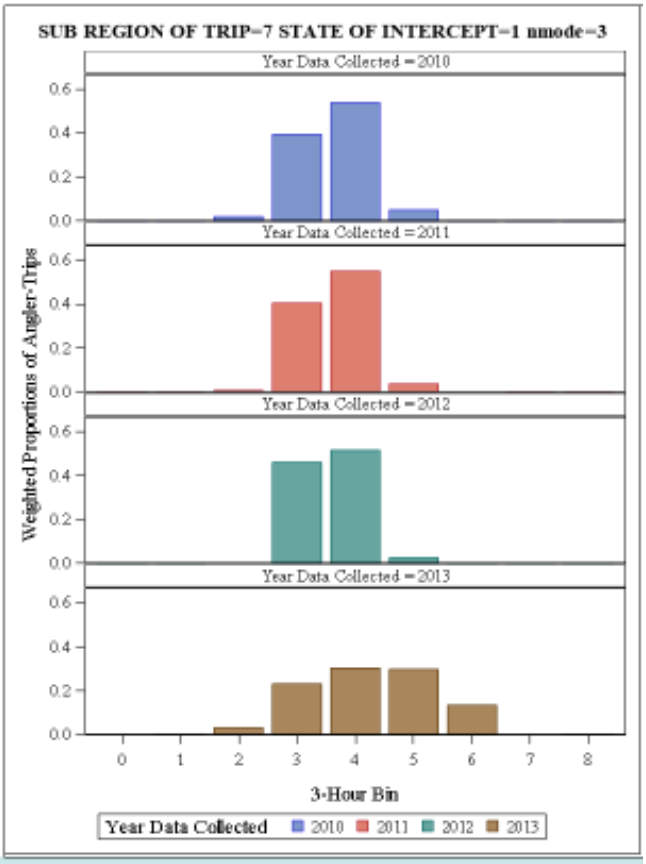
$$\hat{C}^{*}_{Ty} = \hat{R}_{2013|y} * \hat{C}_{Py}\ ,\ y \in \{2004,2005,\ldots,2012\}$$

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 by 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 fo 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