Marine Recreational Information Program (MRIP)

Gulf State Recreational Catch and Effort Surveys Transition: A Workshop of the Gulf of Mexico Subgroup of the MRIP Transition Team

Summary Report

February 23-25, 2022

Prepared by:

Paul Rago

Mid-Atlantic Fishery Management Council

Scientific and Statistical Committee

June 3, 2022

Table of Contents

Executive Summary

Overview. The Marine Recreational Information Program (MRIP) seeks to continuously evolve in response to the changing needs of recreational harvesters, scientists and managers. Within the Gulf of Mexico region, states where MRIP has historically been conducted have implemented alternative surveys for estimation of effort and catch. These surveys have different degrees of temporal and spatial coverage and provide estimates for various fishing modes, dispositions, and species. These survey designs have been certified to be unbiased in expectation. However, owing to different objectives and sampling designs, these surveys produce estimates that differ from those derived in the MRIP survey. Such differences can create problems for stock assessments and subsequent management because the estimates are of different magnitudes, or have different "currencies." As per NMFS Policy Directive 04-114, a Transition Plan will allow full use of the alternative surveys' estimates while at the same time satisfying the scientific and management requirements under the MSA to use the Best Scientific Information Available (BSIA). Development of the Transition Plan has required the efforts of multiple partners from the federal and state agencies as well as the Gulf Council and Gulf States Marine Fisheries Commission.

This report summarizes the results of a virtual meeting, held Feb. 23-25, 2022, to address critical short- and long-term needs of the Transition Plan. It represents the latest in a series of meetings that have addressed the issue of comparability of alternative estimates. Upcoming assessments for gag grouper and red snapper in the Gulf create additional urgency for this task. This report is the proceedings of that meeting—it is designed to faithfully capture the essence of presentations, and more importantly, the ensuing discussions and recommendations. More than 100 individuals attended the meeting and 50 participated directly in the discussions. Notably, five expert statistical consultants provided recommendations in response to presentations, questions and discussions during the meeting. In addition, the Consultants met after the meeting to craft more synthetic responses to the suite of meeting topics. Their findings are included as an appendix to clearly distinguish topics that were addressed in plenary sessions from those that were addressed outside the meeting.

The meeting was supported by the outstanding efforts of the MRIP Transition Team. Prior to the meeting, webinars were held to update the Consultants on current progress towards responding to recommendations since their initial peer review and certification. The webinars also summarized work that is being done to investigate sources of error and reduce potential bias. The advance meetings gave participants the opportunity to brief MRIP consultants on specific improvements, issues, and ongoing work related to each survey before the Feb 23-25 meeting.

An agenda was distributed before the meeting. Individual presentations set the stage for each topic and were followed by series of clarifying questions and open discussion. To ensure the best use of Consultants' expertise, Consultants' questions and comments were addressed first.

However, all participants were afforded the opportunity to speak. In most instances, the issues could be fully vetted within the allotted time although some topics overflowed to the next day.

The need for a Transition Plan is further motivated by a request from the House Committee on Appropriations (2021) to:

- Assess the accuracy and precision of all programs in the Gulf,
- Make recommendations for improvements to all the programs and
- Suggest how best to calibrate the programs to a common currency.

The Transition Plan requires a pragmatic path to meet short-term assessment and management needs and a research path to address specific issues such as non-sampling error. The Transition Plan must balance scientific innovation with stock assessment needs. The Research Track of the process will require improvements to all surveys. No timeline is specified in the congressional directive, but a period of about five years is likely.

Sound stock assessments require a full accounting of landings and discards, and consistent historical data. Sound management requires consistent, complete and accessible data in a common currency. In-season monitoring of harvests is a primary responsibility of federal managers to avoid overfishing limits. Meeting the demands of BSIA is essential for both sound science and management.

A Research Roadmap was outlined by the Consultants. Specific recommendations for each state and for MRIP are listed in the body of the report and in the Consultants' report in Appendix 3. All parties recognized that further in-depth consultations are required. Information about private dock usage would benefit all programs because it is assumed that patterns of landings and reporting between anglers at public and private access points are similar. There are no guarantees that such information will be sufficient to establish agreement among survey estimates. Consultants cited examples from health care surveys where estimates did not overlap, even though all surveys were peer-reviewed and had validated methods.

Recreational catches in the Gulf states are derived either from 1) capture-recapture surveys with a ratio estimator, or 2) probability surveys with a product estimator (i.e., angler intercepts x effort surveys). Each survey has a different sampling frame such that the target population of anglers differs among states. All the surveys may share many common vulnerabilities to non-sampling error such as coverage error, nonresponse error, and measurement error. Capture-recapture surveys are vulnerable to matching errors and failure of the assumption of independence or reports from anglers on intercepted vs. non-intercepted trips. A detailed list of research recommendations from the Consultants may be found in the body of report and Appendix 3.

Database Storage and Data Management: In the short run, all parties should provide both raw and processed data in a standardized format along with associated metadata[1]. Development of a model-based calibration estimate will require high-resolution data rather than aggregated data typically available. Survey estimates, calibration factors, the raw data tables on angler intercepts and biological data, the metadata for survey design, and in some cases the algorithms used for computation of weights, are required. Not all demands of the data are known in advance, so having the ability to reconstruct estimates as appropriate was judged to be a desirable feature. Provisions for public access to summarized data is also required.

Proper timing of data deliveries is essential for use by scientists and particularly managers. The Southeast Fisheries Science Center (SEFSC) endorsed the concept of expanding the role of the Gulf of Mexico Fisheries Information Network (GulfFIN) for this purpose but noted that timetables must be realistic with respect to available resources. GulfFIN representatives emphasized it could not be responsible for full-scale auditing of each submission. State concerns were similar—they are focused primarily on executing the designs and not sufficiently staffed to handle some of the proposed analyses. All participants agreed that improvements to the data standards process are both desirable and valuable. Sound foundational principles such as the NMFS recreational fishing survey and data standards and the Southeast Data, Assessment and Review (SEDAR) stock assessment guidance documents are excellent starting points.

Calibration: The goals of research for accuracy vs. calibration will not necessarily be the same. Nonetheless, calibration is a requirement for stock assessments to assess the status of populations and the basis for subsequent monitoring of catches by jurisdiction. If side-by-side comparisons reveal large differences in catch estimates among and within states, calibration is required to establish a common currency.

Five different calibration methods were actively considered by the participants. The first two come from the 2020 "White Paper", which identified four primary options for calibration. Two variants of a hybrid approach that built upon the White Paper recommendations were also considered. Ratio methods applied to date have been based on empirical comparisons among surveys but have not attempted to address why the estimates were different. All of the surveys have an angler intercept component and this might be a useful place to look for commonalities. Correlations among surveys at the intercept level could help identify predictor variables. Such patterns could lead to a "new currency" which does not use any single survey (e.g., MRIP) as the frame of reference. This approach was called a model-based approach and will necessarily be part of a longer-range effort.

---

[1] Metadata includes all data related to the collection of information. This broader definition of metadata is known as paradata.

Simple ratio comparisons between MRIP estimates and state survey estimates for years when both surveys were conducted vary widely by state, ranging from 1.8x for Louisiana to 11x for Texas. The range of ratios have practical implications for calibration:

- Do we use state, MRIP or composite metrics of harvest?
- Are the years used for calibration fixed, or amenable to future changes?
- Do calibration decisions vary by species?
- Will calibration decisions be updated regularly?

Participants acknowledged that such ratio calibrations could, and possibly should, evolve annually. Regardless of the calibration standard used, developing a consistent time series is perhaps the most important criteria. Participants agreed that multiple conversions between assessment results and monitoring estimates could be time consuming, hard to communicate, and possibly error prone. Some states may be able to update their calibration ratios (e.g., FL) but in other states where MRIP is no longer conducted (e.g., TX), the ratios will remain fixed unless auxiliary information can fill gaps.

Participants recognized the need to pursue both short and long-term solutions and affirmed their willingness to work towards imperfect, but necessary, short-term solutions to meet upcoming stock assessment needs. The diversity and complexity of the survey methods heightens the need for detailed specification of data conversion and increases the chances of miscommunication among stakeholders. Participants noted that consistency is a central tenet of BSIA.

Ratio-based calibrations were reviewed during the 2020 Red Snapper V Workshop, but there are nuances of calibrating back in time and handling different fishing modes, particularly shore-based fishing activity. Consultants have previously advised state and MRIP representatives on methods for calibration of means and standard errors. A similar follow-up meeting to discuss Terms of References for stock assessments was proposed. The upcoming gag and red snapper assessments should be illustrative of ratio calibration limitations and considerations. The gag fishery is dominated by harvests from Florida whereas red snapper are harvested by all states in the Gulf.

Calibration Recommendation: After considerable discussion, a proposal to use Option 1b from the White Paper for red snapper was advanced. Participants recognized that the complexities of adjusting five fishing modes within states and varying degrees of overlap with MRIP estimates would make this a difficult task. Additional concerns about the use of Louisiana and Texas data were expressed. Representatives from Texas were not present for this discussion and concerns were expressed about the implications of rescaling the entire time series of red snapper catches to MRIP equivalents. Since the consultants were not provided the opportunity to visit and discuss the details of the Texas survey, the consultants affirmed that they have not reviewed the procedures used in Texas.

Following some additional clarifications, the group consensus agreed to:

- Use 1b for red snapper.

- Use option 2 for gag [see 2020 White Paper for these options].
- Convene a group to set Terms of References and identify individuals to conduct a formal review of the application of the ratio-based calibrations for gag and red snapper at minimum, and others as appropriate. Owing to the complexity of the underlying surveys, it is anticipated that expertise outside the SEDAR process would be necessary for review.

Reaching agreement on a methodology for calibration was viewed as an important accomplishment of the Workshop.

Transition and Communications Plans: The remainder of the meeting was devoted to outlining the components of a Transition Plan and Communication Plan. The Transition Plan outline will be fleshed out and circulated to the Transition Team as soon as possible for review and comment. Participants expressed concerns that the Transition Plan should be finalized as soon as possible.

The overall goals of the Communications Plan are to facilitate development of the Transition Plan and its implementation. The plan will help ensure that workshop outcomes are clear, easy to find and promptly distributed to target audiences. Target audiences include transition team members, leaders of their respective agencies and external stakeholders such as Council members, congressional representatives, anglers and various advocacy groups. Participants should nominate members for the Communications Working Group.

Proceedings of Three-Day Meeting

Day 1: February 23, 2022

Overview

This report is a summary of a virtual workshop to develop a transition plan for integrating the various surveys for recreational catches (See Appendix 1: Agenda) The workshop included representatives from various state, regional and federal programs as well as members of the Gulf States Fishery Management Councils (See Appendix 2: List of Participants).

Recreational fisheries in the Gulf States are monitored by state or federal surveys. In some states, both approaches have been used in the past and continue to the present. In other states, contemporary monitoring is done solely with a state survey (LA) while in Texas, the MRIP survey has never been used comprehensively. This mixture of methodologies creates problems when catch data are used in stock assessments. Every sampling design has varying levels of bias and inherent variability of estimates. Thus, every sampling design imposes constraints on the subsequent use of the data.

Stock assessments attempt to measure population abundance by quantifying known removals and relating these removals to trends in various indices. Estimated removals in stock assessments may be under or over-estimated if one or more of survey estimates included in the sum of recreational catch is biased. The consequences of bias are also important for sector allocations

and monitoring of harvests that are required under the Magnuson-Stevens Act (MSA). Hence, it is essential to understand the potential sources of bias and to develop a sound basis for contemporary science and management.

The purpose of this workshop was to continue the progress toward development of a common currency for short-term stock assessment and management needs, and to emphasize the need for continued efforts toward fully comparable estimates among programs. The workshop builds upon earlier efforts to develop technical guidance. Five statistical consultants with expertise in survey design participated in the workshop to identify candidate approaches for comparing surveys and to develop plans for future research efforts.

This report constitutes the proceedings of the workshop and a summary of key conclusions. The statistical consultants met shortly after the meeting to synthesize their recommendations for future work. Their report (submitted March 28, 2022) is presented verbatim in Appendix 3.

The meeting was convened by Gregg Bray and Richard Cody who gave some opening remarks, expressing appreciation to all participants for their ongoing commitment to the Transition Process. Paul Rago provided an overview of the Workshop purpose, desired outcomes and agenda. The transition process began with the National Academy of Sciences report in 2006, which identified a wide range of needed improvements in the federal survey. MRIP surveys were developed between 2008 and 2015 and fully replaced the earlier MRFSS surveys in 2018. In the Gulf Region, new state programs (LA Creel, AL Snapper Check, FL SRFS, and MS Tails and Scales) were initiated between 2014-2015. These surveys were developed fully or in part to support timely in-season quota monitoring for state-level management of red snapper and are tailored to state-specific fishing patterns and geography.

Comparisons between catch estimates derived from these surveys and MRIP have revealed substantial differences in some cases. As noted above, such differences are problematic for stock assessments. Such differences are mostly attributable to non-sampling errors including differences in coverage. Thus, it is necessary to simultaneously address the underlying survey design differences and need for stock assessment data. The first path (track) requires continuation of long-term research efforts; the second path (track) requires development of interim methods for combining various catch estimates into a reasonably consistent and precise total. Both paths require a transition plan.

Transition Plan
The overarching goal of the meeting was to agree on the elements of a transition plan for both a research path, and a "pragmatic" path. The pragmatic path requires agreement on a flexible calibration approach for all assessed species and allows for different methods among species. All calibration approaches must meet BSIA standards (https://www.ecfr.gov/current/title-50/chapter-VI/part-600/subpart-D/section-600.315). Within this report, the paths are referred to as tracks. Together, they constitute the "hybrid" plan.

Richard Cody presented a summary of milestones to be achieved by the Transition Plan. Notably, NMFS has received a congressional directive from the House Committee on Appropriations (2021), which calls for an independent review that 1) assesses the accuracy and precision of all programs in the Gulf, 2) makes recommendations for improvements to all the programs and 3) suggests how best to calibrate the programs to a common currency.   To meet this directive, the following steps (milestones) were identified:

- Develop a research plan to understand differences across all surveys and contract for an independent review.
- Finalize calibration methods, e.g., ratio estimation.
- Recommend interim measures.
- Identify and solve unmet certification issues for state and federal surveys.
- Develop a common database for partners.
- Improve communication plans.
- Make final decisions on calibration methods for future assessments.

Participants noted that the transition plan must balance scientific innovation with stock assessment needs. The research track of the process will require improvements to <u>all</u> surveys. No timeline is specified in the congressional directive, but a period of about 5 years is likely. Following the meeting, the Consultants will be making specific comments on both tracks within the Transition Plan. Conducting longer-term side-by-side survey comparisons may ultimately yield diminishing returns. Participants agreed that such concerns will be important as the Transition Plan is implemented.

Questions were also raised about unmet certification issues.   State-specific projects with the Consultants are ongoing and refinement of these issues is expected as more data become available.   A number of participants suggested this workshop should focus on the central issue of identifying the primary drivers of differences among surveys, but acknowledged such efforts would likely entail more detailed discussions than might be possible in the plenary meeting. Consultants cautioned there are many instances at other agencies where differences among survey estimates could not be resolved. All agreed that commitment by all parties was essential for moving forward.

The Consultants noted two competing goals under discussion: estimation of overall catch and allocation of shares of that catch to Gulf states. Further discussion on the relative importance of these goals revealed that the overall goal was to obtain the best possible estimates by state rather than specific allocations. Consultants noted that integration of disparate monitoring systems is much more challenging because of different underlying assumptions and potential biases.

Data Requirements for Stock Assessments and Management
Management of fish stocks under the MSA requires recreational catch data for not only stock assessments but also for a suite of regional management responsibilities. Moreover, the data must satisfy requirements under the BSIA. These topics generated considerable discussion among all participants.

Andy Strelcheck, SERO, outlined the multiple recreational data needs for management. Sound management requires consistent, complete and accessible data in a common currency. In-season monitoring of harvests is a primary responsibility of federal managers to avoid overfishing limits. Additional complexity arises when monitoring must track catches from more than one fishing mode (e.g., private, charter boats). The analyses that support harvest control measures rely on raw intercept data to evaluate effects of changes to bag limits, trip limits and size limits. In turn, these support evaluation of economic tradeoffs. A common database (or warehouse) to support these analyses is viewed as essential.

Consultants noted the complexity of this challenge, especially the difficulties of assigning weights to raw data for a complex survey. Procedures to transfer data processing programs to the warehouse organization/operation would greatly assist post processing of data. Participants asked if data requirements would vary by species. Strelcheck emphasized again the need for consistency and accessibility for all surveys

Recreational data needs for stock assessments were described by Kate Siegfried, SEFSC. Landings and discards, in both numbers and weight, are key ingredients in modern stock assessment models. When combined with additional information such as discard survival rates, and age-length keys, catch data enhance our understanding of population dynamics. Proper estimates of variability in catch facilitate identification of appropriate models for crafting overfishing limits. Siegfried further emphasized the need for consistent historical data. Most models are useless if bias in historical catch data is not accounted for.

Participants noted that some surveys do not estimate out-of-season discards for red snapper, but the importance of this omission is unknown. Ostensibly, this could be a major problem, but other participants noted that closure of the red snapper season also typically signals the end of most reef fishing. These comments helped identify another common thread—surveys are designed to address specific objectives and it is challenging to meet additional objectives after the survey was designed and launched for another purpose. The need for increased coordination among partners was reiterated by participants who also cited a similar conclusion by the 2021 NASEM report on in-season management. Concerns were raised regarding differences between modes within the MRIP as examples of a "mixed currency" system. While such differences in protocols may result in currency differences, these are thought to be small relative to differences in methodologies across states. Participants commented that differences between estimates from private and charter boats were among the motivations for creating alternative state-specific survey methods in the first place.

Under National Standard #2, all <u>recreational data must meet standards under the Best Scientific Information Available (BSIA)</u>. Patrick Lynch highlighted these requirements of a six-step process that guides stock assessments, peer review, revisions, SSC deliberations, Council catch specifications and finally NMFS determination of BSIA. No single element is sufficient to define BSIA. The process for evaluating BSIA compliance should not be viewed as competitive among states and MRIP because all programs have strengths and weaknesses. In fact, the utility of various surveys may vary across species (Gag Grouper was cited as an example in which 95% of the stock is sampled by SRFS). BSIA relies heavily on the peer review process within stock assessments. Reviews outside of the stock assessment process might be advantageous particularly in circumstances where time is insufficient for an in-depth review of underlying data.

Survey Research Roadmap (Part 1)

Prior to the Feb 23-25, 2022 meeting, each of the Gulf States (except Texas) and NMFS provided the Consultants with an overview of progress made in response to peer review recommendations and other research efforts to identify non-sampling error. The state-specific, one to two-hour meetings were appreciated by the Consultants. Gregg Bray provided an overview of the major findings of these informational meetings.

> <u>Florida:</u> No major concerns were raised during the earlier peer-review and the certified design was expanded statewide in 2020. SRFS was designed to minimize sources of bias and FL has ongoing research to verify accuracy and identify additional sources of bias

> <u>Louisiana</u>: LA Creel was designed as a full replacement for MRIP. About 73% of their private angler trips return to publicly accessible docks, and approximately 95% of private angler trips return within sampling timeframe.

> <u>Mississippi:</u> Non-sampling errors could be occurring in Tails 'n Scales (e.g., private access sites). Concerns were expressed regarding trends in MRIP general survey harvest estimates that occurred during periods with low survey counts. MS feel this should be addressed and resolved prior to calibrating. MS also welcomes proposed projects to aid in effort estimation, private dock stats and others that will aid in drawing estimates across surveys closer together.

> <u>Alabama:</u> A large portion of the annual red snapper fishing effort occurs during first few weekends of the red snapper season. The timing of this period of higher activity relative to timing of MRIP Fishing Effort Survey (FES) questionnaires may contribute to higher fishing effort estimates. Nonresponse rates have decreased since inception of program but they are at or above 50%. In terms of ongoing research, the state is exploring use video cameras and artificial intelligence software to determine offshore fishing activity. In addition, AL is starting a program to subsample on the water – data will be used to compare Snapper Check to the MRIP Access Point Angler Intercept Survey (APAIS).

> <u>Texas:</u> No information available.

> NMFS: MRIP has responded to certification peer reviews and 2017 NASEM review. These include research into effects of nonresponse and recall error in the FES, and APAIS coverage error. A sensitivity analysis (Foster et al. 2021) suggest non-sampling errors can account for differences in red snapper landing estimates for private boat mode, but understanding of the drivers is incomplete. It was noted the disparity of estimates across states increases with differences in designs and underlying assumptions.

The extended discussion among participants followed Bray's presentation is summarized below.

The Consultants noted that despite the pre-meeting presentations by the States and NMFS, there was insufficient time for the formulation of detailed questions in the one- to two-hour format. Preparation of a list of questions to each state by the Consultants was proposed to address this problem, but that responses would not likely alter their recommendations from this meeting.

Access to private docks could help estimate effort outside the sampling frame. Mississippi for example, attempts to include enforcement data to help estimate license compliance rates. This raised questions about the inclusion of non-probability samples in creel survey analyses as well as other disciplines. It was noted that the IRS has used non-probability sampling designs to audit high-income earners, which could be blended with their regulatory audits.

Collectively, the Consultants provided additional background on the Research Roadmap. A key point was that differences between MRIP and state surveys are not unique in the statistical literature. A study in 2013 compared five national surveys of health characteristics and found little agreement despite each survey being a rigorous, peer-reviewed design.   The Consultants highlighted that seemingly unimportant aspects of survey implementation (question ordering for instance) could have major impacts. So called "gate-keeper effects" are thought to account for non-sampling error in telephone surveys, notably the Coastal Household Telephone Survey (CHTS) previously used in MRFSS. The Consultants' experience suggests that any detailed comparative survey analyses will require access to the raw sampling data and extensive modeling. Further discussions suggested that comparisons should also consider the varying uses of catch data for assessment and management, particularly within a season.

Day 2: February 24, 2022

Day 2 began with a review of progress on Day 1 and continuation of Survey Research Roadmap discussion. The Consultants then provided an overview of non-sampling errors in the Gulf. This was followed by plenary discussions and some final recommendations. Details of those recommendations and discussions follow.

Survey Research Roadmap (Part 2)

Overview of Non-Sampling Error in Gulf Surveys

Recreational catches in the Gulf states are derived either from 1) capture-recapture surveys with a ratio estimator or 2) probability surveys with a product estimator (i.e., angler intercepts x effort surveys). Each survey has a different sampling frame such that the target population of angler differs among states. While they share many common vulnerabilities to non-sampling error, they differ in others. Shared non-sampling errors include:

- <u>Coverage error</u> occurs when some sampling units are outside the sampling frame
- <u>Nonresponse error</u> occurs when sampling units (e.g., anglers, mail survey recipients) are unwilling to cooperate or unavailable/unable to respond. The key question is whether the behavior of non-respondents is equal to respondents. Effort surveys with product estimators are particularly vulnerable to this error. Consultants noted that while standard formulas are available to estimate the magnitude of nonresponse bias, it is not easy to do so because the required parameters are not available without special experiments.
- <u>Measurement error</u> arises from sources like questionnaire formats, subsequent data handling and other causes. Both the intercept and effort surveys are thus vulnerable to this source in both estimation systems. Note however, that errors in the angler report data are not a source of measurement error for the capture-recapture estimator, even if angles misreport. This is because these data are not used directly as responses, but rather as auxiliary variables.

Capture-recapture surveys are vulnerable to the following sources of non-sampling error:

- <u>Matching error</u> between the capture and recapture units can be either false positives, false negatives or mismatches. All types will inflate bias, but in different directions.
- <u>Failure of the Independence Assumption</u> occurs when anglers on intercepted trips report differently than non-intercepted anglers.

Key questions from Participants included:

- How is nonresponse characterized in capture-recapture compared to the other probability sampling methods?
  - Intercept surveys typically have a response rate around 80% because of the direct interactive nature of the interview. In contrast, self-administered effort surveys have much lower response rates. Increases in the nonresponse rate increase potential bias. Concerning comparability between compliance rates (capture-recapture) and response rates (probability sample), it was noted that the reports used as the second set of data in capture-recapture survey are not a probability sample. Instead, those data are used to calibrate, rather than measure catch. Inaccurate or incomplete data simply reduce the correlation between reporting and the actual trip. In turn, the variance increases. The nonresponse and measurement error that afflicts the capture-recapture estimator comes from the intercept survey.
- What is the impact of over-coverage, e.g., inclusion of non-fishing households in the FES?

o Over-coverage doesn't matter unless respondents are falsely reporting trips not taken. The focus is the percentage of people taking a trip such that nonresponse bias can come from zeros that should actually be non-zero trips. If none of the FES sampled non-fishing households respond, then the calibrated weights (control totals from the address-based sampling frame) for the responding fishing households would yield totals that are biased high. Estimation of proper sample weights is critical. The FES weighting process uses population-based totals and sample aligned with those calibration totals. There is no "disconnect" in that estimation system.

Discussion about measurement error contrasted the higher level of detail in the FL questionnaire with the FES questionnaire. Such differences enhance the opportunity to understand differences between the surveys (e.g., recall bias) but do not necessarily imply one is better than another. Given sound objectives of the survey program, focus groups can indicate how people respond to multiple data collection methods and facilitate choices consistent with those objectives. The Consultants described the cognitive interviewing process, which usually begins with validated questions from other surveys and progress to more complex topics. An iterative process used for developing the FES required several iterations and revealed the critical importance of the ordering of questions related to fishing activity. People generally want to tell surveyors that they fished in the recent past. A survey that does not address that tendency might be incorrectly filled out if the questionnaire narrowly restricts the period for reporting fishing activity. Incentives for responding (i.e., a small reward) have been shown to be useful, even for long questionnaires. Consultants recommended the use of the American Association of Public Opinion Research (AAPOR) response rate method 1 (AAPOR 2016) for all state and federal programs.

Specific Studies Suggested by Consultants
The Consultants made the following recommendations to the Transition Team:

1. All surveys should review estimator calculation. Consider weighting methodologies used to create final analysis weights. Review them and the program code or software to see if they are correctly applied to account for the sampling design and data being analyzed.
2. All states should review QA/QC procedures
   a. Does each state have a rigorous QA/QC program in place?
   b. Is there consistent training of all interviewers? Consistent within and across all the states?
   c. Is QA done on interviewers assigned to a site, e.g., a supervisor subsampling interviewees to ensure data is collected at the site during the assigned time frame on a regular basis?
   d. Does each state use the same formula for response rates (e.g. AAPOR codes)?
3. All programs that use an offsite effort survey should:
   a. Estimate effort using the intercept data and compare that with effort estimates obtained by the state surveys (FL and LA) and by MRIP through the FES data.
   b. The difference between the intercept-only effort and the offsite survey effort should be an estimate of the public access effort share. Is this ratio plausible? Can it be checked against other sources?

4. MRIP should consider conducting simulations to assess whether the standard formulas for variance estimation work for rare event species with low sample sizes and highly variable observations.

5. FL and LA surveys could be examined through use of split experiments to compare questionnaires with different samples. This could reveal if the difference in questionnaires are contributing to the differences observed in estimates. An example would be to use the MRIP questionnaire on the angler license frame, and the LA Creel/SRFS questionnaire on the FES frame. That could tease out if the differences are due to the sample frame (i.e., coverage error) or the questionnaire (measurement error).

6. Split experiment studies for FL could help address differences between the one-month surveys in SRFS with the two-month surveys in MRIP.

7. The MS program would benefit from an in-depth review of the weighting estimation procedures and use of a split experiments recommended for FL and LA. Further investigation of the utility of law enforcement data could be useful for determining origins of trips launched from public vs. private site. This might also inform differences in reporting rate between on-frame and off-frame anglers.

8. Recommendations for AL include cessation of using catch as a matching variable as this may be biasing the estimates. New matching criteria should consider testing variables to make the trips more distinctive from each other. One experiment that might be useful for future matching would be to ask intercepted anglers if they were aware that they are required to report, if they have already reported their trip or not, and when they became aware that they became aware that samplers were collecting data for this trip.

9. The primary recommendation for TX was to use a probability sample to collect data rather than the non-probability sampling currently in place.

10. All the surveys should assess potential differential reporting rates between public and private access anglers. If the catch per unit of fishing effort (CPUE) estimates are equal, then the potential for bias is reduced. Capture-recapture designs may be especially useful in this regard. Advanced remote sensing analyses (e.g., satellite tracking) might help identify the fractions of trips on and off data collection period.

Participant discussions included questions about weighted estimation, question sensitivity, social desirability bias and methods for estimating coverage error. These discussions later broaden into concerns about next practical steps for the roadmap itself.

The Consultants urged all participants to review the weighting procedures used in their surveys. Weights for site selections and meticulous records of trip counts at each sampling time are essential. Importantly, the Consultants noted that estimating inclusion probabilities (the inverse of weights) is formulaic. Weighting is more difficult for intercept surveys than for effort surveys because of the added complexity of selecting not only the sites but also the time period to visit the sites. Intercept surveys also provide valuable ancillary information. Florida, for example, uses supplemental intercept data to predict fishing effort within season for red snapper.

"Question sensitivity and social desirability bias" can affect the validity of catch estimates. Examples include state of residence, license status, and proper species identification for unobservable catches. In MRIP, the state of residence questions in APAIS affect estimates of

coverage within the effort survey. Hence, bias in one survey can induce additional bias in the effort survey.

The problem of private vs. public access sites affects all survey types. For capture-recapture surveys, biases in estimators can occur due to lack of coverage of private sites unless reporting rates and CPUEs are the same in public and private sites. To examine whether or not this is a problem, special studies must be conducted. Consultants noted that collection of date, time and precise location of interview sites can allow application of more advanced techniques. For example, Alabama's monitoring collects information on whether trips end at private or public docks.   Satellite tracking or unique angler ID devices could be valuable for future improvements. These topics led to some discussion of novel effort monitoring studies such as the camera studies in Ocean City, NJ and Alabama ports.

Mississippi reported that marine patrols collect information on private vs. public launch sites at "random" (actually haphazardly), but consultants cautioned that "haphazardly" is not the same as random, and therefore will not produce a probability sample. Haphazard selections cannot be adjusted to perform as a probability sample does without better information about how the selected units were obtained or compare to the unselected ones.   A probability sampling design should be adopted prior to the data collection.

Questions were raised about ongoing simulation studies to help refine estimators for rare species. Initial evidence suggests such estimates are more variable than once thought. Sampling theory is based on large samples so explorations of estimator behavior at small sample sizes are valuable. An overall message from this session was that more examination of existing data, coupled with simulation studies, would be valuable for measuring deviations between theory and practice.

Future research efforts are most obviously constrained by funding and time but also by the roles of the partners. A key consideration is whether research should focus on accuracy or comparability questions. It is not clear if improving individual surveys will make surveys more comparable or not. Despite this uncertainty, all partners expressed in interest in executing the research plan.

NMFS has been mandated by Congress to reallocate $2M. Options include funding for consultants, independent reviews, and potentially additional studies. Given the wide disparity in costs of various studies, it was suggested that states identify their highest priorities. States may also be able to leverage funding from other sources.

Methods for Calibration (Part 1)
As the Consultants noted, the goals of research for accuracy vs. calibration will not necessarily be the same. Nonetheless, calibration is a requirement for stock assessments to assess the status of populations and the basis for subsequent monitoring of catches by jurisdiction. If side-by-side comparisons reveal large differences in catch estimates among states, calibration is required to establish a common currency.   Richard Cody's presentation noted that calibration is a

longstanding issue with five workshops for red snapper over the past nine years. Owing to the complexity of full model-based comparisons, simple ratio methods have been the preferred/feasible method to date. These were reviewed by the Gulf Council's SSC in 2020 and recommended for implementation in 2023. In October 2021, NMFS and State Directors met to finalize plans for transition. Key decisions included recommendations for a common database, improved understanding of the differences among surveys, and improved communications.

Ratio methods applied to date have been based on empirical comparisons among surveys but do not explain the differences in estimates. Current surveys all have an angler intercept component. This might be a useful place to look for commonalities. Correlations among surveys at the intercept level could help identify predictor variables. Such patterns could lead to a "new currency" which does not use any single survey (e.g., MRIP) as the frame of reference. This approach was used by one of the Consultants to calibrate two versions of the 2016 Fishing, Hunting and Wildlife Associated Recreation (FHWAR) survey collected using different modes (see Erciulescu et al, 2021). A primary difference between the Consultant's referenced calibration work and the current calibration needs for stock assessment purposes is the need for historical comparability. While such methodology might be generalizable to all species, the present focus will be on red snapper. Only MRIP, LA Creel and Texas Coastal Creel Survey estimate catch for all species.

An earlier workshop report (GSMFC & NMFS 2020, hereinafter referred to as the "White Paper" identified four primary options for calibration. Discussants noted that Options 1a and 1b were the most viable given time and resource constraints. There was some additional discussion about the details of currency conversion. In particular, everyone agreed that consistent application of conversion factors to different modes within surveys is essential.

A Hybrid Approach
The needs for stock assessments within the SEDAR process were presented by Kate Siegfried. Catch estimates affect every aspect of status determination, so there are no provisions for competing perceptions of harvest removals—an agreed upon time series of catch is essential. Theory and simulations have confirmed however that an arbitrarily scaled time series of catch is sufficient for status determination and management, as long as the estimation methodology remains the same.

John Foster continued the presentation with a series of examples. Simple ratio comparisons between state surveys and MRIP estimates for years when both surveys were conducted revealed a wide range of ratios (MRIP:State) from 1.8x for Louisiana to 11x for Texas. States with comparable programs, such as Florida and Louisiana had similar ratios, 2.7x and 1.8x respectively. Florida's surveys showed similar trends to MRIP. Alabama had a similar increasing trend, but with more variation. Mississippi also had reasonable agreement but the scaling ratio was 5.6x. For Texas and Louisiana there is only one year of overlap with MRIP so no trend comparisons are possible. Calibrations to alternative state surveys have relatively little effect on the trends, but huge differences in the totals. Moreover, the relative allocations among

states will vary with different currencies. These considerations are far from academic. Practical questions include:

- Do we use state, MRIP or composite metrics of harvest?
- Are the years used for calibration fixed, or amenable to future changes?
- Do calibration decisions vary by species?
- Will calibration decisions be updated regularly?

Participants acknowledged that calibrations could and possibly should evolve annually. Some insights into the magnitude of such changes might be found by examining ratios for individual years in instances where multiple years of overlap occur. Regardless of the calibration standard used, developing a consistent time series is perhaps the most important criteria. Currency conversion can be problematic if changes in calibrations induce changes in stock status.

Other participants discussed the importance of accuracy of the surveys in creating a standard for comparison. Comparisons with the recent independent biomass estimate of red snapper (The Great Red Snapper Count) were a possibility but there was limited discussion.   Consultants asked if the Hybrid Approach used weight trimming to reduce outliers. Such trimming is used to exclude weight values above the 95th to 99th percentile. Trimmed estimates are particularly important in management because premature fishery closure based on extreme observations are problematic for all stakeholders.

Further discussions of the Hybrid Method led participants to ask how it differed from Option 1b in the White Paper. MRIP staff explained that while the methods are similar, the Hybrid Method relaxed some of the assumptions about which surveys were more biased. Regardless of the method, participants agreed that multiple conversions between assessment results and monitoring estimates could be time consuming, hard to communicate, and possibly error prone.   Moreover, if present patterns continue, some states may be able to update their calibration ratios (e.g., FL) but in other states where MRIP is no longer conducted (e.g., TX), the ratios will remain fixed unless auxiliary information can fill gaps. A variety of scenario permutations were proposed but the primary calibration options discussed boiled down to:

- Conduct a detailed, model-based analysis to generate an aggregate measure of total catch
- Use the Hybrid Method discussed above
- Use Option 1a or 1b from the White Paper.

Consultants noted that the detailed model-based analyses would not be possible in the short term but should be considered in a longer-term research plan.

Time did not permit sufficient discussion of the remaining calibration options on Day 2 of the workshop. Instead, the topic was deferred to Day 3.

Database Storage and Data Management

Day 2 of the workshop concluded with presentations on the development of a common database and warehouse for handling regular updating. Gregg Bray's (GSMFC) presentation highlighted that in the short run, all parties should provide both raw and processed data with associated metadata. GulfFIN already serves this function for LA and TX data sans the metadata. Expansion of their role to the total set of estimates was proposed as a starting point. In the longer term, development of a standardized database and minimal key data requirements are necessary.   Development of a model-based calibration estimate would require high-resolution data rather than aggregated data typically available. Gregg Bray repeated their commitment to handling whatever level of resolution partners were willing to provide. Others acknowledged that a regional approach to calibration was best but it would require a "burn-in" period to understand the relationships among the survey data.

Matt Nuttall (SEFSC) reported the Center's need for the finest scale resolution possible including survey estimates, population totals used to calibrate the survey weights, the raw data tables on angler intercepts and biological data, and finally, the metadata for survey design. Not all demands of the data known in advance, so having the ability to reconstruct estimates as appropriate was judged a desirable feature. Therefore, all the information used to create sampling, nonresponse and calibration weights is needed. Finally, it was stressed that timing of data deliveries was essential for use by scientists and particularly managers. SEFSC endorsed the concept of expanding the role of GulfFIN for this purpose.

In the ensuing discussion, the Consultants emphasized the ancillary data associated with the design including the weights, stratum information, PSU data and so forth. At the finest scale, it is important to have the actual processing code used by the partners. Proper weighting of observations is perhaps the most important aspect of survey design; having sufficient information for auditing of this process was acknowledged difficult but essential. Proper use of multiple databases requires a deep understanding of the algorithms used for distillation of estimates. It was acknowledged that this aspect of data storage is beyond current mandates and expertise. No single group would likely have the resources to implement such a comprehensive system.

Lisa Hollensead (GMFMC) emphasized that Gulf Council's needs for management are linked to timely decision making. Moreover, such data should be readily available to all parties including the general public. The focus here is the end products rather than the atomic components of the estimates.

Day 3: February 25, 2022

Discussions about data management needs continued on Day 3. GulfFIN representatives emphasized again that it could conduct some auditing responsibilities but could not be responsible for full-scale auditing of each submission. Consultants noted that the intensive auditing was a precursor to having a useful integrated database. Without such auditing, the warehouse would become a collection of disparate, hard to compare databases. Some states

echoed concerns similar to GulfFIN—they are focused primarily on executing the designs and not sufficiently staffed to handle some of the proposed analyses.

MRIP is currently responsible for detailed documentation of the design, weighting methodology, estimation methodology, data processing designs, and quality control. The purpose of this documentation is to ensure that data are being collected according to the design. Current standards (NMFS 2020) also require annual reports which summarize data collection procedures and within year changes for each year.

All participants agreed that improvements to the data standards process are both desirable and valuable. Sound foundational principles such as the Recreational Fishing Survey and Data Standards and SEDAR stock assessment guidance documents are excellent starting points.

SERO highlighted its responsibilities for in-season monitoring and support for development of regulations for landings and discards. These data include daily catch rates, size limits, bag limits and economics of individual trips. SERO noted that many of its responsibilities spanned the entire Gulf of Mexico rather that state specific estimates.

Methods for Calibration (Part 2)

Participants and Consultants returned to the unfinished business from Day 2 to select a calibration method. The primary options under consideration are summarized in the table below. Viable options included two from the 2020 review (Options 1a and 1b, White Paper), and three proposed at this meeting. Foster's two proposals address the alternative currency issue by either converting all state-based catch estimates upward by the ratio of recent MRIP to state estimates. Conversely, one could convert historical MRIP estimates by state to their new values by multiplying them by the ratio of recent state to MRIP estimates (i.e., the inverse transformation). Model outputs would be expressed in either MRIP currency or summed state currencies. The option based on Opsomer's proposal (Erciulescu et al. 2021) to conduct an in-depth model-based analysis of each survey would be an important advance. However, because it would likely take several years to complete, this option was not considered further for short-term assessment needs. Instead, the Opsomer proposal should be considered as part of a research track for longer-term needs.

Option	Model Inputs by State		Model Outputs	Monitoring	Comments
	MRIP Data	State Data			
1a	Fully calibrated(FES, APAIS) series	La Creel converted to MRIP units. TX not addressed	MRIP units	MRIP units	LA only adjusted by ratio of overlap between years
1b	Fully calibrated(FES, APAIS) series	La Creel converted to MRIP units. TX not addressed	Converted to State units	State units	Adjusted by ratio of overlap between years
Foster: Calibrate to MRIP	Fully calibrated(FES, APAIS) series	**Converted to MRIP currency.** Used for to improve estimates for 2017+	MRIP units	MRIP units in FL, AL, MS. Ratio adjusted catches in LA, TX	State data raised by ratio in overlap period. Limited info for LA & TX. Population estimate greater.
Foster: Calibrate to State	Fully calibrated (FES, APAIS) series **converted to state currency.** Used to improve estimates for 2017+.	State only in 2017+. MRIP converted to State units before 2017. Total Catch= sum over all states	Sum of scaled State Units	State units	MRIP data reduced by ratio in overlap period. Model output is linear combination of MRIP adjusted catches. Population estimate lower.
Opsomer, AFWA	Model adjusted	Model Adjusted	New Units	New Units	New currency. Time>2 yrs

Participants recognized the need to pursue both short- and long-term solutions and affirmed their willingness to work towards imperfect but necessary short-term solutions to meet upcoming stock assessment needs. The diversity and complexity of the survey methods heightens the need for detailed specification of data conversion and increases the chances of miscommunication among stakeholders. Discussion among participants focused on identifying specific recommendations in light of these overarching concerns.

Participants noted that consistency is a central tenet of BSIA. It was argued that if state currencies were to be used over the entire assessment period, a peer review would be needed. The inability to meet this requirement in the short run pushes the group toward option 1a or 1b. Even option 2 in the White Paper document would require a peer review. Similar sentiments were expressed about the options proposed by Foster. The current meeting was not designed to be a review of the methodology. If either of the Foster approaches were to be used, an additional peer review would be necessary.

Further discussions helped clarify the scope of needed reviews. Ratio based calibrations have been reviewed in the White Paper, but there are nuances of calibrating back in time and handling different fishing modes, particularly shore-based fishing activity. Consultants have previously advised state and MRIP representatives on methods for calibration of means and standard errors. A similar follow up meeting to discuss Terms of References for stock assessments was proposed. The upcoming gag and red snapper assessments should be illustrative of the problems of calibration. The gag fishery is dominated by harvests from Florida whereas red snapper are harvested by all states in the Gulf.

After considerable discussion, a proposal to use Option 1b from the White Paper for red snapper was advanced. Participants recognized that the complexities of adjusting five state-specific

survey estimates that have varying degrees of overlap with MRIP estimates would make this a difficult task. Some surveys provide year-round estimates of total landings and discards, whereas others only provide in-season harvest estimates, making it difficult to produce a Gulf-wide calibrated time-series for released catch in state survey currencies. Additional concerns about use of Louisiana and Texas calibration ratios based on a single year of overlap with MRIP data were expressed. Presently, there is no agreed upon method for converting Texas catch data into MRIP units. Foster's approaches did include a proposed conversion factor. Representatives from Texas were not present for this discussion and concerns were expressed about the implications of rescaling the entire time series of red snapper catches to MRIP equivalents. Since the consultants were not provided the opportunity to visit and discuss the details of the Texas survey, consultants affirmed that they have not reviewed the survey procedures used in Texas.

Additional discussions by the participants led to further consideration of Option 2 for gag grouper. The above text table (with five options) reflects what was presented to the group at the meeting. Details of Option 2 may be found in the White Paper (2020). Option 2 would rescale the MRIP time series to the different state survey currencies. The White Paper (2020) noted that this premise "would be very difficult to defend on statistical grounds." Gag grouper, however, are caught primarily (~95%) in Florida and monitored by SRFS, so the need for comparable estimates across multiple states is less relevant for this stock. Hence, the group recommended that Option 2 be used for gag grouper.

Following some additional clarifications, the group consensus is as follows:
- Use Option 1b for red snapper.
- Use Option 2 for gag [see 2020 White Paper for these options].
- Convene a group to set Terms of Reference and identify individuals to conduct a formal review of the ratio-based calibration for gag and red snapper at minimum, and others as appropriate.

Components of the Transition Plan

Richard Cody gave a presentation of the Transition planning required under NMFS Policy Directive 04-114. An important aspect is the distinction between certifying a design and implementing it. Plans that are implemented as designed are capable of producing unbiased estimates. An initial draft outline of the Transition Plan was presented to the group. The basic outline is as follows:

I. Executive Summary
II. Introduction and Purpose
III. Description of Approach and Timeline
IV. Potential stock assessment impacts and schedule
V. Identification of Unknowns
VI. Lessons learned
VII. Appendices

Section III outlines the needs for a Transition Track and a Research Track as well as ongoing Communications needs. The Transition Track has both short and long-term needs. Short-term needs are dominated by the responsibilities of getting data and interim calibrations to stock assessment scientists as soon as possible. Longer term needs include convening an independent review of final calibration procedures for use in stock assessment and management, and maintenance of the FIN database. The Research Track involves progressive improvements of survey methodologies and ancillary studies.

Section V on Identification of Unknowns addresses the potential reduction in non-sampling errors, the potential alignment among surveys, and whether full integration of the survey data (e.g., composite estimation) is possible. Funding and resource limitations are a major unknown. Process obstacles such as the Paperwork Reduction Act are non-trivial.

The Plan outline will be fleshed out and circulated to the Transition Team in a timely manner for review and comment. Participants echoed the urgency of finalizing the Transition Plan. Avoiding species- and assessment-specific approaches should also be a goal of the Transition Plan. Owing to the complexity of the underlying surveys, it is anticipated that expertise outside the SEDAR process would be necessary.

Communications Plan

An outline of a communications plan was presented to the group by Catherine Krikstan (OST). Overall goals of the plan are to facilitate development of the transition plan and its implementation. The plan will help ensure that workshop outcomes are clear, easy to find and promptly distributed to target audiences. Progress on milestones will be regularly published. Target audiences include not only transition team members and leaders of their respective agencies but also external stakeholders: Council members, congressional representatives, anglers and various advocacy groups.

Immediate next steps will be development of short-term tactics and development of a formal communications plan for inclusion in the Transition Plan. Participants should nominate members for the Communications Working Group.

Closing Statements

Participants were reminded of a number of remaining tasks. Notably, the Consultants planned to convene following the meeting to develop a list of final recommendations. Their report has been included its entirety in Appendix 3. Consultants noted that not all of future decisions and recommendations will be statistical. Participant consensus will be essential for those non-statistical elements driven by assessment or management needs.

The initial goals of the workshop were largely met. The primary exception was development of a statement of task for the final independent review of methods. This will be addressed later by the transition team after the major tasks and studies have been identified.

Participants and consultants appreciated the planning that occurred before the workshop and the presentations by the states and MRIP staff to the Consultants immediately prior to the workshop. This ensured a better understanding among all participants of the distinctive methods and concerns of each group. Presentations during the meeting were uniformly excellent and stimulated highly engaged discussions among the Participants and Consultants. There was a general sense that the workshop had been successful, especially with respect to agreeing on a framework for satisfying data needs for the upcoming gag and red snapper assessments.

List of Presentations and Useful Links

All of the presentations and a recording of the meeting may be found at:
https://www.fisheries.noaa.gov/event/gulf-state-recreational-catch-and-effort-surveys-transition-workshop

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

National Academies of Sciences, Engineering, and Medicine (2006) Review of Recreational Fisheries Survey Methods. Washington DC: The National Academies Press. https://www.nationalacademies.org/our-work/review-of-recreational-fisheries-survey-methods

National Standard 2 https://www.ecfr.gov/current/title-50/chapter-VI/part-600/subpart-D/section-600.315

References

American Association for Public Opinion Research. 2016. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th edition. AAPOR. https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf

Erciulescu, A., J.D. Opsomer, and F.J. Breidt (2021). "A Bridging Model to Reconcile Statistics Based on Data from Multiple Surveys." *Annals of Applied Statistics,* 15: 1068-1079 (DOI: 10.1214/20-AOAS1437).

Foster, J., Papacostas, K., Krikstan, C., Andrews, R., Kitts-Jensen, R., Cody, R., Colvin, G., Dolinger Few, L. (2021) Evaluating Uncertainty in Gulf Red Snapper Estimates: A Preliminary Sensitivity Analysis of Non-Sampling Errors in the Region's Recreational Fishing Surveys. https://media.fisheries.noaa.gov/2022-02/NOAA-Fisheries-OST-Preliminary-Sensitivity-Analysis-Non-Sampling-Errors-Gulf-Recreational-Fishing-Surveys.pdf

Gulf States Marine Fisheries Commission and National Marine Fisheries Service (2020) GSMFC-NOAA Fisheries Workshop on Red Snapper – Workshop Summary Report. https://gulfcouncil.org/wp-content/uploads/B-8b-Gulf_Calibration_Wrkshp_report_2020_v1.21.pdf.

House Committee on Appropriations (2021) Joint Explanatory Statement of the Committee of Conference on H.R. 133: Title I, Division B. House Rules Committee on Senate Amendment to the Consolidated Appropriations Act of 2021. *Congressional Record* 166:218 (Book III). https://www.govinfo.gov/content/pkg/CREC-2020-12-21/pdf/CREC-2020-12-21-house-bk3.pdf.

Magnuson-Stevens Fishery Conservation and Management Act, P.L. 94-265, as amended through P.L. 109-479. (2006) https://www.congress.gov/bill/109th-congress/house-bill/5946.

National Academies of Sciences, Engineering, and Medicine (2021) Data and Management Strategies for Recreational Fisheries with Annual Catch Limits. Washington, DC: The National Academies Press. https://doi.org/10.17226/26185.

National Marine Fisheries Service (2020) Recommended Use of the Current Gulf of Mexico Surveys of Marine Recreational Fishing in Stock Assessments. Silver Spring, MD. https://media.fisheries.noaa.gov/dam-migration/94100569.pdf.

National Marine Fisheries Service (2020) Recreational Fishing Survey and Data Standards. Silver Spring, MD. https://www.fisheries.noaa.gov/recreational-fishing-data/recreational-fishing-survey-and-data-standards

National Marine Fisheries Service (2021) Policy 04-114: Implementing Recreational Fisheries Catch and Effort Survey Design Changes. https://media.fisheries.noaa.gov/2021-06/04-114_2021.6.9_final%20for%20Doreumus%20Signature_signed.pdf?null

Paperwork Reduction Act, P.L 104-13. (1995) https://www.congress.gov/104/plaws/publ13/PLAW-104publ13.pdf

Appendices

Appendix 1. Agenda

WORKSHOP ON TRANSITION TO USE OF STATE SURVEY CATCH ESTIMATES IN GULF OF MEXICO FISHERIES

Workshop Dates: February 23-25, 2022; all times approximate

AGENDA

Feb. 23: 9:00 am - 5:00 pm EST

I. 9:00 am - 9:15 am: Meeting logistics, introductions: **Gregg Bray**, **Richard Cody**
    A. Virtual connection instructions and where to get help.
    B. Self-introduction of attendees

II. 9:15 am - 9:45 am: Review Workshop Purpose, Desired Outcomes, Agenda: Workshop Chair, **Paul Rago**
    A. Overarching goal: Agree on the elements of a transition plan that will achieve goals 1-5. (This plan will be written and executed following the workshop).
    B. Long-Term Goals of the Transition Plan
        1. Identify any needed design changes to improve the accuracy of all survey programs, thereby minimizing differences in estimates.
        2. Incorporate state data into the federal science and management process while maintaining the needed consistent, regional time series.
        3. Develop a single, publicly accessible, standardized database to house all the recreational fishing data streams in the Gulf of Mexico.
        4. Develop guidelines and best practices to inform future decision-making regarding BSIA when overlapping statistically valid data streams exist (note: while the transition team will be developing guidance, it does NOT have the authority to make BSIA determinations).
        5. Maintain clear and open lines of communication between the Transition Team and all affected stakeholders about progress toward the above goals.
    C. Specific Desired Outcomes of the Workshop
        1. Toward Goal 1:
            a) Adopt a research and analysis plan for understanding the drivers of differing catch estimates among the Gulf red snapper surveys.
            b) Adopt Terms of Reference for a Congressionally directed, independent peer review (that will be commissioned after the research and analysis plan has been implemented).

    2. Toward Goal 2:
        a) Decide on a <u>calibration approach</u> to be used to 1) inform new stock assessments happening this year (e.g. ratio-based methods), and 2) to inform stock assessments and management in the longer term (e.g., model-based methods).
        b) Describe and evaluate a <u>hybrid approach</u> that puts the transition on two tracks: (1) initiate transition using available data for stock assessments in the near term; (2) complete research plan and revise data collection to inform future assessments.
    3. Toward Goal 3:
        a) Determine <u>technical and funding requirements</u> to create the unified, regional database and <u>assign specific roles and responsibilities to NMFS, states, and regional partners</u> to initiate its development.
        b) <u>Identify needed updates to data flow processes</u> for stock assessments.
    4. Toward Goal 4:
        a) <u>Review data requirements</u> of National Standard 2.
    5. Toward Goal 5:
        a) Agree on the goals and components of <u>a communications plan</u> (to be adopted following the workshop).

III.   9:45 am - 10:15 am: Break

IV.   <u>10:15 am - 11:30 am: Transition Plan Content</u>: **Richard Cody**
    A. Present Work Group's list of Milestones that are to be addressed in a Transition Plan;
        1. Include description of hybrid approach (see also VI C below) that puts the transition on two tracks: (1) initiate transition using available data; (2) complete research plan and revise data collection designs/calibration/integration as indicated by the results;
    B. Group discussion facilitated by Workshop chair. Intent is to achieve a general consensus of the Transition Team Subgroup members on the content of the transition plan, and timing for completion of the various milestones.

V.   11:30 am - 2:00 pm: Lunch Break

VI.   2:00 pm - 3:00 pm: Review of Data Requirements to meet NMFS requirements for management and stock assessment:
    A. Management requirements: **Andy Strelcheck**
    B. Science requirements, including SEDAR 74 specific needs: **Kate Siegfried**
    C. BSIA: **Patrick Lynch**

VII. 3:00 pm - 3:30 pm: Break

(Start on item VIII in the afternoon of Day 1, then continue the discussion in the AM on Day 2)

VIII. 3:30 m - 5:00 pm: Survey Research Roadmap: **MRIP Consultants** and Workshop Chair **Paul Rago**
   A. Independent review of surveys called for by Congress
      1. Alignment of the mandate with requirements for transition
   B. Summary of state and NMFS pre-briefings to consultants: **Gregg Bray**

Feb. 24: 9:00 am - 5:00pm EST

(VIII cont.) 9:00 am - 11:00 am
   C. Overview of non-sampling errors and their effects on survey estimates: **MRIP Consultants**
      1. General overview of non-sampling errors and effects with examples from Gulf surveys
   D. **MRIP Consultants** will present a proposed Roadmap that identifies research, pilot studies, and analyses--related to sources of non-sampling error and bases for differences in estimates among the Gulf Surveys (see detailed statement of task for this product appended to this agenda).
   E. Group discussion of Roadmap facilitated by Workshop Chair:
      1. Achieve consensus of attendees regarding scope and needs for completion of identified research.
      2. Agree on roles for partners in executing research plan.
      3. Discuss funding options (limiting the discussion to identifying totals that might be needed to fund research projects), sequencing, and timing.
      4. Discuss Terms of Reference Draft prepared by NMFS/OST

IX. 11:00 am - 11:30 am: Break

X. 11:30 am - 12:30 pm: Calibration I
   A. Review of current status of calibration method development and use: **Richard Cody**
   B. Overview of calibration methods, limitations and future use considerations: **MRIP Consultant(s)**, including requirements for enabling integration of calibrated state estimates into a single Gulf-wide estimate for the affected species: **Jean Opsomer; Lynne Stokes**, **other member(s) of the consultant team**?
   C. Potential Hybrid Approach whereby all available rec catch series are used in assessments while research is ongoing, including different ways the data might be incorporated into the assessment models or how outputs from separate model runs might be integrated into final results (**Foster, Siegfried, Nuttall**).

XI. 12:30 pm - 2:00 pm: Lunch Break

XII. 2:00 pm - 3:30 pm: Calibration II
    A. Facilitated discussion of preferred options for pursuing integration of catch estimates: **Paul Rago**, facilitator
        1. Seek group's consensus on option(s) to implement.
        2. Identify requirements and partner needs for implementation. Note that Reconciliation of completing needs will need to be a primary consideration. There may be options to produce a Gulfwide estimate that considers all of the data and allows us to make progress on transitioning while attending to the research needs that will help identify improvements to be made. Potential options (maybe at the assessment level) beyond standard statistical integration methods and associated assumptions and requirements will need to be evaluated.
        3. Identify calibration requirements for species with state specific or restricted distributions (e.g., gag).
    B. Facilitated discussion of preferred options for calibration, including phased approaches (e.g., using a ratio based approach in the interim while a model based approach is being developed): **Paul Rago**, facilitator
        1. Seek group's consensus on option(s) to implement.
        2. Identify requirements and partner needs for implementation.

XIII. 3:30 pm- 4:00 pm: Break

XIV. 4:00 pm - 4:30 pm: Additional Requirements for Completing Transition
    A. Database Storage and Data Management via FIN
        1. Input and Data Management Requirements: **Gregg Bray**
        2. Output Requirements: SEFSC (**Matthew Nuttall**); SERO: **SERO** requirements for data outputs will be covered in Andy Strelcheck's presentation under VI.A on 2/23--additional questions can be addressed here; Gulf Council (**Lisa Hollensead; Ryan Rindone**?)
        3. Roles and responsibilities of Data Providers/State Survey Managers
    B. Identify and resolve how and when to address any unmet requirements for use of estimates from State Survey Certification Decision Memos **Richard Cody**:
        1. Generally, these include the following for each survey:
            a) integration (see VII.A.2: not completed for any of the Gulf surveys at present,
            b) calibration with the legacy time series,
            c) peer review: Note that, for the Gulf state surveys, this requirement occurs at two levels:
                (1) Initial independent expert review conducted for MRIP Certification, completed for FL, AL, MS, LA surveys;
                (2) independent review of all the current Gulf surveys required by Congress per agenda item V above.

        d) apply to time series
      2. Also address any individual, specific next steps from Certification Memos, including SnapperCheck Conditions of Certification requirements.
    C. Others?

XV.   4:30 pm - 5:00 pm: Components of the Transition Plan
    A. Identify a small group to prepare the Transition Plan Outline, based on the work group's discussion, in an overtime session the evening of Day 2.
    B. Review and discuss the proposed outline

Feb. 25: 9:00am - 1:00pm EST

XV.   9:00 am - 10:30 am: Components of the Transition Plan (continued)
    C. Presentation of the Transition Plan Outline by the assigned group.(XV.B)
    D. Discussion of the proposed components to seek general concurrence with that list and of a process and timing to scale it up to a full Transition Plan for submission to NMFS by the Gulf Subgroup.
    E. Revisit long-term goals and desired outcomes of workshop

XVI.   10:30 am - 11:00 am: Break

XVII.   11:00 am - 12:00 pm: Communications Plan: Discuss, agree on plan components:
Catherine Krikstan
    A. Among Transition Team Members
      1. Goals, purpose of plan
      2. Key tactics
        a) Regular (monthly?) ad hoc State Directors' Meetings
        b) Bimonthly meetings of MRIP CET Regional Work Group
        c) Assemble Gulf Transition Team Communications Work Group consisting of representatives from NMFS, Gulf states, Gulf council, and GulfFIN.
        d) Briefings at Gulf Council and GSMFC meetings
      3. Other?
    B. Stakeholders: same items as above?

XVIII.   12:00 am - 1:00 pm Wrap Up
    A. Final advice and recommendations from MRIP Consultant Team
    B. Closing comments by Transition Team Gulf Subgroup members

Appendix 2: List of Participants

First Name	Last Name	Affiliation
Jason	Adriance	Louisiana Department of Wildlife and Fisheries
Rob	Andrews	NOAA Fisheries OST
Kevin	Anson	Alabama Department of Conservation and Natural Resources
Patrick	Banks	Louisiana Department of Wildlife and Fisheries
Scott	Bannon	Alabama Department of Conservation and Natural Resources
Luiz	Barbieri	Florida Fish and Wildlife Conservation Commission
Dave	Bard	NOAA Fisheries OST
Harry	Blanchet	Louisiana Department of Wildlife and Fisheries
Kevin	Bland	Louisiana Department of Wildlife and Fisheries
Gregg	Bray	Gulf States Marine Fisheries Commission
Mike	Brick	Consultant, Westat
Richard	Cody	NOAA Fisheries OST
Gordon	Colvin	NOAA Fisheries OST
David	Detlor	NOAA Fisheries OST
Jill	Dever	Consultant, RTI International
John	Foster	NOAA Fisheries OST
John	Froeschke	Gulf of Mexico Regional Fishery Management Council
David	Gloeckner	NOAA Fisheries SEFSC
Marie	Head	Alabama Department of Conservation and Natural Resources
Matt	Hill	Mississippi Department of Marine Resources
Lisa	Hollensead	Gulf of Mexico Regional Fishery Management Council
Evan	Howell	NOAA Fisheries OST
Cliff	Hutt	NOAA Fisheries OSF
Catherine	Krikstan	NOAA Fisheries OST
Michael	Larkin	NOAA Fisheries SERO
Virginia	Lesser	Consultant, Oregon State University
Ty	Lindsey	Louisiana Department of Wildlife and Fisheries
Alan	Lowther	NOAA Fisheries SEFSC
Patrick	Lynch	NOAA Fisheries OST
Vivian	Matter	NOAA Fisheries SEFSC
Jessica	McCawley	Florida Fish and Wildlife Conservation Commission
Trevor	Moncrief	Mississippi Department of Marine Resources
Matt	Nuttall	NOAA Fisheries SEFSC
Jean	Opsomer	Consultant, Westat
Katherine	Papacostas	NOAA Fisheries OST

Clay	Porch	NOAA Fisheries SEFSC
Paul	Rago	Workshop Chair, Scientific and Statistical Comm., MAFMC
Ryan	Rindone	Gulf of Mexico Regional Fishery Management Council
Michael	Ruccio	NOAA Fisheries OSF
Bev	Sauls	Florida Fish and Wildlife Conservation Commission
Kate	Siegfried	NOAA Fisheries SEFSC
Carrie	Simmons	Gulf of Mexico Regional Fishery Management Council
Joe	Spraggins	Mississippi Department of Marine Resources
Lynne	Stokes	Consultant, Southern Methodist University
Andy	Strelcheck	NOAA Fisheries SERO
Matt	Titlow	NOAA Fisheries OST
Mike	Travis	NOAA Fisheries SERO
Joe	West	Louisiana Department of Wildlife and Fisheries
Chris	Wright	NOAA Fisheries OSF
Xinan (Adrian)	Zhang	Louisiana Department of Wildlife and Fisheries

Appendix 3: Recommendations of Statistical Consultants

Summary of Recommendations Based on Transition Team Gulf Workshop
Meeting
February 23-25, 2022

Mike Brick, Jill Dever, Virginia Lesser, Jean Opsomer, Lynne Stokes
March 28, 2022

Overall Comments

The goal of the state surveys is to increase the precision of estimates of catch for one or more species of high interest to the state. The state surveys have mostly achieved this goal, in the sense of increasing the sample sizes and greatly reducing the variance of the estimates. However, sizeable non-sampling errors, which vary between the states, may be present in the estimates. Virtually all surveys conducted today are subject to non-sampling errors, so this does not imply that these surveys are not scientifically valid. Nevertheless, the fact that the types and magnitude of the non-sampling errors differ between states and are different from those in MRIP greatly complicates directly comparing the estimates obtained from the surveys as well as obtaining comparable data that when combined across the states allow production of Gulf State-wide estimates.

There are two broad approaches to improving comparability between the surveys: (1) identifying and removing sources of non-sampling error, and (2) harmonizing the methods (questionnaires and other data collection methods, sampling design, sampling frame, estimator construction, etc.) used in the surveys so that the *differences* in non-sampling error are reduced. Approaches (1) and (2) can be pursued in parallel. We believe, however, that implementing (1) alone is unlikely to sufficiently remove differential impacts of non-sampling errors.

One approach to eliminate differences among the states is to adopt a single sampling design for all states. Using an identical design has the best chance of all options available to eliminate the differential errors across states that create challenging comparability problems. In order to also achieve the precision goals of the states, the sample sizes could be increased under this design relative to those currently used in MRIP, while maintaining comparability. While making the designs identical is likely not realistic, we still want to state this option, to make clear that obtaining directly comparable estimates is achievable, at least in principle.

In the remainder of this report, we have assumed this option is not available, and formulated our recommendations accordingly.

Recommendations

1. *Differences in estimates*: As discussed at the meeting, the differences between estimates obtained under various methods used by the states and MRIP is not unusual. Other agencies have seen this issue and have addressed it. Some agencies report their estimates and recognize another agency, with different methods, reports another estimate. Examples of discrepant measures by federal agencies are unemployment estimates from states and federal (BLS) and from the population (Census).

   Another example is one where different federal agencies measure a parameter differently such as poverty measures; sometimes American Community Survey (ACS) estimates have gone in the opposite direction from the Current Population Survey Annual Social and Economic Supplements (CPS ASEC) official estimates. Census has chosen to handle this by advising users to work with the CPS ASEC estimates for national numbers, the ACS for state numbers, and yet a third source, the (modeled) Small Area Income and Poverty Estimates (SAIPE), for substate estimates. SAIPE estimates are used for Title 1 formula funding; ACS estimates are used in some programs; and CPS estimates are not used in funding (although they are the "official" estimates) because the sample size is too small. As indicated, this is an administrative decision.

   Drawing a parallel to the recreational fishery situation, one could use the MRIP estimates to allocate fractions of annual catch to the states (with LA and TX still handled separately), because it is the survey that is most comparable across states, while a different approach is used to determine total annual catch, for instance based on the state surveys. We are not necessarily advocating this particular option, but want to point out that it is possible to "pick and choose" among the estimates for different purposes, as long as it is consistently done and well documented.

2. *Review weighting methodology for creating final weights used in estimation*: This includes reviewing the process for MRIP (both APAIS and FES) and the details used to generate weights for each state, to make sure that they correspond to the sampling design and correctly adjust for nonresponse. This would include both the effort and intercept weights for those states that use both. Simple checks, such as weights should sum to universe size, should be made. Weight trimming for unusually large weights should be considered and documented if not already implemented.

   The population of saltwater anglers is not defined in official statistics. However, MRIP survey weights are adjusted to align with official population-based household estimates, consequently producing estimated population counts by state. In contrast, most states use sampling frames based on licenses and/or permits. Nevertheless, it might be useful to compare estimated population totals from the states and those from MRIP to determine if they align. Substantively large differences that cannot be explained fully by license frame under coverage (e.g., state regulations may not require a license for certain age groups) would suggest an evaluation of the weighting adjustments for the

states.

3. *Variance estimation procedures*: In addition to a review of the weighting methodology, the associated variance estimation procedures should be reviewed, to ensure that they appropriately reflect the stratification and multi-stage nature of the sampling design. The survey estimates and associated measures of precision should be computed using software that accounts for the design. The SAS "survey" procedures or the R survey package are examples of such software.

4. *Study documentation*: Each state should review the NOAA Fisheries reporting standards and provide the corresponding information for the programs used in their state. There were a number of questions we had from the state presentations. Nearly all of our questions would be answered if we saw such details. The standards appear on this website.

   https://www.fisheries.noaa.gov/recreational-fishing-data/recreational-fishingsurvey-and-data-standards

5. *Review of measurement error*: These are questions that would be addressed in the documentation mentioned in #4 above.
   a. Does each state have a QA/QC program in place and is a document available for review? We suggest this documentation, if not already, be available for review.
   b. Are the same methods used by all states to train the interviewers?
   c. Is there a quality assurance check on interviewers assigned to a site? For example, does a supervisor subsample a set of interviews for each interviewer regularly to assure the data are collected at the selected site at the assigned time.
   d. Does each state have a protocol for evaluating and addressing survey responses for extreme values?

   Since we do not know how the procedures differ from state to state, we do not have evidence to suggest whether or not these procedures could be a substantive source of differences. However, if discrepancies do exist, they are examples of the types of procedures that could be harmonized across states to reduce differential non-sampling errors.

6. *Response rates*: All states and MRIP should use the same formula for calculating response rates and be documented. See the American Association of Public Opinion Research website for standard methods. We propose to use RR1. Refer to the following website for this information on response rates.

   https://www.aapor.org/Education-Resources/For-Researchers/Poll-Survey-FAQ/Response-Rates-An-Overview.aspx

7. *Programs with both effort and intercept survey (Florida, Louisiana, and MRIP)*: To further evaluate effort estimates, compare the questionnaires used by each state with the FES questionnaire. Consider splitting a sample into two groups as part of an experiment to examine the effect of questionnaire differences. For example, send the FES questionnaire to half of the sample and the state questionnaire to the other half of the sample. Compare effort estimates for both approaches. Another approach to examine effort is to send the state questionnaires to a sample of households (the current FES frame) and compare the effort estimates (run at the same time) with those households receiving the FES questionnaire. This type of surveys will likely require additional size to ensure sufficient precision of both the original and the experimental results.

8. *Finding comparable components within the state surveys and MRIP*: The estimates produced by the state surveys differ substantially from the MRIP estimates, because of differences in sampling designs, estimation methods and/or survey modes, making it difficult to identify and quantify all the possible sources of non-sampling error. It might be possible to identify *components* of the state surveys that are more comparable, and which could be used to create harmonized estimates. The most promising such component is the intercept surveys, which all the states and MRIP are conducting in person, and which have mostly similar sampling designs for site-days. It is possible to obtain estimates of catch directly from these intercept surveys; these estimates are not as precise as either the capture-recapture or the two-survey estimates, but they are likely to be less impacted by the differential non-sampling errors. We recommend investigating whether this would provide a way to create a consistent benchmark for the relative catch of the states.

9. *Large Pelagic Survey model*: As already mentioned in the general comments above, another option to eliminate the comparability issues is to have a single survey that covers all states. A possible model for this is the Large Pelagic Survey (LPS) on the Atlantic, which is a separate survey targeted at a small number of rare species of high interest that are not adequately addressed by the general MRIP surveys.

10. *Florida*: Because of the similarity of the Florida and MRIP data collections systems on the intercept survey, the state's data provide an excellent testbed to examine the causes of the varying effort estimates. Experiments such as those described in #7 above to compare GRFS and FES should be especially revealing here. Examine each questionnaire and conduct side by side studies changing the questionnaire slightly to examine the impact of a question wording, of the order of questions, etc. Such experiments would provide valuable information on how sensitive the results are to these questionnaire design aspects. Subsequently, it would be possible to select a "standard" questionnaire that can be used across states for eliciting fishing effort.

11. *Mississippi survey*
   a. Review the estimation process to ensure that weights (both base weights and any nonresponse adjustments) are calculated and used correctly for the estimators themselves and the standard error estimates.
   b. One source of non-sampling error that is of concern for areas with large fraction of trips from non-public sites, as Mississippi has, is coverage error. The capture recapture estimator relies for approximate unbiasedness on the assumption that both reporting rate and CPUE are identical for trips from public and private sites. To identify whether or not a violation of these assumptions is causing substantial non-sampling error, it is imperative that they be examined. Consider methods for using the law enforcement data for this purpose. We realize these data are based on a non-probability sample. However, we encourage the state to set up a protocol for law enforcement to collect data from a sample of boats that are as representative as possible for a short period. For example, perhaps they could be instructed to collect data from a systematic sample (every *n*th boat) or a rotating location for a specified data collection period. From these stops, data allowing estimation of reporting rate and average catch should be collected, as well as their return site (public/private). From these data, estimates of differences can be produced to assess the impact on non-sampling error due to coverage error. (See Stokes *et al* (2021) for method).

12. *Alabama*
   a) Do not use catch as a matching variable. When catch is used as a matching variable, it eliminates the possibility of adjustment for misreporting of catch by anglers. Since it is unlikely that anglers never misreport, this provides an inaccurate measure of uncertainty for the final estimates. Evaluate the impact on estimates for removing that criterion as a matching variable.
   b) A source of non-sampling error for capture recapture estimators is matching error. Misclassification of trips as reported or not can have a large effect on the bias, so minimizing the misclassification errors is critical for this methodology. This is not a problem for Mississippi's system because it requires pre-trip registration, so is a unique problem among the Gulf States for Alabama's program. When catch is eliminated, there will be fewer matching variables and therefore less distinguishable trip profiles. If this is a problem, consider adding additional distinguishing variables to the reporting app. Two examples are: (1) which trip of the day (which would also have to be added to the intercept survey); (2) return site (from a dropdown list in the app).
   c) Another source of non-sampling error for capture recapture estimators is lack of independence between the two reports. In the capture recapture implementation, this can only occur if reporters are aware that they are to be intercepted before they report (since the "recapture" sampling units are actually randomly selected). Collecting data about how often this occurs by observation alone is impossible since electronic communication among anglers may be possible. One possible approach is to conduct an experiment in which these data are collected by survey. That is, add

questions to the intercept survey about whether anglers are aware that they are required to report; if they reported their trip prior to interception; and when they became aware that samplers were collecting data for a particular trip. A comparison of the reporting rate and reported catch profile of those with and without prior knowledge of intercept could provide insight into whether this source of non-sampling error is a problem. Since the questions are sensitive, some experimentation to discern what is the most effective way to elicit accurate responses to these questions will be necessary.

d) As noted in #11, the differential between reporting rate and CPUE between public and private sites is a potential source of bias for the capture recapture estimators. Alabama does not have the ability to use law enforcement data to investigate this, so some additional approaches are suggested here. These would also be available approaches for Mississippi.

(1) Ask reporters to provide location of their return in a way that allow analysts to know if the site is or is not on the intercept frame. This would allow comparison of reported CPUE for public and private sites, at least among reported trips. Some experimentation is likely needed to elicit this information in a format that is understandable to the angler and usable by the analyst for categorizing sites. Two possibilities are to request latitude and longitude, or to select Other on a drop down list of sites.

(2) Assessing differential trip reporting rates between public and private sites is more difficult, since data from reporters as suggested in (i) clearly will not allow reporting rates to be compared. To make such a comparison, some alternative source of data is required. One possibility is a passive data collection system for a sample of boats, such as that of the pilot survey conducted by CLS America in the Gulf of Mexico in 2016-17. (See Stokes et al (2021) for discussion of how these data were used for this purpose.) A second possibility is an address-based (or license frame based) survey asking anglers to report return location and reporting status for recent trips.

13. *Texas*

a. Since a review of the sampling design and procedures was not conducted for Texas, we have no details to assess whether the methodology used by the State of Texas is accurately describing a probability sampling design (Nuttall and Matter, 2020). The lack of detailed documentation on the procedures used by Texas and the inability to obtain a statistical review of the Texas methodology is limiting the ability to obtain gulf-wide estimates and the partitioning, since this number is not available. Additionally, we are unclear if data are collected for private areas or shore fishing.

b. Given the inability to assess the methodology, the estimates from Texas are only an index and should not be considered a scientifically based estimate.

c. Since we have no sampling error to evaluate, we also have no non-sampling error to evaluate.

14. *Simulations*:
    a) Estimate effort using the intercept data (associated with public-access only) and its site-day sampling design, and compare it with the effort estimates obtained by the effort surveys. This estimate should link to public access effort. Discuss how different these two approaches are and whether the differences are plausible.
    b) *Rare species*. Define rare species: this should include rare in terms of time, space and sample size.

    > i. For all surveys using the product estimator to obtain catch (e.g., FL, LA and MRIP), determine a rare species. This should be done for one group, so it is all done the same way. Assuming the data and the weights currently being used, produce estimates and the standard error using the formulae adopted in MRIP. Are the estimates subject to large variability? Are the standard errors appropriate for these rare species using the sample sizes typically observed? With the weighting structure currently in place, when does it give "good" and "bad" answers, from the perspective of useability and comparability of the estimates.

    > ii. For states and MRIP that use a product estimator, it is possible to have an outlier for the product that is not an outlier for either one of the surveys. A weight given on one of these estimates is then multiplied by the other estimate which could create an outlier. Investigate the circumstances for this to occur in the simulations.

    > iii. Using the simulations, how well does the product estimation work? Are there conditions which give rise to unstable estimates?

15. *Calibration*. We understand some of the concerns regarding the use of a simple ratio for calibrating state-level estimates. It does not account for the differences in precision and non-sampling errors between the different state surveys. Because of differences in data availability on which to compute adjustment ratios, even the adjusted estimates are not directly comparable: unlike the other states, Louisiana has a single historical year of overlap between MRIP and its state survey and Texas has no overlap at all (nor a valid state survey to begin with). Despite these concerns, a simple ratio adjustment provides a way forward in the short term, and is also relatively transparent to data users. We suggest further examination of this approach, including the development of a variance estimator and, at least for the states continuing with MRIP, an investigation of what time and location scales to use as the basis for the ratio (annual vs. by wave, multi-year or not, state vs. substate).

    It is possible to use a more rigorous approach to develop a set of estimates that fully take advantage of all the available data sources. This involves creating a statistical model that uses the *data* available in the different surveys, instead of relying only on the

final state estimates. This is similar to what was done for the FES-CHTS calibration as well as for a more recent project for the 2016 FHWAR (Erciulescu *et al*, 2021), in which two versions of the survey were combined to create a new set of model-based estimates that were compatible with earlier versions of the survey. Creating such a model is likely a challenging and time-consuming statistical endeavor, so that estimates would not be available for 1-2 years while the method is being developed (this is a guess on our part at this point). However, subsequent estimation would not take as long, since it involves re-fitting the same (or very similar) model and ensuring that its model diagnostics remain acceptable. An advantage of this modeling approach is that it would provide a statistically rigorous way to combine the data from the different surveys, so that the estimates take advantage of the larger sample sizes and precision of the state surveys, while also being available in a single "currency."

References

Erciulescu, A., J.D. Opsomer, and F.J. Breidt (2021). "A Bridging Model to Reconcile Statistics Based on Data from Multiple Surveys." *Annals of Applied Statistics,* 15: 1068-1079 (DOI: 10.1214/20-AOAS1437).

Nuttall, M.A., Matter, V.M. (2020) Texas Parks and Wildlife Department's Marine Sport-Harvest Monitoring Program Metadata. SEDAR70-WP-03. SEDAR, North Charleston, SC. https://sedarweb.org/docs/wpapers/S70_WP_03_TXPWD_metadata.pdf

Stokes, L., B. Williams, R. McShane, and S. Zalsha (2021). "The Impact of non-sampling errors on estimators of catch from electronic reporting systems." *Journal of Survey Statistics and Methodology* 9, 159-184 (2021). (doi.org/10.1093/jssam/smz042)

Appendix 4: List of Abbreviations

AAPOR American Association of Public Opinion Research

APAIS Access Point Angler Intercept Survey

BSIA Best Scientific Information Available

CPUE Catch Per Unit (of Fishing) Effort

FES Fishing Effort Survey

FIN Fisheries Information Network

CHTS Coastal Household Telephone Survey

GMFMC Gulf of Mexico Fishery Management Council

GSMFC Gulf States Marine Fisheries Commission

GulfFIN Line item designation for Fisheries Information Network in Gulf of Mexico

MAFMC Mid-Atlantic Fishery Management Council

MRIP Marine Recreational Information Program

MSA Magnuson Stevens Act

NMFS National Marine Fisheries Service

NOAA National Oceanographic and Atmospheric Administration

OSF NOAA Fisheries Office of Sustainable Fisheries

OST NOAA Fisheries Office of Science and Technology

SEDAR Southeast Data, Assessment, and Review

SERO Southeast Regional Office

SEFSC Southeast Fishery Science Center

SRFS (Florida) State Reef Fish Survey