Pilot Test of a Dual Frame Two-Phase Mail Survey of Anglers in North Carolina

Final Report

November, 2010

This report was prepared for the National Marine Fisheries Service, National Oceanic and Atmospheric Administration by William R. Andrews (NOAA), J. Michael Brick (Westat), Nancy Mathiowetz (University of Wisconsin-Milwaukee) and Lynne Stokes (Southern Methodist University) with the assistance of Dong Lin (Southern Methodist University). Data collection for the project was completed by ICF/Macro, Seth Muzzy, Project Director.

Introduction and Study Objectives

The National Marine Fisheries Service of the National Oceanic and Atmospheric Administration (NOAA Fisheries) administers several ongoing data collection efforts designed to estimate saltwater fishing participation (number of people who went marine recreational fishing at least once within the calendar year), fishing effort (number of angler trips), and catch (numbers of finfish caught, harvested, and released) in the U.S. The Marine Recreational Fisheries Statistics Survey (MRFSS) is a nationwide program with two independent components, a Coastal Household Telephone Survey (CHTS) to assess fishing effort, and an access-point intercept survey to assess catch per unit effort. Data from the two surveys are combined to estimate total fishing effort, participation, and catch by species.

In a review of the MRFSS conducted by the National Research Council (NRC, 2006), panel members suggested major revisions of the methods used in data collection. In particular, the CHTS, a random digit dial (RDD) survey of households, was criticized because of its under-coverage and inefficiency. The CHTS design suffers from inefficiency, due to the low rate of saltwater angler participation among the general population, as well as potential coverage bias, due to its sampling only coastal county residences and landline-based telephone numbers. The NRC report endorsed mandatory registration of all saltwater anglers. In the absence of a complete registry, the NRC recommended dual-frame procedures, and suggested sampling from incomplete lists of saltwater anglers (e.g. state saltwater license databases) and state resident or household frames (e.g. RDD frames or address-based sample frames).

The three major sources of under-coverage in the current CHTS are (1) households that do not reside in the coastal counties, (2) coastal county households without landline telephone service

(Blumberg and Luke (2010) estimate this at 26.5% of U.S. households at the end of 2009), and (3) coastal county households with landline numbers that are excluded in standard RDD list-assisted samples (Fahimi, Kulp, and Brick (2008) estimate about 20% of all landline telephone households are not in the standard RDD frame). The current survey approach accounts for under-coverage of the CHTS sample frame by adjusting estimates upward using expansion factors derived through the independent access-point intercept survey. The NRC (2006) indicated that these expansion factors are susceptible to a variety of errors.

Besides its potential coverage error, the CHTS is inefficient, as a small percentage of households participate in marine recreational fishing. As noted by the NRC report:

> Random digit dialing, even limited to coastal county residences, is not the most efficient way to gather angler effort information. In urban areas, less than 1 in 20 of the telephone intercepts reaches an angler. Improving the process whereby anglers are identified and contacted would not only improve the quality of the estimates but should also reduce costs. Remedies exist for other inefficiencies as well. For example, under the current sampling regime, identifying an angler costs more than the taking of information once the angler has been identified. (NRC, 2006, p. 30)

To compensate for the shortcomings of the CHTS, NOAA Fisheries has developed a dual-frame telephone survey approach that integrates the CHTS with surveys that sample from lists of licensed anglers. These angler license directory surveys (ALDS) are more efficient than the CHTS in terms of identifying saltwater anglers, but are susceptible to coverage error since state licensing programs exempt anglers in certain categories (for example minors or disabled) from licensing requirements

The dual-frame telephone survey approach provides better coverage than either the CHTS or ALDS alone. However, the methodology is limited by the quantity and quality of telephone numbers included in ALDS sample frames. During the most recent waves of fielding, nearly 25% of cases in the study area resulted in non-contacts due to “bad telephone numbers” (Not in Service, Business Phone, Wrong Number or Missing Number). In addition, determination of the overlap of the frames (households that could be selected from both sample frames) is difficult in telephone surveys due to the occurrence of bad numbers and cell phone numbers on the license frames. Knowing whether a unit is in the overlap is essential for calculating selection probabilities of sampled units. The dual-frame telephone survey attempts to overcome this shortcoming by asking respondents questions aimed at determining whether they are in the overlap. An inability or unwillingness to answer these questions accurately is a potential source of measurement error that could result in biased estimates. A final concern with the dual-frame telephone survey approach is the decline in response rates to telephone surveys in general, and the CHTS in particular. Since 2003, CHTS response rates in NC have decreased from 39% to 25%[1]. Response rates for the ALDS have not been much better, hovering around 30% over the past two years.

Given these concerns, an alternative dual-frame survey using mail rather than telephone was proposed. The pilot study of this alternative is the focus of this paper. Mail surveys have several potential benefits over telephone surveys in a dual-frame approach, including, 1) cost reductions, 2) greater coverage, and 3) an increased likelihood of identifying overlapping frame units through address matching. Recent evidence also suggests that mail self-administered surveys have the potential to improve response rates over comparable telephone surveys (e.g., Link, et al, 2008).

[1] During the same time period, response rates for the CHTS sample for all states along the Atlantic and Gulf coasts have decreased from 31% to 18%.

Recent Use of Address Based Samples

Increased interest in the use of address-based sampling (ABS) in the U.S. for surveys of the general population has been spurred by decline in response rates for telephone surveys (a trend that began in the 1980s) coupled with the increasing cost of attempts to convert non-respondents. In addition, an increasing percentage of households that are “cell phone only” --and thus excluded from standard RDD samples -- have resulted in a downward trend in coverage for standard RDD telephone surveys. At the same time, improvements to databases of U.S. household addresses have facilitated their use for sampling households. A number of studies have examined the feasibility of using address-based sampling in place of listing households in sampled segments prior to sampling for in-person surveys (Iannacchione, Staab, and Redden, 2003; Kennel and Li, 2009). These studies have generally concluded that ABS is a viable alternative for sampling households in the U.S.

Even more recently, several surveys have explored using the ABS to sample households for both mail and telephone data collections. One approach has been to replace RDD samples with an ABS sample, recruit households by telephone (for those that can be matched using commercial lists) or mail, and then conduct data collection in the mode used regularly in the survey. This approach has been used in the U.S. Nielsen TV Ratings Diary Survey (Link, et al, 2009) and by Knowledge Networks (DiSogra, Callegaro, and Hendarwan, 2009). According to internal analyses conducted by Nielson, the ABS method improved coverage from 70% using an RDD design to 98% with the ABS design, and representation of younger adults increased from a penetration rate of 8.8% to 13.5%. The change to the mixed mode approach did not result in any change in the overall response rate to the extended diary survey.

Another approach uses the ABS frame with an all mail mode of data collection. Link et al. (2008) used this method as an alternative to the traditional RDD method for the Behavioral Risk Factor Surveillance Survey. The National Household Education Survey (Montaquila et al. 2010) and the National Survey of Veterans (Han et al. 2010) use a two-phase mail survey to interview subgroups of the population, as does the current study, which follows a first-phase mail screener to identify eligible households by a second-phase mail survey to interview a sample of those that are eligible.

Study Objectives

The pilot test is intended to examine the feasibility of conducting an angler effort survey incorporating an ABS mail approach, with special interest in the dual frame components of the methodology. It uses a mail survey with samples selected from the general household frame (the ABS) and from a license frame. One goal is to assess the response rates that can be achieved using a mail survey for screening and identifying anglers in the general population, and for conducting an extended interview with these anglers. The dual frame nature of the design allows for exploration of potential nonresponse error resulting from households with avid anglers responding at a higher rate than other households.

A second goal is related to the combining of the samples from the two frames to produce efficient estimates. Accuracy of methods to determine if sampled households are on both frames are investigated. The pilot study also provides data about the amount of undercoverage of the CHTS, albeit limited to a small sample in only one state.

Sample and Study Design

Sample Design

The target population for the survey is North Carolina (NC) saltwater anglers, both those living in households in NC as well as those living outside the state. The current CHTS attempts to survey this population by means of an RDD sample of households that live in counties along the coast, while the ALDS attempts to survey this population by means of a telephone survey of licensed saltwater anglers. The address frame used for this pilot is derived from the USPS Delivery Sequence Files (DSF). One of the advantages of using the ABS is the relatively cost efficient sampling from all households in NC, not just coastal county households.[2]

The dual frame approach used in the pilot study samples households that are in the union of the address frame and the license frame, neither of which is limited to coastal counties. The union of the frames consists of three domains: households in the address frame but not in the license frame (S_1), households in the license frame but not the address frame (S_2), and households in both frames (S_{12}). If the address frame were complete, then S_2 would be empty except for licensed anglers who reside outside of NC.

Samples were selected independently from the two frames, and estimates of the total numbers of participants and fishing effort (number of trips) were made for each of the three domains. From the address frame, estimates are made for domains S_1 and S_{12}; from the license frame estimates are made for S_2 and S_{12}. Since both frames estimate the characteristics for the overlap domain (S_{12}), these two are averaged to produce a more precise estimate for S_{12}. The three estimates are then

[2] The CHTS could include non-coastal county households; however, the efficiency of such an RDD design, in which the yield is less than 10% of households with an active angler, results in an extremely cost inefficient design. The use of a mail screening survey offers a cost-efficient means to reach the elusive angler sample in the non-coastal counties.

summed to produce estimates for the total population. In this study, we investigate the similarity of the estimates from the two frames for S_{12}, but do not produce combined estimates.

The Address Frame

A stratified sample was selected from the address frame, with different sampling rates in the strata. Addresses in the coastal counties were in the first stratum, and addresses in the remaining counties were in the second stratum. A total of 900 addresses of the 774,652 on the frame were selected in the coastal stratum, and 900 of the 3,055,903 addresses were sampled in the second stratum. The selected addresses constitute the first-phase sample from the ABS.

The second phase sample included adult anglers (saltwater fished in the previous year) in households that responded to the mail screener. One angler was sampled from each household that reported saltwater fishing by an adult during the previous 12 months. A supplemental sample of anglers was selected by sampling another adult angler in a subset of households that reported saltwater fishing by more than one adult in the previous year.

The License Frame

The license frame, which the state maintains as a part of its administrative records system, is a list of individuals who were licensed to participate in saltwater fishing in NC during the reference period (November – December, 2009). A database containing 551,060 registered anglers was provided by NC's Division of Marine Fisheries. While anyone on this file was licensed for saltwater fishing in NC, some of them may never have fished but held licenses for other types of activities that also bestowed the license for fishing. The types of licenses are discussed later.

Before samples were selected from the license file, it was processed to make it suitable as a sampling frame. The following steps were followed:

- Duplicates (records with the same core data: name, date of birth, and mailing address) were deleted.
- Records without core data were deleted.
- Persons under the age of 18 were deleted.
- Addresses were "normalized" to be in the standard formats used by the postal service.
- Records were stratified by county (coastal, non-costal, or out-of-state strata), and unique household identifiers were assigned to anglers with a common mailing address or telephone number

Frame processing resulted in a total of 456,474 unique angler records, distributed among coastal (184,593), non-coastal (239,450) and non-resident (32,431) strata.

The file was sorted by address and a systematic sample of 450 anglers was selected in each stratum. The ordering was done to minimize the possibility of including unidentified duplicated household listings. As in the ABS, a supplemental sample of anglers was selected. A second angler was selected in every sampled household identified as having more than one licensed angler.

Data Collection Procedures

A screening survey was mailed to all 1800 ABS sample addresses in the fall of 2009[3]. Consistent with the methods suggested in Dillman, Smyth, and Christian (2008), the household was mailed an instrument that included a cover letter and a $1 cash incentive. The household was asked to complete the questionnaire and mail it back in the envelope provided.

Mailing of the screener was split into two batches, with 900 addresses in each batch. The first batch was mailed November 10, 2009 and the second on November 20, 2009. The batches were mailed at different times to examine the effect of delay between the screener and the angler interview in the two phase mail survey. This is an issue that only arises in two phase mail samples and is discussed later in the analysis. Sample units in both batches were exposed to the same treatment: (1) an initial mailing of the screener questionnaire; (2) a reminder postcard mailed 1 week after the initial mailing; and (3) a second mailing of the screener questionnaire to non-respondents two weeks after the mailing of the postcard, accompanied by a non-response conversion letter.

[3] All data collection instruments are included in the attached methodology report.

Randomly selected anglers from each fishing household identified in the ABS screener as well as anglers sampled from the license frame were mailed an Angler Survey, beginning January 4, 2010. Similar to the screening data collection protocol, the Angler Survey data collection consisted of (1) the original mailing of the survey instrument (different letters for the ABS and License frame sample units), including a $1 incentive for participation; (2) a reminder postcard (one week later); (3) a second mailing of the survey two weeks following the postcard reminder (that included a modified cover letter, but no additional incentive); (4) and a final questionnaire, delivered by Federal Express 2-day delivery.

Appendix A shows the sample disposition for the ABS sample screener (Table A-1), the ABS angler survey (Table A-2), and the License sample (Table A-3) by stratum. A detailed report of the methodology used in the study, including detailed information concerning the de-duplication of the sampling frames, is included as an attachment to this report.

Findings

Matching and Domain Identification

A critical issue in the development of estimates from dual frame designs is the accurate identification of elements in each frame as well as those units which appear in both frames (overlap). Often, the identification of overlap between frames relies on data reported by the respondents. This approach is currently being explored in tests estimating fishing effort from telephone dual frame surveys. In the case of the present study, we were able to identify the overlap via matching of ABS addresses to addresses in the license frame. Both methods of identifying overlapping units are subject to errors that affect the quality of the dual frame estimators. We begin by looking at the matching of addresses and then discuss the accuracy of the self-identification of domain membership by respondents.

One way to assess the quality of the matching as a method of identifying overlap is to compare the estimate of the total number of licensed anglers from the ABS sample, both overall and by stratum, to the known number of licensed anglers in NC. A complication is that the matching is done by address, while the units of the license frame are the individuals holding licenses. To compare the number of ABS sampled addresses that are matched to the number from the license frame, we first convert the person-level license frame size to a household level size. To do this, we estimated the average number of adult licensed anglers per household from the license frame by stratum (in the coastal stratum the average was 1.19 and in the non-coastal stratum it was 1.16). The number of anglers in the stratum was divided by this average to estimate the number of angler households in each stratum.

Table 1 shows the estimated number of households with licensed anglers in each stratum for the first phase ABS sample, along with the NC license frame counts, where the license frame estimate is adjusted to be at the household level. The table shows that overall the matching was very close to unity, with the ABS sample estimate of licensed addresses being just 1.06 times the adjusted number from the license frame. This suggests the approach is effective (the 1.06 estimate is not significantly different from unity). The ratio in the coastal stratum is estimated to be 0.88 and is statistically different from unity, while the non-coastal stratum estimate is 1.19. We expected the matching error to be primarily one-directional, with some addresses not matching due to errors in the license frame and vagaries in matching. However, the tabulation suggests that the matching is either of very high quality or the matching error goes in both directions. This result supports the initial rationale of matching addresses and is consistent with the premise that the dual frame domain membership is accurately obtained from this procedure.

Table 1. Estimated number of addresses in the overlap from the ABS first phase sample and from the license frame

Stratum	ABS sample	License frame	Ratio of ABS to license	95% CI lower limit	95% CI upper limit
Total	381,326	360,610	1.06	0.90	1.21
Coastal	136,854	154,975	0.88	0.79	0.98
Non-coastal	244,472	205,635	1.19	0.93	1.45

Response Rates

We begin the analysis by examining weighted response rates[4] for the two frames and across the strata. The response rates are shown in Table 2. The study achieved an overall response to the screener of 45.6% and an extended interview response rate of 72.5% for an overall response rate for the ABS sample of 33.1%. This rate exceeds the comparable CHTS telephone response rate for Wave 6 in NC of 25.4%. Among those sampled from the license frame, we achieved a response rate of 58.2%, also exceeding the ALDS response rate for NC during the same wave of 30.1%[5].

[4] Weighted by the base weight and using AAPOR response rate RR3 (AAPOR, 2009).
[5] Note that the ABS mail survey and the CHTS are limited to NC residents whereas the license mail survey and the ALDS include anglers from out of state who have a NC saltwater fishing license.

These response rates are encouraging. They suggest that the angler population can be reached via a self-administered mail survey, that coverage of the population is possible via an ABS with a self-administered mail questionnaire, and that response rates may improve, especially for the license frame, over those of a telephone survey.

Table 2. Response Rates by Frame and Stratum. Geo-coding, Batch, License Match and Number of Anglers Sampled (all response rates weighted by base weight)

<table>
<tr><th></th><th colspan="3">ABS Frame</th><th rowspan="2">License Frame</th></tr>
<tr><th></th><th>Screener</th><th>Angler Survey</th><th>Overall</th></tr>
<tr><td>Overall</td><td>45.6%</td><td>72.5%</td><td>33.1%</td><td>58.2%</td></tr>
<tr><td>Stratum</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Coastal</td><td>48.4%</td><td>70.1%</td><td>34.0%</td><td>57.3%</td></tr>
<tr><td>Non-Coastal</td><td>44.9%</td><td>73.9%</td><td>33.2%</td><td>57.6%</td></tr>
<tr><td>Out of State</td><td>NA</td><td>NA</td><td>NA</td><td>67.7%</td></tr>
<tr><td>Geo-coding</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Borders Ocean</td><td>48.9%</td><td>73.8%</td><td>36.0%</td><td>53.7%</td></tr>
<tr><td>Coastal, not border</td><td>48.1%</td><td>67.0%</td><td>32.3%</td><td>59.9%</td></tr>
<tr><td>Other</td><td>44.9%</td><td>73.9%</td><td>33.2%</td><td>58.8%</td></tr>
<tr><td>Batch</td><td></td><td></td><td></td><td></td></tr>
<tr><td>First</td><td>46.4%</td><td>75.1%</td><td>34.9%</td><td>NA</td></tr>
<tr><td>Second</td><td>44.8%</td><td>70.1%</td><td>31.4%</td><td>NA</td></tr>
<tr><td>License Match</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Match</td><td>65.5%</td><td>70.1%</td><td>45.9%</td><td>NA</td></tr>
<tr><td>No Match</td><td>43.2%</td><td>73.4%</td><td>31.7%</td><td>NA</td></tr>
<tr><td>Number of Sampled Anglers</td><td></td><td></td><td></td><td></td></tr>
<tr><td>1 Angler HH</td><td>NA</td><td>68.4%</td><td rowspan="2">31.2%</td><td>66.5%</td></tr>
<tr><td>1 Angler/2+ HH</td><td>NA</td><td>74.8%</td><td>NA</td></tr>
<tr><td>2 Anglers/2+HH</td><td>NA</td><td>74.3%</td><td>34.0%</td><td>56.5%</td></tr>
</table>

Response rates from the mail surveys did not vary by stratum for either frame, with the exception of higher rates among anglers from out of state within the license frame. We also examined response rates by a three category geo-code, examining those who live in a county that directly borders the ocean, those in the coastal stratum, but not directly adjacent to the ocean, and all others. These geo-code categories also showed no significant differences in response rates.

As noted, the screening interview for the ABS sample was conducted in two batches, with the initial batch mailed about 10 days before the second. The second phase mailing for both batches was done at the same time, so the first batch respondents had a longer time period between the first and second phase mailings. As expected, the first phase response rates were not significantly different between the two batches (46.4% and 44.8%, respectively). Differences between the second phase response rates were also not significant, perhaps because of the small sample sizes. The direction of the difference, with a higher angler survey response rate for Batch 1 as compared to Batch 2 (75.1% and 70.1%) suggests that a longer lag time between the screening interview and the extended interview may be beneficial with respect to increasing the second phase response rate. This finding warrants further study as we explore the use of ABS for two phase designs.

In households with more than one adult angler, we sometimes sampled two anglers for participation in the second phase angler survey[6]. There was no difference in response rates among anglers in the ABS second phase sample as a function of number of anglers sampled in the household. However, when more than one angler was sampled from a household in the license frame, the response rate was 10 percentage points lower than for the anglers who were the sole recipients of the angler survey request (56.5% vs. 66.5%).

[6]For the ABS we sometimes sampled one and other times sampled two anglers, while in the license frame we always sampled two anglers when there were two present.

Avidity Bias

With surveys that focus on a specific segment of the population, there is always concern about differential nonresponse related to participation in the behavior of interest to the study. Previous studies (e.g., Thomson, 1991; Fisher, 1996; Connelly, Brown, and Knuth, 2000) have demonstrated avidity bias in angler surveys. In this context, avidity bias would result from a higher propensity to respond by avid anglers when surveyed about fishing. To examine this, the ABS sample units were matched to the NC license frame to determine whether those with NC saltwater fishing licenses were more likely to participate in the survey than those without a license. Overall, 12.8% of the ABS sample was matched to the license file, with a higher match rate in coastal counties (17.7%) than non-coastal counties (8.0%). The quality of the matching was very good, as discussed previously.

Table 2 indicates that the screener response rate was over 20 percentage points higher for households that were matched to the license frame than those that were not (65.5% vs. 43.2%). However, the second phase response rate of adults who said they had fished in the last year did not differ significantly by whether they matched to the license frame. Because of the large first phase difference in response rates, the overall response rate did show a significant difference, 45.9% vs. 31.7%.

This is an important finding with respect to the feasibility, as well as the benefits, of using a residential address frame to estimate the total number of anglers and the total number of recreational fishing trips. If the respondents to the first phase sample are adjusted to account for nonresponse without accounting for the avidity bias, then the effect will be to overestimate the number of anglers and angler trips because the avid anglers are over-represented in the sample. Because we were able to successfully match the ABS sample to the license frame, we were able to adjust the first-phase nonresponse weights for the ABS sample to account for differential nonresponse between avid (households with at least one licensed) and non-avid (households with no licensed anglers) households. As described below, this greatly reduces the potential for avidity bias in estimates of the total number of anglers and the total number of fishing trips. However, the adjustment does not account for avid anglers who could not be matched to the license frame, or for differential avidity of licensed anglers. Research on methods to avoid this potential source of nonresponse bias is needed, for example, by examining the effect of screening focusing on a broader range of topics than just angling.

It should be noted that avidity bias may also be present in other surveys that sample from residential household frames, including the Coastal Household Telephone Survey. A research project is currently underway that will attempt to match CHTS sample units to license frames in NC and LA by telephone number and address. Successfully matching the CHTS sample to license frames will help to identify and quantify avidity bias in the CHTS, as well as allow survey managers to develop adjustments to nonresponse weights that will account for avidity bias in the survey.

We also wanted to explore differential response rates among those sampled from the license frame and those sampled from the ABS frame who matched to the license frame to determine if different types of licensure were associated with differential response rates. There are numerous saltwater license types in NC and they may be informative about avidity bias in the license frame. Some licenses are for salt water fishing only, while others are combination licenses that also permit the holder to freshwater fish or hunt. Some types provide lifetime licensure, while others are purchased annually.

Table 3 presents the response rates for the ABS screener and the License frame angler survey by strata and license type. No clear pattern overall emerges. We do see that among the license frame, the combination license holders tend to respond at a lower rate than those who hold other types of licenses. Within each of the strata of the license frame, those who held a 10-day license responded at a higher rate than other license holders. Higher response rates among these respondents with highly targeted licenses are consistent with the hypothesis that anglers who have fished recently have higher propensity to respond to the survey. These findings are evidence that not all anglers on the license frame are equally likely to respond to the survey and also have implications for nonresponse bias from the license frame. With respect to angler response rates among the ABS sample that linked to the license frame, we also see some variability in response rates by license type. However, among this sample, we see the lowest response rates among those with the 10-day license. Although the findings are mixed with respect to response rates by license type, the findings do suggest that greater reduction of nonresponse bias might be obtained by using information about license type in nonresponse adjustment.

Table3. Screener Response Rates (ABS Frame) and Angler Response Rates (License Frame) by Sample Strata and License Type

	ABS Frame: Matched			License Frame			
	Coastal Strata	Non-Coastal Strata	Overall	Coastal Strata	Non-coastal Strata	Out of State Strata	Overall
License Type							
Combo	68.0%	66.1%	66.7%	55.3%	56.6%	59.0%	56.3%
Saltwater	64.8%	57.3%	61.0%	58.7%	59.5%	69.8%	60.3%
10 day		55.2%	55.2%	100.0%	66.7%	70.4%	70.6%
Annual	68.5%	69.5%	69.0%	52.8%	61.3%	69.0%	57.6%
Lifetime	62.8%	60.6%	61.2%	63.7%	55.2%	60.3%	58.1%
Unmatched	44.1%	43.2%	43.4%				

.

Missing Data Rates

In considering a shift away from an interviewer-administered telephone survey to a self-administered mail survey, data quality, specifically missing data rates (associated with incomplete questionnaires or incorrect skip patterns) as well as inconsistent data are a concern. We examined missing data rates for several key variables. We defined missing data rates as either no response or an indication of a “don’t know” response. These rates ranged from a low of less than 2% for questions concerning whether or not the respondent had participated in recreational saltwater fishing during the reference period to over 25% for questions concerning valid recreational saltwater fishing license ownership for the reference period.

Respondents were asked two summary questions concerning fishing effort during the wave: (1) whether they gone saltwater recreational fishing in North Carolina during the wave (November 1 – December 31, 2009) and (2) for those who had gone fishing during the reference period, they were asked to simply circle the dates on a calendar indicating that they had fished that day. The later information was then summarized during data processing to produce a total number of trips taken by the respondent. Instructions following the collection of this summary data requested that the respondent complete detailed trip information for the four most recent angling trips taken.

In light of this two-step process for obtaining effort information, a second form of missing data consists of those cases in which the respondent indicates multiple angling trips during the wave (indicated by circling the dates of the trips on a calendar) but then failing to complete the detailed trip information for the four most recent trips taken. We found that for 1.2% of the cases, the two pieces of information were inconsistent.[7] For 11 of the 884 anglers (0.2%) who recorded no information on the calendar, detailed information for 1 or more trips was provided. Conversely, 15 of the 270 anglers (5.6%) who circled at least one date on the calendar provided no information for the detailed trip questions, and 98 of these anglers (36%) detailed fewer than they reported on the calendar. There was a particularly serious omission rate for those anglers reporting many trips. Of the 139 anglers who reported 4 or more trips in the calendar, 71 (51%) provided detailed information for fewer than 4 trips.

Comparisons for Under-covered Populations

One of the criticisms of the current CHTS estimates is the lack of coverage of persons living in non-coastal counties and coastal residents living in households without landlines. In this section

[7] Obviously, we can only examine inconsistencies to a limited extent since avid anglers could indicate a high number of trips (> 4) on the calendar but then only report details for the most recent four trips. However, avid anglers who reported a high number of trips but then failed to complete the detailed sets of questions for the four most recent trips are classified as inconsistent.

we examine the demographic and behavioral characteristics of anglers as a function of geographic location and landline service among the ABS sample members[8].

We begin by looking at the percentage of anglers who fished during the year that would be excluded by each reason. The mail survey estimated that 64.5% of all anglers who fished during the year resided in non-coastal counties and would be excluded from the CHTS. An estimated 21.4% of anglers reported in the mail survey that they did not have a landline in their home. By examining the size of the union of these two domains, we estimated that 69.3% of anglers reported in the ABS would be excluded from the CHTS (i.e., only 30.7% of NC anglers reside in coastal households with landlines). Similarly, the mail survey provides an opportunity to assess the coverage of state license databases as sample frames. The mail survey estimated that 57% of the anglers who fished during wave 6 did not possess a saltwater fishing license. The source of this undercoverage in NC likely includes minors (<16) and anglers who fished on state fishing piers, both of which are exempted from state licensing requirements. Based on the ABS frame, we estimate that the CHTS and ALDS surveys in North Carolina exclude about 35% of anglers and approximately 38% of trips (that is, noncoastal anglers without licenses or coastal anglers without licenses or landline telephones). Clearly, the ABS mail survey has a great deal to offer to improve coverage compared to the current RDD and ALDS designs[9].

Table 4 presents the estimated demographic, angler licensure, and fishing activity characteristics for the subset of the ABS frame who fished in the last year, for NC as a whole and by stratum..

---

[8] All estimates are weighted to account for the probability of selection and for nonresponse.
[9] We can also examine the percentage of active anglers in the wave who would have been missed in the CHTS. The mail survey estimated that 44.3% of all anglers who fished during the wave resided in non-coastal counties, and 11.4% of the coastal residents who fished during the wave did not have a landline in their home. As a result 52.6% of anglers in the ABS who fished during the wave would be excluded from the CHTS.

Although demographic characteristics of the coastal and non-coastal anglers are similar, the incidence of fishing during the wave for the coastal anglers was 2.5 times the rate of non-coastal anglers (37.4% vs. 14.4%). However, among those who did fish during the wave, we find similar levels of effort; however, small sample sizes for the estimation of effort limits the power to detect differences.

Table 4. Demographic composition, Angler Participation Rates, Licensure, and Average number of days fishing from the ABS sample, by geographic location (standard errors in parentheses)

	ABS Frame: all respondents (n=152)	**Coastal Counties (n=105)**	**Non-Coastal Counties (n=47)**
Gender: Male	76.9% (4.5)	78.7% (4.6)	75.9 (6.5)
Gender: Female	16.4% (3.9)	14.9% (3.7)	17.3% (5.6)
Gender :Missing	6.7% (2.8)	6.4% (2.4)	6.8% (4.1)
Age: 18-44	16.8% (4.3)	15.8% (3.6)	17.3% (6.3)
Age:45 and older	77.2% (4.8)	79.4% (3.8)	75.9% (7.1)
Age :Missing	6.1% (2.8)	4.8% (2.0)	6.8% (4.1)
Anyone in household Salt Water Fishing in 2009? [A]	25.4% (1.9)	40.8% (2.6)	21.4% (2.2)
NC License, past 12 months?	64.2% (5.7)	76.9% (4.6)	57.1% (8.4)
NC License, past 12 months for Saltwater Fishing?	54.7% (5.8)	65.6% (5.0)	48.7% (8.4)
NC Saltwater Fishing License: Valid November, 2009?	38.9% (5.2)	59.4% (5.1)	27.5% (7.4)
Salt Water Fishing During Wave?	24.2% (4.5)	37.4% (5.4)	16.9% (6.2)
Average number of days spent fishing, during wave, per angler[B] **……by boat**	1.66 (0.27)	1.97 (0.26)	1.26 (0.49)
……by shore	1.92 (0.61)	2.66 (0.99)	0.99 (0.48)
……total trips	3.58 (0.73)	4.63 (1.03)	2.26 (0.90)

Note: All estimates limited to those who reported fishing during the 2009, except as noted.
[A] Based on information obtained in the screening interview among all screening respondents; n = 685, 357, 328 for the three columns
[B] Among those anglers who fished during the wave; n = 49, 41, 8 for the three columns

We also compared the demographic and behavioral characteristics between those with and without landlines, once again limited to the ABS sample frame. The results, reported in Table 5, show that 21.4% of those NC residents who saltwater fished in 2009 do not have a landline telephone. Anglers in NC with no landline phones are more likely to be female and younger, as compared to anglers with landline phones. The incidence of fishing during Wave 6 for those with landlines was twice that for those with no landline phones. Across all other measures of angling behavior and licensure, those without landlines are similar to those with landline phones.

The findings from Tables 4 and 5 suggest that the rate of angling among those in non-coastal counties in NC is less than those in coastal counties and that those without landline phones are less likely to have fished during the reference period than those with landline phones. Still, a majority of NC anglers do not reside in coastal county households with landline phones. Once again, small sample sizes limit our ability to draw sharp conclusions about what proportion of fishing effort takes place in households not covered by the current CHTS.

The observed differences in demographic characteristics and fishing incidence between anglers who are and who are not covered by the CHTS do not necessarily indicate that fishing effort estimates derived through the CHTS are biased. The CHTS adjusts for undercoverage by expanding estimates of fishing effort upward by correction factors derived through an access-point angler intercept survey (APAIS) of completed fishing trips. Specifically, intercepted anglers are asked for their state and county of residence, as well as whether or not their household has a landline telephone. CHTS estimates are then expanded by the inverse of the

Table 5. Demographic composition, Angler Participation Rates, Licensure, and Average number of days fishing for ABS frame by landline phone status (standard errors in parentheses)

	ABS Frame: all respondents (n=152)	**Landline Phone (n=123)**	**No Landline Phone (n=27)**
Gender: Male	76.9% (4.5)	80.9% (4.5)	65.4% (12.9)
Gender: Female	16.4% (3.9)	10.7% (2.9)	33.4% (12.9)
Gender:Missing	6.7% (2.8)	8.4% (3.6)	1.2% (1.2)
Age: 18-44	16.8% (4.3)	14.4% (4.6)	26.8% (10.74)
Age:45 and older	77.2% (4.8)	78.0% (5.4)	72.0% (10.8)
Age: Missing	6.1% (2.8)	7.6% (3.5)	1.2% (1.2)
Anyone in household Salt Water Fishing in 2009? [A]	25.4% (1.9)	26.3% (2.2)	25.8% (4.1)
NC License, past 12 months?	64.2% (5.7)	61.8% (6.4)	73.9% (13.4)
NC License, past 12 months for Saltwater Fishing?	54.7% (5.8)	50.2% (6.4)	71.2% (13.4)
NC Saltwater Fishing License: Valid November, 2009?	38.9% (5.2)	37.9% (5.8)	41.0% (12.8)
Salt Water Fishing During Wave?	24.2% (4.5)	28.0% (5.5)	13.1% (5.9)
Average number of days spent fishing, during wave, per angler [B] **……by boat**	1.66 (0.27)	1.59 (0.30)	2.19 (0.41)
……by shore	1.92 (0.61)	1.47 (0.36)	5.45 (4.13)
……total trips	3.58 (0.73)	3.06 (0.53)	7.64 (4.32)

Note: All estimates limited to those who reported fishing during the 2009, except as noted.
[A] Based on information obtained in the screening interview among all screening respondents; n = 685, 516, 149 for the three columns

[B] Among those anglers who fished during the wave; n = 49, 42, 7 for the three columns ratio of CHTS-covered trips (trips taken by anglers in coastal households with landlines) to total trips (CHTS-covered trips, as well as trips taken by anglers from non-coastal counties or households without landline phones)[10]. These expansion factors are unbiased provided the sample of angler trips derived from the APAIS is representative of all angler trips. Sampling from the ABS provides an excellent opportunity to test the assumption that APAIS samples are representative. However, sample sizes in the present pilot study were insufficient to support this analysis.

Dual Frame Considerations

The reasons for using a dual frame design are to improve coverage and reduce the cost for achieving more precise estimation of angler effort. The license frame provides a mechanism for identifying the group of interest efficiently because anglers occur in a small fraction of households. But the license frame is incomplete for saltwater recreational anglers, so it must be used together with the general population ABS to control the bias due to noncoverage.

Here, we describe some of the issues that arise in the dual frame system in the presence of nonsampling errors. The special effects of nonsampling errors on dual frame estimates have only recently been discussed in the sampling literature (see Lohr, 2009; Brick et al. forthcoming). In this section we explore the implications of certain nonsampling errors in the pilot study. In our concluding comments, we describe possible changes to the survey design and implementation that

[10] A similar approach is used to expand effort estimates derived from the ALDS; expansion factors are derived from angler-reported information about the possession of a saltwater fishing license. This approach is also potentially susceptible to reporting error based upon an inability or unwillingness to provide accurate information about license status as discussed in this report.

could alleviate some of the biases that they cause. Statistical adjustments are also being investigated, but design modifications that would eliminate or reduce the errors would be preferable.

As noted before, the overlap is the population of anglers residing in NC who have a license (more specifically, are on the license frame with sufficient information to be eligible for sampling). This assumes that all the licensed anglers in the state are in housing units that are on the ABS, a reasonable assumption based on data on coverage of households using the ABS in NC. The non-overlapping component of the ABS frame is the set of anglers residing in NC who did not have a license; the non-overlapping component of the license frame is the set of NC license holders who reside outside of the state. Our analysis begins by concentrating on the overlap component since this is relevant only in dual frame surveys.

In the pilot study, we can identify and partially quantify two sources of nonsampling errors that could bias estimates for the overlap domain. The first is nonresponse, resulting in bias due to differential response rates associated with avidity. Earlier we showed that ABS addresses matched to the license frame responded at a higher rate than those that did not. We also found that response propensity in the license frame sample depended on the type of license in a way that was consistent with avidity differences. We expand on our earlier discussion focusing on the size of avidity bias for estimates from a dual frame estimator.

A second source of bias in the dual frame estimator is error in matching the ABS sample units to the license frame. Matching is required to determine which units in the ABS are in the overlap. As discussed earlier, one of the rationales for using a self-administered mail survey is that address matching to the license frame is less error-prone than telephone number matching.

Effects of Avidity Bias

We first examine evidence about the magnitude of nonresponse bias in estimation of fishing effort in NC. Estimation of effort, defined as the number of trips, requires accurate assessment of the number of active anglers, as well as the number of trips those anglers make. If active anglers respond to the survey at a higher rate than others, or if anglers who respond take more trips than nonrespondents, then the estimate of number of trips would be biased upward. Though samples from both frames could suffer from this source of nonresponse bias, it would be expected to be more severe in the ABS frame because the variability in avidity is likely to be greater there than in the license frame.

Table 6 shows information about avidity bias in the first of those components, estimation of the number of active anglers. The first three rows of the table present independent estimates from the two frames of the number of licensed anglers who fished in the wave for the overlap, overall and by stratum. License status for both frames is based on being on the license frame rather than the response to the interview questions about license status (for the ABS this required the address match to an address on the license frame). To qualify as having fished in the wave, we also required that the angler responded that they fished during the past year (the data were not fully edited so a few cases did not meet this logical requirement).

For the ABS estimate, we produced a nonresponse adjustment by forming weighting classes that included both geographic information (proximity to the ocean) and match status, both of which should account for some avidity bias (these weights were used in previous analyses). Even with this adjustment, the ratio of the ABS estimate to the license sample estimate is about 1.35 overall and in each stratum, indicating the ABS sample estimates more anglers fished in the wave than is estimated from the license sample.

Table 6. Estimated number of licensed anglers in the overlap who fished in the wave by screener nonresponse adjustment method

	ABS sample	License sample	Ratio of ABS to license	95% CI lower limit	95% CI upper limit
ABS first phase matching adjustment					
Total	102,918	75,391	1.37	0.62	2.11
Coastal	58,801	42,571	1.38	0.60	2.16
Non-coastal	44,117	32,820	1.34	-0.04	2.73
ABS first phase geographic adjustment					
Total	135,595	75,391	1.80	0.80	2.80
Coastal	73,877	42,571	1.74	0.73	2.74
Non-coastal	61,717	32,820	1.88	-0.02	3.78
ABS first phase no cells adjustment method					
Total	138,999	75,391	1.84	0.83	2.86
Coastal	78,098	42,571	1.83	0.77	2.90
Non-coastal	60,901	32,820	1.86	-0.02	3.73

To get some idea of the potential magnitude of the avidity bias, the bottom portion of the table shows the same quantities with the ABS estimates computed using different nonresponse weighting classes; the middle three rows of estimates use cells based on geography but not match status and the last three rows uses no weighting classes at all. The ratios of the estimates that use a nonresponse adjustment based only on geography are closer to 1.80, consistent with greater overestimation of anglers when the nonresponse adjustment procedure does not account for avidity as completely. When no weighting classes are used, the ratios are slightly higher still. Because of the small sample sizes, however, the 95% confidence intervals are very wide. Since the license frame estimates are likely to be subject to some avidity bias as well, the table shows bias from differential nonresponse (as a function of fishing activity) is potentially serious.

Table 7 summarizes information about the size of the second component of potential avidity bias, the estimation of average number of trips per active angler. The table shows that ratios of the estimates of average number of trips per angler from the two frames are in the opposite direction from those shown in Table 4 (i.e., the ratios are less than unity rather than greater than unity). We also observed that the weighting cells have little effect on these estimates of average trips. The three sets of rows in the table by screener adjustment method show this result. It appears that active anglers responding in the ABS sample fish either less frequently or about the same as those from the license frame.

For estimating total trips (the product of the number of anglers and their average number of trips), the ABS and license samples are closer than either of the two components because the ratios of the components partially offset each other. A standard dual frame estimation strategy is to average the two estimates for the overlap to produce a more precise estimate of this population. Since the components of the two frames are different, averaging the two estimates could give a biased estimate of the number in the overlap. In theory, the two estimates of the overlap are assumed to both be unbiased, so a question arises about the appropriateness of simply combining the estimates from the two frames given these results. We discuss the dual frame estimators and their biases and variances below.

Table 7. Estimated average number of trips per active angler in the overlap who fished in the wave by screener nonresponse adjustment method (standard errors in parentheses)

	ABS sample n=25	License sample n=117	Ratio of ABS to license	95% CI lower limit	95% CI upper limit
ABS first phase matching adjustment method					
Average shore trips	1.01 (0.31)	2.31 (0.28)	0.44	0.14	0.73
Average boat trips	1.83 (0.44)	2.09 0 (0.30)	0.88	0.38	1.37
Total trips	2.84 (0.59)	4.40 (0.37)	0.65	0.36	0.93
ABS first phase geographic adjustment method					
Average shore trips	0.99 (0.31)	2.31 (0.28)	0.43	0.14	0.72
Average boat trips	1.80 (0.44)	2.09 0 (0.30)	0.86	0.37	1.35
Total trips	2.79 (0.60)	4.40 (0.37)	0.63	0.34	0.93
ABS first phase no cells adjustment method					
Average shore trips	1.00 (0.31)	2.31 (0.28)	0.43	0.14	0.72
Average boat trips	1.82 (0.44)	2.09 0 (0.30)	0.87	0.38	1.36
Total trips	2.82 (0.59)	4.40 (0.37)	0.64	0.35	0.93

Before leaving this subject, it is interesting to note that by using the match status in nonresponse adjustment reduced the avidity bias in the ABS sample and made the estimates of the overlap more similar. An adjustment of this type is more difficult to implement in a dual frame telephone approach because the matching is subject to greater error. One way to do this with a telephone frame is to attempt to match the telephone numbers from the telephone sample to the telephone numbers on the license frame. The two main problems with this approach are: (1) the telephone numbers on many license frames are incomplete and out-of-date making matching difficult, and (2) many people may be reached by telephone on multiple telephone numbers (cell numbers and landline numbers) so that the telephone sampled might not be the telephone number included in the license frame. Another way of accomplishing the matching is to rely on the angler to indicate whether they have a license or not and consider this response to determine license status. As we discuss in this report, anglers may not report their license status accurately (as discussed later there are substantial errors of omission and commission).

Next we assess the accuracy of determining overlap membership from data reported by respondents. The overlap consists of licensed anglers residing in NC, and the only characteristic required from the respondent (in the absence of a method of matching) is possession of a valid saltwater license for the wave. As noted earlier, this is a method that is currently being used in dual frame telephone samples and might be more precise than matching by telephone number.

While only the quality of self-reported information about licensure among the ABS respondents is in question, the respondents in both frames were given the same questionnaire in the pilot. Thus, respondents from both frames provide information about the error rates of overlap identification. The sample from the license frame and the matched ABS frame both provide estimates of the false negative rate for the licensure question (i.e., the proportion of validly licensed respondents who claim they do not have a license). The ABS sample also provides an estimate of the false positive rate for the licensure question (i.e., the proportion of respondents who do not have a valid license but claim they do); this quantity may not be estimated as accurately as the false negative rate because of the small sample size and issues in matching addresses.

Table 8 provides estimates for three licensure questions (if the respondent has a NC fishing license, has a NC recreational saltwater fishing license, and has a NC recreational saltwater fishing license for the reference period, Wave 6, November –December 2009). First we examine the false negative rates estimated from both frames. The first column shows the estimated percentage claiming they have a valid license for the license frame respondents who reside in NC. The second column shows the same percentage for the ABS respondents who match to the license frame. All respondents in these two columns should report “yes” to all three questions. The upper half of the table shows estimates for the overlap population who saltwater fished in 2009, and the bottom half for those who fished in the wave. From the license frame, we estimate that about 15% of those who fished during the year and 10% of those who fished during the wave reported erroneously that they did not have a license to do so. The comparable estimates from the ABS frame were about 30%. This suggests that there may be a higher false negative rate from the ABS sample than from the license sample, although the sampling errors are so large that the difference is not significant.

Table 8. Percent of Respondents in from both frames reporting ownership of various NC fishing licenses (standard errors in parentheses)

	License frame: Resident licenses	ABS frame: match to license frame	ABS frame: not match to license frame
Among Respondents Who Fished during 2009:	**(n=435)**	**(n=60)**	**(n=92)**
NC Fishing License	95.5 (1.1)	85.5 (7.4)	---
NC Saltwater License	90.0 (1.6)	72.2 (4.9)	---
NC Saltwater:Wave	85.3 (1.9)	69.8 (4.3)	27.9 (6.4)
Respondents Who Fished in Wave 6, 2009:	**(n=122)**	**(n=25)**	**(n=22)**
NC Fishing License	94.4 (2.1)	70.8 (13.8)	---
NC Saltwater License	90.4 (2.8)	70.8 (13.8)	---
NC Saltwater:Wave	89.5 (2.9)	70.8 (13.8)	46.0 (17.0)

The last column of the table gives estimates from the ABS sample for those who did not match to the license frame, providing information about false positive rates. It shows that 28% of the anglers who fished during the year and 46% of the anglers who fished during the wave and did not have a license (at least they did not match by address) erroneously reported having a license. With so few respondents in these cells, the sampling errors are very large.

Given the small sample sizes, the estimates of error rates for some subpopulations are very tentative, but there are some mechanisms that might support higher error rates in the matched ABS sample than in the license frame. The data collection procedures varied somewhat between the samples in subtle ways that may have influenced responses. For example, in the license frame sample, the mail was addressed to the angler by name and there was no first phase mailing. In addition, the matching from the ABS is by address not by angler, and households with more than one angler may have both licensed and unlicensed anglers, leading to the appearance of more error. Additional studies with larger sample sizes are needed to more adequately determine the magnitude and sources of the errors. However, the pilot does show that relying on respondents to self-identify their domain membership is a source of error that can be greatly reduced by the address matching in the ABS approach. This finding suggests that the current approach used to match sample frames in the dual-frame telephone survey design is insufficient.

Because we have additional information as to the nature of the license held by respondents in the license frame, we can also examine factors that may be related to the quality of reporting about licensure. Table 9 shows estimates of false negative rate by the type of saltwater fishing license held by the individual made from the license frame. Overall, the highest rate of accurately reporting licensure was for the broad category of “NC Fishing License,” with the poorest reporting for the wave specific saltwater fishing license. This is as expected, since the wave-specific reporting requires the respondent to retrieve information not only about the type of license, but the valid dates for that license. Those anglers who held licenses specific to saltwater fishing tended to be more accurate than those who held combination licenses. Once again, this is not an unexpected finding given that the question wording for holders of recreational saltwater fishing license closely matches the nature of the license they hold, making the reporting task cognitively easier than for those with combination licenses. Non-lifetime licenses must have been purchased sometime during the past 12 months, making the reporting task more salient and of higher quality for respondents with those types of licenses than for lifetime license holders. Finally, we see that among all respondents, non-resident license holders were more likely to report accurately than NC residents. We might speculate that the nature of the licenses for the non-resident groups differs (e.g., one week vs. one year) and by definition, requires travel from outside the state, once again adding to the saliency of the license.

Table 9. Estimated Percentage of saltwater license holders from the license frame who claim they do not have a valid license, by type of license (standard errors in parentheses)

	Total	**Saltwater Only**	**Combo**	**Lifetime**	**Not a Lifetime**	**Resident**	**Non-Resident**
Respondents who say they don't have a…	**Respondents Who Fished during 2009**						
	(n=718)	(n=527)	(n=191)	(n=99)	(n=619)	(n=435)	(n=383)
NC Fishing License	4.5 (1.0)	4.0 (1.1)	5.3 (1.8)	1.7 (1.7)	5.2 (1.2)	4.5 (1.1)	5.6 (1.4)
NC Saltwater License	9.9 (1.5)	6.0 (1.3)	15.8 (3.0)	12.5 (5.0)	7.1 (1.3)	10.0 (1.6)	8.6 (1.7)
NC Saltwater Wave	15.1 (1.7)	13.0 (1.9)	18.3 (3.2)	22.9 (5.0)	13.4 (1.8)	14.7 (1.9)	19.0 (2.4)
	Respondents Who Fished in Wave 6, 2009						
	(n=227)	(n=179)	(n=48)	(n=17)	(n=219)	(n=122)	(n=105)
NC Fishing License	5.3 (1.9)	5.0 (2.2)	5.8 (3.4)	0 (NA)	6.6 (2.1)	5.6 (2.1)	3.2 (1.7)
NC Saltwater License	9.1 (2.5)	6.6 (2.5)	13.5 (5.3)	22.9 (11.8)	7.2 (2.3)	9.6 (2.8)	5.3 (2.2)
NC Saltwater Wave	10.9 (2.6)	9.4 (2.8)	13.6 (5.3)	22.9 (11.8)	9.3 (2.4)	10.5 (2.9)	13.8 (3.4)

Dual Frame Estimators

Above, we explored some of the key error components for dual frame estimators, and found the domain identification (with and without a license) among the ABS to be of relatively high quality but the response patterns from the two frames to be somewhat different. In the overlap, the respondents from the ABS sample appear to be more likely to have fished in the wave but to have gone on fewer trips than the respondents from the license sample. As a result, the consequences for the bias and variance for estimating total trips from a dual frame estimator are not clear.

To better understand the consequences for the dual frame estimators we created three dual frame estimators. The estimators were all of the simple form of averaging the overlap estimates from the two frames to produce an overlap estimate, and then adding the non-overlap estimates from the separate frames. . More specifically, let $\hat{y}_{12}^{1}$ and $\hat{y}_{12}^{2}$ be the weighted estimates of the overlap domain from frame 1 (the ABS frame) and frame 2 (the license frame), respectively, then an average or composite dual frame estimator is $\hat{y}_{ave} = \hat{y}_1 + \hat{y}_2 + \lambda \hat{y}_{12}^{1} + (1-\lambda)\hat{y}_{12}^{2}$, with $0 \le \lambda \le 1$, where the subscript 1 denotes the non-overlap component from the ABS frame and 2 is the non-overlap component from the license frame. Lohr (2009) provides a good discussion of these estimators.

.

The typical assumption is that $\hat{y}_1$ and $\hat{y}_b$ $\hat{y}_2$ are unbiased for the totals in the two nonoverlapping domains, and $\hat{y}_{12}^{1}$ and $\hat{y}_{12}^{2}$ are both unbiased for the total in the overlap domain. If this set of assumptions holds, then $\hat{y}_{ave}$ is an unbiased estimator of the total. To produce estimates of characteristics using weights, the weights for units in the overlap that are sampled from frame 1 are multiplied by λ and the weights for overlap units sampled from frame 2 are multiplied by $(1-\lambda)$.

Our main concern is that the assumption that $\hat{y}_{12}^{1}$ and $\hat{y}_{12}^{2}$ are both unbiased for the total of the overlap domain may not hold, since the estimated number of anglers and average trips per angler from the overlap differ by frame. The assumption of unbiasedness for the non-overlap component estimates is also a concern, but we do not have any evidence from the survey to evaluate it.

As a simple method of evaluating the effect the choice of the compositing factor might have on the bias and standard errors of the estimates, we created three dual frame estimators with $\lambda = 0.2$, 0.5, and 0.8. The standard choice might have been to choose $\lambda = 0.2$, since about this percentage of the overlap cases were from the ABS frame (25 of the 142 who fished in the wave). Because the weights were so much larger for the ABS cases and their contribution to the variance might be large, another reasonable choice might have been closer to $\lambda = 0.5$. The choice of $\lambda = 0.8$ was used to investigate a compositing factor that was very different from these more reasonable factors.

Table 10 gives estimates of the number of anglers, the percent of anglers, the number of trips (boat, shore and total), and the mean number of trips by stratum and overall for the three estimators. The first two columns give the estimates and their standard errors computed using $\lambda = 0.5$. All of the estimates of standard errors were computed using replication methods. The next two columns give the ratio of the estimates for $\lambda = 0.2$ and $\lambda = 0.8$ to the estimates to $\lambda = 0.5$. When these estimates equal unity, it means the choice of λ did not affect the magnitude of the estimates. Scanning over the column shows the effect on the magnitude from the choice of the λ is not very large, with only the few bolded estimates outside of the range (0.95 to 1.05). The last two columns show the effect on the precision of the estimates by taking the ratio of the standard errors of the estimators using $\lambda = 0.2$ and $\lambda = 0.8$ to the standard errors to $\lambda = 0.5$. Once again, there are few ratios outside of the range (0.95 to 1.05) and those are in bold. It appears that the standard errors for the estimators using $\lambda = 0.8$ are somewhat more affected by the choice of the compositing factor than the other estimators, as might be suspected.

In general, the estimators and standard errors seem to be fairly robust to the choice of the compositing factor, especially for the two more reasonable choices of $\lambda = 0.2$ and $\lambda = 0.5$. One explanation for this robustness is the fact that the overlap only has 37% of all the anglers who fished during the last year. (This estimated percentage varies slightly depending on the choice of the compositing factor). The non-overlap component from the license frame is about 5% of the total angler estimate while the non-overlap component from the ABS is about 58%. If the license frame were more complete, then the overlap would be a larger component of the total and the dual frame estimators and standard errors might be less robust; i.e., changes in the compositing factor might be more important to bias and standard errors of the dual frame estimators. However, under the current circumstances we can be fairly confident that the compositing factor can be chosen using standard methods without introducing large biases or inefficiencies.

Table 10. Dual Frame Estimates of Anglers and Trips By Compositing Factor.

		Computations with Composite Factor λ=0.5		Ratio of Estimates to Composite with λ=0.5		Ratio of std Error to Composite with λ=0.5	
Estimate	**Stratum**	estimate	std err	λ=0.2	λ=0.8	λ=0.2	λ=0.8
Number of Anglers	**Coastal**	122,625	18,562	0.96	1.04	**0.94**	**1.12**
	Non-Coastal	101,894	37,497	0.97	1.03	0.97	**1.06**
	Out of State	10,225	939	1.00	1.00	1.00	1.00
	Overall	234,743	42,035	0.96	1.04	0.97	**1.07**
Percent of Anglers	**Coastal**	52.24	9.82	1.00	1.00	1.01	1.03
	Non-Coastal	43.41	10.37	1.00	1.00	1.01	1.02
	Out of State	4.36	0.89	1.03	0.97	1.05	0.99
Number of Boat Trips	**Coastal**	246,294	45,182	0.97	1.03	1.00	**1.09**
	Non-Coastal	131,565	68,746	0.98	1.02	0.98	1.04
	Out of State	14,432	3,142	1.00	1.00	1.00	1.00
	Overall	392,291	83,215	0.98	1.02	1.00	1.05
Number of Shore Trips	**Coastal**	342,487	144,812	0.98	1.02	1.00	1.01
	Non-Coastal	151,760	68,235	**1.18**	**0.82**	1.01	1.00
	Out of State	35,094	4,623	1.00	1.00	1.00	1.00
	Overall	529,341	159,098	1.04	0.96	1.00	1.01
Number of Total Trips	**Coastal**	588,781	165,348	0.98	1.02	1.00	1.02
	Non-Coastal	283,325	132,638	1.09	0.91	1.00	1.02
	Out of State	49,525	6,813	1.00	1.00	1.00	1.00
	Overall	921,631	211,898	1.01	0.99	1.00	1.02
Mean Number of Boat Trips	**Coastal**	2.01	0.237	1.01	0.99	1.04	**1.09**
	Non-Coastal	1.29	0.583	1.02	0.99	1.03	0.99
	Out of State	1.41	0.246	1.00	1.00	1.00	1.00
	Overall	1.67	0.265	1.01	0.99	1.03	1.02
Mean Number of Shore Trips	**Coastal**	2.79	1.041	1.03	0.98	1.04	0.97
	Non-Coastal	1.49	0.596	**1.22**	**0.79**	**1.09**	0.97
	Out of State	3.43	0.298	1.00	1.00	1.00	1.00
	Overall	2.25	0.622	**1.08**	**0.93**	1.04	0.98
Mean Number of Total Trips	**Coastal**	4.8	1.066	1.02	0.98	1.03	0.98
	Non-Coastal	2.78	1.103	**1.12**	**0.88**	1.05	0.99
	Out of State	4.84	0.406	1.00	1.00	1.00	1.00
	Overall	3.93	0.748	1.05	0.95	1.04	0.98

Magnitude of clustering

In the pilot survey, data were collected from more than one angler in some households in both frames, when they were identified. This allowed for an investigation of the similarity between the responses obtained from two anglers in the same household. In addition, a previous study (Lin, 2009) had shown that in the CHTS, the responses for multiple anglers in the same household have such high correlation that there is some question about whether or not attempts to obtain information from multiple anglers is even worthwhile. We wanted to see if that remains true with the self-administered mail survey. We believed that the within-household correlation might be reduced in the mail survey, due to the fact that the responses for multiple anglers are often obtained from a single household respondent in the telephone survey, and in the mail survey each individual angler received his or her own questionnaire. In this section, we describe how we estimated the level of clustering for both angler and trip characteristics within a household from the mail survey. Then we compare those estimates to similar estimates for the telephone survey from the same time period.

Clustering of angler behavior within household

Because *ICC* (intra-cluster correlation) is defined only for clusters of equal size, we use a more general measure of clustering, the adjusted R^2, denoted R_a^2, to describe the effect. This parameter is defined (Lohr 2010, p. 175) as

$$R_a^2 = 1 - \frac{MSW}{S^2}, \qquad (1)$$

where MSW and S^2 are defined as in an analysis of variance; i.e.,

$$MSW = SSW/(K-N) = \sum_{i=1}^{N}\sum_{j=1}^{M_i}(y_{ij}-\bar{y}_{iU})^2 \Big/ (K-N)\,, \tag{2}$$

$$S^2 = SST/(K-1) = \sum_{i=1}^{N}\sum_{j=1}^{M_i}(y_{ij}-\bar{y}_{U})^2 \Big/ (K-1)\,, \tag{3}$$

$K = \sum_{i=1}^{K} M_i$ = # of secondary sampling units (anglers) in the population, N = number of psu's (households) in the population, and M_i = # of ssu's in the i^{th} psu. Because these parameters are to be estimated from a complex design, weights are needed, and each frame and variable requires its own estimator due to differences in the designs.

First we consider estimation of R_a^2 for number of shore trips, boat trips, and total trips in the license frame. In this case, we actually don't know for sure the number of licensed anglers within each household. However, the sample from the license frame was matched to the total license frame, and whenever an address match was found, the second angler was also sampled. The angler-level weighting of this sample then assumed that exactly two licensed anglers were present in every household in which a match was found. Thus we assume M_i = 1 or 2 for all households in the license frame. Note that households with only one angler make no contribution to SSW, but they do make a contribution to SST. There are two reasonable ways to estimate R_a^2 for this frame. One is that we use all households, with the one-angler households contributing only to S^2 but not *MSW*. This would also require estimating N, the number of households represented on the license frame.[11] A second approach is to compute R_a^2 only for that subset of the

[11] This could be done using the method in the JOS paper. That is, we could estimate the average number of licensed anglers per household for each stratum and divide the total number of anglers on the frame in each stratum by this quantity, and sum them over strata.

population that contains multiple licensed angler households; i.e., those for which $M_i = 2$. We take this approach, since it makes the results from the two frames more comparable, due to the fact that the proportion of households having multiple licensed anglers on the license frame may differ from the proportion of households having multiple anglers in the address frame. Thus we estimate

$$M\hat{S}W = S\hat{S}W/(2\hat{N}_m - \hat{N}_m) = \sum_{i=1}^{n_m}\sum_{j=1}^{2} w_{ij}(y_{ij} - \bar{y}_{iU})^2 \Big/ \hat{N}_m \qquad (4)$$

and

$$\hat{S}^2 = \frac{S\hat{S}B + S\hat{S}W}{2\hat{N}_m - 1}, \qquad (5)$$

(see Lohr 2010, p. 177) where n_m is the number of sampled households with 2 licensed anglers, $\hat{N}_m = \hat{K}_m/2 = \sum_{i=1}^{n_m}\sum_{j=1}^{2} w_{ij}/2$ is the estimate of the number of households in the population with 2 licensed anglers, and $S\hat{S}B = \sum_{i=1}^{n_m}\sum_{j=1}^{2} w_{ij}(\bar{y}_{iU} - \hat{\bar{y}}_U)^2$, where $\hat{\bar{y}}_U$ is the estimate of population mean from the complex design.[12] The first three rows of Table 11 show the components of these estimators, as well as the resulting estimateestimates of R_a^2 for the effort variables in the license frame.

Recall that differences in the design of the license and ABS sample caused differences in estimates of effort. In the ABS frame, only those anglers who fished during the past year were sampled in the second phase, which made the 2nd phase ABS anglers potentially more avid than the anglers sampled from the license frame. To make the two samples

[12] The estimates for the sums of squares for the complex design can be obtained using SAS PROC SURVEYREG's ANOVA table.

more comparable, we produced estimates of effort for the license frame that first filtered on the flag indicating whether or not the angler had fished in the last year. It seems reasonable that the same difference in design might cause different estimates of the clustering parameter as well. Therefore, we also made estimates of R_a^2 for the population of households containing two licensed anglers who have fished in the last year. These estimates were calculated from (4) and (5), but this time for the population of households containing two *active* licensed anglers. The results for these estimates are shown in rows 4 through 6 of Table 11. The differences in the correlations for the two populations are slight.

Table 11. Computation of R_a^2 for angler effort variables for License frame

Population	Variable	n_m	$\hat{K}$	$\hat{N}_m$	$S\hat{S}W$	$S\hat{S}B$	R_a^2
All licensed anglers in multiple angler hh's	Shore trips	134	82,540	41,270	82,145	158,867	0.32
	Boat trips	134	82,540	41,270	129,636	194,110	0.20
	Total trips	134	82,540	41,270	330,029	446,974	0.15
All active licensed anglers in multiple angler hh's	Shore trips	102	59,345	29,672	79,284	150,573	0.31
	Boat trips	102	59,345	29,672	129,636	186,443	0.18
	Total trips	102	59,345	29,672	327,167	418,003	0.12

Next we consider estimation of R_a^2 for the ABS frame. A different estimation method is required due to a difference in the design that was used to sample anglers within households, and the information available about the size of the household clusters. In the ABS frame, two anglers were sampled from a subset of the multiple angler households in the sample, and a single angler was sampled from the rest. In all cases, the number of anglers in the household was known. The angler weights that were calculated for the

ABS sample used the information about the number of anglers in the household, and so varied from one household to another, even within the same stratum and non-response weighting class. As with the license frame, we can use respondents in all households to estimate S^2 defined in (3), or only those respondents who contribute to estimation of MSW; i.e., those in households from which we sampled two members. As before, we chose the latter method. Thus the parameters being estimated will again be for the subset of the ABS frame residing in households with at least two adult active anglers. The estimators of the parameters in (2) and (3) are thus

$$M\hat{S}W = S\hat{S}W/(\hat{K}_m - \hat{N}_m) = \sum_{i=1}^{n_m}\sum_{j=1}^{2} w_{ij}(y_{ij} - \hat{\bar{y}}_{iU})^2 \Big/ (\hat{K}_m - \hat{N}_m) \qquad (6)$$

and

$$\hat{S}^2 = \frac{S\hat{S}B + S\hat{S}W}{\hat{K}_m - 1} \qquad (7)$$

(see Lohr 2010, p. 177) where n_m is the number of sampled households with 2 licensed anglers, $\hat{\bar{y}}_{iU}$ is the estimate of mean for household i, $\hat{K}_m = \sum_{i=1}^{n_m}\sum_{j=1}^{2} w_{ij}$, and

$\hat{N}_m = \hat{K}_m / \hat{\bar{M}}_m$, where $\hat{\bar{M}}_m = \sum_{i=1}^{n_m} w_i M_i / \sum_{i=1}^{n_m} w_i$ is an estimate of the average number of anglers in households with multiple anglers, and w_i is a household weight computed from the angler weights (w_i = (2/ # of adult anglers in hh)*w_{ij}). These estimates are used to form an estimate of R_a^2 as shown in (1). The results are shown in Table 12. Note that the sample size is much smaller in this case than the license sample; only 17 households in the sample had responses from 2 active anglers, so the estimates have high variability.

Table 12. Computation of R_a^2 for angler effort variables for ABS frame

Population	**Variable**	n_m	$\hat{K}$	$\hat{N}_m$	$S\hat{S}W$	$S\hat{S}B$	R_a^2
All active licensed anglers in multiple angler hh's	Shore trips	17	177,747	78,270	438,928	453,380	0.12
	Boat trips	17	177,747	78,270	63,512	86,729	0.24
	Total trips	17	177,747	78,270	549,682	591,967	0.14

The estimates of R_a^2 for boat trips and total trips are very similar to those for the license frame, while the estimate of R_a^2 for shore trips is slightly lower, though the small sample size for the ABS frame may be the cause of this.

Clustering of trip-level characteristics within angler

Next we consider estimation of R_a^2 for trip-level characteristics. There are two levels of clustering for trips: within angler and within household clustering. The analysis here estimates the correlation of trip characteristics within angler. As noted earlier, the respondents were asked to profile only their four most recent trips. This does provide some information about the clustering within angler on characteristics such as public/private access or time of return. However, the profiled trips are not a probability sample of trips made in the wave. Despite this, we did use the data to make estimates of R_a^2. To the extent that the four recalled trips have similar characteristics to a random sample of trips made by the angler, the estimates will be valid.

We estimated R_a^2 as shown in (1), but this time the two mean squares must be defined differently:

$$MSW = SSW/(T-K) = \sum_{i=1}^{N}\sum_{j=1}^{M_i}\sum_{k=1}^{T_{ij}} (y_{ijk} - \bar{y}_{ijU})^2 \Big/ (T-K)\,, \tag{8}$$

$$S^2 = SST/(T-1) = \sum_{i=1}^{N}\sum_{j=1}^{M_i}\sum_{k=1}^{T_{ij}} (y_{ijk} - \bar{y}_U)^2 \Big/ (T-1)\,, \tag{9}$$

where T is the total number of trips in the population, y_{ijk} is a characteristic of the k^{th} trip made by the j^{th} angler in household i (referred to henceforth as the $(i,j)^{\text{th}}$ angler), and $\bar{y}_{ijU}$ is the mean of all trips made by that angler. To estimate MSW and S^2, we used only those anglers who made at least two trips in estimation of both sums of squares. Thus

$$M\hat{S}W = S\hat{S}W/(\hat{T} - \hat{K}) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{t_{ij}} w_{ijk}(y_{ijk} - \hat{\bar{y}}_{ijU})^2 \Big/ (\hat{T} - \hat{K}) \tag{10}$$

and

$$\hat{S}^2 = \frac{S\hat{S}B + S\hat{S}W}{\hat{T} - 1}\,, \tag{11}$$

where n and m_i are the number of households and anglers in the subsample of anglers with multiple trips, t_{ij} is the number of trips reported by the $(i,j)^{\text{th}}$ angler, $\hat{\bar{y}}_{ijU}$ is the estimate of mean for the trips of the (i,j)$^{\text{th}}$ angler, $S\hat{S}B = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{t_{ij}} w_{ijk}(\bar{y}_{ijs} - \hat{\bar{y}}_U)^2$,

$\hat{T} = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{t_{ij}} w_{ijk}$, and $\hat{K} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} w_{ij}$. The weight w_{ijk} was constructed by assuming that the profiled trips are a random sample of all trips made by the angler, yielding

$$w_{ijk} = w_{ij} * (\text{\# of trips made by angler } (i,j))/(\text{\# of trips profiled by angler } (i,j))\,. \tag{12}$$

We also made a second estimate of R_a^2 only for that subset of anglers who reported all their trips, to see if the (untrue) assumption that the sampled trips were a random sample of all the angler's trips made a substantial difference in the estimate.

For completeness, we present in Table 13 a summary of the four variables we will be examining for within angler correlation: the number of anglers on each trip reported (TOT), whether or not the trip was (or ended) at a public site (PUB), whether or not it ended between 6:00 p.m. and 6:00 a.m. (LATE), and whether it included an accompanying child (CHILD). The trips accessed through the two frames do appear to be quite different, with those from the license frame more likely to be at public sites and less likely to end during night hours, which suggests they are more closely aligned with the trips profiled by the intercept survey. The trips accessed through the license frame appear to be less likely to include additional family members than those encountered through the ABS frame.

Table 13. Estimates of trip characteristics for the two frames

Sample from:	**Mean # of anglers on trip (sd)**	**Proportion of trips in public site (sd)**	**Proportion of trips ending at night (sd)**	**Proportion of trips including a child (sd)**
License frame	1.6 (0.1)	0.81 (0.04)	0.21 (0.04)	0.15 (0.03)
ABS frame	3.0 (0.7)	0.67 (0.12)	0.41 (0.15)	0.50 (0.14)

Table 14 summarizes the calculations (using (1), (8) and (9)) for estimating R_a^2 for the 4 variables described in Table 13. These calculations were carried out for the samples from each frame.

Table 14. Computation of R_a^2 for angler effort variables for two frames

Sample from…	Variable	$\hat{T}$	$\hat{K}$	$S\hat{S}W$	$S\hat{S}B$	R_a^2
Trips accessed through LIC frame for domain of anglers reporting >1 trip	TOT	258,551	48,780	87,477	378,619	0.72
	PUB	258,551	48,780	16,609	26,028	0.52
	LATE	258,551	48,780	12,732	28,631	0.62
	CHILD	258,551	48,780	7,310	30,448	0.76
Trips accessed through ABS frame for domain of anglers reporting > 1 trip	TOT	828,266	216,378	35,201	3,597,963	0.99
	PUB	828,266	216,378	7,574	175,568	0.94
	LATE	828,266	216,378	11,757	189,238	0.92
	CHILD	828,266	216,378	14,085	192,896	0.91

The results show that the trips made by an angler tend to be quite similar. This is especially true for the trips taken by anglers in the ABS frame.

Comparison with Telephone Frame

Simultaneously with the mail survey experiment, a dual frame telephone survey was conducted, which collected similar data about anglers and their fishing trips. The two frames were an RDD frame (CHTS) and the license frame (ALDS). The CHTS chose telephone numbers only from coastal households, while the ALDS sample drew from all licensees whose telephone numbers could be discerned from the license frame. We used the data from that survey to make estimates of R_a^2, for total number of trips and for two of the trip characteristic variables (whether or not the trip ended at a public site, PUB, or between 6:00 p.m. and 6:00 a.m., LATE), which we compared with those from the mail survey.

There were differences in the sample designs of mail and telephone that make the measures of correlation apply to different populations, and therefore which may not be directly comparable. In the telephone survey, information was collected about every angler in the household, so that clusters of more than two anglers were possible. Since there was no matching to the license frame, there was no way to identify who was licensed and who was not in the CHTS, so correlations were computed for all anglers in the household, whether they were licensed or not. In the collection of trip characteristics, anglers were required to recall all the trips he/she took, rather than the four most recent ones[13]. So by design, there should have been no sampling at the last stage, but rather a complete observation of trips within anglers. However, many anglers did not provide information for all trips. During Wave 6, 2009 in NC, 71.51% of trip records were imputed (i.e. not profiled).Instead, weights were created to account for the missing trips, based on the number of trips reported by the angler, and effectively the trips that were reported were treated as though they were a random sample of trips for the purpose of estimating R_a^2.

The estimator of R_a^2 that we used for angler characteristics in both frames was the same as that shown in (6) and (7), except that the upper limit of the inner sum in (6) can be larger than 2, since data was collected about all the anglers in a household. The estimator of R_a^2 that we used for trip characteristics in both frames was the same as that shown in (10) and (11). The weight associated with the $(i,j,k)^{th}$ trip is defined as in (12), though the absent profiles were due to nonresponse, rather than from the instruction to profile only the most recent trips.

Results are shown in Table 15 for the angler characteristic, total number of trips, for both frames. A comparison of Tables 9, 10, and 15 shows that the correlation of effort within household is much larger for the telephone than for the mail survey, as expected.

[13] Note that proxy reported information is accepted for the telephone surveys if the individual angler(s) can not be interviewed. In these cases, a respondent may be reporting about his or her own trips as well as those of other household members.

Table 15: Computation of R_a^2 for angler effort for the two telephone frames:

Sample From...	Variable	n_m	$\hat{K}$	$\hat{N}_m$	$S\hat{S}W$	$S\hat{S}B$	R_a^2
All anglers in multi-angler hh in ALDS	Total trips	82	93,697	41,994	53,501	1,703,411	0.945
All anglers in multi-angler hh in CHTS	Total trips	30	12,184	5423	67.99	1,284,853	0.999

For completeness, summary data for the two trip characteristic variables for the two telephone frames is shown in Table 16.

Table 16. Estimates of trip characteristics for two frames

Sample From:	Proportion of trips in public site (sd)	Proportion of trips ending at night (sd)
Anglers with > 0 trips in ALDS	0.80 (0.04)	0.11 (0.02)
Anglers with > 0 trips in CHTS	0.69 (0.07)	0.11 (0.04)

Table 17 displays the estimated R_a^2 for the trip characteristic variables for the telephone frames. They are much smaller than the correlation among anglers in the same households. Comparison with Table 14 shows that the correlations of characteristics among trips by the same angler is similar for the mail and telephone frames in the license frame, but not for the CHTS/ABS frame. The anomaly seems to be the correlation for the ABS frame, which is unusually high, and much higher than the correlation for the CHTS frame. One could imagine that tourists to the coast from non-coastal counties may take more similar trips, especially since they were instructed to report only their last 4 (consecutive) trips. The CHTS would contain no such non-coastal anglers in its sample, while the ABS does contain such anglers. Still the magnitude of this difference is hard to explain.

Table 17. Computation of R_a^2 for angler effort variables for two frames

SAMPLE FROM…	VARIABLE	$\hat{T}$	$\hat{K}$	$S\hat{S}W$	$S\hat{S}B$	R_a^2
Trips accessed through ALDS frame for domain of anglers reporting > 0 trips	LATE	132,775	20,226	5,065	8,270	0.55
	PUB	132,775	20,226	10,797	17,627	0.55
Trips accessed through CHTS frame for domain of anglers reporting >0 trips	LATE	537,895	126,644	21,578	29,947	0.45
	PUB	537,895	126,644	22,698	67,284	0.67

These findings about correlation suggest that a design which attempts to sample more than one angler from the same household is more cost effective for the mail survey than the telephone survey, for estimating effort. For trip characteristics, this does not appear to be so. However, the latter finding comes with the caveat that the method of sampling trips for an angler differed by mode. The mail survey asked respondents to describe their four most recent trips, which may explain why the trip characteristics would be more similar to each other than the characteristics of all trips made during the wave, which were requested of telephone respondents.

Discussion

The primary goal of the pilot study was to examine whether a self-administered two phase study could be successfully implemented to estimate fishing effort among NC anglers in the fall of 2009, with an eye toward improving both the coverage and the response rates currently achieved via telephone surveys. With respect to response rate and the feasibility of conducting a two-phase self administered survey among anglers, the response rates presented in Table 2 clearly indicate that such a design is feasible and offers a potential alternative to the RDD design currently used by MRIP. Both the two-phase approach used with an ABS frame as well as the single-phase approach based on a license frame yielded response rates that exceed the current response rates achieved via telephone data collection (CHTS and ALDS, respectively). But the response rates from the ABS sample also raise concerns about avidity bias, an issue in angler surveys regardless of the mode and method of data collection. We also see a pattern (albeit not significant) similar to findings from other studies (Montaquila et al., 2010) that a longer lag time between the screener survey field period and the mailing of the extended survey instrument may be beneficial with respect to response rates. The small sample for the field test limits our ability to draw additional conclusions or recommendations with respect to the details of fielding a two-phase dual frame study by mail, but does provide sufficient positive findings to motivate further research in this area.

Other indications of data quality, specifically missing data rates or data inconsistencies did not signal a red flag. We saw relatively low levels of missing data, with the exception of detailed trip reports for avid anglers. However, both the CHTS and ALDS telephone surveys are plagued with similar problems, with respondents either not providing detailed reports for each trip or opting for the response option that all trips are similar. Regardless of the mode of data collection, attempting to collect detailed trip level information for a two month recall period for avid anglers is difficult and may require a reconsideration of the data elements to be collected for these anglers.

With respect to the dual frame nature of the study, the study had two goals: to estimate the improvement in coverage the two frames provide and to examine means by which to identify the overlap among elements across the two frames. Here too we found significant gains via the use of a dual frame design consisting of a license frame and an addressed-based frame in comparison to the current CHTS and ALDS sample designs. The findings support the improvement in the identification of frame overlap via the use of addresses as compared to self-reported fishing licensure. Thus, the use of a self-administered mail survey (based on addresses from an ABS frame and a license frame) facilitates improved identification of overlapping sample members as compared to what is possible for a dual-frame telephone survey.

The findings clearly support empirical results that have been well established in the literature, namely the presence of avidity bias in surveys of recreational anglers. We are planning to test a revised household screener that allows respondents to provide information about other recreational activities besides fishing. The goal of the revised instrument is to reduce the fishing avidity bias in the ABS sample. For both the ABS and license sample, we plan to use the type of license in nonresponse adjustment to reduce nonresponse bias. We hope that both of these steps may reduce the differences in the estimates for the overlap domain.

The major limitation of the study is its small sample sizes for active anglers during the wave. This limitation makes it difficult to precisely estimate fishing effort and reduces our ability to understand differences in fishing behavior as a function of geography and ownership of landline phones. In addition, the small sample size makes it impossible to assess the degree to which the current approach to coverage adjustment in the MRIP, that is, the use expansion factors based on the APAIS, is fully representative of all fishing trips. While we see indications of differences in these population subgroups which have traditionally been under-covered in the telephone surveys, we cannot address the extent to which their actual fishing behavior differs.

As is true for many exploratory pilot studies, the goal was not to be able to provide the definitive answer with respect to a redesign of the current MRIP telephone surveys. Rather, the pilot was successful in examining the feasibility of moving away from the telephone to a self-administered two-phase survey. It also clearly demonstrated the utility of this design in the context of a dual frame sample. The success of the two-phase mail survey, especially with respect to the dual frame design, shows the substantial potential for improving future angler surveys.

Recommendations

The 2009 two phase dual frame study conducted in North Carolina was a first step toward exploring sample and design options to address coverage, efficiency, and other issues that were raised in the report of the National Research Council. As noted above, the size of the sample limits our ability to offer definitive recommendations for a full scale redesign of the MRIP program, but the findings do suggest the following:

1. This study, as with other empirical studies, clearly indicates that surveys of anglers are subject to avidity bias. As noted above, we recommend further experimental studies to reduce avidity bias (e.g. broaden the base of the screener questionnaire) and further examination of how to reduce avidity bias through the use of license type information in nonresponse adjustments.

2. We suspect that the avidity bias evident in the mail survey also exists for CHTS and ALDS. We recommend implementing studies to test for avidity bias in CHTS and ALDS.

3. Matching household sample frames to license frames, regardless of whether using a dual-frame approach or a single-frame approach is a good approach to adjust for avidity bias. We would recommend this for surveys conducted by either the telephone or mail; however, telephone surveys would need to reverse link to addresses to facilitate this matching.

4. Conduct follow-up studies with sufficient sample sizes to test the assumption that the APAIS (intercept) survey is representative of all trips (e.g. do the trips that we can cover in the APAIS (public access) adequately represent all trips?). Sufficient sample sizes would also facilitate more robust estimation of trip-level information and comparisons of the effort levels and characteristics of trips by frame and for subgroups not currently covered by the CHTS.

5. Not addressed in the present study is the need for timely data. Clearly a shift toward self-administered mail surveys comes at the potential cost of longer field periods than comparable telephone surveys. This is particularly true

when the survey involves the need to screen households (the two phase ABS frame design). Future studies should examine the relative speed of the CHTS/ALDS design compared to a mail mode (or possibly mixed mode).

References

American Association for Public Opinion Research (AAPOR) (2009). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions&Template=/CM/ContentDisplay.cfm&ContentID=1819

Blumberg, Stephen and Luke, Julian. (2010). Wireless substitution: Early release of estimates from the National Health Interview Survey, July-December 2009 *National Center for Health Statistics*. Available from: http://www.cdc.gov/nchs/nhis.htm.

Brick J. M., Cervantes, I. F., Lee, S., and Norman, G. (forthcoming). Nonsampling Errors in Dual Frame Telephone Surveys, *Survey Methodology*.

Connelly, Nancy A., Brown, Tommy, and Knuth, Barbara (2000) Assessing the Relative Importance of Recall Bias and Non-response Bias and Adjusting for Those Biases in Statewide Angler Surveys. *Human Dimensions of Wildlife*, 5:19-29.

Dillman, Don .A., Smyth, Jolene. D., and Christian, Leah. M. (2008). *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Wiley & Sons, New York.

DiSogra, Charles, Callegaro, Mario, and Hendarwan, Erlina (2009) Recruiting Probability-Based Web Panel Members Using an Address-Based Sample Frame: Results from a Pilot Study Conducted by Knowledge Networks. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 5270- 5283.

Fahimi, M., Kulp, D., and Brick, J.M. (2009). A reassessment of list-assisted RDD methodology. *Public Opinion Quarterly*, 73, 751-760.

Fisher, Mark (1996) Estimating the Effect of Nonresponse Bias on Angler Surveys. *Transactions of the American Fisheries Society* 125: 118-126.

Han, Daifeng, Cantor, David, Brick, Pat Dean, and Aponte, Maribel. (2010). Findings from a Two-Phase Mail Survey for a Study of Veterans. *Presentation at the American Association of Public Opinion Research Meetings,* Chicago, Il.

Iannacchione, Vince, G., Stabb, Jennifer, M., and Redden, David T. (2003) Evaluating the Use of Residential Mailing Addresses in a Metropolitan Household Survey. *Public Opinion Quarterly,* 67:202-210.

Kennel,T. and Li, M. (2009). Content and coverage quality of a commercial address list as a national sampling frame for household surveys. *Proceedings of the Survey Research Methods Section, American Statistical Association,* 2364-2378.

Lin, Dong (2009) "CHTS and ALDS Data Summary." Report submitted to the National Oceanic and Atmospheric Administration.

Link, Michael W., Battaglia, Michael P., Frankel, Martin R., Osborn, Larry, and Mokdad, Ali H. (2008) A comparison of Address-Based Sampling (ABS) Versus Random Digit Dialing (RDD) for General Population Surveys. *Public Opinion Quarterly,* 72(1):6-27.

Link, Michael W., Daily, Gail, Shuttles, Charles, D., Yancey, L. Tracie, Burks, Anh Thu, and H. Christine Bourquin (2009) Building a New Foundation: Transitioning to Address Based Sampling After 30 Years of RDD. Paper presented at the 64th Annual Meeting of the American Association for Public Opinion Research, Hollywood, FL.

Lohr, S. (2009). Multiple Frame Surveys. **Chapter 4** in D. Pfeffermann and C. R. Rao. (Eds.) *Sample Surveys: Design, Methods and Applications, Vol. 29A,* (pp. 71-88) Elsevier, The Netherlands: North-Holland.

Montaquila, Jill M., Brick, J. Michael, Hagedorn, Mary, C., Williams, Douglas, W. (2010). Maximizing Response in a Two-Phase Survey with Mail as the Primary

Mode. *Presentation at the American Association of Public Opinion Research Meetings*,Chicago,Il.

National Research Council (2006) *Review of Recreational Fisheries Survey Methods*. Washington, D.C.: National Academies Press.

Thomson, Cynthia (1991) Effects of the Avidity Bias on Survey Estaimtes of Fishing Effort and Economic Value. *American Fisheries Society Symposium*, 12:356-366.

Appendix A: Disposition of ABS and License Sample Units

Table A-1. Screener Disposition for ABS Sample, Both Waves, by Stratum

Disposition	Coastal Counties	Non-Coastal Counties	Total
Total Completes	357	328	685
With Anglers	154	74	228
Without Anglers	203	254	457
Refusals	14	8	22
Bad Address	78	79	157
Unknown/No Response	451	485	936
Totals	900	900	1800

Table A-2. Angler Survey Disposition for ABS Sample, by Stratum

Disposition	Coastal Counties	Non-Coastal Counties	Total
Total Completes	130	57	187
With Trips	43	8	51
Without Trips	87	49	136
Refusals	1	3	4
Bad Address	5	3	8
Unknown/No Response	47	16	63
Totals	183	79	262

Table A-3. Angler Survey Disposition of License Sample, Both Waves, by Stratum

Disposition	Coastal Counties	Non-Coastal Counties	Out of State	Total
Total Completes	316	307	343	966
With Trips	76	43	108	227
Without Trips	240	264	235	739
Refusals	5	9	5	19
Bad Address	51	41	47	139
Unknown/No Response	164	167	107	438
Totals	536	524	502	1562

Dual Frame Mail Survey of Fishing Effort: Project Documentation

Issued under Contract No. DG133F-09-RQ-0666

Conducted on behalf the National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Office of Science and Technology

ICF Macro
an ICF International Company

116 John Street
New York, NY 10038
Seth Muzzy, Principal
sMuzzy@icfi.com

Contents

Introduction

Since 1981, the Federal government has relied upon telephone-based general population interviews to estimate fishing effort and catch by marine recreational anglers. However, increasing issues with telephone frame coverage has caused the National Marine Fisheries Service (NMFS) to investigate alternate methodologies which may lead to increased efficiency and reduced coverage error.

As a part of this effort, the Dual-Frame Mail Survey of Fishing Effort pilot study was awarded to ICF Macro under the Blanket Purchase Agreement DG133F-09-RQ-0666.

Project Background

Historically, recreational fisheries estimates have been developed through two main components:

- An access-site intercept study (the Atlantic Coast Access Point Angler Intercept Survey, APAIS) which documents angler activity and catch; and
- Telephone surveys of fishing effort such as the Coastal Household Telephone Survey (CHTS) which primarily operate as a weighting factor to expand angler data to represent activity across all recreational fisheries.

One of the key statistics derived from the CHTS is the incidence of saltwater recreational anglers living in the coastal regions of the country. This is obtained through a relatively efficient random digit dialing (RDD) methodology targeting relevant coastal counties. However, coverage errors may weaken the integrity of resulting statistics. Specifically:

- The CHTS only incorporates traditional land–line telephone numbers in its sample frame. The National Center for Health Statistics estimated that, at the end of 2008, about one-in-seven American households received all or most calls using cellular telephones. The demographics of these households are statistically unique from those which can be contacted using a traditional landline telephone number (Blumberg & Luke, 2009).
- The CHTS limits the sample frame to areas with the highest concentrations of anglers. Specifically, non-coastal anglers and anglers active in northern states during winter months do not have a probability for selection.

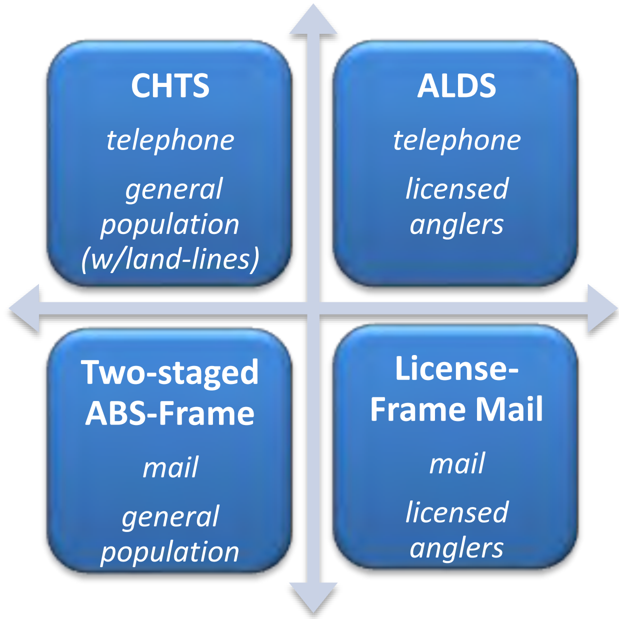
In addition, the RDD effort lacks the ability to efficiently profile adequate numbers of anglers needed to produce effective fisheries management information. A significant investment is required to produce precise figures regarding a wide variety of fishing behaviors.

License-based angler frames promise to be a primary component to resolving MRFSS' methodological issues as the program is refined as part of the Marine Recreational Information Program (MRIP) initiative. Since early 2007, the Angler Directory License

Survey has supplied data similar to the CHTS, economically providing additional details about fishing behaviors by utilizing state-based registration databases as sample frames. However, this dual-frame approach does not resolve all MRFSS coverage issues. Specifically, registration laws provide exemptions to some anglers and not all active anglers register with the state, thereby weakening state databases.

Survey Design

A dual-frame multi-stage collection methodology has been designed to mirror current CHTS and ALDS activity, adapting the current telephone methodology to a mail-based approach.



- Similar to the CHTS, a general population survey identifies households with residents recently participating in fishing activities.
- A follow-up survey sent to anglers identified in the household survey provides detail of recent activity.
- The same follow up survey sent to select registered anglers efficiently increases the amount of angler data.

During analysis, data resulting from each mail effort may be assessed in relation to the data collected via its analogous sampling frame or collection procedure.

Sample Design: Delivery Sequence File

In order to obtain an accurate estimate of the incidence of anglers in the general population, sampling was conducted using the Delivery Sequence File (DSF) from the United States Postal Service (USPS). The DSF includes addresses with both single-family style addresses and multi-unit residential property addresses such as for apartments, condominiums, and trailer properties. Non-city style addresses (i.e. post office boxes) are not included. The Census Bureau reports that in areas where city-style addresses are prominent, people who receive mail at post office boxes will often also receive postal mail at their city-style address. This assertion has been backed by other researchers who have concluded that most people who maintain a post office box also receive postal mail at their physical residence (Iannacchione, Staab, & Redden, 2003).

Records selected from the DSF were limited to households in the State of North Carolina. Addresses were stratified into Coastal/Non-Coastal classifications consistent with CHTS sampling during November and December in North Carolina. A total of 1,800 households were selected, split evenly between the two strata.

Coastal Counties		Non-Coastal Counties			
013 Beaufort	103 Jones	001 Alamance	051 Cumberland	101 Johnston	159 Rowan
015 Bertie	107 Lenoir	003 Alexander	057 Davidson	105 Lee	161 Rutherford
017 Bladen	117 Martin	005 Alleghany	059 Davie	109 Lincoln	165 Scotland
019 Brunswick	129 New Hanover	007 Anson	063 Durham	111 McDowell	167 Stanly
029 Camden	131 Northampton	009 Ashe	067 Forsyth	113 Macon	169 Stokes
031 Carteret	133 Onslow	011 Avery	069 Franklin	115 Madison	171 Surry
041 Chowan	137 Pamlico	021 Buncombe	071 Gaston	119 Mecklenburg	173 Swain
047 Columbus	139 Pasquotank	023 Burke	075 Graham	121 Mitchell	175 Transylvania
049 Craven	141 Pender	025 Cabarrus	077 Granville	123 Montgomery	179 Union
053 Currituck	143 Perquimans	027 Caldwell	081 Guilford	125 Moore	181 Vance
055 Dare	147 Pitt	033 Caswell	085 Harnett	127 Nash	183 Wake
061 Duplin	149 Polk	035 Catawba	087 Haywood	135 Orange	185 Warren
065 Edgecombe	155 Robeson	0 Chatham	089 Henderson	145 Person	189 Watauga
073 Gates	163 Sampson	039 Cherokee	093 Hoke	151 Randolph	193 Wilkes
079 Greene	177 Tyrrell	043 Clay	097 Iredell	153 Richmond	197 Yadkin
083 Halifax	187 Washington	045 Cleveland	099 Jackson	157 Rockingham	199 Yancey
091 Hertford	191 Wayne				
095 Hyde	195 Wilson				

Sample Design: Angler Registry Frame

In order to conduct the Licensed Angler Study, a database containing approximately 551,060 million registered anglers was provided by North Carolina's Division of Marine Fisheries. In order to prepare the sample file for sampling, the following steps were completed.

- Duplicate records matching on core information such as name, date of birth, and mailing address were also deleted.
- Records lacking fundamental information such as name, date of birth, and mailing address were eliminated from the file.
- Anglers under the age of 18 were excluded.
- Addresses were "normalized" using Satori Software's "Mailroom Toolkit" which is designed to correct minor deviations from standard formats used by the USPS.
- Records were classified into appropriate coastal, non-coastal, or out-of-state strata groups.

- Unique household identifiers were assigned to anglers who share a common mailing address or telephone number.

The sample draw for the license-frame survey involved an n^{th} selection procedure for each stratum. A file listing households was sorted by address in order to minimize the possibility of including unidentified duplicate household listings. 450 records were selected from each stratum and designated as "original sample."

Supplemental sample was obtained from other anglers living in the same households as the original sample. At most one additional angler was selected for each household, with up to 100 secondary anglers permitted per stratum. Counts are listed in Table 1.

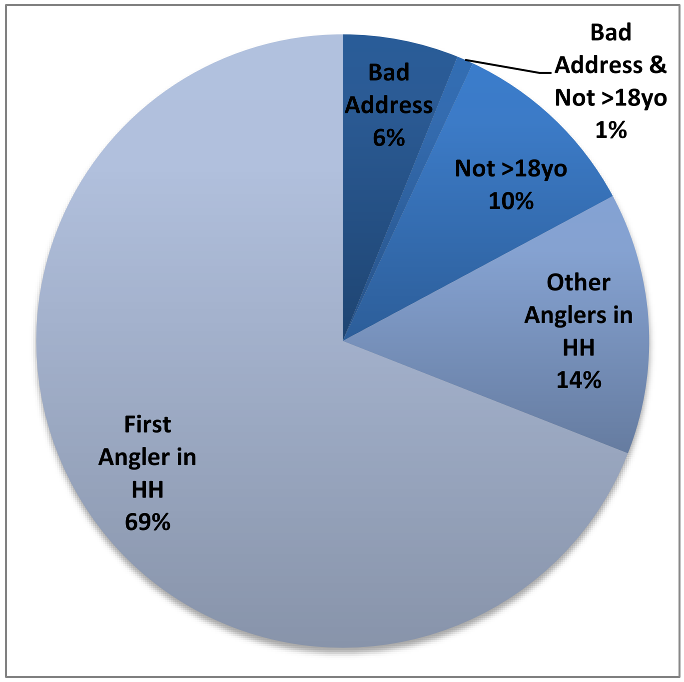


Table 1

Supplemental Records	Count
Coastal	86
Non-Coastal	74
Out-of-State	52

Questionnaire Design

Three primary data forms were designed for the study:

- An initial household screener for the ABS sample
- An angler survey for individuals identified through the household screener or using the North Carolina Angler Registry
- A trip form associated with the angler survey which captured details regarding up to four recent outings.

Questions in the mail survey were selected from key measures in the CHTS instrument. Wording modifications were required to adapt an interviewer guided telephone survey script to a self administered paper form.

Design of the initial household screener and the angler survey involved:

- Printing on 11" x 17" white paper later folded into a four page 8.5" x 11" booklet.
- A front cover incorporating the study name, NOAA logo, OMB approval number and expiration date, and informed consent information including an assurance of confidentiality. The front cover was printed in color.

- A back cover printed in color listing commonly asked questions including items involving sampling procedures, study purpose, anticipated time burden, and contact information for the survey sponsor.
- An interior spread clustering questions on the right-hand page. Major question groups were presented in shaded text boxes with response areas appearing in white. The booklet's control number appeared vertically as a form number along the crease of the booklet where it was protected from mutilation.
- The only design element appearing on the obverse of the front cover was a bar code of the respondent control number. The bar code was overlaid with a stencil of a fish, transforming it into a graphical element unlikely to be tampered with by a respondent.

A supplemental form for recording details of up to four recent trips was included with the angler questionnaire. The 11" x 17" page was printed on tan paper with black ink and folded so that all questions about each of the trips appeared independently on one 8.5" x 11" page.

Other designed components of the survey efforts included:

- A 10" x 13" white outbound envelope. The return address referenced "A Study of Fishing in NC" with the ICF Macro office location listed in the return address. NOAA's logo was prominently displayed next to the return address. The envelope was clearly marked with a "Return Service Requested" stamp to facilitate accurate classification of undeliverable pieces. Adhesive labels showing the respondent's address incorporated a unique numeric identifier to help ensure survey materials were properly matched to envelopes.
- A 9" x 12" business reply envelope (BRE). This BRE directed returns to "A Study of Fishing in NC" at the same ICF Macro office location printed on the outbound envelope.
- Cover letters. Five different cover letters were designed to motivate:
    - Households receiving an initial survey instrument,
    - Non-responding households receiving a replacement form,
    - Anglers receiving an initial angler activity survey instrument,
    - Non-responding anglers receiving a replacement forms, and
    - Non-responding anglers receiving a third and final form.

  An electronic letterhead included the NOAA logo, address, telephone number, and web address printed in color. Each motivational message displayed the signature of the NOAA's Fisheries Statistics Division's Chief, David Van Voorhees. Letters were personalized with an inside address (including the respondent's name if known).

- Postcards. Approximately one week after receiving an initial household or angler survey packet, respondents received a postcard reiterating the importance of response. Postcards were printed on white cardstock and prominently displayed the NOAA logo.

Images of survey material can be found in the appendices.

Data Collection

Assembly protocols

Household survey packets sent to the ABS participants included a cover letter, survey booklet, and business reply envelope. Initial surveys to households also included a one dollar bill clipped to the front of the packet. Outbound envelopes were stuffed with the BRE flap at the bottom of the envelope, cradling other components to ensure their orderly removal by the respondent.

Angler survey packets were assembled in a similar manner. A personalized cover letter, angler questionnaire and trip detail form were stacked and tucked into the lip of a BRE. As with the household study, initial mailings also included a dollar bill clipped to the front of the packet.

Survey materials for each mailing were sorted and printed in order of a process control number. Pieces were batched in groups of 100 and released to assembly staff by a process supervisor. If any materials were left over after assembling a batch of 100, the cause of the discrepancy was investigated and corrected. A supervisor performed a quality assurance check on approximately one out of every 10 envelopes noting proper nesting of materials and matching of all control numbers.

After assembly, packets were sealed and metered. A first-class postage rate was used in order to generate a positive impact on response rates (Fox, Crask, & Kim, 1988) and avoid possibly delays in delivery associated with second-class, third-class, or bulk mail postage rates.

Mailing protocols (issuance)

Household Sample

In an effort to optimize the timing between the household screener and angler follow-up surveys, ABS sample was split into two equal groups. Initial surveys for the first group were sent eight weeks prior to the start of the angler effort. Fielding to the remainder of the ABS sample was completed in a compressed timeline of only six weeks.

	Group 1 Extended Fielding	Group 2 Compressed Fielding
Count	900	900
Date of initial mailing	November 10, 2009	November 20, 2009
Date of postcard mailing	November 16, 2009	November 30, 2009
Date of replacement form	November 30, 2009	December 14, 2009
Fielding Window	8 weeks	6 weeks
Number of Completes	360	351

Households selected for the ABS survey were sent packets containing a $1 incentive for participation. Approximately one week later, the same households received a postcard with a reminder to complete the survey. The status of returned questionnaires were checked into a process control system using various codes including completed interview, refusal to participate, and unable to be delivered by the Postal Service. Non respondents were sent replacement survey packets including an updated cover letter but no dollar bill.

Data from all returned surveys were entered to permit the creation of a list of identified anglers. The data file was compared to the check in system to ensure a complete file for sampling.

Angler Sample

The second stage of the project involving the sampling of anglers used the same mailing procedures for anglers identified in the ABS household survey and in the North Carolina licensed angler frame. Because multiple anglers were sampled in some households, materials were personalized to include the names of anglers. If the names of anglers were not provided in the ABS household study, name fields were hand edited to include specifications such as "male angler" or "eldest female angler".

Initial packets were sent with a $1 incentive for participation. All sampled anglers received a postcard reminder to complete the survey approximately one week later. If a form had not been returned within 3 weeks, a replacement packet using a modified cover letter was sent without the monetary incentive. Those who did not return a survey within seven weeks were sent a second replacement form with a final request for participation. This last appeal was sent using Federal Express 2-day delivery.

	Anglers from ABS sample frame	Anglers from NC License Frame
Count	262	1562
Date of initial mailing	January 4, 2009	
Date of postcard mailing	January 12, 2009	
Date of replacement form	January 25, 2009	
Date of final replacement	February 18, 2009	
Number of Completes	191	985

Process Control Procedures

The mailing of all survey items and the receipt of all survey forms (regardless of completion status) were logged into a process tracking system. When available, bar code readers were used to automatically enter control numbers and minimize errors in documentation. Status codes included specific actions (e.g. mailing of initial survey packet) as well as outcome codes consistent with guidelines set by the American Association for Public Opinion Research (AAPOR).

All returned BREs were opened and grouped into batches corresponding to the day's receipts. An initial check of surveys ensured reasonable completeness and blank forms were logged into the tracking system as "refused interview". Each survey was scanned for errors or inconsistencies. Directive clarifications for data entry staff were written directly on the survey, initialed and dated by the reviewer in a distinguishable colored pencil.

Data Entry

A data entry program was created using specialized research software and incorporated range and logic checks. These checks can be described as hard edits, soft edits, and consistency checks:

- Hard Edits represent a finite permissible range for the response and trigger an error message if an unallowable value is entered into the program.
- Soft Edits represent response values that may be valid, but are viewed as extreme. These values trigger an "unlikely" message when entered by the data entry person. Data entry personnel review these responses for verification prior to entering them as data.
- Internal Consistency Edits represent programmed checks to ensure responses are consistent throughout the survey. Since these contradictions may reflect data

recorded on the form by the respondent, consistency checks operate like soft edits, flagging the data entry personnel to possible errors but not preventing the recording of data.

Standard codes for illegible or missing values were incorporated for each question. Each survey was entered into the system twice. Inconsistencies between data records were rectified to ensure digitized files accurately reflected the information provided by the respondent on the paper survey. In the case that coding decisions were not immediately clear to the data entry staff person, project management would clarify guidance directly on the survey form along with their initials and the date.

Data Cleaning Procedures

ICF Macro employed limited data cleaning on data files:

ABS Household Survey

- If the number of anglers with recent activity was detailed in Q2, Q1 may be coded to indicate the presence of anglers.

Angler Survey

- Given an indication of recent participation (e.g. in Q7 or Q8), Q1 may be marked to indicate 2009 recreational saltwater fishing activity
- If dates of trips were marked in the Q8 calendars, Q7 could be marked to indicate recreational saltwater fishing in North Carolina during November and/or December
- If valid trips were detailed, the following assumptions could be made:
    - Q1: Respondent participated in recreational saltwater fishing
    - Q7: Indication of recreational saltwater fishing in North Carolina during November and/or December
    - Q8a & Q8b: dates of saltwater activity

Trip Detail

- It was required that non-missing dates of trips must occur during November or December. Trips from other months were considered invalid.
- Missing trip dates may be transcribed from Q8 of the angler survey provided the angler made four or fewer trips and the mode of trips (boat, shore) were sequenced as expected.
- Fishing on a boat (Q2) could be assumed if details of a boating trip were provided in Q2a and Q2b.
- Fishing from the shore (Q3) could be assumed if details of a shore trip were provided in Q3a, Q3b, and Q3c.

- Additional household anglers for the trip (Q6) could be assumed if the additional fishers were described in Q6a and Q6b.

Production of the Data File

Data files were constructed with one record per selected piece of sample. Questionnaire variables for non-respondents appear as missing values within the data file. Final files were checked for consistency with process control databases. Values exceeding logical and reasonable tolerances were compared to original forms to ensure the fidelity of information.

Final data files were built to include all data from the dual-frame mail survey with one record for each sampled unit. In addition to data from the survey instrument, the following were provided:

- A unique record ID assigned to anglers,
- A household identification numbers,
- Angler number,
- Sample source (ABS-frame or license-frame),
- Stratum,
- AAPOR-based outcome codes,
- Original/supplemental record classification, and
- Reverse-matched telephone number.

A complete data dictionary can be found in *Appendix H: Data Dictionary* on page 36.

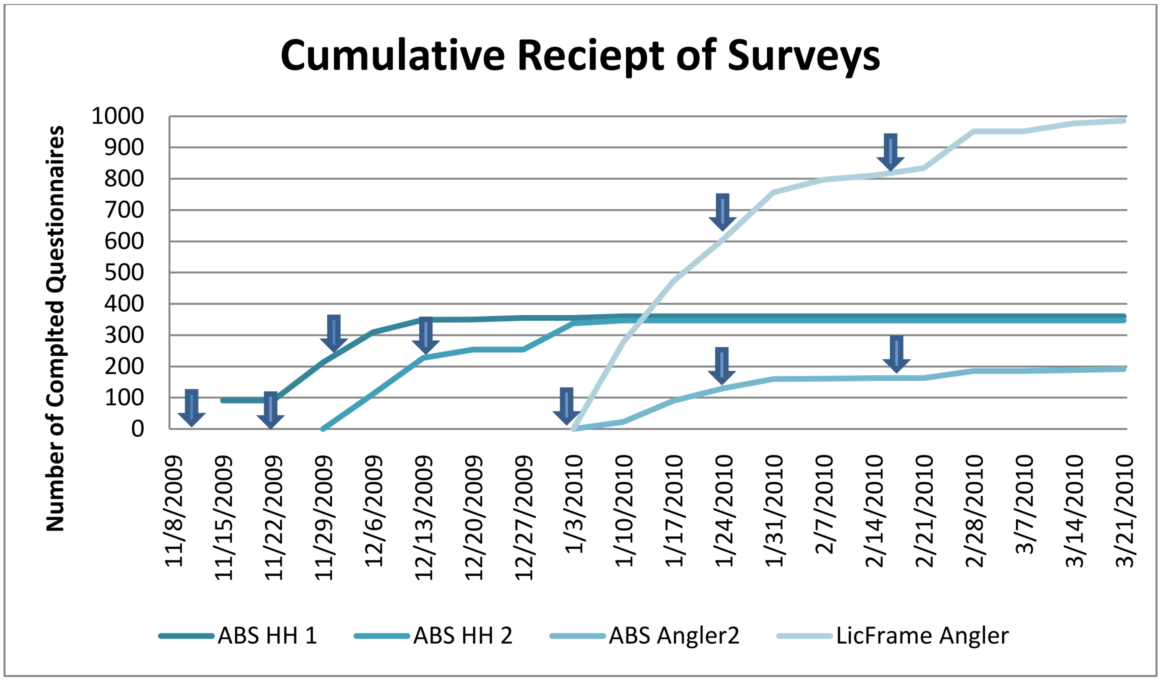
The data file will be delivered in SAS format with final content, coding, formatting, and naming conventions developed in conjunction with NMFS.

Survey Response

The survey protocol for ABS Household study resulted in a 42% response rate (measured in completes over presumably delivered surveys). The rates for the extended and compressed fielding periods were near identical. It appears that most respondents sent back forms within four weeks of the initial mailing.

A 74% response rate was achieved when contacting anglers identified in the household survey. The same survey administered to anglers identified in the license frame produced a response rate of 68%. While the majority of respondents returned forms within four to five weeks, a third mailing via Federal Express produced a swell of returns at the end of the fielding period. Approximately 10% - 15% of total returns resulted from the third mailing.

The graph below shows the cumulative receipt of surveys from each of the four efforts. Arrows mark the dates of questionnaire mailings.



Final Status of Records and Response Rates

The following tables account the final outcomes of the sample associated with each survey effort.

	ABS HH Screener	Group 1 Extended Fielding	Group 2 Compressed Fielding
Total Sent	1,800	900	900
Complete: HHs with anglers	229	113	116
Complete: HHs w/o anglers	456	235	221
Refusal	22	12	10
Undeliverable	157	77	80
Unknown outcome	936	463	473
$\frac{completed}{(total - undeliverable)}$	42%	42%	41%

	ABS Angler Study	License Frame Angler Study
Total Sent	262	1,562
Complete: Recent activity	51	227
Complete: no recent activity	137	739
Refusal	3	19
Undeliverable	8	139
Unknown outcomes	63	438
$\frac{completed}{(total - undeliverable)}$	74%	68%

Limitations of the Study

There are several inherent sources of error commonly recognized in mail-based research.

Language

According to the 2006-2008 American Community Survey 3-Year Estimates, 10% of North Carolina residents speak a language other than English at home (U.S. Census Bureau). Printed materials were in English only creating a barrier to those who cannot read the language.

Coverage

Although the ABS frame contains a comprehensive set of mailing addresses, coverage issues may result through sources such as illegal housing units or households that only receive mail through a post office box.

Because the fishing activity of households in the ABS sample frame is collected using a two stage design, the completeness of the angler data file is dependent on responses to the household screening study. Non-respondents and those who go fishing for the first time in a year after completing the household screener reduce the coverage of the angler study.

Coverage issues associated with the Licensed Angler frame come from several key sources. Minors under the age of 18 are excluded from sample through license exemptions and filtering of the sample frame. Members of the Armed Forces on temporary military leave are not required to obtain a license and therefore will not appear in the registry. Illegal activity performed by those without a fishing license cannot be captured using this sample frame. Issues with the same frame, such as incorrectly entered mailing information, may be associated with specific licensing sites and could precipitate exclusion from the sampling frame. Anglers who have recently moved may be less likely to be included in final data files.

Non-Response

As with other research studies that attempt to provide close measures of representative samples, refusal rates are of concern for this study. It is commonly cited that response rates for surveys have been dropping significantly in recent years. While weighting of data will minimize many distortions, it is commonly accepted that there will be distinct differences between the attitudes and opinions of those who complete the study verses those who refuse to do so. Therefore, any response rate less than 100% indicates some level of inaccuracy in the final data. In the same vein of reasoning, the refusal of any specific question during a survey compromises the precision of its measure.

Limited protocols

The ABS Household study received two questionnaire mailings while the Angler studies received three questionnaire mailings. The final distributions for each stage resulted in significant levels of response suggesting additional completes could be obtained through additional outreach. However, this is not to say that the cost of efforts would create a proportionate benefit.

For most respondents, fishing activity will be fully documented using the current form detailing the most recent four trips. However, earlier trips of more avid anglers may not be captured. Errors could result if undocumented trips were distinct or imputed values do not match actual activity.

Response bias

Respondents can also control the accuracy of the data depending on the level of consideration and seriousness to which they approach answering the questions. Although the questionnaire forms were designed to aid cognitive processing (e.g. through the display of a calendar to mark dates of fishing activity), ultimately the respondent controls how accurate their responses are in representing their recent activities. While the added delay between activity and reporting may cause greater immediate recall issues when compared to the telephone survey, the format of a paper self administered survey should ultimately make it easier for a respondent to verify event details (e.g. by reviewing schedules, though discussions with other members of a trip, etc.).

Other sources of error involve the design of the questions themselves. Although questions originated from the long-standing CHTS, wording needed to be adjusted to accommodate a paper-based methodology. Questions and response categories should be relatively easy for most individuals to comprehend, however some respondents could have difficulty accurately responding to some questions. Unlike the CHTS, this is a self-administered questionnaire which prohibits clarification of items.

Considerations for Future Data Collection Efforts

The following may be considered for future iterations of the project:

- Continued testing of household screener fielding schedule. The number of anglers from the ABS sample qualified to receive an angler survey may quickly change, especially during springtime months. The impact of a compressed fielding period should continue to be investigated.
- Cognitive interviewing to improve the questionnaire. Topic areas might include:
    - Methods for insuring better matches between dates on the angler survey and trip detail, possibly by listing months on the trip detail as close-ended responses.
    - Clarification for reporting in-state and out-of-state trips. Currently, Q7 in the angler survey specifies trips taken in North Carolina. Respondents may inconsistently provide information about out-of-state trips in following questions.
    - Improved ways to indicate county of trip. This may include displaying a county-level map of the state where the respondent may fill in the location of the trip.
- A non-response telephone follow-up. It is common to see over 50% of mailing addresses matched to a telephone number. A large percentage of records drawn from the licensed angler registry include a telephone number. In order to maximize response, respondents could receive a reminder call requesting that they complete the survey, allowing the respondent to complete by telephone. This option could be implemented economically given the fact that the CHTS and ALDS provide the basis for the CATI system.

Works Cited

Blumberg, S. J., & Luke, J. V. (2009, June 17). *NCHS - NHIS - Wireless Substitution, July-December 2008 (Released 5/2009)*. Retrieved September 7, 2009, from Centers for Disease Control and Prevention, National Center for Health Statistics: http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless200905.htm

Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Pratap, S., Wentz, R., et al. (2002). Increasing Response Rates to Postal Questionnaires: Systematic Review. *British Medical Journal* , 1183-1185.

Fox, R. J., Crask, M. R., & Kim, J. (1988). Mail Survey Response Rate: A Meta-Analysis of Selected Techniques for Inducing Response. *Public Opinion Quarterly* , 467-491.

Harrison, R. A., Holt, D., & Elton, P. J. (2002). Do Postage-Stamps Increase Response Rates to Postal Surveys? A Randomized Controlled Trial. *International Journal of Epidemiology* , 872-874.

Iannacchione, V., Staab, J., & Redden, D. (2003). Evaluating the Use of Residential Mailing Addresses in a Metropolitan Household Survey. *Public Opinion Quarterly, Vol 67* , 202-210.

U.S. Census Bureau. (n.d.). *North Carolina - Fact Sheet - American FactFinder*. Retrieved April 30, 2010, from http://factfinder.census.gov/servlet/ACSSAFFFacts?_event=Search&geo_id=&_geoContext=&_street=&_county=&_cityTown=&_state=04000US37&_zip=&_lang=en&_sse=on&pctxt=fph&pgsl=010

Yammarinio, F. J., Skinner, S. J., & Childers, T. L. (1991). Understanding Mail Survey Response Behavior. *Public Opinion Quarterly* , 613-639.

Appendix A: Key Dates (timeline)

Event	Date
ABS HH Group 1: Initial Survey Packet	November 10, 2009
ABS HH Group 1: Postcard	November 16, 2009
ABS HH Group 1: Replacement Packet	November 30, 2009
ABS HH Group 2: Initial Survey Packet	November 20 , 2009
ABS HH Group 2: Postcard	November 30, 2009
ABS HH Group 2: Replacement Packet	December 14, 2009
Angler Survey: Initial Survey Packet	January 4, 2010
Angler Survey: Postcard	January 12, 2010
Angler Survey: Replacement Packet	January 25, 2010
Angler Survey: FedEx Replacement	February 15, 2010
End of Collection	March 26, 2010

Appendix B: Disposition Report

ABS Household

Outcome	Extended Fielding	Compressed Fielding	Combined
1.1 Complete (net)	348	341	689
1.1.1: Complete with Anglers	113	117	230
1.1.2: Complete without Anglers	235	224	459
2.1 Refusals	12	10	22
3.3 Undeliverable addresses	78	82	160
TOTAL COUNT	438	433	871

Angler Survey

Outcome	ABS Sample	Licensed Based Frame	Combined
1.2 Complete (net)	188	966	1,154
1.2.1: Complete with Anglers	51	227	278
1.2.2: Complete without Anglers	137	739	876
2.1 Refusals	3	19	22
3.3 Undeliverable addresses	8	139	147
TOTAL COUNT	199	1,124	1,323

Appendix C: Material for Household Questionnaire Packets

Household questionnaire packets were comprised of:

- A customized cover letter from NOAA,
- A booklet style questionnaire, and
- A business reply envelope (BRE).

Initial mailings also included a dollar bill.

November 30, 2009

North Carolina Resident
«street»
«city», «state» «zip5»

Dear North Carolina Resident,

I am writing to ask you for your help in a study being conducted for the National Oceanic and Atmospheric Administration. This study is part of an effort to learn more about recreational fishing activities in North Carolina.

The purpose of this questionnaire is to identify people who fish. However, it is important for us to obtain responses from people who do fish as well as those who do not participate in recreational saltwater fishing. Your address was randomly selected from a list of all home addresses in North Carolina. Your household represents thousands of other households like yours. Only with participation by everyone selected will the findings from the study represent everyone in North Carolina. If there are people who fish in your household, we may send them a second questionnaire to learn about their recreational saltwater fishing experiences. We have enclosed a small token of appreciation as a way of saying thanks for your help.

Your answers are completely confidential and will be used for statistical purposes only in accordance with the Privacy Act of 1974. You are not required to answer any question that you feel is an intrusion of your privacy.

If you have any questions or comments about this study, we would be happy to talk with you. Please contact Rob Andrews at his number (301-713-2328) or you can write to us at the address at the bottom of this letter.

Thank you very much for your help with this important study. Please return your completed questionnaire in the postage paid envelope provided.

Sincerely,

Dave Van Voorhees,
Chief, Fisheries Statistics Division

«ID»

1315 East-West Hwy, Silver Spring, Maryland 20910
Phone: 301-713-2328 Internet: www.st.nmfs.gov

November 30, 2009

North Carolina Resident
«street»
«city», «state» «zip5»

Dear North Carolina Resident,

About three weeks ago we sent a questionnaire to your household in an effort to learn more about recreational fishing activities in North Carolina. According to our records, your household has not yet returned a completed questionnaire.

Your address was randomly selected from a list of all home addresses in North Carolina. Only with participation by everyone selected will the findings from the study represent everyone in North Carolina. Many North Carolinians have told us about their fishing behaviors. Your completed questionnaire will contribute to our understanding of the state's fishing activity.

This study being conducted for the National Oceanic and Atmospheric Administration. It is important for us to obtain information from people who do fish as well as those who do not participate in recreational saltwater fishing. If there are people who fish in your household, we may send them a second questionnaire to learn more about their recreational saltwater fishing experiences.

Your answers are completely confidential and will be used for statistical purposes only in accordance with the Privacy Act of 1974. You are not required to answer any question that you feel is an intrusion of your privacy.

If you have any questions or comments about this study, we would be happy to talk with you. Please contact Rob Andrews at his number (301-713-2328) or you can write to us at the address at the bottom of this letter.

We hope that you will take a few minutes to fill out and return the questionnaire soon. If, for any reason, you prefer not to answer it, please let us know by returning a note or blank questionnaire in the postage paid envelope provided.

Sincerely,

Dave Van Voorhees,
Chief, Fisheries Statistics Division

«ID»

1315 East-West Hwy, Silver Spring, Maryland 20910
Phone: 301-713-2328 Internet: www.st.nmfs.gov

Commonly Asked Questions

Q How did you get my address?

A *Your address was randomly selected from among all of North Carolina's addresses. It was selected using scientific sampling methods to represent other households in your part of the state.*

Q Nobody in my household fishes. Should I respond to the survey?

A *Yes. It is important that we gather information about households that do not fish as well as those who do. Once we receive your completed questionnaire, you will not be sent any additional mailings such as replacement questionnaires.*

Q Why can't you interview another household instead of mine?

A *In order to make sure final results of the study are accurate, receiving information about your household is important to us. Households selected for this study were chosen using scientific sampling methods and your responses cannot be replaced by others.*

Q How will the information I provide be used?

A *Information from this study will be used to improve the monitoring of North Carolina's fishing activity. All information will be kept confidential; It will be combined with information from other households to produce statistical summaries and reports.*

Q How much time will this survey take?

A *On average, it should take less than five minutes for you to respond, including the time for reviewing instructions, and completing and reviewing the collection of information.*

Q Who is sponsoring the survey?

A *This study is sponsored by the National Oceanic and Atmospheric Administration. Questions regarding the study can be directed to Rob Andrews at National Oceanic and Atmospheric Administration by calling 301-713-2328.*

OMB No: 0648-0052 Approval Expires 04/30/2011

A Study of Fishing In North Carolina

This study is being conducted with the assistance of the National Oceanic and Atmospheric Administration in accordance with the Magnuson-Stevens Fishery Conservation and Management Act of 2006. Your participation is voluntary. All responses will be kept confidential under the Privacy Act of 1974.

The information you provide will be combined with information provided by other participants to produce statistical summaries and reports.

A Study of Fishing in North Carolina

This form is for households with and without people who fish. It just takes a few minutes and your answers will be confidential.

1. In the last 12 months has anyone living in your household, including children and adults, been recreational saltwater fishing in the U.S. or a U.S. territory?

☐ Yes
☐ No ➞ *(please go to question 6)*

Recreational saltwater fishing refers to fishing primarily with hook and line for pleasure, amusement, relaxation, or home consumption in oceans, bays, inlets, canals, intra-coastal waterways, and brackish portions affected by the tides. Inland saltwater bodies includes sounds, passes, inlets, bays, estuaries, brackish portions of rivers, and other areas of salt or brackish water like bayous and canals.

2. How many people in this household (including children and adults) have participated in recreational saltwater fishing during the past 12 months?

☐☐ *Number of people who have fished*

Please tell us the following for each member of the household who has gone saltwater fishing during the past 12 months. If more than four household members have participated in recreational saltwater fishing during the past 12 months, please list the four oldest participants.

	Fisher 1	Fisher 2	Fisher 3	Fisher 4
3. What is the person's first name?				
4. What is the person's sex?	☐ Male	☐ Male	☐ Male	☐ Male
	☐ Female	☐ Female	☐ Female	☐ Female
5. How old is the person?	☐ Less than 16	☐ Less than 16	☐ Less than 16	☐ Less than 16
	☐ 16-18	☐ 16-18	☐ 16-18	☐ 16-18
	☐ 18-24	☐ 18-24	☐ 18-24	☐ 18-24
	☐ 25-34	☐ 25-34	☐ 25-34	☐ 25-34
	☐ 35-44	☐ 35-44	☐ 35-44	☐ 35-44
	☐ 45-54	☐ 45-54	☐ 45-54	☐ 45-54
	☐ 55-64	☐ 55-64	☐ 55-64	☐ 55-64
	☐ 65 or older	☐ 65 or older	☐ 65 or older	☐ 65 or older

6. For statistical purposes, we are interested in learning about your telephone usage. Is there at least one phone inside your home that is currently working and not a cell phone?

☐ Yes
☐ No

Thank you! Please return this form in the postage paid envelope provided or mail to:
A Study of Fishing in North Carolina
116 John Street, Suite 800
New York, NY 10038

Appendix D: Household Survey Reminder Postcard

Date

Last week a questionnaire was sent to you on behalf of the National Oceanic and Atmospheric Administration. Your household was randomly selected from a list of all households addresses in the state of North Carolina.

If you have already completed and returned the questionnaire to us, please accept our sincere thanks. If not, please do so today. Information collected in this study will be used to learn more about recreational fishing activities in North Carolina.

If you did not receive a questionnaire, or if it was misplaced, please call Rob Andrews at his number (301-713-2328).

Thank you.

Dave Van Voorhees
Chief, Fisheries Statistics Division

NORTH CAROLINA RESIDENT
Street
Town, City, #####

Appendix E: Material for Angler Questionnaire Packets

Questionnaire packets for anglers were comprised of:

- A customized cover letter from NOAA,
- A booklet style questionnaire for detailing angler activity,
- A booklet-style questionnaire for detailing up to 4 recent trips,and
- A business reply envelope (BRE).

Initial mailings also included a dollar bill.

January 6th, 2009

«NAME»
«ADDRESS»
«CITY», «ST» «ZIP»

Dear «Name__Proper»

I am writing to ask you for your help in a study being conducted for the National Oceanic and Atmospheric Administration. This study is part of an effort to learn more about recreational fishing activities in North Carolina.

Earlier this year, your address was randomly selected from among all of the home addresses in the state of North Carolina. From among those households that responded to our earlier questionnaire, we selected a random sample of recreational saltwater anglers to learn about the frequency of their fishing trips and some information about those trips, such as where and when the fishing trip occurred. Information collected in this study will be used to evaluate the impact of recreational fishing on natural fishing resources and help improve fisheries management policies. We have enclosed a small token of appreciation as a way of saying thanks for your help.

Your answers are completely confidential and will be used for statistical purposes only in accordance with the Privacy Act of 1974. You are not required to answer any question that you feel is an intrusion of your privacy.

If you have any questions or comments about this study, we would be happy to talk with you. Please contact Rob Andrews at his number (301-713-2328) or you can write to us at the address at the bottom of this letter.

Thank you very much for your help with this important study. Please return your completed questionnaire in the postage paid envelope provided.

Sincerely,

Dave Van Voorhees,
Chief, Fisheries Statistics Division

<ID>

1315 East-West Hwy, Silver Spring, Maryland 20910
Phone: 301-713-2328 Internet: www.st.nmfs.gov

January 5th, 2009

«NAME»
«ADDRESS»
«CITY», «ST» «ZIP»

Dear «Name__Proper»

I am writing to ask you for your help in a study being conducted for the National Oceanic and Atmospheric Administration. This study is part of an effort to learn more about recreational saltwater fishing activities in North Carolina.

Your name was randomly selected using scientific sampling methods from among a list of persons who has purchased a saltwater fishing license in the state of North Carolina during the past year. We are contacting a random sample of recreational saltwater anglers to learn about the frequency of their fishing trips and some information about those trips, such as where and when the fishing trip occurred. Information collected in this study will be used to evaluate the impact of recreational fishing on natural resources and help improve fisheries management policies. We have enclosed a small token of appreciation as a way of saying thanks for your help.

Your answers are completely confidential and will be used for statistical purposes only in accordance with the Privacy Act of 1974. You are not required to answer any question that you feel is an intrusion of your privacy.

If you have any questions or comments about this study, we would be happy to talk with you. Please contact Rob Andrews at his number (301-713-2328) or you can write to us at the address at the bottom of this letter.

Thank you very much for your help with this important study. Please return your completed questionnaire in the postage paid envelope provided.

Sincerely,

Dave Van Voorhees,
Chief, Fisheries Statistics Division

«ID»

1315 East-West Hwy, Silver Spring, Maryland 20910
Phone: 301-713-2328 Internet: www.st.nmfs.gov

Commonly Asked Questions

Q How did you get my address?

A *Your address was either randomly selected from a database of licensed anglers in North Carolina or you completed a similar mail survey in the past two months. Addresses were selected using scientific sampling methods to represent other households in your part of the state.*

Q Nobody in my household fished in the past few months. Should I respond to the survey?

A *Yes. It is important that we gather information about households that do not fish as well as those who do. Once we receive your completed questionnaire, you will not be sent any additional mailings such as replacement questionnaires.*

Q Why can't you interview another household instead of mine?

A *In order to make sure final results of the study are accurate, receiving information about your household is important to us. Households selected for this study were chosen using scientific sampling methods and your responses cannot be replaced by others.*

Q How will the information I provide be used?

A *Information from this study will be used to improve the monitoring of North Carolina's fishing activity. All information will be kept confidential; It will be combined with information from other households to produce statistical summaries and reports.*

Q How much time will this survey take?

A *On average, it should take less than five minutes for you to respond, including the time for reviewing instructions, and completing and reviewing the collection of information.*

Q Who is sponsoring the survey?

A *This study is sponsored by the National Oceanic and Atmospheric Administration. Questions regarding the study can be directed to Rob Andrews at National Oceanic and Atmospheric Administration by calling 301-713-2328.*

A Study of Fishing In North Carolina

Angler Survey

This study is being conducted with the assistance of the National Oceanic and Atmospheric Administration in accordance with the Magnuson-Stevens Fishery Conservation and Management Act of 2006. Your participation is voluntary. All responses will be kept confidential under the Privacy Act of 1974.

The information you provide will be combined with information provided by other participants to produce statistical summaries and reports.

OMB No: 0648-0052 Approval Expires 04/30/2011

A Study of Fishing in North Carolina: Angler Survey

This form will take just a few minutes to complete and your answers will be confidential.

Thank you for completing this survey. Here are some terms we will be using:

- Saltwater fishing includes fishing in oceans, sounds, or bays, or in brackish portions of rivers. This does not include fishing in freshwater, or for shellfish, such as crabbing
- Recreational fishing means the primary purpose of the fishing is for fun or relaxation, as opposed to providing income from the sale of fish
- Private boat refers to any boat owned by an individual where no fee is paid to the individual for the use of the boat

1. **During 2009, did you participate in *recreational saltwater fishing*?**
   - ☐ Yes
   - ☐ No

2. **Are you…?**
   - ☐ Male
   - ☐ Female

3. **How old are you?**
   - ☐ Less than 16
   - ☐ 16 – 17
   - ☐ 18 – 24
   - ☐ 25 – 34
   - ☐ 35 – 44
   - ☐ 45 – 54
   - ☐ 55 – 64
   - ☐ 65 or older

4. **During the <u>past 12 months</u>, did you have a fishing license for the state of North Carolina?**
   - ☐ Yes
   - ☐ No → *(PLEASE GO TO QUESTION 7)*

5. **Was this particular license for *recreational saltwater fishing*?**
   - ☐ Yes
   - ☐ No → *(PLEASE GO TO QUESTION 7)*

6. **Was this license valid for the month of November this year?**
   - ☐ Yes
   - ☐ No

7. **Between November 1st, 2009 and December 31st, 2009, did you go *recreational saltwater fishing* in North Carolina or from a *private boat* launched in North Carolina?**
   - ☐ Yes
   - ☐ No

Even if you did not go saltwater fishing during November or December, it's important for us to get a response from every person selected for this study. Please return your now completed questionnaire to us in the enclosed prepaid return envelope.

8a. **Please circle the day(s) when you went *recreational saltwater fishing* using a *private boat* on the calendar below.**

MONTH	S	M	T	W	R	F	S
NOVEMBER	01	02	03	04	05	06	07
	08	09	10	11	12	13	14
	15	16	17	18	19	20	21
	22	23	24	25	26	27	28
	29	30					
DECEMBER			01	02	03	04	05
	06	07	08	09	10	11	12
	13	14	15	16	17	18	19
	20	21	22	23	24	25	26
	27	28	29	30	31		

8b. **Please circle the day(s) when you went *recreational saltwater fishing* from the shore, pier, dock, or any other land based fishing on the calendar below.**

MONTH	S	M	T	W	R	F	S
NOVEMBER	01	02	03	04	05	06	07
	08	09	10	11	12	13	14
	15	16	17	18	19	20	21
	22	23	24	25	26	27	28
	29	30					
DECEMBER			01	02	03	04	05
	06	07	08	09	10	11	12
	13	14	15	16	17	18	19
	20	21	22	23	24	25	26
	27	28	29	30	31		

Please use the enclosed tan colored booklet to provide information about fishing trips you took during November and December. The booklet contains room for reporting your most recent four trips.

4th Most Recent Trip

The following questions ask information about fishing trips taken during the previous two months. If you fished between November 1, 2009 and December 31, 2009 please fill out information for your four most recent saltwater recreational fishing trips. Once completed, please return your questionnaire to us in the enclosed prepaid return envelope.

1. Please enter the date of your most recent trip:

Month Day

2. Did you fish from a boat on this date?
- ☐ Yes
- ☐ No → (PLEASE GO TO QUESTION 3)

2a. What state and county were you in when the boat landed and you stepped off of the boat?

State

County, Parish, Island or Municipality Name

2b. Does the public have access to the place where the boat landed?
- ☐ Yes, Public Access
- ☐ No, Private Access

3. Did you fish from the shore on this date?
- ☐ Yes
- ☐ No → (PLEASE GO TO QUESTION 4)

3a. Where were you fishing?

State

County, Parish, Island or Municipality Name

3b. Was that from a...?
(CHECK ALL THAT APPLY)
- ☐ Pier
- ☐ Dock
- ☐ Jetty or Breakwater
- ☐ Bridge or Causeway
- ☐ Other man-made structure
- ☐ Bank or Beach

3c. Does the public have access to this place?
- ☐ Yes, Public Access
- ☐ No, Private Access

4. Was most of your fishing effort this day in ...?
(CHECK ONLY ONE)
- ☐ Ocean, within 3 miles from the shore
- ☐ Ocean, more than 3 miles from the shore
- ☐ Sound
- ☐ River
- ☐ Bay
- ☐ Inlet
- ☐ Someplace else

5. At approximately what time did this fishing trip end?
(CHECK ONLY ONE)
- ☐ Midnight – 3:00 a.m.
- ☐ 3:00 a.m. – 6:00 a.m.
- ☐ 6:00 a.m. – 9:00 a.m.
- ☐ 9:00 a.m. – Noon
- ☐ Noon – 3:00 p.m.
- ☐ 3:00 p.m. – 6:00 p.m.
- ☐ 6:00 p.m. – 9:00 p.m.
- ☐ 9:00 p.m. – Midnight

6. Did anyone else from your household, including children and adults, fish with you on this date?
- ☐ Yes
- ☐ No

6a. Who else from your household fished with you on this date?
(CHECK ALL THAT APPLY)
- ☐ Spouse/partner
- ☐ Child(ren)
- ☐ Other household member

6b. Including you, how many household members went saltwater fishing on this date?

Number of household members

1. Please enter the date of your most recent trip:

Month Day

2. Did you fish from a boat on this date?
- ☐ Yes
- ☐ No → (PLEASE GO TO QUESTION 3)

2a. What state and county were you in when the boat landed and you stepped off of the boat?

State

County, Parish, Island or Municipality Name

2b. Does the public have access to the place where the boat landed?
- ☐ Yes, Public Access
- ☐ No, Private Access

3. Did you fish from the shore on this date?
- ☐ Yes
- ☐ No → (PLEASE GO TO QUESTION 4)

3a. Where were you fishing?

State

County, Parish, Island or Municipality Name

3b. Was that from a...?
(CHECK ALL THAT APPLY)
- ☐ Pier
- ☐ Dock
- ☐ Jetty or Breakwater
- ☐ Bridge or Causeway
- ☐ Other man-made structure
- ☐ Bank or Beach

3c. Does the public have access to this place?
- ☐ Yes, Public Access
- ☐ No, Private Access

4. Was most of your fishing effort this day in ...?
(CHECK ONLY ONE)
- ☐ Ocean, within 3 miles from the shore
- ☐ Ocean, more than 3 miles from the shore
- ☐ Sound
- ☐ River
- ☐ Bay
- ☐ Inlet
- ☐ Someplace else

5. At approximately what time did this fishing trip end?
(CHECK ONLY ONE)
- ☐ Midnight – 3:00 a.m.
- ☐ 3:00 a.m. – 6:00 a.m.
- ☐ 6:00 a.m. – 9:00 a.m.
- ☐ 9:00 a.m. – Noon
- ☐ Noon – 3:00 p.m.
- ☐ 3:00 p.m. – 6:00 p.m.
- ☐ 6:00 p.m. – 9:00 p.m.
- ☐ 9:00 p.m. – Midnight

6. Did anyone else from your household, including children and adults, fish with you on this date?
- ☐ Yes
- ☐ No → (PLEASE CONTINUE TO YOUR NEXT MOST RECENT TRIP)

6a. Who else from your household fished with you on this date?
(CHECK ALL THAT APPLY)
- ☐ Spouse/partner
- ☐ Child(ren)
- ☐ Other household member

6b. Including you, how many household members went saltwater fishing on this date?

Number of household members → (PLEASE CONTINUE TO YOUR NEXT MOST RECENT TRIP)

2nd Most Recent Trip

1. Please enter the date of your most recent trip:

Month Day

2. Did you fish from a boat on this date?
☐ Yes
☐ No → (PLEASE GO TO QUESTION 3)

2a. What state and county were you in when the boat landed and you stepped off of the boat?

State

County, Parish, Island or Municipality Name

2b. Does the public have access to the place where the boat landed?
☐ Yes, Public Access
☐ No, Private Access

3. Did you fish from the shore on this date?
☐ Yes
☐ No → (PLEASE GO TO QUESTION 4)

3a. Where were you fishing?

State

County, Parish, Island or Municipality Name

3b. Was that from a...?
(CHECK ALL THAT APPLY)
☐ Pier
☐ Dock
☐ Jetty or Breakwater
☐ Bridge or Causeway
☐ Other man-made structure
☐ Bank or Beach

3c. Does the public have access to this place?
☐ Yes, Public Access
☐ No, Private Access

4. Was most of your fishing effort this day in ...?
(CHECK ONLY ONE)
☐ Ocean, within 3 miles from the shore
☐ Ocean, more than 3 miles from the shore
☐ Sound
☐ River
☐ Bay
☐ Inlet
☐ Someplace else

5. At approximately what time did this fishing trip end?
(CHECK ONLY ONE)
☐ Midnight – 3:00 a.m.
☐ 3:00 a.m. – 6:00 a.m.
☐ 6:00 a.m. – 9:00 a.m.
☐ 9:00 a.m. – Noon
☐ Noon – 3:00 p.m.
☐ 3:00 p.m. – 6:00 p.m.
☐ 6:00 p.m. – 9:00 p.m.
☐ 9:00 p.m. – Midnight

6. Did anyone else from your household, including children and adults, fish with you on this date?
☐ Yes
☐ No → (PLEASE CONTINUE TO YOUR NEXT MOST RECENT TRIP)

6a. Who else from your household fished with you on this date?
(CHECK ALL THAT APPLY)
☐ Spouse/partner
☐ Child(ren)
☐ Other household member

6b. Including you, how many household members went saltwater fishing on this date?

Number of household members

→ (PLEASE CONTINUE TO YOUR NEXT MOST RECENT TRIP)

3rd Most Recent Trip

1. Please enter the date of your most recent trip:

Month Day

2. Did you fish from a boat on this date?
☐ Yes
☐ No → (PLEASE GO TO QUESTION 3)

2a. What state and county were you in when the boat landed and you stepped off of the boat?

State

County, Parish, Island or Municipality Name

2b. Does the public have access to the place where the boat landed?
☐ Yes, Public Access
☐ No, Private Access

3. Did you fish from the shore on this date?
☐ Yes
☐ No → (PLEASE GO TO QUESTION 4)

3a. Where were you fishing?

State

County, Parish, Island or Municipality Name

3b. Was that from a...?
(CHECK ALL THAT APPLY)
☐ Pier
☐ Dock
☐ Jetty or Breakwater
☐ Bridge or Causeway
☐ Other man-made structure
☐ Bank or Beach

3c. Does the public have access to this place?
☐ Yes, Public Access
☐ No, Private Access

4. Was most of your fishing effort this day in ...?
(CHECK ONLY ONE)
☐ Ocean, within 3 miles from the shore
☐ Ocean, more than 3 miles from the shore
☐ Sound
☐ River
☐ Bay
☐ Inlet
☐ Someplace else

5. At approximately what time did this fishing trip end?
(CHECK ONLY ONE)
☐ Midnight – 3:00 a.m.
☐ 3:00 a.m. – 6:00 a.m.
☐ 6:00 a.m. – 9:00 a.m.
☐ 9:00 a.m. – Noon
☐ Noon – 3:00 p.m.
☐ 3:00 p.m. – 6:00 p.m.
☐ 6:00 p.m. – 9:00 p.m.
☐ 9:00 p.m. – Midnight

6. Did anyone else from your household, including children and adults, fish with you on this date?
☐ Yes
☐ No → (PLEASE CONTINUE TO YOUR NEXT MOST RECENT TRIP)

6a. Who else from your household fished with you on this date?
(CHECK ALL THAT APPLY)
☐ Spouse/partner
☐ Child(ren)
☐ Other household member

6b. Including you, how many household members went saltwater fishing on this date?

Number of household members

→ (PLEASE CONTINUE TO YOUR NEXT MOST RECENT TRIP)

Appendix F: Angler Survey Reminder Postcard

Date

Last week a questionnaire was sent to you on behalf of the National Oceanic and Atmospheric Administration. Your household was randomly selected from a list of all households addresses in the state of North Carolina.

If you have already completed and returned the questionnaire to us, please accept our sincere thanks. If not, please do so today. Information collected in this study will be used to learn more about recreational fishing activities in North Carolina.

If you did not receive a questionnaire, or if it was misplaced, please call Rob Andrews at his number (301-713-2328).

Thank you.

Dave Van Voorhees
Chief, Fisheries Statistics Division

NORTH CAROLINA RESIDENT
Street
Town, City, #####

Appendix G: Coding of Text Questions

All responses to questions in the survey were pre-coded with the exception of location of fishing trip. Responses were coded to the county level using Federal Information Processing Standards (FIPS) codes. FIPS codes for North Carolina are provided below.

37001 Alamance County	37051 Cumberland County	37101 Johnston County	37151 Randolph County
37003 Alexander County	37053 Currituck County	37103 Jones County	37153 Richmond County
37005 Alleghany County	37055 Dare County	37105 Lee County	37155 Robeson County
37007 Anson County	37057 Davidson County	37107 Lenoir County	37157 Rockingham County
37009 Ashe County	37059 Davie County	37109 Lincoln County	37159 Rowan County
37011 Avery County	37061 Duplin County	37111 McDowell County	37161 Rutherford County
37013 Beaufort County	37063 Durham County	37113 Macon County	37163 Sampson County
37015 Bertie County	37065 Edgecombe County	37115 Madison County	37165 Scotland County
37017 Bladen County	37067 Forsyth County	37117 Martin County	37167 Stanly County
37019 Brunswick County	37069 Franklin County	37119 Mecklenburg County	37169 Stokes County
37021 Buncombe County	37071 Gaston County	37121 Mitchell County	37171 Surry County
37023 Burke County	37073 Gates County	37123 Montgomery County	37173 Swain County
37025 Cabarrus County	37075 Graham County	37125 Moore County	37175 Transylvania County
37027 Caldwell County	37077 Granville County	37127 Nash County	37177 Tyrrell County
37029 Camden County	37079 Greene County	37129 New Hanover County	37179 Union County
37031 Carteret County	37081 Guilford County	37131 Northampton County	37181 Vance County
37033 Caswell County	37083 Halifax County	37133 Onslow County	37183 Wake County
37035 Catawba County	37085 Harnett County	37135 Orange County	37185 Warren County
37037 Chatham County	37087 Haywood County	37137 Pamlico County	37187 Washington County
37039 Cherokee County	37089 Henderson County	37139 Pasquotank County	37189 Watauga County
37041 Chowan County	37091 Hertford County	37141 Pender County	37191 Wayne County
37043 Clay County	37093 Hoke County	37143 Perquimans County	37193 Wilkes County
37045 Cleveland County	37095 Hyde County	37145 Person County	37195 Wilson County
37047 Columbus County	37097 Iredell County	37147 Pitt County	37197 Yadkin County
37049 Craven County	37099 Jackson County	37149 Polk County	37199 Yancey County

Appendix H: Data Dictionary

ABS Household Screener

There is one record for every sampled address, regardless of the final outcome associated with the record.

Question	Field Name	Description	Coding Scheme
	HH_ID	Unique household identifier	
	MATCH_FLG	Was the household address successfully matched to the license frame? Is the household on both sample frames?	Yes=1, No=0
	STRATUM	Coastal, non-coastal, out-of-state	Coastal=1, Non-Coastal=2, Out-of-state=3
	RES_ST	State of residence	37 = North Carolina
	RES_CNTY	County of residence	
	RES_ADDRESS	Address of residence	
	HH_STATUS	Disposition of sample (complete with anglers, complete no anglers, refuse, non-contact, bad address)	1.1.1 = Household with angler 1.1.2 = Household with no angler 2.1 = Refused 3.3 = Mailing returned undelivered
Q1	FISH12_FLG	Fishing household flag. Did anyone in the household fish during previous 12 months?	Yes=1, No=0, 8 = Missing
Q2	FF12	How many people in HH fished during previous 12 months?	
Q6	HH_PHN_FLG	Does HH have a landline telephone?	Yes=1, No=0, 8 = Missing
	REC_DATE	Date questionnaire was received by contractor	
	MAIL_DATE	Date questionnaire was mailed by contractor (initial mailing)	
	SURV_YEAR	Survey year	
	SURV_WAVE	Survey wave	
	SAMP_WT	Sample weight (N/n)	
	FRM_SIZE	Number of HH units on sample frame for stratum (N)	
	BATCH	Wave 1 or Wave 2	

ABS Angler

There is one record for every angler identified in the household screener, regardless of whether or not the angler was sampled or returned a questionnaire.

Question	Field Name	Description	Coding Scheme
	HH_ID	Unique household identification number	
	ANG_ID	Unique identification for anglers within a household	
	HH_ANGLERS	Number of anglers uniquely identified in screener questionnaire (screener Q3).	
Q4	GENDER		Male=1, Female=2
Q5	AGE		Less than 16 = 1 16 -- 17 = 2 18 -- 24 = 3 25 -- 34 = 4 35 -- 44 = 5 45 -- 54 = 6 55 -- 64 = 7 65 or older = 8
	SAMP_FLG	Identifies anglers that were sampled from angler frame.	Primary angler=1, Supplemental angler=2, Not sampled=3
	ANG_STATUS	Final disposition of second-stage sample (complete with trips, complete no trips, refusal, non-contact, etc.)	1.2.1 = Trips taken in the 2 month period 1.2.2 = No trips taken in the 2 month period 2.1 = Refused 3.19 = Nothing ever returned 3.3 = Mailing returned undelivered
	REC_DATE	Date questionnaire was received by contractor	
	MAIL_DATE	Date questionnaire was mailed by contractor (initial mailing)	
Q1	FISH_YEAR_FLG	Did angler fish during 2009?	Yes=1, No=0, 88 = Missing
Q4	LICENSE_FLG	Did angler have a NC fishing license during previous 12 months (Y/N)?	Yes=1, No=0, 88 = Missing
Q5	SALT_LIC_FLG	Was license for recreational saltwater fishing? (Y/N)	Yes=1, No=0, 88 = Missing
Q6	WAVE_LIC_FLG	Was license valid during	Yes=1, No=0, 88 = Missing

		November 2009? (Y/N)	
Q7	FISH_WAVE_FLG	Did angler fish during the wave (wave 6, 2009)?	Yes=1, No=0, 88 = Missing
Q8	BOAT_TRPS	Number of private boat trips during the wave	
Q9	SHORE_TRPS	Number of shore trips during the wave	

License Angler

There is one record for every angler identified in the household screener, regardless of whether or not the angler was sampled or returned a questionnaire.

Question	Field Name	Description	Coding Scheme
	SURV_YEAR	Survey year	2009
	SURV_WAVE	Survey wave	6
	LIC_ST	License state (will be 37 (NC) in all cases for pilot study)	CRFL Infant = 1 CRFL Youth = 2 Res CRFL = 3 Res CRFL 10-Day = 4 Res CRFL Adult = 5 NonRes CRFL = 6 NonRes CRFL 10-Day = 7 NR CRFL Adult = 8 Age 65 CRFL = 9 Disabled Vet CRFL = 10 Totally Disabled CRFL = 11 Perm Disabled State Fish w CRFL = 12 Uni Adlt Care Hme Inland/CRFL = 13 Uni Blind Inland/CRFL = 14 Unified Inland/CRFL = 15 Unified Sptm/CRFL = 16 Ltime Unified Inland/CRFL = 17 Ltime Comp Inland Fish w/CRFL = 18 Subsis Inland/CRFL Waiver = 19 Disabled Combo H/F/CRFL Basic = 20 Sportsman Infant w CRFL = 21 Unified Sptm/CRFL Infant = 22 Sportsman Youth w CRFL = 23 Unified Sptm/CRFL Youth = 24 Res Sportsman Adult w CRFL = 25 Res Uni Sptm/CRFL Adult = 26 NonRes Sportsman Adult w CRFL = 27 NR Uni Sptm/CRFL Adult = 28 Unified Age 65 Sptm/CRFL = 29 Res Ltime Over 70 Sportsman w CRFL = 30 Ltime H/F/Trap/CRFL Disabled Vet = 31 Lifetime Comp Over 70 Fish w CRFL = 32 Disabled Sportsman w CRFL = 33 Uni Disabled Vet Sptm/CRFL = 34 Uni Totally Disabled Sptm/CRFL = 35
	REC_DATE	Date questionnaire received	
	MAIL_DATE	Date questionnaire was sent (initial mailing)	
	ANG_ID	Unique angler identification	

	HH_ID	Unique household identifier	
	ANG_STATUS	Final disposition of sample (complete with trips, complete no trips, refusal, non-contact, etc.)	1.2.1 = No trips taken in the 2 month period 1.2.2 = Trips taken in the 2 month period 2.1 = Refused 3.19 = Nothing ever returned 3.3 = Mailing returned undelivered
	STRATUM	Coastal, non-coastal, out-of-state	Coastal=1, Non-Coastal=2, Out-of-state=3
	RES_ST	State of residence	
	RES_CNTY	County of residence	
Q1	FISH_YEAR_FLG	Did angler fish during 2009?	Yes=1, No=0, 88 = Missing
Q2	GENDER		Male=1, Female=2
Q3	AGE		Less than 16 = 1 16 -- 17 = 2 18 -- 24 = 3 25 -- 34 = 4 35 -- 44 = 5 45 -- 54 = 6 55 -- 64 = 7 65 or older = 8
Q4	LICENSE_FLG	Did angler have a NC fishing license during previous 12 months?	Yes=1, No=0, 88 = Missing
Q5	SALT_LIC_FLG	Was license for recreational saltwater fishing?	Yes=1, No=0, 88 = Missing
Q6	WAVE_LIC_FLG	Was license valid during November 2009?	Yes=1, No=0, 88 = Missing
Q7	FISH_WAVE_FLG	Did angler fish during the wave (wave 6, 2009)?	Yes=1, No=0, 88 = Missing
Q8	BOAT_TRPS	Number of private boat trips during the wave	
Q9	SHORE_TRPS	Number of shore trips during the wave	
	SAMP_WT	Sample weight (N/n)	
	FRM_SIZE	Number of anglers on sample frame for stratum (N)	

Trip Information

Anglers provide detailed trip information for up to four recent trips. There is one record per trip.

Question	Field Name	Description	Coding Scheme
	SURV_YEAR		
	SURV_WAVE		
	HH_ID		
	ANG_ID		
	TRIP_ID	Unique identifier for each trip within an angler	
Q3B	MODE		Pier=1 Dock = 2 Jetty or Breakwater = 3 Bridge or Causeway = 4 Other man-made structure = 5 Bank or beach = 6
Q2/Q3	MODE_FX	Shore or private boat	Yes=1, No=0
Q1	TRIP_DATE	Date of trip	11/1 - 12/31
	FRAME	Is trip for an angler sampled from the license frame or the ABS frame?	ABS=1, License=2
2A/3A	TRIP_ST	State of trip	North Carolina
2A/3A	TRIP_CNTY		
2B/3B	ACCESS	Private/public	Yes, public access = 1 No, private access = 2
Q4	AREA		Ocean, within 3 miles from the shore = 1 Ocean, more than 3 miles from the shore = 2 Sound = 3 River = 4 Bay = 5 Inlet = 6 Someplace else = 7
	AREA_X		

5	RTN_TIME	Return time (time trip ended)	Midnight -- 3:00 am = 1 3:00 am -- 6:00 am = 2 6:00 am -- 9:00 am = 3 9:00 am -- Noon = 4 Noon -- 3:00 pm = 5 3:00 pm -- 6:00 pm = 6 6:00 pm -- 9:00 pm = 7 9:00 pm -- Midnight = 8
6	ADD_ANG_FLG	Did anyone else from your household fish with you (Y/N)	Yes=1, No=0
6A_1	SPOUSE_FLG	Did sampled angler fish with spouse in this trip?	6a = 1, 6a = 2,3
6A_2	CHILD_FLG	Did sampled angler fish with child on this trip?	6a = 2, 6a = 1,3
6A_3	OTHER_FLG	Did sampled angler fish with an other household member?	6a = 3, 6a = 1,2
6B	TOT_ANG	Total number of household members fishing on trip	

Appendix I: Tabulations of Key Variables

Household Questionnaire

Disposition of Sample		
	Frequency	Percent
Household with angler	228	12.7%
Household with no angler	457	25.4%
Refused	22	1.2%
Mailing returned undelivered	157	8.7%
No response	936	52.0%
Total	1800	100%

Did Anyone in the Household Fish During the Previous 12 Months?		
	Frequency	Percent
Yes	228	32.2%
No	457	64.6%
Missing	22	3.1%
Total	707	100.0%

Does the Household have a Landline Telephone?		
	Frequency	Percent
Yes	516	75.3%
No	149	21.8%
Missing	20	2.9%
Total	685	100.0%

How Many People in Household Fished During the Previous 12 Months?		
	Frequency	Percent
1	80	35.1%
2	89	39.0%
3	32	14.0%
4	14	6.1%
5	5	2.2%
6	4	1.8%
7	1	.4%
8	2	.9%
10	1	.4%
Missing	228	12.7%

Angler Questionnaire

Disposition of Sample				
	License		ABS	
	Frequency	Percent	Frequency	Percent
Trips taken in the 2 month period	227	14.5%	51	14.0%
No trips taken in the 2 month period	739	47.3%	137	37.6%
Refused	19	1.2%	3	.8%
Nothing ever returned	438	28.0%	63	17.3%
Mailing returned undelivered	139	8.9%	8	2.2%
No response			102	28.0%
Total	1562	100%	364	100%

Gender of the Respondent				
	License		ABS	
	Frequency	Percent	Frequency	Percent
Female	164	16.6%	34	17.8%
Male	735	74.6%	141	73.8%
Missing	86	8.7%	16	8.4%

Did the Respondent Perform in Recreational Saltwater Fishing in 2009?				
	License		ABS	
	Frequency	Percent	Frequency	Percent
Yes	718	23.4%	152	79.6%
No	230	72.9%	31	16.2%
Missing	37	3.8%	8	4.2%

License Type		
	Frequency	Percent
Residential CRFL	366	23.4
Residential CRFL 10-day	11	.7
Residential CRFL Adult	4	.3
Non-residential CRFL	282	18.1
Non-residential CRFL 10-day	115	7.4
Age 65 CRFL	87	5.6
Disabled Vet CRFL	4	.3
Totally Disabled CRFL	6	.4
Perm Disabled State Fish w CRFL	15	1.0
Uni Blind Inland / CRFL	1	.1
Unifed Inland / CRFL	40	2.6
Unified Sptm / CRFL	87	5.6
Lifetime Unifed Inland / CRFL	1	.1
Lifetime Comp Inland Fish w CRFL	15	1.0
Subsidized Inland / CRFL Waiver	35	2.2
Disabled Combo H/F/CRFL Basic	11	.7
Sportsman Infant w CRFL	39	2.5
Sportsman Youth w CRFL	29	1.9
Residential Sportsman Adult w CRFL	145	9.3
Residential Uni Sptm / CRFL Adult	6	.4
Non-residential Sportsman Adult w/ CRFL	26	1.7
Non-residential Uni Sportsman / CRFL Adult	2	.1
Unified Age 65 Sportsman / CRFL	54	3.5
Residential Lifetime Over 70 Fish w/ CRFL	122	7.8
Lifetime Comp Over 70 Fish w/ CRFL	45	2.9
Disabled Sportsman w/ CRFL	10	.6
Uni Disabled Vet Sptm / CRFL	4	.3
Total	1562	100.0

Age of the Respondent				
	License		ABS	
	Frequency	Percent	Frequency	Percent
Less than 16	2	.2%	1	.5%
16 – 17	1	.1%		
18 – 24	42	4.3%	13	6.8%
25 – 34	86	8.7%	19	9.9%
35 – 44	136	13.8%	33	17.3%
45 – 54	206	20.9%	37	19.4%
54 – 64	196	19.9%	45	23.6%
65 or older	226	22.9%	28	14.7
Missing	90	9.1%	15	7.9%

Has the Respondent Fished in NC During the Past 12 Months?				
	License		ABS	
	Frequency	Percent	Frequency	Percent
Yes	871	88.4%	132	69.1%
No	31	3.1%	43	22.5%
Missing	83	8.4%	16	8.4%

Was the License for Recreational Saltwater Fishing?				
	License		ABS	
	Frequency	Percent	Frequency	Percent
Yes	731	10.9%	106	55.5%
No	107	74.2%	22	11.5%
Missing	147	14.9%	63	33.0%

Was the License Valid During November 2009?				
	License		ABS	
	Frequency	Percent	Frequency	Percent
Yes	675	68.5%	84	44
No	51	5.2%	23	12
Missing	259	26.3%	84	44

Did the angler fish during Wave 6, 2009?				
	License		ABS	
	Frequency	Percent	Frequency	Percent
Yes	227	23%	49	25.7%
No	738	74.9%	136	71.2%
Missing	20	2.0%	6	3.1%

Trip Questionnaire

Was the Trip from the Shore or Private Boat?		
	Frequency	Percent
Shore	385	59.8%
Boat	252	39.1%
Missing	7	1.1%
Total	644	

Collapsed Mode of Fishing		
	Frequency	Percent
Ocean, less than 3 miles from the shore	306	47.5%
Ocean, more than 3 miles from the shore	48	7.5%
Inland trip	231	35.9%
Missing	8	1.2%
More than one response checked	51	7.9%

Public Access for Boat and Shore Trips				
	Boating Trips		Shore Trips	
	Frequency	Percent	Frequency	Percent
Yes, public access	179	71.0%	340	86.7%
No, private access	60	23.8%	38	9.7%
Missing	13	5.2%	14	3.6%
Total	252		392	

Mode of Shore Trip		
	Frequency	Percent
Pier	74	16.5
Dock	31	6.9
Jetty or Breakwater	18	4.0
Bridge or causeway	26	5.8
Other man-made structure	8	1.8
Bank or beach	276	61.5
Missing	16	3.6
Total	449	

Fishing Area of Trip		
	Frequency	Percent
Ocean, within 3 miles from the shore	306	47.5%
Ocean, more than 3 miles from the shore	48	7.5%
Sound	79	12.3%
River	66	10.2%
Bay	5	.8%
Inlet	59	9.2%
Someplace else	22	3.4%
Missing	8	1.2%
More than one response checked	51	7.9%

Time the Trip Ended		
	Frequency	Percent
Midnight – 3:00 am	7	1.1%
3:00 am – 6:00 am	6	.9%
6:00 am – 9:00 am	27	4.2%
9:00 am – Noon	75	11.6%
Noon – 3:00 pm	121	18.8%
3:00 pm – 6:00 pm	286	44.4%
6:00 pm – 9:00 pm	82	12.7%
9:00 pm – Midnight	11	1.7%
Missing	6	.9%
Invalid answer (multiple responses)	23	3.6%

Was Anyone in the Household who was also an Angler?		
	Frequency	Percent
Yes	332	47.0%
No	303	51.6%
Missing	9	1.4%

Did the Angler's Spouse Fish with the Respondent on this Trip?		
	Frequency	Percent
Spouse	156	45.1%
Child / Children	121	35.0%
Other	69	19.9%
Total	346	100.0%

Total Number of Household Members Fishing on the Trip		
	Frequency	Percent
1	346	53.7%
2	190	29.5%
3	57	8.9%
4	22	3.4%
5	6	.9%
7	4	.6%
8	4	.6%
10	2	.3%
12	1	.2%
Missing	12	1.9%
Total	644	100.0%

Development and Testing of Recreational Fishing Effort Surveys

Testing a Mail Survey Design

Final Report

Project Team Members:

Rob Andrews, NOAA Fisheries
J. Michael Brick, Westat
Nancy A. Mathiowetz, University of Wisconsin-Milwaukee

July 31, 2014

Acknowledgements: Westat completed all phases of the data collection, including developing and programming survey instruments, sampling, delivering and receipting survey materials, administering a non-response survey, quality control and data editing. The project team specifically thanks Howard King, who managed the project on behalf of Westat, as well as Pat Clark, Janice Ricketts, Pat Dean Brick and Shelley Brock. Staff from the NOAA Fisheries Service, Office of Science and Technology provided a substantive review of this report.

1. Executive Summary

The mail survey design evaluated in this project is the culmination of several years' worth of testing and analysis to develop an alternative to the Coastal Household Telephone Survey (CHTS) for estimating marine recreational fishing effort. The objectives of the project were to: 1) test the feasibility of the design for collecting recreational fishing effort data and estimating fishing effort for shore and private boat anglers, 2) compare mail survey and CHTS results, including metrics of survey quality and estimates of marine recreational fishing activity, 3) describe, to the greatest extent possible, differences between mail survey and CHTS estimates in terms of sources of survey error, and 4) provide recommendations for follow-up action, including implementation of improved survey methods.

This report is intended to summarize the findings for a non-technical audience. For those interested in the details related to sampling, estimation, or instrument design, we have provided references to detailed reports and publications.

Results from the study continue to demonstrate that mail survey designs are a feasible alternative to telephone surveys for collecting recreational fishing data and producing population estimates in a timely manner. Overall, final mail survey response rates were nearly three times higher than CHTS response rates, and preliminary estimates, derived from partial data collected within two weeks from the end of the reference wave, were not significantly different from final estimates, demonstrating that a mail survey can generate stable fishing effort estimates within the current estimation schedule for the CHTS. In addition, the sampling design, which includes over-sampling of households with licensed anglers, is more efficient for collecting fishing data than simple random sampling currently used for the CHTS.

Overall, the mail survey estimate of total fishing effort was 4.1 times larger than the corresponding CHTS estimate. Differences between mail survey and CHTS estimates, which were relatively consistent among the states included in the study, can largely be attributed to differences in fishing prevalence – households in the mail survey sample were more likely to report fishing than households in the CHTS sample. We explored these differences within the context of survey error and conclude that the mail survey design is less susceptible than the CHTS to bias resulting from nonresponse and non-coverage. We also suggest that the nature of the mail survey mode results in more accurate responses to questions about fishing activity than the CHTS, which expects respondents to answer questions on-the-spot, without the benefit of aided recall or memory cues. Finally, we demonstrate that the CHTS sampling levels and estimation strategy may introduce biases, particularly in low-activity waves, and we suggest that CHTS coverage correction factors, derived from a complementary onsite site survey of completed fishing trips to compensate for the geographic limitations of the CHTS, may result in biases in fishing effort estimates due to the exclusion of private access fishing sites from the onsite survey sample frame.

Given the potential for bias in the CHTS, we conclude that the mail survey design is a superior approach for monitoring recreational fishing effort. We also encourage continued testing and evaluation to assess additional sources of survey error and ensure that evolving advancements in survey methodology are considered and customer needs are satisfied.

2. Background

Traditionally, marine recreational fishing effort data for the U.S. Atlantic Coast and the Gulf of Mexico have been collected by NOAA Fisheries through the Coastal Household Telephone Survey (CHTS). The CHTS utilizes a list-assisted, random digit dialing (RDD) telephone survey approach to contact residents of coastal county households and collect information on fishing activities that occurred within a two-month reference period (wave). Specifically, households are screened to determine if any household members participated in marine recreational fishing during the previous 2 months, and each active angler is asked to recall, episodically, the number of saltwater fishing trips that were taken during the wave, as well as provide details about each trip.

In recent years, the efficiency and effectiveness of RDD surveys in general, and the CHTS specifically, have been questioned due to declining rates of coverage and response (Curtin et al. 2005; Blumberg and Luke 2013). A 2006 review by the National Research Council (NRC 2006) noted that the CHTS design suffers from inefficiency due to the low rate of saltwater angler participation among the general population, as well as potential coverage bias due to the survey's limitation to coastal county residences and landline-based telephone numbers (NRC 2006). In addition, response rates to the survey have declined considerably over the past decade, increasing the potential for nonresponse bias. To address these shortcomings, the NRC review recommended the development of and subsequent sampling from a comprehensive list of registered saltwater anglers or, in the absence of such a list, implementation of dual-frame procedures that include sampling from both lists of licensed saltwater anglers and residential household frames.

The Marine Recreational Information Program (MRIP) has designed and tested several different data collection alternatives to address concerns about the CHTS. Below, we outline the various approaches to collecting fishing effort data that have been studied by MRIP. More detailed descriptions of the data collection designs and comparisons of estimates and metrics of survey quality, such as response rates and coverage rates, are documented elsewhere (Brick et al. 2012a; Andrews et al. 2013).

2.1 Angler License Directory Telephone Survey

As noted by the NRC, a more efficient approach for surveying anglers is to sample directly from lists of individuals who are licensed to participate in saltwater fishing. Working collaboratively with the Gulf States Marine Fisheries Commissions, the Gulf Coast states, and the North Carolina Division of Marine Fisheries, MRIP has designed and tested Angler License Directory Telephone Surveys (ALDS), which sample from state databases of licensed anglers. The ALDS was implemented as a pilot project in Florida (FL), Alabama (AL), Mississippi (MS) and Louisiana (LA) in 2007 and expanded to North Carolina (NC) in 2008. The survey was most recently administered in in 2012.

The data collection procedures for the ALDS are nearly identical to the CHTS, with the exception of the screening portion of the survey; the ALDS requests to speak with the individual licensed angler by name and then proceeds to determine if the angler, or any other individuals who reside in the same household as the angler, fished during the wave. As with the CHTS, trip details are collected through episodic recall beginning with the most recent trip.

As predicted, the ALDS is more efficient than the CHTS at identifying anglers – in a recent reference wave, 46% of ALDS respondents reported fishing, while only 6.5% of CHTS respondents reported fishing during the same wave. However, exemptions to state licensing requirements and unlicensed fishing activity, as well as incomplete and inaccurate contact information for individuals included on the sample frames, create gaps in the coverage of the ALDS. Subsequent studies (Brick et al. 2012a; Andrews et al. 2013) have suggested that undercoverage due to unlicensed fishing activity may be as high as 70% in some states for certain types of fishing activity, and that as many as 20% of frame entries may be unreachable due to “bad” (missing, nonworking, wrong number) telephone numbers. In addition, response rates for the ALDS are only marginally higher than CHTS response rates.

2.2 Dual-Frame Telephone Survey

As noted above, the CHTS and the ALDS, considered individually, do not provide complete coverage of the angler population. To compensate for potential sources of coverage error in the CHTS and ALDS, MRIP developed an estimation design that integrates CHTS and ALDS sampling in a dual-frame design (Lai and Andrews 2008). The union of the CHTS and ALDS sample frames defines three domains: 1) anglers who can only be sampled from the CHTS frame (unlicensed anglers with landline phones who reside in coastal counties covered by the CHTS); 2) anglers who can only be sampled from the ALDS frame (licensed anglers who reside outside of the coverage area of the CHTS); and, 3) anglers who can be sampled from both the CHTS and ALDS frames (licensed anglers who reside in coastal counties). A fourth domain includes anglers who cannot be sampled by either the CHTS or ALDS (unlicensed anglers without landline telephones within the CHTS coverage area and unlicensed anglers residing outside the coverage area of the CHTS).

While the dual-frame telephone survey design increases the coverage over either the CHTS or the ALDS, the methodology is not without limitations. As mentioned, the union of the CHTS and ALDS sample frames excludes a segment of the angling population, creating a potentially significant gap in coverage - up to 38% of fishing trips in NC are taken by anglers who are excluded from either the CHTS or ALDS (Andrews et al. 2010). In addition, partitioning anglers into the appropriate domains and subsequently adjusting sample weights is based upon the survey respondents' willingness and ability to classify themselves as licensed or unlicensed anglers. This has been demonstrated to be an unreliable approach for defining dual-frame domains (Andrews et al. 2010) and results in survey weights that may produce biased estimates. Finally, the dual-frame telephone survey approach is susceptible to nonresponse bias due to the low response rates of the component surveys.

2.3 Dual-Frame Mail Survey

An alternative to the dual-frame telephone survey is to identify and contact anglers through a dual-frame mail survey design. MRIP initially tested the feasibility of a dual-frame mail survey design in NC in 2009, and conducted a follow-up study aimed at enhancing response rates and the timeliness of responding in NC and LA in 2010.

The specific details of the dual-frame mail survey design tested in 2009 and 2010 are described elsewhere (Andrews et al. 2010; Brick et al. 2012b). Briefly, anglers are sampled both from state databases of licensed saltwater anglers and residential address frames maintained and made commercially available by the United States Postal Service. The address-based sample (ABS) is matched to the license databases by searching the license frame for the same address and/or telephone number (for the cases in which a telephone number can be located through a commercial service for the ABS sampled address). This matching identifies those households that could be sampled from both frames.

For the studies conducted in 2009 and 2010, anglers were sampled from the license frame in a single phase, and the sampled anglers were mailed a brief questionnaire asking them to report the number of days fished from the shore and from a boat during a two-month reference wave. The ABS sampling was conducted in two phases. In the first phase, residential addresses in the state were sampled and mailed a screening questionnaire to identify individuals who fished during the previous twelve months. In the second phase, anglers identified in the screening phase were sent a second-phase questionnaire that was identical to the one sent to those sampled from the license frame.

Results of these pilot studies were encouraging; sampling from the ABS frame provides nearly complete coverage of the U.S. population, and response rates to the mail survey were greater than either the ALDS or CHTS (Andrews et al. 2010; Brick et al. 2012a). In addition, the ability to match ABS sample to license frames *a priori* by address matching provides a more effective means for defining domain membership that is not susceptible to recall error or inaccurate reporting. Frame matching also provides supplemental information for assessing nonresponse error for the ABS sample and for nonresponse weighting adjustment.

The dual-frame mail survey design provides many benefits over telephone survey approaches. However, frame matching is not 100% accurate, resulting in misclassification of domain membership for some sample units; generally frame units that could have been sampled from both frames are excluded from the overlapping domain due to a failure to match. Consequently, dual-frame weights are not down-weighted appropriately, resulting in an overestimation of fishing effort (Brick et al. 2012a). In addition, there were concerns that a mail survey design could not satisfy customer needs for timely estimates, although comparisons between early survey returns and later survey returns showed little difference in terms of fishing activity, suggesting that preliminary effort estimates could be produced within the timeframe required by customers.

2.4 Dual-Frame, Mixed-Mode Survey

To further address concerns about timeliness, as well as explore differences between mail and telephone data collection modes, MRIP administered a dual-frame, mixed-mode survey in 2012 (Andrews et al 2013). The sample design for the survey was nearly identical to the dual-frame mail survey – anglers were sampled from angler license frames and households were sampled from residential address frames. As with the dual-frame mail survey, the ABS sample was mailed a screening questionnaire to identify anglers at the sampled addresses. The methodology differed from the dual-frame mail survey in that anglers identified through household screening, as well as anglers sampled from the state license databases, were randomly allocated into telephone and mail treatment groups – anglers in the telephone treatment group were contacted and asked to provide information about recent recreational fishing trips through a telephone interview, and anglers in the mail treatment group were mailed a questionnaire that asked about recent recreational fishing activity. If no phone number for the sampled household was available, then the second phase was done by mail.

Results from the study continued to demonstrate that mail survey designs are feasible for collecting recreational fishing data and estimating fishing effort. Final response rates for the mail survey component of the study were higher than the telephone component and eclipsed telephone survey response rates after about three weeks of data collection (Andrews et al., 2013). In addition, preliminary estimates derived from early mail survey returns were not significantly different from final estimates, demonstrating that a mail survey can generate valid preliminary estimates within the current estimation schedule for the CHTS.

The impact of data collection mode on survey measures required further investigation. We hypothesized that differences between telephone and mail estimates were the result of differential recall and coverage errors, and suggested that telephone samples are more susceptible to bias resulting from these errors (Andrews et al. 2013).

3. Mail Survey with Screening Prior to Data Collection

The pilot tests described in the previous section were very informative and provided the basis for a revised design that is the focus of this report. The revised design again uses a mail questionnaire to collect data from households, but also addresses weaknesses identified in the prior studies. For example, the design uses the license frame in a way that eliminates biases resulting from inaccurate matching to the address frame. Furthermore, the mail data collection scheme and the questionnaire were revised to attempt to further increase response rates. These and other features of the design are described below.

The new design was tested in MA, NY, NC and FL beginning in wave 5 (Sep/Oct), 2012 and continuing through wave 6 (Nov/Dec), 2013. The objectives of the study are to assess the feasibility of the design in terms of response rates, timeliness, and efficiency, as well as examine the impact of different sources of survey error on estimates of fishing prevalence and total fishing effort.

3.1 Methods

The survey employed a dual-frame design with non-overlapping frames; residents of the target states - states included in the pilot study - were sampled from the United States Postal Service computerized delivery sequence file (CDS), and non-residents - individuals who were licensed to fish in one of the target states but lived in a different state - were sampled from state-specific lists of licensed saltwater anglers.

Sampling from the CDS utilized a stratified design in which households with licensed anglers were identified prior to data collection (Lohr 2009). The address frame for each state was stratified into coastal and non-coastal strata defined by geographic proximity to the coast[1]. For each wave and stratum, a simple random sample of addresses was selected from the CDS and matched to addresses of anglers who were licensed to fish within their state of residence[2]. Augmenting address samples in this manner effectively screened the sample into strata defined by the presence (matched) or absence (unmatched) of at least one licensed angler at an address. All matched addresses were retained in the sample and unmatched addresses were subsampled at a rate of 30%. Initial addresses samples were sufficiently large to support subsampling from the unmatched stratum. Screening the address sample prior to data collection and subsampling the resulting sub-populations at different rates (e.g., sampling addresses with licensed anglers at a higher rate) was expected to increase the efficiency of the design while maintaining the coverage of the address frame – two concerns identified by the NRC Review. Furthermore, because the

[1] For waves 1, 2 and 6 the coastal strata included all addresses in counties that were within 25 miles of the coast. For waves 3-5, the coastal strata included all addresses in counties that were within 50 miles of the coast.
[2] Matching was by exact address and/or telephone number when available.

matching was only used to determine the sampling rate, matching errors (e.g., not identifying some addresses with licensed anglers due to matching errors) will only impact the efficiency of data collection. This approach was a fairly substantive departure from the dual-frame sampling designs tested in prior pilot studies.

Non-resident anglers were sampled directly from state license databases. The sample frame for each of the targeted states consisted of unique household addresses that were not in the targeted state but had at least one person with a license to fish in the targeted state during the wave. For each state and wave, a simple random sample of addresses was selected.

For both the resident and non-resident samples, a questionnaire was developed to measure fishing activity within the targeted state. Household members that did not fish were asked to indicate that they had no trips. The questionnaire was totally revised from previous pilot studies and required only one step of data collection (previous pilots included two phases of data collection; a household screening phase to identify anglers and a second phase to collect detailed fishing information from anglers). In the new questionnaire, any adult in the household could respond for all household members. The mail survey collected fishing effort data for all household residents, including the number of saltwater fishing trips by fishing mode (shore and private boat), for two-month reference waves, beginning with wave 5, 2012 and continuing through wave 6, 2013. The single phase of data collection was designed to increase the timeliness and the response rates to levels above those observed in the earlier pilots.

The data collection procedures for residents and non-residents were identical. One week prior to the end of each wave, sampled addresses were mailed a survey packet including a questionnaire[3] (Appendix A), a cover letter stating the purpose of the survey, a cash incentive[4] and a business reply envelope. One week after the initial mailing, all households received either an automated telephone reminder call or a postcard reminder, depending on whether or not a telephone number could be matched by a commercial vendor to the sampled address[5]. A final survey packet, excluding the cash incentive, was sent to all nonrespondents three weeks after the initial mailing.

[3] The questionnaire included as Appendix A is the final version of the questionnaire that was tested in the study.
[4] Cash incentives are discussed in more detail below.
[5] All addresses for which a telephone number could be matched received the automated telephone reminder.

Cognitive interviews of both anglers and non-anglers were conducted at the outset of the study to explore respondent reactions to different versions of the survey instrument. The interviews resulted in multiple versions of the questionnaire, which were subsequently tested in an experimental design. In addition to the questionnaire experiments, we tested the impact of different levels of prepaid cash incentives on response rates and survey measures. The design and results of the questionnaire and incentive experiments are described in Appendix B. Based upon the results of the incentive experiment, we included a $2.00 prepaid cash incentive in the initial survey mailing for subsequent waves (Wave 1, 2013 – Wave 6, 2013). The comparisons to the CHTS presented below are for waves 4-6, 2013, after the initial questionnaire and incentive experiments were completed, and are based on the fielding of one version of the questionnaire with the use of the $2 incentive.

4. Findings

This section compares the outcomes from the pilot test of the mail survey design to the outcomes from the production CHTS, which was fielded concurrently in the pilot test states. The first outputs are related to survey quality and the second outputs are survey estimates. Unless otherwise noted, all estimates presented have been weighted. For the CHTS, the survey weights are the regular production weights, and for the mail survey, the weights include the base weights, nonresponse adjustments, and adjustments to control totals of the number of households within each study state.

4.1 Quality Metrics

Overall, the response rate for the mail survey was 40.4% (Table 1). Response rates ranged from 32% in NY to 45.4% in FL. Overall, the mail survey response rate was 2.8 times higher than the CHTS response rate of 14.1% for the same states and waves. The overall response rate for the license sample (nonresident anglers) was 47.5% and ranged from 46.7% in FL to 55.8% in MA.

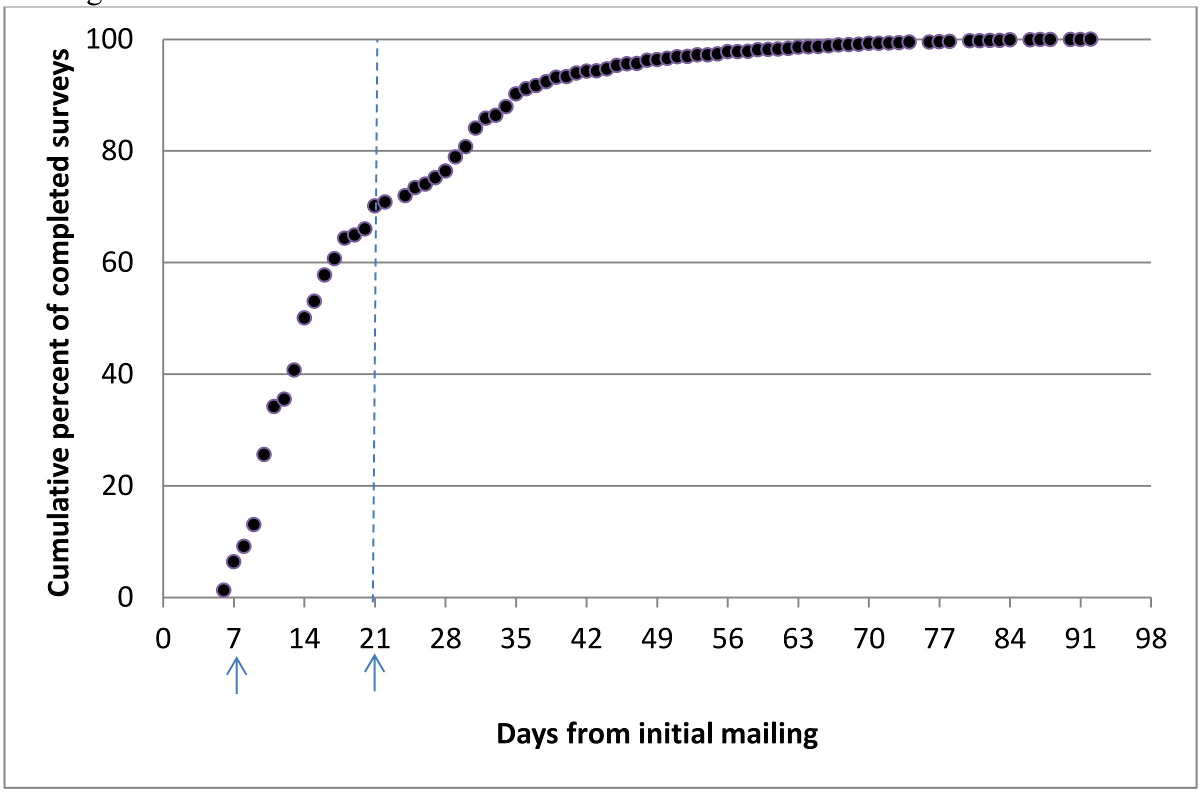
Table 1. Response rates, by state, from the CHTS and mail survey, for coastal counties and waves 4, 5, and 6, 2013.

	Mail		CHTS	
State	%	n	%	n
Florida	45.4	7,460	14.5	2,588,115
Massachusetts	40.6	6,279	13.1	275,967
New York	32.0	4,908	11.6	421,636
North Carolina	41.7	6,203	16.4	332,934
All	40.4	24,850	14.1	3,618,652

Note: American Association for Public Opinion Research Response Rate 3 (AAPOR RR3). Response rate formula excludes ineligible addresses and estimates the proportion of unknown cases that are actually eligible based upon known sample dispositions. Sample sizes reflect the total number of addresses and telephone numbers sampled for the mail survey and CHTS, respectively, regardless of eligibility.

The median response time for the resident mail survey was 14 days. Median response times were consistent among states. Approximately 72% of surveys were returned within 21 days of the initial survey mailing or within two weeks following the conclusion of the reference wave (Figure 1), resulting in a preliminary response rate of approximately 30%. This corresponds with the timing of CHTS data collection, which is conducted during the first two weeks following the end of the reference wave.

Figure 1. Cumulative distribution of mail survey returns from the timing of the initial survey mailing. The dashed vertical line represents the completion of data collection for the CHTS (2 weeks following the end of each wave). The arrows show the timing of the IVR/post-card reminder and mailing of the second questionnaire at 7 and 21 days, respectively, after the initial mailing



To assess the feasibility of generating mail survey estimates within the timeframe for producing CHTS estimates, we compared preliminary estimates of fishing prevalence (percent of household that reported fishing during the reference wave), derived from mail surveys returned within two weeks of the end of the reference wave, to final estimates, derived from complete survey data collected over a 12-week period (Table 2). Overall, the relative difference between preliminary and final estimates of fishing prevalence was approximately 3% (9.7% vs. 10.0%), and there were no significant differences between preliminary and final estimates, overall, at the state level or by fishing mode. These results demonstrate that preliminary estimates are consistent with final estimates, and that a mail survey is a feasible alternative to telephone surveys for producing recreational fishing statistics in a timely manner.

Table 2. Preliminary and final estimated fishing prevalence, by state, from the mail survey, for waves 4, 5, and 6, 2013.

	Preliminary		Final		*p*-value
State	%	SE	%	SE	
Florida	16.4	0.9	16.3	0.7	0.9124
Massachusetts	8.2	0.8	8.2	0.6	0.9630
New York	5.0	0.6	5.5	0.5	0.2123
North Carolina	8.4	0.8	8.7	0.7	0.4799
All	9.7	0.4	10.0	0.4	0.1916

Note: Significance based upon results of a z-test where the standard deviation of the difference was computed taking into account the correlation due to the estimates containing a common subset of observations.

One of the goals of this study was to assess the effectiveness of the design for sampling saltwater anglers, a relatively rare population. Overall, addresses that matched to a license list were more likely than unmatched addresses to both respond to the survey (48.6% vs 34.1%) and report fishing during the reference wave (42.1% vs. 8.1%)[6,7]. These results suggest that matching was effective at defining sub-populations that were distinct with respect to fishing activity. We quantified the benefits of the design by comparing weighted and unweighted estimates of fishing prevalence. Overall, the unweighted estimate (16.0%), which reflects the relative occurrence of fishing households within the sample, was 1.6 times higher than the weighted estimate (10.0%), which reflects fishing activity within the population as a whole. In other words, the design was 1.6 times more likely to result in a survey completed by a fishing household than one would expect from a simple random sample of households. This factor can be further adjusted by changing the subsampling rate for the unmatched households, but this feature of refining the design was not an objective of this feasibility study.

We also calculated the design effect for estimates of fishing prevalence by comparing the estimated sample variance to the variance which would have been obtained from a simple random sample of the same size. For estimates of fishing prevalence, the overall design effect was 0.90, which suggests that the mail survey design can achieve the same precision as simple random sampling (i.e., the same effective sample size) with 10% less sample. A design effect of

[6] The impact of differential response between matched and unmatched households is discussed below.
[7] Response rates and prevalence rates are for both coastal and non-coastal residents.

less than 1.0 indicates that a sample design, including stratification, weighting, non-response adjustment, etc., is more efficient than simple random sampling.

4.2 Estimate Comparisons

While the CHTS is the basis for estimating total fishing effort for all anglers, the data collection of the survey is limited to counties within a specified distance of the coast – the CHTS estimates fishing effort by sampling residents of coastal counties[8]. Consequently, we limit direct comparisons between the CHTS and mail survey estimates to the coastal region. We also explore the impact of CHTS geographic coverage by comparing mail survey estimates to CHTS coverage correction factors. These factors are derived from the Access-Point Angler Intercept Survey (APAIS), an independent dockside survey of completed recreational fishing trips, and are used to expand the CHTS estimates to the full population.

Table 3 compares mail survey and CHTS estimates for several measures of interest. In the coastal counties covered by both surveys, the mail survey estimate of total fishing effort is approximately 4.1 times larger than the CHTS estimate (63,082,000 trips vs. 15,510,000 trips). The direction of differences between CHTS and mail survey estimates of total effort is consistent among states, although the magnitude of the differences varies from a factor of approximately 3.4 in NC to a factor of over 5 in NY. The direction of differences between CHTS and mail survey estimates is also consistent between fishing modes (private boat fishing and shore fishing), although differences are much more pronounced for shore fishing, where the mail estimate is larger than the CHTS estimate by a factor of 6.1 (40,425,000 vs. 6,642,000), than for private boat fishing, where the mail estimate is 2.6 times larger than the CHTS estimate (22,658,000 vs. 8,868,000).

We first examine the differences between CHTS and mail survey estimates of total effort by comparing the components of effort estimates. One component is fishing prevalence, or the percentage of households that reported fishing during a reference wave, and the other component is mean trips per fishing household. Among those households that reported fishing during a

[8] Generally, a coastal county is defined as a county that is within 25 miles of the coast. However, there are exceptions to this definition, including FL where all counties are considered coastal and NC, where the coastal region is expanded to 100 miles during periods of high fishing activity (June-October).

reference wave, CHTS and mail survey estimates of mean trips per household are similar – overall, mail survey estimates of mean trips are larger than CHTS estimates by a factor of 1.2 (11.2 trips vs. 9.0 trips). Estimates are also similar for households that reported fishing in a specific mode. For mean shore trips per household, mail estimates are larger than CHTS estimates by a factor of 1.1 (9.0 trips vs. 8.0 trips), and for mean boat trips per household, CHTS estimates are larger than mail estimates by a factor of 1.1 (8.3 trips vs. 7.7 trips).

In contrast, the mail survey estimate of overall fishing prevalence is 2.7 times larger than the CHTS estimate (12.8% vs. 4.8%). Collectively, these results suggest that households in the mail sample are much more likely to report fishing during a reference wave than households in the CHTS sample, but fishing behavior in the two samples is similar for those households that reported at least one fishing trip.

Consequently, we focus on exploring differences between the two surveys in estimated fishing prevalence – i.e., why do more households report fishing in the mail survey than the CHTS? There are several substantive differences between the CHTS and the mail survey designs that likely contribute to differences in estimated prevalence, notably the sample frames and data collection modes. In the following section, we examine the impact of these design features on survey estimates and describe the impacts in terms of survey error. We also explore the impact of CHTS geographic coverage on estimates of total state fishing effort, as well as the impact of stratification and sampling levels on CHTS estimates.

Table 3. Recreational fishing effort estimates by state, from the mail survey and CHTS, for coastal residents and waves 4, 5 and 6, 2013.

State	Mode and Method of Data Collection	Percent of Households Fishing In Wave	Mean Number of Anglers per Fishing Household	Total Trips	Mean Number of Trips per Household	Total Trips by Shore	Mean Trips by Shore per Household	Total Trips by Boat	Mean Trips by Boat per Household
Florida	ABS	16.31	1.78	39,244	11.3	25,973	9.05	13,271	7.29
	CHTS	6.22	1.78	9,730	9.01	4,042	8.06	5,688	8.17
Massachusetts	ABS	9.2	1.60	5,152	10.27	3,090	8.3	2,062	7.34
	CHTS	3.18	1.56	1,403	9.49	525	8.08	879	9.16
New York	ABS	7.9	1.70	11,784	11.24	6,807	8.99	4,977	9.38
	CHTS	2.4	1.58	2,319	9.66	1,131	9.54	1,188	8.24
North Carolina	ABS	14.48	1.57	6,903	11.38	4,555	9.13	2,348	7.1
	CHTS	6.73	1.78	2,058	8.34	944	6.58	1,114	8.4
All	ABS	12.77	1.73	63,082	11.21	40,425	8.98	22,658	7.65
	CHTS	4.8	1.74	15,510	9.04	6,642	8.03	8,868	8.29

5. Discussion

5.1 Sample Frames

The sample frame for the CHTS is comprised exclusively of landline telephone numbers. The NRC Review (2006) identified the increasing penetration of cell phones and subsequent abandonment of landline telephones as a potential source of bias in the CHTS. Since publication of the NRC report, landline use has continued to decline (Blumberg and Luke 2013). In contrast, the address frame used to sample residents of coastal states in the mail survey design includes all residential addresses serviced by the USPS, providing nearly complete coverage of U.S. households (Iannacchione 2011).

Based upon data collected through the mail survey, we estimate that 26.8% of coastal county households within the study states do not have landline telephone service (wireless-only households)[9] and are excluded from the CHTS sample frame. The percent of wireless households ranged from approximately 20% in MA and NY to approximately 31% in FL and NC. Non-coverage of wireless-only households will result in biased estimates of fishing activity if residents of wireless-only households fish more or less than residents of landline households.

Table 4 shows household fishing prevalence, estimated from mail survey data, by the type of telephone service. Overall, estimated fishing prevalence was 1.3 times higher for wireless-only households than landline households (15.2% vs. 11.9%). Higher fishing prevalence for wireless-only households is consistent, though not necessarily significant, among all states included in the study. These results demonstrate that non-coverage of wireless-only households from the CHTS sample frame is a source of bias resulting in an underestimate of fishing prevalence and total fishing effort.

[9] Addresses that could be matched to a telephone number by a commercial vendor were assumed to have landline telephone service regardless of survey responses to questions about type of household telephone service. This may result in an under-estimate of wireless only households.

Table 4. Estimated fishing prevalence, by state and type of telephone service, from the mail survey, for coastal counties and waves 4, 5, and 6 of 2013.

State	Landline		Wireless Only		
	%	n	%	n	*p*-value
Florida	15.3	1,926	18.4	696	0.0669
Massachusetts	9	1,796	9.2	357	0.9372
New York	7.9	1,045	8.3	217	0.8645
North Carolina	13.4	1,703	16.9	529	0.0809
Overall	11.9	6,470	15.2	1,799	0.0024

Note: Landline includes households that reported having landline telephone service as well as households that could be matched by a commercial vendor to a telephone number, regardless of reported telephone service. Significance based upon the results of a logistic regression model predicting the effect of type of household telephone service on reported fishing activity.

The impact of non-coverage bias in the CHTS is consistent with the direction of observed differences between CHTS and mail survey estimates of prevalence. However, non-coverage of wireless-only households in the CHTS can explain only a portion of the difference. Table 5 compares fishing prevalence for the full address sample within coastal counties, the portion of the address sample that either reported having a landline telephone or could be matched to a landline telephone number – i.e., households that would also be covered by the CHTS, and the CHTS. Comparisons between the full address sample and the "covered" address sample demonstrate the impact on survey estimates of non-coverage bias resulting from the exclusion of wireless-only households – estimated prevalence is approximately 8% higher for the full sample than the "covered" sample. Comparisons between CHTS estimates and the "covered" address sample, which coincide with the same population – households with landline telephone service, demonstrate that mail survey estimates of fishing prevalence are still 2.5 times larger than CHTS estimates (11.9% vs. 4.8%). Residual differences after accounting for non-coverage bias must be attributed to other sources of survey error.

Table 5. Estimated fishing prevalence, by state, from the full mail survey sample, the portion of the mail survey sample that would also be covered by the CHTS (households with landline telephones), and the CHTS, for coastal counties and waves 4, 5, and 6, 2013.

	Full Address Sample		"Covered" Addresses		CHTS	
state	%	SE	%	SE	%	SE
Florida	16.3	0.7	15.3	0.8	6.2	0.2
Massachusetts	9.2	0.7	9	0.8	3.2	0.7
New York	7.9	0.8	7.9	0.9	2.4	0.8
North Carolina	14.5	0.9	13.4	0.9	6.7	0.5
All	12.8	0.4	11.9	0.5	4.8	0.2

5.2 Survey Mode

The choice of survey mode can have different and sometimes substantial impacts on survey estimates. We use mode as a term to cover a diverse set of effects associated with the data collection such as differences in questionnaires and context. Dillman et al. (2009) and de Leeuw (2005) suggest that different data collection modes can result in very different responses to survey questions, particularly when comparing visual vs. aural or interviewer-administered vs. self-administered modes. The amount of time available to provide a response, visual or aural memory cues, and respondent interpretation of survey questions can all contribute to differential measurement between survey modes.

For residents of coastal counties, the largest differences between CHTS and mail survey estimates were for fishing prevalence. This finding is consistent with results from previous studies that measured higher fishing prevalence in mail surveys than telephone surveys (Brick et al. 2012a; Andrews et al. 2013). These studies suggested that differences in screening approaches between telephone and mail survey designs contributed to the observed differences in prevalence. Specifically, differences are partially attributed to a “gatekeeper effect”, where the initial respondent to a household telephone interview, who is asked a series of screener questions to determine if anyone in the household fished during the reference wave, may give inaccurate responses. The gatekeeper hypothesis is based upon the observations that the initial household respondent to the CHTS interview is overwhelmingly female, and households in which a female is the initial respondent are much less likely to report fishing than households in which a male is the initial respondent[10]. This hypothesis suggests a systematic bias in under-reporting of prevalence.

Andrews (unpublished) documented a gatekeeper effect in a telephone survey experiment, where the odds that a household reported fishing during the wave were 37% higher when household-level fishing questions were administered specifically to the sampled angler than when they were administered to the person who initially answered the phone (39.7% prevalence vs. 32.5%)[11]. The magnitude of the effect was likely minimized by the fact that the sample frame used for the study included cell phone numbers, which increased the likelihood that the person who initially answered the phone was also the sampled angler. The impact of the gatekeeper effect may be much larger in a RDD landline telephone survey such as the CHTS. Regardless of the magnitude, a gatekeeper effect in the CHTS is likely to result in underestimates of fishing prevalence, and consequently total fishing effort. The direction of the difference is consistent with the direction of differences between CHTS and mail survey estimates. While not tested, we assume that a gatekeeper effect is less problematic for household mail surveys, where the household has more time to consider the survey request, determine who should respond to the survey, and consult personal records or discuss the survey with other members of the household.

The gatekeeper effect may result from the tasks imposed upon the CHTS respondent. For example, the CHTS contacts households without prior notice, and the initial household respondent is expected to describe household-level fishing activity immediately, without the benefit of memory cues. This may result in cursory cognitive processing and failure to recall past events, particularly if those events are not especially memorable (de Leeuw 2005). As described, the recall error results from the nature of the CHTS interview and should produce under-reporting of household fishing activity at the screener stage. This hypothesis also suggests that the impact of recall error should be more pronounced for shore fishing, which is presumably less memorable than private boat fishing (Andrews et al. 2013). Lower salience of shore fishing could impact reporting at both the screener phase – e.g., households with only shore anglers may

[10] For example, during a recent CHTS wave, 62% of initial respondents were female, of which 3.3% reported household fishing during the wave. In contrast, 10.9% of male respondents reported household fishing activity.
[11] Estimated odds ratio of 1.37 (1.167,1.609) resulting from logistic regression model predicting the effect of screener respondent on reported fishing activity.

be more susceptible to the gatekeeper effect – and the topical phase – e.g., active anglers may be more likely to recall and describe boat fishing trips than shore fishing trips.

The impact of recall error and under-reporting of shore fishing trips at the topical phase may be exacerbated by the nature of the CHTS interview. Specifically, the CHTS interview consists of a series of household-level screening questions to identify fishing households, followed by individual interviews with each active angler to first estimate the total number of fishing trips taken by each angler and then sequentially characterize each individual fishing trip. In an experiment to assess recall error in the CHTS, Mathiowetz and Andrews (paper read at the Annual Meeting for the American Fisheries Society, 2011) observed that anglers provided details, including fishing mode, for fewer than 60% of reported trips, and that the percentage of estimated trips that are profiled decreases dramatically as the number of trips increases[12,13,14]. Given the financial and time commitments required for boat fishing, we hypothesize that anglers are more likely to recall and report details for boat fishing trips, resulting in under-representation of shore fishing activity in the CHTS data relative to boat fishing. This hypothesis is supported by the fact that differences between mail and CHTS estimates are considerably larger for shore fishing than private boat fishing (Table 3).

An alternative explanation for differential measurement between the CHTS and mail survey may be related to respondent interpretation and understanding of survey questions. Cognitive interviewing initiated prior to implementation of the mail survey demonstrated that anglers were very eager to provide information about fishing activity, even when that information was inconsistent with the questions being asked. For example, participants in cognitive interviews described fishing activity that occurred prior to the reference wave, outside of the reference state or in freshwater. The questionnaire was designed and modified to minimize reporting of out-of-scope fishing activity, and follow-up testing of different questionnaire versions suggests that these modifications were at least partially successful[15]. However, it is likely that some residual

[12] Reasons for incomplete trip profiling include mid-interview refusals, an inability to remember trip details, and volunteered reports that all trips are the same.
[13] 93% of reported trips were profiled for anglers who initially reported a single trip, while only 47% of trips were profiled for anglers who reported 5 trips.
[14] The CHTS compensates for incomplete trip information through a hot deck imputation process in which trip details for missing trips are imputed from a donor dataset comprised of complete trip records.
[15] mail survey estimates of fishing prevalence were lower in questionnaire versions that highlighted the scope of the survey request and/or provided space for respondents to document trips that were prior to the reference wave.

reporting error continues. This type of reporting error may be less likely in the CHTS, where the interviewer can confirm trip details. Reporting error resulting from misinterpretation of mail survey questions may contribute to differences between CHTS and mail survey estimates. A follow-up study, in which mail survey respondents will be re-interviewed via telephone, will be implemented during the spring of 2014 to assess the level that reported information is within the scope of the survey.

5.3 Nonresponse

In addition to impacting measurement, different survey modes may result in very different response rates. For example, mail survey response rates in the present study were nearly 3 times higher than CHTS response rates. While nonresponse rate is a poor predictor of nonresponse bias (Groves 2006), a higher nonresponse rate increases the risk for nonresponse bias. Consequently, the risk of nonresponse bias is higher in the CHTS than the mail survey design.

Nonresponse will result in bias if respondents and nonrespondents are different with respect to what is being measured. Previous mail surveys of anglers (Andrews et al. 2010, 2013; Brick et al. 2012a) have demonstrated that households with licensed anglers are both more likely to respond to a mail survey about fishing and more likely to report fishing activity during the reference period than households without licensed anglers. We observed similar results in the present study. Failure to account for this differential response between households with and without licensed anglers will result in nonresponse bias. By matching address samples to state license databases in the mail survey design, we effectively stratify the sample into sub-populations that are more similar with respect to fishing activity and response propensity than the sample as a whole. This formation of strata mitigates the impact of differential response between the two groups. Consequently, any nonresponse bias in the mail survey design will be residual after accounting for the population of licensed anglers.

The CHTS is also susceptible to nonresponse bias resulting from differential response between anglers and non-anglers. W.R. Andrews (paper read at the Annual Meeting for the American Fisheries Society, 2011) demonstrated that differential response between households with and without anglers resulted in an overestimation of fishing effort by as much as 17% in the CHTS.

Unlike the mail survey design, the CHTS does not account for differential response between sub-populations, resulting in nonresponse bias. However, the bias does not explain differences between CHTS and mail survey estimates, as it results in an overestimate of fishing effort in the CHTS.

We attempted to assess nonresponse bias in the mail survey design by conducting a nonresponse follow-up study. Each wave, a sample of 320 nonresponding addresses[16] was randomly selected and mailed a follow-up questionnaire[17]. The survey mailing, which resulted in a response rate of approximately 40%, was delivered via FedEX and included a $5.00 cash incentive. Table 6 compares fishing prevalence for the initial address samples and the follow-up study samples. Overall, estimates of fishing prevalence for the initial sample are approximately 1.1 times larger than estimates from the nonresponse sample (13.9% vs. 12.7%)[18]. There are no systematic differences between initial sample estimates and nonresponse sample estimates among states. Based upon these results, we have no evidence to suggest that nonresponse in the mail survey design results in nonresponse bias.

The combined mail survey response rate, including both the initial sample and the nonresponse follow-up sample, is approximately 64% (40% for the initial sample and 40% for the nonresponse follow-up sample). While we have not observed nonresponse bias in the mail survey, we can estimate the maximum possible nonresponse bias if we assume that all nonrespondents are non-anglers. In this scenario, the estimated prevalence is 7.76%, which corresponds to a maximum bias of approximately 5 percentage points (12.77% vs. 7.76%). This is not trivial (approximate 40% relative difference) considering the relatively low overall magnitude of fishing prevalence. However, even in this extreme case, the estimated prevalence for the mail survey is still 1.6 times larger than the CHTS estimate (7.76% vs. 4.8%), which suggests that factors other than nonresponse bias must contribute to the differences between CHTS and mail survey estimates.

[16] Nonresponse samples were distributed equally among states (80 addresses per state and wave).
[17] The questionnaire used for the nonresponse study was identical to questionnaire included in the initial mailings.
[18] The Full Sample estimates are the fully weighted estimates used in the rest of this section. The Nonresponse sample estimates are based on weights that account for the original sampling and for subsampling for the nonresponse bias study.

Table 6. Estimated fishing prevalence for the full mail survey sample and the nonresponse follow-up sample by state.

	Full Sample		Nonresponse Sample	
State	%	n	%	n
Florida	21.5	11,767	18.4	203
Massachusetts	11.0	11,094	13.2	216
New York	8.6	8,479	9.2	172
North Carolina	11.4	13,570	9.8	248
All	13.9	49,910	12.7	839

Note: Estimates are based upon data collected from 7 waves (wave 5, 2012-wave 5, 2013) and include information collected through multiple versions of the survey instrument. Consequently, estimates may differ from those reported elsewhere in the report.

Based upon the results of this and previous studies, we suspect that differential bias resulting from measurement errors contributes significantly to the observed differences between CHTS and mail survey estimates. While nonresponse is a concern, particularly for the CHTS, we do not believe that bias resulting from nonresponse contributes to the observed differences in estimates between survey designs.

5.4 Stratification and Sample Size

The previous sections explored potential impacts on survey estimates of non-sampling errors – coverage error, measurement error and nonresponse error - resulting from survey design features. We also considered the extent to which sample design and estimation strategies may impact survey estimates.

Within each coastal state, the CHTS is stratified by county, and the sample is allocated among counties in proportion to the square root of the number of occupied housing units within each county. While this strategy assures that sample is distributed among all coastal counties within a state, it also results in small sample sizes in some counties during some survey waves. Because recreational saltwater fishing is a relatively rare occurrence among the general population (<10%), small sample sizes can result in situations in which the likelihood of contacting at least one fishing household is extremely small. This is especially true during off-peak waves when fishing activity is particularly low (<1-2%). Because CHTS estimates are produced at the stratum level (i.e., county) and then aggregated to state estimates, we hypothesized that low sample sizes in the CHTS during low-activity waves result in a systematic underestimate of state-level fishing effort.

We tested this hypothesis by comparing base CHTS estimates to independent estimates derived from the CHTS methodology but with much larger samples in New York and North Carolina during wave 6, 2013, an historically low-activity fishing period. Table 7 provides results for the base and experimental CHTS samples. Overall, base sampling levels resulted in 10 counties with no reported fishing activity, while only a single county was classified as non-fishing at the larger, experimental sample sizes. Similarly, the experimental estimate of fishing prevalence was 13.6% larger than the base estimate, and experimental estimates were more than 10% greater than base estimates in both New York and North Carolina. While differences in estimated prevalence between base and experimental sample sizes are not significant, they are in the direction that supports the hypothesis as well as the suggestion that differences between mail survey and CHTS estimates may be partially the result of insufficient sampling levels to support the stratification and estimation design of the CHTS.

The CHTS estimation design – stratified random sampling with separate ratio estimates – is unbiased when sample sizes in each stratum are large (Cochran 1953). However, in practice, sample sizes in some strata may be insufficient to produce unbiased state-level estimates of fishing activity. A combined ratio estimate may be more appropriate when stratum sample sizes are small. In addition, county-level stratification and low fishing prevalence result in very high probabilities of not encountering a single fishing household at the sample sizes allocated to some counties. For rare populations, such as fishing households, relatively small random samples are likely to result in a distribution of estimates that is highly skewed with zero occurrences of the rare event (Christman 2009).

.

Table 7. Comparison of survey results between base CHTS and experimental sampling levels by state for wave 6, 2013.

	Base Sample				Experimental Sample			
State	n	Avg. County n	No Fish County	Prevalence (%)	n	Avg. County n	No Fish County	Prevalence (%)
New York	920	92	4	1.26	4,299	430.1	0	1.45
North Carolina	1,578	43.9	6	5.86	3,994	111	1	6.52
All	2,498	68	10	1.98	8,293	270.6	1	2.25

Note: No Fish County is the number of counties in which no fishing households were contacted, and Avg. County n is the average sample size per county.

5.5 Geographic Coverage

Geographic coverage of the CHTS is limited to counties that are within a specified distance of the coast. This is done to maximize interviews with anglers and minimize data collection costs, as fishing activity is generally assumed to be more common for residents of coastal counties than noncoastal counties. To account for geographic non-coverage, CHTS estimates of coastal resident fishing effort are expanded by correction factors derived from the Access-Point Angler Intercept Survey (APAIS), an onsite survey of completed recreational fishing trips conducted at publicly accessible fishing or access sites (e.g. fishing piers, beach access sites, boat ramps, marinas, etc.)[19]. These correction factors attempt to account for fishing trips taken by residents of non-coastal counties within coastal states, as well as residents of non-coastal states. The correction factor has its own problems, especially since the sample frame for the APAIS is limited to publicly accessible sites. Estimates derived from the APAIS, including the residency correction factors, are susceptible to bias resulting from non-coverage of fishing trips taken at or returning to non-public sites.

In contrast to the CHTS, the sample frame for the mail survey includes all residential addresses within coastal states, so we assume that non-coverage bias for residents of coastal states is minimal (Iannacchione 2011). Non-resident anglers in the mail survey design are sampled exclusively from state database of licensed saltwater anglers, which are potentially susceptible to non-coverage resulting from license exemptions and unlicensed fishing activity among non-resident anglers.

[19] Within each state, CHTS estimates are expanded by the ratio of completed angler intercepts to completed angler intercepts with residents of coastal counties.

We examined the impact of geographic exclusions to the CHTS by comparing APAIS correction factors to analogous estimates derived from the mail survey (Table 8). Overall, the APAIS estimates that 76% of saltwater fishing trips in the study states are taken by residents of coastal counties who are "covered" by the CHTS, resulting in a correction factor ("Coastal ratio") of 1.3 (1.0/0.76). In contrast to the APAIS, the mail survey estimates that 88% of saltwater fishing trips in the study states are taken by residents of coastal counties, which corresponds to a coastal ratio of 1.14. Differences between APAIS and mail survey estimates in the relative distribution of effort by residency are highly variable among states – APAIS coverage correction factors are larger than analogous mail survey estimates in Florida and Massachusetts and smaller in New York and North Carolina.

Table 8. Percent of total saltwater fishing trips by residents of coastal counties and the ratio of total effort, including coastal, non-coastal and non-resident anglers, to coastal resident effort.

	Single Phase Mail			APAIS		
State	n	% Coastal	Coastal Ratio	n	% Coastal	Coastal Ratio
Florida	2,829	96.5	1.04	9,759	76.9	1.30
Massachusetts	2,684	87.5	1.14	3,203	75.6	1.32
New York	2,146	83.4	1.20	1,494	95.7	1.04
North Carolina	3,058	61.4	1.63	8,260	62.5	1.60
All	10,717	87.7	1.14	22,716	76.3	1.31

Note: Sample sizes reflect the combined number of completed surveys for the sample of resident addresses and the sample of non-resident licensees. The coastal ratio is the ratio of total angler trips to trips by residents of coastal counties. Coastal ratios derived from the APAIS are used to expand CHTS estimates to account for trips by non-resident anglers and residents of non-coastal states.

We further examine differences between APAIS correction factors and mail survey estimates separately for residents and non-residents of coastal states. Of saltwater fishing trips taken by residents of coastal states, the mail survey estimates that approximately 78% are by residents of coastal counties (Table 9). Among states, effort by coastal county residents varies from 64.3% in NC to 89.2% in MA. In contrast, the APAIS estimates that nearly 90% of trips are taken by coastal county residents; coastal residents accounted for 91.2%, 99.3% and 78.9% of total resident effort for MA, NY and NC, respectively. Assuming that other potential sources of bias in the mail design are uniform between coastal and non-coastal residences, these results suggest that the APAIS underestimates fishing activity by residents of non-coastal counties. This would result in an underestimate of total fishing effort. The magnitude of bias varies by state; APAIS samples provide a reasonable representation of anglers in MA along this dimension, but under-represent non-coastal residents in NY and NC.

The mechanism for this bias is not intuitive. Because the APAIS sample frame is limited to publicly accessible fishing sites, one may expect the sample to over-represent trips by residents of non-coastal counties, whose primary access to saltwater fishing is from public-access sites[20]. An alternative explanation for the difference between the mail survey and APAIS in the distribution of resident fishing effort is that mail survey respondents may be including in their counts fishing activities that are outside the scope of the survey, such as freshwater fishing. The distinction between saltwater and freshwater fishing can be subtle, particularly in inland water bodies such as estuaries and the brackish portions of rivers. New York provides an example of how difficult it can be to distinguish between fresh and saltwater fishing. New York anglers are required to register as saltwater anglers if fishing for saltwater species in marine or coastal regions of the state or for "migratory fish of the sea" in the tidal Hudson River and its tributaries[21] (http://www.dec.ny.gov/permits/54950.html). The tidal portion of the Hudson River extends to north of Albany, which is more than 100 miles beyond the most upstream fishing site on the APAIS sample frame. While fishing on much of the Hudson River does not qualify as saltwater fishing by the APAIS definition, anglers who fish on the Hudson River may report these trips a saltwater because they are required to register as saltwater anglers and they're fishing for saltwater species[22]. The reporting of fishing activities on water bodies such as the Hudson River, which extends well into the noncoastal portion of the state, could skew the distribution of effort toward noncoastal residents and explain differences between the mail survey and APAIS in the distribution of effort among types of residence.

[20] In contrast to coastal county residents who may have direct access to saltwater fishing via personal or community beaches, docks and/or boat slips that are inaccessible to APAIS interviewers.
[21] New York does not have a saltwater fishing license but does require saltwater anglers to enroll in a free registry.
[22] Anecdotal evidence collected during follow-up telephone interviews suggests that some anglers distinguish between salt and freshwater fishing based upon the species targeted, not the geographic location.

Table 9. Percent of total resident fishing trips by residents of coastal counties, estimated by the mail survey and the APAIS[23].

	Single Phase Mail		APAIS	
State	% Coastal	n	% Coastal	n
Massachusetts	89.2	2,629	91.2	3,203
New York	83.9	1,973	99.3	1,494
North Carolina	64.3	2,876	78.9	8,260
All	78.0	7,966	89.2	12,957

Table 10 shows the estimated percentage of total trips taken by non-resident anglers for the mail survey and the APAIS. Overall, the APAIS estimates that 19.8% of fishing trips in the study states are taken by non-resident anglers. In contrast, the mail design estimates that only 2.9% of trips are by non-resident anglers. These results suggest that either the license frames used to sample non-resident anglers are incomplete (i.e. many non-resident anglers fish without a license), or APAIS samples over-represent non-resident anglers. Both explanations are plausible, if not likely. For example, previous studies (Brick et al. 2012; Andrews et al. 2013) suggested that state license databases are incomplete as the result of license exclusions and illegal fishing activity. It is not clear if these omissions are as serious for non-residents and they are for residents. These sources of non-coverage will result in underestimation of total fishing effort for non-resident anglers. Similarly, the APAIS sample frame excludes private residences (e.g., private docks and boat slips, private marinas, etc.), which are likely to have a much higher proportion of resident anglers than public-access fishing sites. Over-representation of non-resident anglers in the APAIS will result in over-estimation of total fishing effort.

[23] Florida is excluded from the table because all counties are considered coastal and are included in the coverage of the CHTS.

Table 10. Percent of total fishing trips by non-resident anglers, estimated by the mail survey and the APAIS.

	Single Phase Mail		APAIS	
State	% Non-resident	n	% Non-resident	n
Florida	3.5	2,829	23.1	9,759
Massachusetts	1.9	2,684	17.1	3,203
New York	0.7	2,146	3.6	1,494
North Carolina	4.5	3,058	20.8	8,260
All	2.9	10,717	19.8	22,716

The consequences of limiting the CHTS to coastal counties are still somewhat unclear. We expect non-sampling errors in the mail design to be relatively uniform between coastal and non-coastal residences within a state, suggesting that estimates of the distribution of effort between coastal and non-coastal residents are unbiased. This implies that APAIS samples over-represent trips by coastal resident anglers, resulting in under-estimates of fishing effort. The impact of non-resident angling is less clear as both APAIS and mail survey estimates are susceptible to non-coverage bias – non-coverage of private access fishing sites in the APAIS and unlicensed anglers in the mail survey. Regardless of the source of differences, the APAIS attributes a larger proportion of total effort to non-resident anglers, resulting in larger correction factors and larger estimates of total fishing effort. The overall net differences between the APAIS and mail survey in the estimated distribution of effort by residency are variable among states, likely reflecting differences in the coverage of both state license databases and APAIS sample frames.

6. Conclusions and Recommendations

The mail survey design tested in this study is a feasible alternative to the CHTS and has numerous substantive advantages over the CHTS design. Overall, response rates for the mail survey were 2-3 times higher than the CHTS, and the design produced stable preliminary estimates within the current data collection and estimation schedule for the CHTS. Furthermore, matching household address samples to state license databases and over-sampling matched households effectively increased the likelihood of contacting fishing households.

In terms of survey error, we conclude that the mail survey design is less susceptible than the CHTS to bias resulting from nonresponse and non-coverage. We also found that the nature of the mail survey mode results in more accurate responses to questions about fishing activity than the CHTS, which expects respondents to answer questions on-the-spot, without the benefit of aided recall or memory cues. Furthermore, we have demonstrated that insufficient sampling in the CHTS in conjunction with the estimation scheme creates a functional bias that results in underestimates fishing activity. Table 11 summarizes sources of survey error, as well as the observed and/or hypothesized impact of bias on survey estimates for the CHTS and mail survey design.

Table 11. Summary of sources of error in the CHTS and mail survey designs.

Error Source	Direction of Bias Mail	CHTS	Comment
Non-Coverage	NA	↓	Results from mail survey demonstrate that residents of wireless-only households are more likely to fish than residents of landline households.
Nonresponse	↑	↑	Based upon response rates, the risk for nonresponse bias is greater in the CHTS than the mail survey. Differential response between households with and without licensed anglers is mitigated in mail survey by treating populations as separate strata - there is no such adjustment in the CHTS. A nonresponse follow-up study did not identify nonresponse bias in the mail survey design. However, any nonresponse bias in the mail survey design is likely to result in an over-estimate of fishing effort.
Measurement	↑	↓	A "gatekeeper effect", resulting in under-reporting of household fishing activity, has been documented in telephone surveys of licensed anglers. We suggest that this source of measurement bias is greater in landline RDD telephone surveys. We also suggest that the mail mode facilitates recall of past fishing activity. The lack of interviewers in the mail survey may result in reports of fishing activity that are beyond the intended scope of the mail survey.
Sample Size	NA	↓	County-level stratification in the CHTS results in insufficient sample size to detect fishing activity in some strata during low-activity waves. This source of error would also impact the mail survey at small sample sizes.

In addition to direct comparisons between the CHTS and the mail survey in the geographic regions where the survey overlapped, we also explored the impact of geographic limitations of the CHTS on total effort estimates and determined that coverage correction factors, derived from the APAIS, are likely biased due to the exclusion of private access fishing sites from APAIS sample frames. Comparisons between the APAIS and mail survey of the distribution of effort between coastal and non-coastal resident anglers suggest that the APAIS sample over-represents trips by coastal resident anglers, which would result in under-estimates of total resident fishing effort. Comparisons between the two designs of the magnitude of non-resident angling are less clear and confounded by potential coverage bias in the mail survey resulting from unlicensed fishing activity by non-resident anglers.

Given the potential for bias in the CHTS, we conclude that the mail survey design is a superior approach for monitoring recreational fishing effort. Other designs, including dual-frame telephone surveys that sample from both landline and cell phone frames, were also considered as alternatives to the CHTS. However, these designs were not tested due to the expected low response rates, prohibitive costs, and the need to target anglers within specific geographic regions (AAPOR Cell Phone Task Force 2010).

The mail survey design described above also improved upon weaknesses identified in previous tests of mail surveys. For example, the response rate for the new design was considerably higher than previous mail surveys largely because it eliminated the screening mail instrument. The new design also eliminated the potential bias due to matching errors in the earlier dual-frame designs.

We believe the results reported here demonstrate the utility of the mail survey design. Nonetheless, we encourage continued development and testing. For example, additional questionnaire testing and varying the length of the reference period (e.g., one-month waves) could provide additional assessments of measurement errors. Similarly, testing alternative data collection modes, such as email and web surveys, could improve response rates and potentially provide cost savings. These types of evaluations will help ensure that advancements in survey methodology are considered and customer needs are satisfied.

7. References

AAPOR Cell Phone Task Force (The American Association for Public Opinion Research). 2010. New considerations for survey researchers when planning and conducting RDD telephone surveys in the U.S. with respondents reached via cell phone numbers. Available: http://www.aapor.org/Cell_Phone_Task_Force_Report.htm#.U07cS_ldV8E.

Andrews, W.R., J.M. Brick, N.A. Mathiowetz, and L. Stokes. 2010. Pilot test of a dual frame two-phase mail survey of anglers in North Carolina. Available: https://www.st.nmfs.noaa.gov/mdms/public/finalReport.jsp?ReportID=355.

Andrews, W.R, J. Foster, J.M. Brick, N.A Mathiowetz, and L. Stokes. 2011. Nonresponse weighting adjustment in a telephone survey of recreational fishing activity, presentation at Annual Meeting for the American Fisheries Society in Seattle, Washington.

Andrews, W.R., J.M. Brick, and N.A. Mathiowetz. 2013. Continued development and testing of dual-frame surveys of fishing effort: testing a dual-frame, mixed-mode survey design. Available: https://www.st.nmfs.noaa.gov/mdms/public/finalReport.jsp?ReportID=831.

Blumberg, S. J. and J.V. Luke. 2013. Wireless substitution: Early release of estimates from the National Health Interview Survey, July–December 2012, National Center for Health Statistics. Available: www.cdc.gov/nchs/nhis.htm.

Brick. J.M., W.R. Andrews, and N.A. Mathiowetz. 2012a. A Comparison of recreational fishing effort survey designs. Available: https://www.st.nmfs.noaa.gov/mdms/doc/08A_Comparison_of_Fishing_Effort_Surveys_Report_FINAL.pdf.

Brick, J.M., W.R. Andrews, P.D. Brick, H. King, N.A. Mathiowetz, and L. Stokes. 2012b. Methods for improving response rates in two-phase mail surveys. Survey Practice 5(3). Available: http://surveypractice.org/index.php/SurveyPractice/article/view/17/pdf.

Christman, M.S. 2009. Sampling of rare populations. Pages 109-124 *in* D. Pfeffermann, editor. Handbook of statistics: sample surveys design, methods and applications, volume 29A. Elsevier, Amsterdam.

Cochran, W.G. 1953. Sampling techniques. Wiley and Sons, New York.

Curtin, R., S. Presser, and E. Singer. 2005. Changes in telephone survey nonresponse over the past quarter century. Public Opinion Quarterly 69(1): 87-98.

deLeeuw, E.D. 2005. To mix or not to mix data collection modes in surveys. Journal of Official Statistics 21: 233-255.

Dillman, D.A., J.D. Smyth, and L.M. Christian. 2009. Internet, mail, and mixed-mode surveys: The tailored design method. Wiley and Sons, New York.

Groves, R.M. 2006. Nonresponse rates and nonresponse bias in household surveys. Public Opinion Quarterly 70(5): 646-675.

Iannacchione, V. 2011 . The changing role of address-based sampling in survey research. Public Opinion Quarterly 7: 556-575.

Lai, H-L. and R. Andrews. 2008. Development of a dual-frame methodology for estimating marine recreational fishing effort. ICES Conference and Meeting Documents 2008 K:06.

Lohr, S. 2009. Multiple frame surveys. Pages 71-88 *in* D. Pfeffermann, editor. Handbook of statistics: sample surveys design, methods and applications, volume 29A. Elsevier, Amsterdam.

Mathiowetz, N.A and W.R. Andrews. 2011. The effects of length of recall on reports of fishing effort, presentation at Annual Meeting for the American Fisheries Society in Seattle, Washington.

NRC (National Research Council). 2006 . Review of recreational fisheries survey methods. National Academies Press, Washington, D.C.

Please complete for all members of your household. Include those who fish and those who do not fish.

Household Member 4

11. What is this person's gender?
   - ☐ Male
   - ☐ Female

12. How old is this person?
   If less than 1 year, mark 0 years.
   ☐☐ Age in years

13. Is this person of Hispanic, Latino, or Spanish origin?
   - ☐ Yes, of Hispanic origin
   - ☐ No, not of Hispanic origin

14. What is this person's race?
   Mark one or more boxes.
   - ☐ White ☐ Black, African American
   - ☐ Asian ☐ American Indian or Alaska Native
   - ☐ Native Hawaiian or Other Pacific Islander

Please think only about recreational saltwater fishing in <Merged State>.

15a. How many days did this person go recreational saltwater fishing from the shore in <Merged State>? The shore includes docks, bridges, causeways, beaches, banks, or any other shore-based place or area.
   - ☐ No shore recreational saltwater fishing in last 12 months → Go to question 15b
   - ☐☐ Number of days in July and August of 2013
   - ☐☐ Total number of days in last 12 months

15b. How many days did this person go recreational saltwater fishing from a private or rental boat that returned to shore in <Merged State>? Do not include charter boats or commercial boats that have a captain or crew who help locate and catch fish.
   - ☐ No private boat recreational saltwater fishing in last 12 months → Go to the next Household Member
   - ☐☐ Number of days in July and August of 2013
   - ☐☐ Total number of days in last 12 months

Go to Household Member 5

Household Member 5

11. What is this person's gender?
   - ☐ Male
   - ☐ Female

12. How old is this person?
   If less than 1 year, mark 0 years.
   ☐☐ Age in years

13. Is this person of Hispanic, Latino, or Spanish origin?
   - ☐ Yes, of Hispanic origin
   - ☐ No, not of Hispanic origin

14. What is this person's race?
   Mark one or more boxes.
   - ☐ White ☐ Black, African American
   - ☐ Asian ☐ American Indian or Alaska Native
   - ☐ Native Hawaiian or Other Pacific Islander

Please think only about recreational saltwater fishing in <Merged State>.

15a. How many days did this person go recreational saltwater fishing from the shore in <Merged State>? The shore includes docks, bridges, causeways, beaches, banks, or any other shore-based place or area.
   - ☐ No shore recreational saltwater fishing in last 12 months → Go to question 15b
   - ☐☐ Number of days in July and August of 2013
   - ☐☐ Total number of days in last 12 months

15b. How many days did this person go recreational saltwater fishing from a private or rental boat that returned to shore in <Merged State>? Do not include charter boats or commercial boats that have a captain or crew who help locate and catch fish.
   - ☐ No private boat recreational saltwater fishing in last 12 months → Thank you!
   - ☐☐ Number of days in July and August of 2013
   - ☐☐ Total number of days in last 12 months

Thank you!

18989

OMB #0648-0652
Exp. Date 10/31/2015

<Merged State> Weather and Outdoor Activity Survey

Public reporting burden for this collection of information is estimated to average 10 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other suggestions for reducing this burden to Rob Andrews, NOAA Fisheries Service, 1315 East-West Hwy., Silver Spring, MD 20910.

This is a voluntary survey, and responses are kept confidential as required by section 402(b) of the Magnuson-Stevens Act and NOAA Administrative Order 216-100, Confidentiality of Fisheries Statistics, and will not be released for public use except in aggregate statistical form without identification as to its source. Notwithstanding any other provisions of the law, no person is required to respond to, nor shall any person be subjected to a penalty for failure to comply with, a collection of information subject to the requirements of the Paperwork Reduction Act, unless that collection of information displays a currently valid OMB Control Number.

<Merged State> Weather and Outdoor Activity Survey

START HERE

- Please use a black or blue pen to complete this form.
- Mark ☒ to indicate your answer.
- If you want to change your answer, darken the box ■ and mark the correct answer.

1. How do members of this household obtain information about the weather, including current weather conditions, forecasts, and warnings? *Mark all that apply.*
   - ☐ Television
   - ☐ Radio
   - ☐ Newspaper
   - ☐ Internet
   - ☐ Other
2. During the past 12 months, has anyone in this household had to evacuate or seek shelter due to a severe weather event such as a tornado, hurricane, or thunderstorm?
   ☐ Yes ☐ No
3. In your area, how often do the advanced warnings you get for severe weather events allow you enough time to prepare properly?
   - ☐ All the time
   - ☐ Some of the time
   - ☐ Rarely
   - ☐ Never
4. During the past 12 months, has anyone in this household visited a public beach, national seashore, coastal state park, or other coastal nature reserve or protected area?
   ☐ Yes ☐ No
5. During the past 12 months, has anyone in this household been backpacking, climbing, or hiking?
   ☐ Yes ☐ No
6. During the past 12 months, has anyone in this household been hunting?
   ☐ Yes ☐ No
7. Which category best describes the telephone service for you and members of the household?
   - ☐ Regular or Landline phone only
   - ☐ Cellular phone only
   - ☐ Both Landline and Cellular phone
   - ☐ No working phone service
8. Which of the following best describes this house, apartment, or mobile home?
   - ☐ Owned with a mortgage or loan
   - ☐ Owned (without a mortgage)
   - ☐ Rented
   - ☐ Occupied without payment or rent
9. How long have you lived at this address?
   - ☐ 1 year or less
   - ☐ Less than 5 years, more than 1 year
   - ☐ 5 years or more
10. How many people, including all adults and children, live in this household?
    ☐☐ → Please continue and complete the questions in the next section for EACH HOUSEHOLD MEMBER, whether or not that household member saltwater fishes.

S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

Please use the calendars to answer question 15.

S	M	T	W	T	F	S
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

Please complete for all members of your household. Include those who fish and those who do not fish.

Household Member 1

11. What is this person's gender?
    - ☐ Male
    - ☐ Female
12. How old is this person?
    If less than 1 year, mark 0 years.
    ☐☐ Age in years
13. Is this person of Hispanic, Latino, or Spanish origin?
    - ☐ Yes, of Hispanic origin
    - ☐ No, not of Hispanic origin
14. What is this person's race?
    Mark one or more boxes.
    - ☐ White ☐ Black, African American
    - ☐ Asian ☐ American Indian or Alaska Native
    - ☐ Native Hawaiian or Other Pacific Islander

Please think only about recreational saltwater fishing in <Merged State>.

15a. How many days did this person go recreational saltwater fishing from the shore in <u><Merged State></u>? The shore includes docks, bridges, causeways, beaches, banks, or any other shore-based place or area.

☐ No shore recreational saltwater fishing in last 12 months → Go to question 15b

☐☐ Number of days in July and August of 2013

☐☐ Total number of days in last 12 months

15b. How many days did this person go recreational saltwater fishing from a private or rental boat that returned to shore in <u><Merged State></u>? Do not include charter boats or commercial boats that have a captain or crew who help locate and catch fish.

☐ No private boat recreational saltwater fishing in last 12 months → Go to the next Household Member

☐☐ Number of days in July and August of 2013

☐☐ Total number of days in last 12 months

Go to Household Member 2

Household Member 2

11. What is this person's gender?
    - ☐ Male
    - ☐ Female
12. How old is this person?
    If less than 1 year, mark 0 years.
    ☐☐ Age in years
13. Is this person of Hispanic, Latino, or Spanish origin?
    - ☐ Yes, of Hispanic origin
    - ☐ No, not of Hispanic origin
14. What is this person's race?
    Mark one or more boxes.
    - ☐ White ☐ Black, African American
    - ☐ Asian ☐ American Indian or Alaska Native
    - ☐ Native Hawaiian or Other Pacific Islander

Please think only about recreational saltwater fishing in <Merged State>.

15a. How many days did this person go recreational saltwater fishing from the shore in <u><Merged State></u>? The shore includes docks, bridges, causeways, beaches, banks, or any other shore-based place or area.

☐ No shore recreational saltwater fishing in last 12 months → Go to question 15b

☐☐ Number of days in July and August of 2013

☐☐ Total number of days in last 12 months

15b. How many days did this person go recreational saltwater fishing from a private or rental boat that returned to shore in <u><Merged State></u>? Do not include charter boats or commercial boats that have a captain or crew who help locate and catch fish.

☐ No private boat recreational saltwater fishing in last 12 months → Go to the next Household Member

☐☐ Number of days in July and August of 2013

☐☐ Total number of days in last 12 months

Go to Household Member 3

Household Member 3

11. What is this person's gender?
    - ☐ Male
    - ☐ Female
12. How old is this person?
    If less than 1 year, mark 0 years.
    ☐☐ Age in years
13. Is this person of Hispanic, Latino, or Spanish origin?
    - ☐ Yes, of Hispanic origin
    - ☐ No, not of Hispanic origin
14. What is this person's race?
    Mark one or more boxes.
    - ☐ White ☐ Black, African American
    - ☐ Asian ☐ American Indian or Alaska Native
    - ☐ Native Hawaiian or Other Pacific Islander

Please think only about recreational saltwater fishing in <Merged State>.

15a. How many days did this person go recreational saltwater fishing from the shore in <u><Merged State></u>? The shore includes docks, bridges, causeways, beaches, banks, or any other shore-based place or area.

☐ No shore recreational saltwater fishing in last 12 months → Go to question 15b

☐☐ Number of days in July and August of 2013

☐☐ Total number of days in last 12 months

15b. How many days did this person go recreational saltwater fishing from a private or rental boat that returned to shore in <u><Merged State></u>? Do not include charter boats or commercial boats that have a captain or crew who help locate and catch fish.

☐ No private boat recreational saltwater fishing in last 12 months → Go to the next Household Member

☐☐ Number of days in July and August of 2013

☐☐ Total number of days in last 12 months

Go to Household Member 4 on the next page

Please use other side for additional household members

27138

A mail survey design was implemented in Massachusetts, New York, North Carolina and Florida in October, 2012 to test a revised data collection design for monitoring marine recreational fishing effort. The survey, which collects information for two-month reference waves, included two experiments during the first two study waves, wave 5 (Sept-Oct 2012) and wave 6 (Nov-Dec, 2012), to test different survey design features aimed at maximizing efficiency and minimizing nonresponse error. Specifically, the experiments tested two versions of the survey instrument and four levels of cash incentives. Details of the experiments are provided below.

Instrument Testing

The study included an experiment to test two versions of the survey instrument. The objective of the experiment was to identify the instrument that maximized overall response rates while minimizing the potential for nonresponse bias resulting from differential nonresponse between anglers and non-anglers. One version of the instrument (Saltwater Fishing Survey) utilized a "screen out" approach that quickly identifies anglers (and non-anglers) and encourages participation by minimizing the number of survey questions, particularly for non-anglers. Person-level information, including details about recent fishing activity and limited demographic information, is collected for all household residents, but only if someone in the household reported fishing during the reference wave. The second version (Weather and Outdoor Activity Survey) utilized an "engaging" approach that encourages response by broadening the scope of the questions to include both fishing and non-fishing questions. This version collects person-level information for all residents of sampled households, regardless of whether or not household residents participated in saltwater fishing. Each wave, sampled addresses were randomly assigned to one of the two questionnaire types, which were evaluated in terms of response rates and reported fishing activity.

Table 1 provides the weighted response rates (AAPOR RR1 after excluding undeliverable addresses) and estimated fishing prevalence (percentage of households with residents who reported fishing during the wave) for the two versions of the instrument. Overall, the Weather and Outdoor Activity Survey achieved a significantly higher response rate than the Saltwater Fishing Survey, and there was no significant difference between instruments in estimated prevalence. The lack of a significant difference between instruments for estimated prevalence suggests that the gain in response for the engaging instrument cannot be attributed to increased survey participation by either anglers or non-anglers, but that both groups are more likely to respond to the Weather and Outdoor Activity Survey than the Saltwater Fishing Survey.

We also compared response rates and prevalence between instruments both among and within subpopulations defined by whether or not sampled addresses could be matched to state databases of licensed saltwater anglers – subpopulations expected to distinguish between households with anglers and households with no anglers or less avid anglers. As expected, both response rates and estimated prevalence were higher in the matched subpopulation than the unmatched subpopulation, confirming that a population expected to be interested in the survey topic - households with licensed anglers - is more likely to respond to a fishing survey and report fishing activity than a population that excludes licensed anglers[1]. Because we can identify household license status prior to data collection, we can account for differential nonresponse between matched and unmatched households in the estimation design by treating matched an unmatched domains as strata (Lohr 2009).

[1] The classification of sample into domains is dependent upon matching ABS sample to license databases by address and telephone number. This process is unlikely to be 100% accurate, so the unmatched domain is likely to include some households with licensed anglers. The unmatched domain also includes households with residents who fish without a license.

Table 1. Weighted response rates and estimated prevalence overall and by domain for two versions of the survey instrument.

	Saltwater Fishing Survey		Weather and Outdoor Activity Survey	
	(%)	(n)	(%)	(n)
Response Rate				
Overall	31.1 (0.4)	17,511	34.7 (0.4)*	17,510
Matched	45.4 (1.1)	3,160	45.0 (1.0)	3,247
Unmatched	30.3 (0.4)	14,351	34.0 (0.5)*	14,263
Prevalence				
Overall	13.4 (0.5)	5,943	14.1 (0.5)	6,498
Matched	49.9 (1.7)	1,491	48.5 (1.6)	1,552
Unmatched	11.2 (0.6)	4,452	12.2 (0.6)	4,946

Notes: (1) standard errors are in parentheses. (2) Domains are defined by matching ABS samples to state databases of licensed saltwater anglers.
*Significantly different from Saltwater Fishing Survey ($p<0.05$).

There were no significant differences between instruments for either response rate or prevalence within the matched domain, suggesting that the inclusion of non-fishing questions in the Weather and Outdoor Activity Survey did not have an impact on response by either anglers or non-anglers. In the unmatched domain, the response rate was significantly higher for the Weather and Outdoor Activity Survey than the Saltwater Fishing Survey. However, the higher response rate did not translate to lower or higher estimates of prevalence; estimates of prevalence were not significantly different between instruments within the domain. This suggests that the engaging instrument uniformly increased the probability of response for anglers and non-anglers within the unmatched domain.

Differential nonresponse to a survey request between subpopulations will result in nonresponse bias if the subpopulations are different with respect to the survey topic. In the tested design, we account for differential nonresponse between matched and unmatched households during sampling – matched and unmatched subpopulations are treated as independent strata. Consequently, the potential for nonresponse bias is limited to differential nonresponse between anglers and non-anglers within the matched and unmatched subpopulations. While the Weather and Outdoor Activity Survey achieved a higher response rate than the Saltwater Fishing Survey, both overall and within the unmatched subpopulation, the gains in response do not appear to result from a higher propensity to respond to the survey by either anglers or non-anglers. As a result, we cannot conclude that one of the instruments is more or less likely to minimize differential nonresponse between anglers and non-anglers. However, higher response rates decrease the risk for nonresponse bias and either lower data collection costs (for a fixed sample size) or increase the precision of estimates (for a fixed cost)[2]. Consequently, we conclude that the Weather and Outdoor Activity Survey is superior to the Saltwater Fishing Survey and recommend that the instrument be utilized for subsequent survey waves. Because it collects person-level information for all residents of all sampled households, the Weather and Outdoor Activity Survey also supports post-stratification of survey weights to population controls, which is an additional benefit of this recommendation.

Incentive Testing

The study included an experiment to test the impact of modest, prepaid cash incentives on survey response and survey measures. Each wave, sampled addresses were randomly allocated to incentive treatment groups of $1, $2, and $5, as well as a non-incentive control group. Incentives were only included in the initial survey mailing. As in the instrument experiment, the objective of the incentive testing was to identify an optimum level of incentive that maximizes overall response while controlling costs and minimizes the potential for nonresponse bias resulting from differential nonresponse between anglers and non-anglers. Response rates, estimated fishing prevalence and relative costs of completing an interview were compared among incentive treatments to quantify the impacts of incentives.

Table 2 shows weighted response rates and the results of a logistic regression model predicting the effects of incentives on the odds of obtaining a completed survey. Including an incentive in the initial survey mailing significantly increased the odds of receiving a completed survey, and the odds increased significantly as the incentive amount increased. Cash incentives of $1, $2,

[2] Assuming that fixed costs are the same for the two instruments, which was the case in the experiment.

and $5 increased the odds of receiving a completed survey by 63%, 93% and 137%, respectively.

Table 2. Weighted response rates and odds of receiving a completed survey by incentive amount.

Incentive	Response Rate (%)	n	Odds Ratio	95 % CI
$0	22.6	8,760	1.00	
$1	32.2	8,737	1.63*	(1.51, 1.77)
$2	36.0	8,738	1.93*	(1.78, 2.09)
$5	40.8	8,786	2.37*	(2.18, 2.56)

*Significantly different from the $0 control ($p<0.05$). Results of pairwise comparisons are as follows: $1>$0 ($p<0.05$), $2>$1 ($p<0.05$), $5>$2 ($p<0.05$).

Previous studies (Groves et al. 2006) have demonstrated that prepaid cash incentives can motivate individuals with little or no interest in a survey topic to respond to a survey request. Consequently, we hypothesized that incentives would have a larger impact on non-anglers than anglers, minimizing differential nonresponse between the two populations. We initially explored this hypothesis by comparing estimated fishing prevalence among incentive conditions, expecting that gains in response in the incentive conditions would translate to lower estimates of fishing prevalence. The results do not support this hypothesis; there were no significant differences in prevalence among incentive conditions (Table 3).

Table 3. Overall estimated fishing prevalence by incentive amount.

Incentive	Prevalence (%)	n
$0	12.8	2,154
$1	14.1	3,065
$2	13.6	3,415
$5	14.1	3,807

Note: Differences in prevalence among treatments are not significant ($p=0.05$)

We further explored the interaction of topic salience and incentives by examining response rates and estimated fishing prevalence for the incentive conditions within domains defined by whether or not sampled addresses could be matched to databases of licensed saltwater anglers. We expected incentives to have a more pronounced effect in the unmatched domain, a population less likely to have an interest in the survey topic, than in the matched domain. Table 4 shows that incentives increased the odds of receiving a completed survey in both the matched and unmatched subpopulations. However, the value of the incentive seems to be more important in the unmatched domain, where the odds of receiving a completed survey increased uniformly and significantly as the value of the incentive increased ($0<$1<$2<$5). In contrast, the incentive amount was less significant in the matched domain, where the odds of receiving a completed survey were relatively flat among incentive conditions. These results are consistent with our expectations and suggest that a population with a low propensity to respond to a fishing survey can be motivated to participate by cash incentives, and that the motivation may increase as the incentive amount increases.

Table 4. Odds of receiving a completed survey by level of incentive for sample that could and could not be matched to state databases of licensed anglers.

Comparison Pair	Subpopulation Matched OR	Unmatched OR
$1 vs. $0	1.75**	1.63**
$2 vs. $0	2.01**	1.93**
$5 vs. $0	2.11**	2.39**
$2 vs. $1	1.15	1.18**
$5 vs. $1	1.21*	1.46**
$5 vs. $2	1.05	1.24**

Notes – The second value in the comparison pair is the reference value.
Significance: *p<0.05, **p<0.0001

As noted previously, we expected that the gains in response in the incentive conditions would translate to lower estimates of fishing prevalence, particularly in the unmatched subpopulation. Once again, the results are not consistent with expectations; differences in fishing prevalence among treatments were not significant in either the matched or unmatched domain (Table 5). The lack of an effect of incentives on fishing prevalence suggests that the gains in response associated with increasing incentive amounts are uniform between anglers and non-anglers. However, it's also possible that the gains in response are accompanied by an increase in measurement error; non-anglers may be more likely to report fishing behavior than anglers when an incentive is provided. This hypothesis was not tested and requires further investigation.

Table 5. Estimated fishing prevalence by incentive amount for a population of anglers (matched) and non-anglers (unmatched).

	Subpopulation			
	Matched		Unmatched	
Incentive	(%)	(n)	(%)	(n)
$0	49.2	533	10.7	1,621
$1	50.3	779	12	2,286
$2	48.6	837	11.6	2,578
$5	48.2	894	12.4	2,913

Note: Within subpopulations differences in prevalence among treatments are not significant (p=0.05).

We also examined the effect of cash incentives on overall data collection costs, specifically the direct costs of printing, postage, and the cash incentives themselves. Table 6 shows that the $5 incentive provided the largest gain in response, but the gain came at a relative cost of approximately $0.15 per completed interview. In contrast, the additional costs of the $1 and $2 incentives (20% and 38% higher cost than the $0 control, respectively) are more than offset by the associated gains in the number of completed surveys (42% and 58%, respectively). In other words, including a $1 or $2 cash incentive in the initial survey mailing actually decreased the cost of receiving a completed survey by 22% and 20%, respectively. These cost savings, which are conservative[3], could be used to lower overall data collection costs (for a fixed sample size) or increase the precision of survey estimates (for a fixed cost).

[3] The cost comparison assumes that the non-incentive direct costs (postage and printing) are the same for all survey treatments and does not reflect the fact that incentive conditions may not require as many follow-up mailings.

Table 6. Effect of incentives on data collection costs

Incentive Amount	Relative Cost Difference	Relative Difference in Completed Surveys	Relative Cost per Completed Survey
$0	1.00	1	$1.00
$1	1.20	1.42	$0.78
$2	1.38	1.58	$0.80
$5	1.90	1.75	$1.15

Note: relative differences reflect the ratio of quantities (cost, completes) in the experimental treatments to the zero dollar control.

Including a modest prepaid cash incentive in survey mailings clearly has a positive effect on survey response rates; the odds of receiving a completed survey increased significantly as the incentive amount increased. We expected the incentives to have a greater effect on non-anglers than anglers and decrease the potential for nonresponse bias by minimizing differential nonresponse between these two populations. However, the results of the experiment suggest that incentives increase response propensities for non-anglers and anglers equally. While this result does not support our hypothesis, it does demonstrate that incentives can increase the quantity of data without having a negative impact on survey measures. The experiment also demonstrated that incentives can decrease overall data collection costs. Based upon these findings, we conclude that a $2 incentive is optimal in terms of both maximizing response rates and minimizing data collection costs.

References

Groves, R.M., M.P. Couper, S. Presser, E. Singer, R. Tourangeau, G.P. Piani, N. Lindsay. 2006. Experiments in producing nonresponse bias. Public Opinion Quarterly 70: 720-736.

Lohr, S. 2009. Multiple frame surveys. Pages 71-88 *in* D. Pfeffermann, editor. Handbook of statistics: sample surveys design, methods and applications, volume 29A. Elsevier, Amsterdam.

Project Team Response
7/31/2014

Peer Review Report for

"Development and Testing of Recreational Fishing Effort Surveys, Testing a Single-Phase Mail Survey Design"

Reviewed by

Dr. Michael Cohen, American Institutes for Research
Dr. Ronald Fesco, Ernst & Young LLP
Dr. Phillip Kott, RTI International

Introduction

This document combines the comments provided by three different peer reviewers of the MRIP Project Report entitled "*Development and Testing of Recreational Fishing Effort Surveys, Testing a Single-Phase Mail Survey Design*." The document provides verbatim reviewer comments without identifying the source of each comment.

Reviewer 1

This review of the report entitled "Development and Testing of Recreational Fishing Effort Surveys: Testing a Single-Phase Mail Survey Design" provides comments and suggestions on the methods, results and conclusions found in the report. The review does not include any working with the original data and thus does not encompass any validation of data or primary calculations with the data. The review examines only summary calculations found in the report and, accepting those as shown, assesses the reasonableness of methods, approach and use of results to reach conclusions about aspects of Recreational Fishing Effort Surveys (RFES), especially the recommendation to move to a mail survey design.

The report presents the results of an evaluation of a single phase mail survey design as an alternative to the Coastal Household Telephone Survey (CHTS) for estimating marine recreational fishing effort. The objectives identified in the report were to:

1) test the feasibility of a mail survey design for collecting recreational fishing effort data and estimating fishing effort for shore and private boat anglers,

2) compare single phase mail survey and CHTS results, including metrics of survey quality and estimates of marine recreational fishing activity,

3) describe, to the greatest extent possible, differences between single phase mail survey and CHTS estimates in terms of sources of survey error, and

4) provide recommendations for follow-up action, including implementation of improved survey methods.

This review will discuss the objectives in order and provide several other insights to conclude.

Generally, the analysis is done very well with considerable thought about identifying and measuring sources of differences between the surveys. I find no meaningful issues in the methodology used or the analyses and therefore provide brief comments on the 4 objectives above and I do not reiterate the various findings. Finally, I will discuss some ideas for future consideration.

OBJECTIVE 1) test the feasibility of a mail survey design for collecting recreational fishing effort data and estimating fishing effort for shore and private boat anglers

The authors (Andrews, Brick and Mathiowetz) describe a well conceived experimental approach to providing metrics to lead to decisions on survey approaches. They describe problems with the existing survey, especially low response rates, and identify issues that can further degrade quality of the existing design e.g., declining landline use. They make reasoned and convincing arguments, supported by the metrics, that response rates and response error are less of a problem with mail surveys and those improvements also reduce bias problems. The authors also show that the quality improvements can be achieved within the time frame required of the survey operations. I agree with their conclusion that a mail survey design is feasible and preferred.

The use of a $2 incentive was clearly justified by the analysis of experiments found in appendix B. Often incentive experiments fail to discuss overall cost relative to effect. Here, the authors provide a fair comparison taking cost into consideration. Further analysis of the impact on broader survey costs including the typically expensive follow up of nonrespondents for incremental incentives from $2 to $5 would add to the understand, but the gains in response at the $2 level would typically be cost effective, making the use in the design reasonable.

Response: We appreciate the reviewer's encouraging comments regarding the feasibility of the mail survey design.

OBJECTIVE 2) compare single phase mail survey and CHTS results, including metrics of survey quality and estimates of marine recreational fishing activity

The research appropriately examines design features that may impact differences between survey approaches. The analysis indicates that mail survey methods result in larger estimates of percent of households fishing while mean numbers of within household statistics vary with mean trips larger for mail and other items not particularly different. Reasons for the differences are hypothesized and explored in a balanced and fair manner.

While "quality" is not specifically defined in the report, most methodologists would consider cost, timeliness and relevance along with the usual focus on error sources. The authors have exhibited some cost improvements in the mail survey approach and that it meets timeliness needs. The authors explore various thoughts on response differences and bias sources (geographic, unlicensed anglers, etc.) finding that the mailing methods perform well and the responses may be more in line with the concepts desired.

Response: We appreciate the reviewer's positive comments regarding the analyses described in the report.

OBJECTIVE 3) describe, to the greatest extent possible, differences between single phase mail survey and CHTS estimates in terms of sources of survey error

As mentioned above, survey error is one of the quality dimensions. The report explores usual sources of error for the survey types. Identifying sources of error is an intuitive and experience based endeavor. The authors were creative and explored a commendable range of ideas. The range of finding are sufficient to support their conclusions regarding survey methodology changes.

Response: We appreciate the reviewer's positive comments regarding the findings described in the report.

OBJECTIVE 4) provide recommendations for follow-up action, including implementation of improved survey methods

The matching of ABS sample to license frames (p. 8) is a good idea and can be effective for stratification and sample allocation.

The main recommendation, using a single-phase mail survey, covers many potential improvements. This recommendation is supported and reasonable. The suggestion for continued development and testing (p. 32) is reasonable because there usually are changes to consider when moving to full scale implementation.

With the evolution of e-mail and web collection modes, the recommendation to explore such methods is reasonable. Methodologists such as Don Dillman are conducting current research that should be examined for applicability.

Response: We appreciate the reviewer's positive comments regarding the conclusions and recommendations described in the report.

COMMENTS

Bottom line, I can find nothing of concern in the methods, analyses or conclusions in the paper. That said, identifying error sources in surveys is difficult, but the authors explored a wide and thoughtful set of issues and make appropriate suggestions for further research. As such, I find no reason to be concerned about their suggestion to move to a mail survey approach and believe it would be a reasonable thing to do.

Response: We appreciate the reviewer's positive comments.

IDEAS

Consider development of a bridging survey approach. Estimates will be changing with a move to mail and the research is based on a subset of areas to be sampled. Methodology will likely evolve a bit as well. A bridge helps to keep the time series of estimates usable.

Response: We agree that a bridging approach would help transition from the CHTS to a new survey design.

The may be a number of co varying attributes related to response and fishing. Age comes to mind as it is likely related to landline or cell use. It may also be something that increases with age to a point at which infirmity reduces fishing. The age distribution in the study states may be impacting some of the results. FL and NC are more destination states for retirees from the north. Thus, age may be influencing some of the state differences found (e.g. Table 4) and mail could reduce the impact in states with an older population.

Response: We appreciate the suggestion to explore co-variates to fishing effort. Person- and household-level demographic information is collected in the mail survey instrument. We will continue to examine differences in fishing activity among sub-populations and explore ways to incorporate this information into the estimation design (e.g., raking survey weights to control totals).

The analysis of difference from APAIS should consider the non-coastal travelers reason to travel and method of travel. Someone driving can take poles for surf fishing and avoid piers etc. Those flying have a much more difficult time taking equipment. This could influence the APAIS results. Also some areas are more known for travel to surf fish - NC - and travel there may be more by personal vehicle and with gear. Other areas like Florida may be more by air travel.

Response: Neither the APAIS nor mail survey collects information about the method of travel or reason for travel. We will continue to explore differences in residency distributions between the mail survey and APAIS by state and fishing mode.

I'm not sure that I agree with footnote 15. I've never had a problem finding a non-APAIS place to surf fish near the hotel or condo wherever we stay. It may be instructive to look at differences by state for domain estimates for in-state vs. out-of-state people in the APAIS data.

Response: We will continue to explore differences in residency distributions between the APAIS and mail survey.

Another factor to consider may be the proportion of the state's population living near the coast. If large cities are coastal, surf fishing may dominate.

Response: We will continue to explore differences in residency distributions between the APAIS and mail survey.

The thought in the above comments is that other characteristics may be useful in further improving the survey design and information useful to collect. Exploring how fishing responses compare to other characteristics collected in the survey may provide more ideas.

Response: We appreciate this constructive suggestion.

Pay pier is not specifically mentioned in the questionnaire in Q 15a or b. Dock etc of 15a may not draw the memory out. I might not have considered the fishing pier experience when answering 15a and then it is not a part of 15b.

Response: We appreciate this constructive suggestion and will consider modifications to the survey instrument to improve the accuracy of reporting.

Reviewer 2

"Developing and Testing of Recreational Fishing Effort Survey Testing a Single Phase Mail Survey Design" reports on research designed to improve the way estimates of recreational fishing effort are made with an emphasis on the last test conducted in four states using what the authors call a "single-phase dual-frame mail survey." The research itself is sturdy and the results (that the new estimation strategy is far superior to what is done now) convincing. The report itself, however, has a number of flaws.

One flaw that afflicts many research reports is the inconsistent use of tense. This is understandable given that the research has already been done but the methods used can be repeated, so describing them in the present tense makes some sense. What makes the tense-use problem particularly acute here is that some of the methods described were tested before the method on which the report focuses. The reader would have an easier time understanding what is old and what is new if the past perfect where used ("anglers *had been* mailed") in describing previous methods tested. Instead, the present is used to describe a method that had been tested before the single-phase dual-frame mail survey, while single-phase dual-frame mail survey is later described in the past tense.

Response: The text was modified to more clearly distinguish between the current pilot study and previous pilot studies.

A second flaw is that the authors' single-phase dual-frame mail survey, although a mail survey, is not single phase (there is subsampling in certain strata) and only technically dual frame. There *are* two frames in a state, an address-based resident frame and a frame containing non-resident licensed saltwater anglers, but since these frames do not overlap, dual-frame methodology is not employed. Instead, these separate frame as used in creating disjoint strata.

Response: References to a single-phase were intended to reflect the fact that data were collected in a single phase. However, we agree that this description is confusing and contradicts with the sample design, which includes sub-sampling in certain strata. We eliminated references to the "single-phase design" and explicitly state that data were collected in a single phase.

References to the dual-frame design were not changed as the survey employed a dual-frame design with non-overlapping frames (the ABS frame and the non-resident license frames are the two non-overlapping frames).

There is much discussion of stratification, but not enough to satisfy this reader. What exactly were the strata in each state, the targeted stratum sampling rates, and the actual stratum response rates? Readers are lead to believe that weights were equal within strata and reflected both the within-stratum sampling and response rates but are never told so explicitly. Consequently, that reasonable approach to handling nonresponse is never justified. (The lack of details carries over to Appendix B, where readers are given very little information about a logistic regression used to draw many conclusions.)

Response: We appreciate the suggestion to include more technical details in the report. However, the intended audience for the report includes managers and administrators. Consequently, we did not want to overwhelm the audience with technical details. Technical details about the survey design will be documented elsewhere.

There is one minor technical error (excusing the use of "single-phase" because there is only a single phase of data collection) and a somewhat larger technical embarrassment in the report. The minor technical error is the suggestion on page 25 that the expectation operator on probability-sampling theory breaks down for very small prevalences. It does not, estimates remain unbiased. The problem is that they are not very accurate. Their relative variances are high, and their nonnormality makes coverage-interval construction from their variance estimates dubious.

Response: We agree the language about this bias was confusing. We have revised the text to indicate the bias is that of separate stratum ratio estimators (the poststratified estimator in this case at the county level). When stratum sample size is small in the denominator of a ratio estimator, it is biased. A combined rather than separate ratio estimator would avoid this bias but is not used in CHTS. Furthermore, because saltwater fishing is a relatively rare event among the general population, repeated samplings from the general population will result in a distribution of estimates that is skewed with zero occurrences of reported fishing activity – so the bias of the ratio estimator results in underestimation. We revised the report to more clearly state the impact of small sample sizes on CHTS estimates.

The somewhat larger embarrassment is that, contrary to the authors' assertion, the fraction of respondents engaged in fishing is not a reasonable measure of the efficiency of the single-phase-dual-frame-mail-survey estimation strategy because targeted anglers are down- weighted in the estimation. Good measures of the strategy's relative statistical efficiency are the design effects of the estimates it produces. The only design effect the authors report is, unfortunately, close to 1. Others, especially for estimates of the anglers themselves, are likely to be smaller (if correctly computed for the purpose of evaluating the design).

Matching address samples to lists of licensed anglers proved to be an effective way to sample anglers, a relatively rare population. The key statistic from the survey is a characteristic of anglers (the number of trips taken) and by having a larger sample of anglers we are able to increase the statistical efficiency of this estimate. A much larger address sample would have been required to achieve the same effective sample of fishing households if license matching (i.e., screening prior to data collection) was not possible. This would have required additional mailings and would have resulted in substantially higher costs. In this sense, the design was more efficient that simple random sampling. We revised the text to more clearly characterize the benefits of the design. We did include some design effects in the revision, but that measure is not related to cost efficiency in that the same design effect can be achieved with different costs.-

Ultimately, however, these criticisms of the report are minor. As I wrote earlier, I found the report's conclusions convincing. I very much like what I can make out of the sampling and estimation strategy that the authors' recommend. The flaws in the report are statistical in nature. On the survey-methodology side, the report contains a commendable treatment of the problems and limitations involved in collecting the information desired.

Response: We appreciate the reviewer's positive comments about the report.

Reviewer 3

This well written and thoughtful report makes its main case overwhelmingly. The single phase mail survey (SPMS) is the clear winner when compared to the Coastal Household Telephone Survey (CHTS).

Response: We appreciate the reviewer's positive comments about the report.

Given the stark differences in marine fishing activity reported by the two surveys, there will be keen interest in how the differences break out by age, racial/ethnic, and sex groups. Are the young and elderly fishing off piers sometimes being missed? Are women and girls sometimes regarded as participants in marine fishing and other times just thought of as on- lookers? Do we know that racial/ethnic minorities are being represented fairly? There doubtless will be great interest in such questions.

Response: We will continue to examine the demographic characteristics of the sample and explore ways to incorporate this information into the estimation design.

Specific Comments:

Page 12, lines 5-7 from bottom: "median" is not explained correctly. It means that half the responses were received before the 14th day (or possibly on the 14th day, depending on the specifics of the definition).

Response: We have revised the report to accurately describe median response times.

On page 13, Figure 1, I did not understand the dots. There are many more dots after 20 days than before.

Response: Each dot represents a point in time. There are more dots after 20 days because the data collection continued for several additional weeks beyond 20 days. The figure shows the cumulative percentage of completed mail surveys over time and demonstrates that the vast majority (>70%) of completed surveys are returned within about three weeks of the initial mailing.

The last paragraph on page 23 makes perfect sense right up to the final "i.e.". The phrase "i.e., only individuals in households without licensed anglers could have contributed to nonresponse bias resulting from differential response between anglers and non-anglers" does not seem to me to follow from the rest of the paragraph nor do I think it is true. On rereading this some time after I wrote the previous two sentences, the point may be that unlicensed anglers mess up the nonresponse adjustment. I still do not think the quoted sentence is the right way to say it.

Response: We modified the sentence to more clearly articulate the benefit of frame matching on nonresponse weighting adjustment.

I disagree with the argument at the end of the first complete paragraph on page 25: "…we hypothesized that low sample sizes in the CHTS during low-activity waves result[s] in underestimates of state-level fishing effort." Small sample sizes will increase variance but not cause bias. It could happen that one would get a larger than average number (e.g. 2) of anglers, and they would have large weights.

Response: We address the impact of small sample sizes on CHTS estimates above.

I kept wanting to see discussion of possible measurement bias, and finally there is an excellent discussion in the paragraph beginning on page 28. But measurement bias could affect the earlier analyses so should be introduced sooner.

Response: We agree that measurement bias is a likely source of differences between mail survey and CHTS estimates. However, the discussion of measurement bias is largely hypothetical and based upon the results from previous pilot studies. The assessment of non-coverage bias is more direct and quantifiable. Consequently, we chose to discuss the impacts of non-coverage bias first.

It is remarkable (page B8, Table 6) that the $1 and $2 incentives lead to lower relative costs per completed survey compared to no incentive or $5 incentive. But I do not think one can conclude that the $5 incentive is sub-optimal (last line on page B8). It depends on the relative value one puts on maximizing response rates versus minimizing data collection costs. Even though (page B7, Table 5) the prevalence rate estimates do not differ significantly among the incentive levels, other estimates may be enhanced by a higher response rate.

Response: We agree that assigning a value to survey incentives involves a trade-off between cost considerations and data quality. For the purposes of this study, we determined that a $2.00 incentive had a greater relative value than the other incentive amounts. A $5.00 incentive would have resulted in a higher response rate, but the gains in additional sample would have been outweighed by the additional data collection cost. The $2.00 incentive resulted in the largest effective sample for a fixed data collection cost.

Editorial Comments:
Executive Summary, line 4: Either delete semi-colon or replace with colon.

Response: The semi-colon has been replaced with a colon.

On page 18, line 3 of second paragraph: I would change "(wireless households)" to "(wireless only households)".

Response: "Wireless households" has been replaced with "wireless-only households".

Page 25, last line of first complete paragraph: Change "results" to "result".

Response: "Results" has been changed to "result".

Page 33, second reference: I think the %20s in the URL should be spaces. Some systems changes spaces to %20s.

Response: The URL has been updated.

Page B5, Table 2, $2 Incentive line: Change "36" to "36.0".

Response: "36" has been replaced with "36.0".

A Small Area Estimation Approach for Reconciling Mode Differences in Two Surveys of Recreational Fishing Effort

DRAFT: PLEASE DO NOT CITE OR DISTRIBUTE

F. Jay Breidt
Teng Liu
Jean D. Opsomer
Colorado State University

June 10, 2017

Abstract

For decades, the National Marine Fisheries Service has conducted a telephone survey of United States coastal households to estimate recreational effort (the number of fishing trips) in saltwater. The effort estimates are computed for each of 17 US states along the coast of the Gulf of Mexico and the Atlantic Ocean, during six two-month waves (January-February through November-December). Recently, concerns about coverage errors in the telephone survey have led to implementation of a mail survey of the same population. Results from the mail survey are quite different from those of the telephone survey, due to coverage differences and mode effects, and a means of "calibrating" or reconciling the two sets of estimates is needed by fisheries managers and stock assessment scientists. We develop a log-normal model for the estimates from the two surveys, accounting for temporal dynamics through regression on population size and state-by-wave seasonal factors, and accounting in part for changing coverage properties through regression on wireless telephone penetration. Using the estimated design variances, we develop a regression model that is analytically consistent with the log-normal mean model. Finally, we use the modeled design variances in a Fay-Herriot small area estimation procedure to obtain empirical best linear unbiased predictors of the reconciled effort estimates for all states and waves.

1 Introduction

For decades, the National Marine Fisheries Service (NMFS) has conducted the Coastal Household Telephone Survey (CHTS) to collect recreational saltwater fishing effort (the number of fishing trips) from shore and private boat anglers in 17 US states along the coasts of the Atlantic Ocean and the Gulf of Mexico: Alabama, Connecticut, Delaware, Florida, Georgia, Louisiana, Maine, Maryland, Massachusetts, Mississippi, New Hampshire, New Jersey, New York, North Carolina, Rhode Island, South Carolina, and Virginia. Data collection occurs during a two-week period at the end of each two-month sample period (or “wave”), yielding six waves for each year. However, samples are not obtained for every wave in every state; for example, many states have no wave 1 sample, reflecting minimal fishing effort during January and February in those states.

The CHTS uses random digit dialing (RDD) for landlines of households in coastal counties. RDD suffers from several shortcomings in this context, such as the inefficiency at identifying anglers (National Research Council, 2006), the declining response rate for telephone surveys (Curtin et al., 2005), and the undercoverage of anglers due to the increase in wireless-only households (Blumberg and Luke, 2013). Thus, after some experimentation, NMFS implemented the new Fishing Effort Survey (FES) that involves mailing questionnaires to a probability sample of postal addresses (Andrews et al., 2014).

The telephone-based CHTS and the mail-based FES have obvious methodological differences. The two surveys have different coverage properties, because they use very different frames: RDD of landlines for CHTS versus address-based sampling, with oversampling of addresses matched to licensed anglers, for FES. They have different nonresponse patterns, with overall FES response rates nearly three times higher than CHTS response rates (Andrews et al., 2014). Finally, the measurement processes are fundamentally different, due to the differences in asking about angling activity over the phone versus a paper form.

Due at least in part to these methodological differences, there is a large discrepancy between the effort estimates from the CHTS and the FES estimates. Whatever the reasons for the discrepancy, it is of interest to fisheries managers and stock assessment scientists to be able to convert from the "units" of the telephone survey estimates to those of the mail survey estimates, and vice versa. This conversion is known as "calibration" in this context, and is not to be confused with the calibration method common in complex surveys. The calibration allows construction of a series of comparable estimates across time.

The data used for the calibration exercise come from the CHTS for most states and waves from 1982 to 2016, and from the FES for states and waves from 2015 to 2016. For each survey, the data consist of estimated total effort for shore fishing and for private boat fishing, along with estimated design variances and sample sizes, for each available state and wave.

The methodology described here uses effort estimates transformed via natural logarithms, for either shore or private boat fishing. Let $\widehat{M}_{st}$ denote the estimated log-effort based on the mail survey in state s and year-wave t and let $\widehat{T}_{st}$ denote the estimated log-effort based on the telephone survey. We build a model that assumes that both mail and telephone estimates target a common underlying time series of true effort, but that each survey estimate is distorted both by sampling error and non-sampling error. The true effort series is further described with a classical time series model consisting of trend, seasonal, and irregular components. The sampling error series have properties that are well-understood based on features of the corresponding sampling designs, including well-estimated design variances. The non-sampling error cannot be completely disentangled from the true effort series. But given the overlap of mail and telephone estimates for some states and waves, the difference in the non-sampling errors can be estimated, and can be modeled with available covariates to allow extrapolation forward or backward in time. This extrapolation is a key part of the calibration procedure.

The combined model for the two sets of estimates and the underlying true effort series is a linear mixed model of a type that commonly appears in the context of area-level small area estimation, where it is known as the Fay-Herriot model (Fay and Herriot, 1979). In Fay-Herriot, it is standard to treat design variances as known. Our design variances are based on moderate to large sample sizes (minimum size $n = 39$) in each state and wave and so are well-estimated by the standards of small area estimation. A complication is that our design variances are on the original effort scale rather than the log scale. As an alternative to standard Taylor linearization, we develop a novel approach to transforming the estimated design variances that ensures analytic consistency between our mean model and our variance model.

The Fay-Herriot methodology leads to empirical best linear unbiased predictors (EBLUP's) of the mail target or the telephone target, and these constitute our calibrated effort series. Unlike the standard Fay-Herriot context, the EBLUP's require prediction at new sets of covariates. We adapt standard mean square error (MSE) approximations and estimates to this non-standard situation, and evaluate their performance via simulation. Finally, we apply the methods to the problem of calibrating past telephone survey estimates to the mail survey.

2 Model

2.1 Mean model

We fix attention on one type of fishing behavior, either shore or private boat: the model development is identical in both cases. We assume that the telephone effort estimate $\widehat{T}_{st}$ is a design-unbiased estimator of the "telephone target" T_{st}, which includes both the true effort and survey mode effects due to the telephone methodology, while the mail effort estimate $\widehat{M}_{st}$ is a design-unbiased estimator of the "mail target" M_{st}, which includes both the true effort and survey mode effects due to the mail methodology. That is,

$$\widehat{T}_{st} = T_{st} + e_{st}^{T} \text{ and } \widehat{M}_{st} = M_{st} + e_{st}^{M}$$

where the sampling errors $\{e_{st}^{T}\}$ and $\{e_{st}^{M}\}$ have zero mean under repeated sampling.

We assume that both the telephone target and the mail target contain the true effort series, which is further assumed to contain state-specific trends, due in part to changing state population sizes, state-specific seasonal effects that vary wave to wave, and irregular terms that are idiosyncratic effects not explained by regular trend or seasonal patterns. We model state-specific trends by using annual state-level population estimates from the US Census Bureau US Census Bureau (2016) on a log scale. We model a general seasonal pattern via indicators for the two-month waves, and allow the seasonal pattern to vary from state to state. The remaining irregular terms, denoted $\{\nu_{st}\}$ below, represent real variation not explained by the regular trend plus seasonal pattern, and are modeled as independent and identically distributed (iid) random variables with mean zero and unknown variance, ψ.

The survey mode effects present in the telephone and mail targets are non-sampling errors, including potential biases due to coverage error (population $\neq$ sampling frame), nonresponse error (sample $\neq$ respondents), and measurement error (true responses $\neq$ measured responses). These effects may have their own trend and seasonality: for example, due to changes in the quality of the frame over time, changes in response rates over years or waves, changes in implementation of measurement protocols over time, etc. These non-sampling errors thus cannot be completely disentangled from the true effort series (a problem in every survey).

Because of the availability of overlapping effort estimates, however, the difference in the effort estimates is an unbiased estimator of the difference in the survey mode effects. These differences can then be modeled and extrapolated to other time points that do not have overlapping data, allowing calibration from the telephone target to the mail target, and vice versa. The extrapolation requires a model and suitable covariates, which in this setting means covariates that explain the change in measurement error, nonresponse error, or coverage error over time. The calibration thus relies critically on extrapolation, with the usual caveat that the calibrated values may be badly wrong if the model does not hold over the full range of time.

The changing proportion of wireless-only households is a potential covariate for explaining changes in coverage error over time for the landline-only telephone survey. Accordingly, we obtained June and/or December wireless-only proportion estimates for each state from 2007–2014 from the National Health Interview Survey, conducted by the National Center for Health Statistics (Blumberg and Luke, 2013). We transformed these proportions via empirical logits and fitted the transformed values as state-specific lines with a slope change in 2010. The fitted model has an adjusted R^2 value of 0.9948. Transforming back to proportions and extrapolating backward in time yields a series $\{w_{st}\}$ that is approximately zero prior to the year 2000.

Either trend or seasonal could contain survey mode effects. Accordingly, we allow for the possibility that trend and seasonal are different for mail versus telephone, and in particular we allow for the possibility that either trend or seasonal can change with the level of wireless.

Our combined model then assumes

$$\begin{aligned}
\widehat{T}_{st} &= T_{st} + e_{st}^{T} \\
T_{st} &= \boldsymbol{a}_{st}'\boldsymbol{\alpha} + 0 \cdot \boldsymbol{b}_{st}'\boldsymbol{\mu} + w_{st}\boldsymbol{c}_{st}'\boldsymbol{\gamma} + \nu_{st} \\
&= [\boldsymbol{a}_{st}', \mathbf{0}', w_{st}\boldsymbol{c}_{st}']\,\boldsymbol{\beta} + \nu_{st} \\
&= \boldsymbol{x}_{Tst}'\boldsymbol{\beta} + \nu_{st} \\
\widehat{M}_{st} &= M_{st} + e_{st}^{M} \\
M_{st} &= \boldsymbol{a}_{st}'\boldsymbol{\alpha} + 1 \cdot \boldsymbol{b}_{st}'\boldsymbol{\mu} + 0 \cdot \boldsymbol{c}_{st}'\boldsymbol{\gamma} + \nu_{st} \\
&= = [\boldsymbol{a}_{st}', \boldsymbol{b}_{st}', \mathbf{0}']\,\boldsymbol{\beta} + \nu_{st} \\
&= \boldsymbol{x}_{Mst}'\boldsymbol{\beta} + \nu_{st}, \qquad (1)
\end{aligned}$$

where

- $\boldsymbol{a}_{st}$ is a vector of known covariates, including intercept, log(population), state indicators, wave indicators, and state by log(population) and state by wave interactions;
- $\boldsymbol{b}_{st}$ and $\boldsymbol{c}_{st}$ are subvectors from $\boldsymbol{a}_{st}$;
- $\boldsymbol{\beta}' = [\boldsymbol{\alpha}', \boldsymbol{\mu}', \boldsymbol{\gamma}']$ is a vector of unknown regression coefficients;
- the sampling errors $\{e_{st}^{T}\}$ are independent $\mathcal{N}(0, \sigma_{Tst}^{2})$ random variables, with known design variances σ_{Tst}^{2};
- the sampling errors $\{e_{st}^{M}\}$ are independent $\mathcal{N}(0, \sigma_{Mst}^{2})$ random variables, with known design variances σ_{Mst}^{2};
- the irregular terms $\{\nu_{st}\}$, representing real variation not explained by the regular trend plus seasonal pattern, are independent and identically distributed (iid) $\mathcal{N}(0, \psi)$ random variables, with unknown variance ψ;
- $\{e_{st}^{T}\}$, $\{e_{st}^{M}\}$ and $\{\nu_{st}\}$ are mutually independent.

The assumed independence of the sampling errors is justified by independent samples drawn state-to-state and wave-to-wave, and the assumed normality is justified by central limiting effects of moderate to large-size stratified samples in each state and wave. Further, we assume that because the mail and telephone surveys are selected and conducted independently, the sampling errors $\{e_{st}^{T}\}$ and $\{e_{st}^{M}\}$ are independent of one another. We use simulation to assess the sensitivity of some of our methods to the normality assumption on the random effects in §4.1 below. The design variances $\{\sigma^2_{Tst}\}$ and $\{\sigma^2_{Mst}\}$ are on the log scale, while the available design variance estimates $\{\widehat{V}_{Tst}\}$ and $\{\widehat{V}_{Mst}\}$ are on the original scale; we address this discrepancy in §2.2 below.

2.2 Design variance model

Under the log-normal effort models (1), the variances of the sampling errors are given by

$$\begin{aligned} V_{Tst} &= \mathrm{Var}\left(\exp(\widehat{T}_{st}) \mid T_{st}\right) \\ &= \left\{\exp(\sigma^2_{Tst}) - 1\right\} \exp\left\{2T_{st} + \sigma^2_{Tst}\right\} \end{aligned} \tag{2}$$

and

$$\begin{aligned} V_{Mst} &= \mathrm{Var}\left(\exp(\widehat{M}_{st}) \mid M_{st}\right) \\ &= \left\{\exp(\sigma^2_{Mst}) - 1\right\} \exp\left\{2M_{st} + \sigma^2_{Mst}\right\}. \end{aligned} \tag{3}$$

We need to estimate σ^2_{Tst} and σ^2_{Mst}, incorporating the approximately design-unbiased estimates $\widehat{V}_{Tst}$ and $\widehat{V}_{Mst}$ of V_{Tst} and V_{Mst}, respectively.

We follow an approach related closely to generalized variance function estimation (e.g., Ch. 7 of Wolter (2007)). Assume that given T_{st} and M_{st}, the empirical coefficients of variation (CV's) are log-normally distributed, independent of the effort estimates $\widehat{T}_{st}$ and $\widehat{M}_{st}$:

$$\ln\left(\frac{\widehat{V}_{Tst}}{\exp(2\widehat{T}_{st})}\right) = \boldsymbol{d}'_{Tst}\boldsymbol{\delta}^T_0 + \delta^T_1 \ln(n_{Tst}) + \eta^T_{st}, \quad \eta^T_{st} \sim \mathcal{N}(0, \tau^2_T) \tag{4}$$

where $\boldsymbol{d}_{Tst}$ is a vector of known covariates (including state, wave, and state by wave interaction), and

$$\ln\left(\frac{\widehat{V}_{Mst}}{\exp(2\widehat{M}_{st})}\right) = \boldsymbol{d}'_{Mst}\boldsymbol{\delta}^M_0 + \delta^M_1 \ln(n_{Mst}) + \eta^M_{st}, \quad \eta^M_{st} \sim \mathcal{N}(0, \tau^2_M), \tag{5}$$

where $\boldsymbol{d}_{Mst}$ is a vector of known covariates. These models can be rewritten as regression models for the design variance estimates, with known offsets:

$$\ln\left(\widehat{V}_{Tst}\right) = 2\widehat{T}_{st} + \boldsymbol{d}'_{Tst}\boldsymbol{\delta}^T_0 + \delta^T_1 \ln(n_{Tst}) + \eta^T_{st}, \quad \eta^T_{st} \sim \mathcal{N}(0, \tau^2_T)$$

and

$$\ln\left(\widehat{V}_{Mst}\right) = 2\widehat{M}_{st} + \boldsymbol{d}'_{Mst}\boldsymbol{\delta}_0^M + \delta_1^M \ln(n_{Mst}) + \eta_{st}^M, \quad \eta_{st}^M \sim \mathcal{N}(0, \tau_M^2).$$

Empirically, each of these models fits very well: 94.54% adjusted R^2 value for telephone, and 98.01% adjusted R^2 value for mail.

These empirical models may be of independent interest as generalized variance functions for variance estimation on the original scale: by plugging the point estimate, state, wave, and sample size into the fitted versions of (4) or (5), one obtains excellent point estimates of the coefficient of variation.

Assuming that $\widehat{V}_{Tst}$ is exactly unbiased for V_{Tst}, we then have from the log-normal CV model (4) and the assumed conditional independence of $\widehat{V}_{Tst}$ and $\widehat{T}_{st}$ given T_{st} that

$$\begin{aligned}
&\exp\left\{\boldsymbol{d}'_{Tst}\boldsymbol{\delta}_0^T + \delta_1^T \ln(n_{Tst}) + \frac{\tau_T^2}{2}\right\} = \mathrm{E}\left[\left.\frac{\widehat{V}_{Tst}}{\exp\left(2\widehat{T}_{st}\right)}\right| T_{st}\right] \\
&= \mathrm{E}\left[\widehat{V}_{Tst} \mid T_{st}\right] \mathrm{E}\left[\exp\left(-2\widehat{T}_{st}\right) \mid T_{st}\right] \\
&= V_{Tst} \exp\left(-2T_{st} + 2\sigma_{Tst}^2\right), \qquad (6)
\end{aligned}$$

and similarly

$$\begin{aligned}
&\exp\left\{\boldsymbol{d}'_{Mst}\boldsymbol{\delta}_0^M + \delta_1^M \ln(n_{Mst}) + \frac{\tau_M^2}{2}\right\} = \mathrm{E}\left[\left.\frac{\widehat{V}_{Mst}}{\exp\left(2\widehat{M}_{st}\right)}\right| M_{st}\right] \\
&= \mathrm{E}\left[\widehat{V}_{Mst} \mid M_{st}\right] \mathrm{E}\left[\exp\left(-2\widehat{M}_{st}\right) \mid M_{st}\right] \\
&= V_{Mst} \exp\left(-2M_{st} + 2\sigma_{Mst}^2\right). \qquad (7)
\end{aligned}$$

Thus, we have from (2) and (6) that

$$\begin{aligned}
&\exp\left\{\boldsymbol{d}'_{Tst}\boldsymbol{\delta}_0^T + \delta_1^T \ln(n_{Tst}) + \frac{\tau_T^2}{2}\right\} \\
&= \left\{\exp(\sigma_{Tst}^2) - 1\right\} \exp\left\{2T_{st} + \sigma_{Tst}^2\right\} \exp\left(-2T_{st} + 2\sigma_{Tst}^2\right) \\
&= \exp(4\sigma_{Tst}^2) - \exp\left(3\sigma_{Tst}^2\right) \qquad (8)
\end{aligned}$$

and from (3) and (7) that

$$\begin{aligned}
&\exp\left\{\boldsymbol{d}'_{Mst}\boldsymbol{\delta}_0^M + \delta_1^M \ln(n_{Mst}) + \frac{\tau_M^2}{2}\right\} \\
&= \left\{\exp(\sigma_{Mst}^2) - 1\right\} \exp\left\{2M_{st} + \sigma_{Mst}^2\right\} \exp\left(-2M_{st} + 2\sigma_{Mst}^2\right) \\
&= \exp(4\sigma_{Mst}^2) - \exp\left(3\sigma_{Mst}^2\right). \qquad (9)
\end{aligned}$$

The left-hand-side parameters of (8) can be estimated from (4) and the left-hand-side parameters of (9) can be estimated from (5). The resulting estimates of σ_{Tst}^2 and σ_{Mst}^2 can then be obtained by solving the equations (8) and (9), which are quartic polynomials in $\exp(\sigma_{Tst}^2)$ and $\exp(\sigma_{Mst}^2)$. Using Descartes' rule of signs, it can be shown that each of these quartic equations has one negative real root, two complex conjugate roots, and one positive real root. The solutions for σ_{Tst}^2 and σ_{Mst}^2 are then the logarithms of the unique, positive real roots, which can be obtained via standard numerical procedures. While these solutions are in fact estimates, we will treat them as fixed and known in what follows, as is standard in the small area estimation techniques which we will apply in subsequent sections.

The resulting design variances on the log scale, σ_{Tst}^2 and σ_{Mst}^2, are strongly correlated with the estimated variance approximations from Taylor linearization, $\widehat{V}_{Tst} \exp\left(-2\widehat{T}_{st}\right)$ and $\widehat{V}_{Mst} \exp\left(-2\widehat{M}_{st}\right)$: 0.798 and 0.803, respectively. But they are not identical (see Figure 1), and the method described forces analytical consistency between the mean model and the variance model.

2.3 Fay-Herriot small area estimation model

Define

$$\boldsymbol{x}'_{st} = \begin{cases} \boldsymbol{x}'_{Tst}, & \text{if no mail estimate is available;} \\ \boldsymbol{x}'_{Mst}, & \text{if no telephone estimate is available;} \\ (\boldsymbol{x}_{Tst} + \boldsymbol{x}_{Mst})'/2, & \text{otherwise.} \end{cases}$$

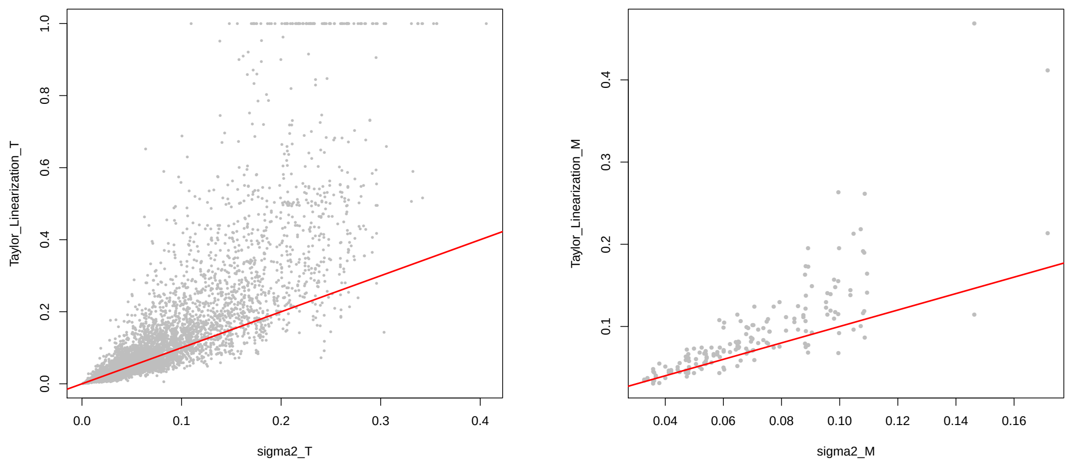


Figure 1: Estimated design variances for log-effort via Taylor linearization versus solution of the quartic polynomial equations (8) for telephone (left panel) and (9) for mail (right panel).

Then it is convenient to write

$$\begin{aligned} Y_{st} &= \begin{cases} \widehat{T}_{st}, & \text{if no mail estimate is available;} \\ \widehat{M}_{st}, & \text{if no telephone estimate is available;} \\ \left(\widehat{T}_{st} + \widehat{M}_{st}\right)/2, & \text{otherwise;} \end{cases} \\ &= \begin{cases} \boldsymbol{x}'_{Tst}\boldsymbol{\beta} + \nu_{st} + e^T_{st}, & \text{if no mail estimate is available;} \\ \boldsymbol{x}'_{Mst}\boldsymbol{\beta} + \nu_{st} + e^M_{st}, & \text{if no telephone estimate is available;} \\ (\boldsymbol{x}_{Tst} + \boldsymbol{x}_{Mst})'\boldsymbol{\beta}/2 + \nu_{st} + (e^T_{st} + e^M_{st})/2, & \text{otherwise;} \end{cases} \\ &= \boldsymbol{x}'_{st}\boldsymbol{\beta} + \nu_{st} + e_{st}. \end{aligned} \tag{10}$$

This model then follows exactly the linear mixed model structure of Fay and Herriot (1979), with direct estimates Y_{st} equal to regression model plus random effect ν_{st} plus sampling error with “known” design variance, given by

$$D_{st} = \begin{cases} \sigma^2_{Tst}, & \text{if no mail estimate is available;} \\ \sigma^2_{Mst}, & \text{if no telephone estimate is available;} \\ \frac{1}{4}\left(\sigma^2_{Tst} + \sigma^2_{Mst}\right), & \text{otherwise.} \end{cases}$$

Averaging the telephone and mail estimates results in a small loss of information, since we are replacing two correlated observations with one observation, but allows the use of standard software for estimation.

3 Methods

3.1 Estimation for the Fay-Herriot model

Define $\mathcal{A} = \{(s,t) : Y_{st} \text{ is not missing}\}$ to be the set of all state by year-wave combinations for which we have an estimate from either survey. Let m denote the size of the set $\mathcal{A}$. Define $\boldsymbol{X} = [\boldsymbol{x}'_{st}]_{(s,t)\in\mathcal{A}}$, $\boldsymbol{Y} = [Y_{st}]_{(s,t)\in\mathcal{A}}$, and

$$\boldsymbol{\Sigma}(\psi) = \mathrm{Var}\,(\boldsymbol{Y}) = \mathrm{diag}\{\psi + D_{st}\}_{(s,t)\in\mathcal{A}}.$$

Then

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + [\nu_{st}]_{(s,t)\in\mathcal{A}} + [e_{st}]_{(s,t)\in\mathcal{A}}.$$

If ψ were known, the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ would be

$$\widetilde{\boldsymbol{\beta}}_\psi = \left\{\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{X}\right\}^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{Y}. \tag{11}$$

Since ψ is not known, we replace it by a consistent estimator to obtain

$$\widehat{\boldsymbol{\beta}} = \left\{\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\hat{\psi})\boldsymbol{X}\right\}^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\hat{\psi})\boldsymbol{Y}. \tag{12}$$

We will use the Restricted Maximum Likelihood (REML) estimate $\hat{\psi}$ unless otherwise indicated.

3.2 Prediction

In the classical Fay-Herriot context, it is of interest to predict

$$\boldsymbol{x}'_{st}\boldsymbol{\beta} + \nu_{st}$$

from (10). In our setting, however, we seek to predict

$$\phi_{st} = \boldsymbol{z}'_{st}\boldsymbol{\beta} + \nu_{st}, \tag{13}$$

where $\boldsymbol{z}_{st}$ may not equal $\boldsymbol{x}_{st}$. For example, for a past time point with a telephone survey estimate but no mail survey estimate, we may want to use

$$\boldsymbol{z}'_{st} = \boldsymbol{x}'_{Mst} = [\boldsymbol{a}'_{st}, \boldsymbol{b}'_{st}, \boldsymbol{0}']$$

to predict the mail target M_{st}, while for a future time point with a mail survey estimate but no telephone, we may want to use

$$\boldsymbol{z}_{st} = [\boldsymbol{a}'_{st}, \mathbf{0}', \mathbf{0}']$$

to predict the telephone target, corrected for the wireless effect: $T_{st} - w_{st}\boldsymbol{c}'_{st}\boldsymbol{\gamma} = \boldsymbol{a}'_{st}\boldsymbol{\alpha} + \nu_{st}$.

Let $\boldsymbol{\lambda}_{st}$ denote a $m \times 1$ vector with a one in the (s,t)th position and zero elsewhere. Under normality, it is well-known that the best mean square predictor of ϕ_{st} in (13) is

$$\phi_{st}\left(\boldsymbol{\beta}, \psi\right) = \boldsymbol{z}'_{st}\boldsymbol{\beta} + \psi\boldsymbol{\lambda}'_{st}\boldsymbol{\Sigma}^{-1}(\psi)(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}), \tag{14}$$

which is feasible only if both $\boldsymbol{\beta}$ and ψ are both known. If only ψ is known, the best linear unbiased predictor (BLUP)

$$\phi_{st}\left(\widetilde{\boldsymbol{\beta}}_{\psi}, \psi\right) = \boldsymbol{z}'_{st}\widetilde{\boldsymbol{\beta}}(\psi) + \psi\boldsymbol{\lambda}'_{st}\boldsymbol{\Sigma}^{-1}(\psi)(\boldsymbol{Y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}(\psi)) \tag{15}$$

is obtained by plugging the BLUE from (11) into (14). Finally, if neither $\boldsymbol{\beta}$ nor ψ is known, then the empirical best linear unbiased predictor (EBLUP) can be obtained by substituting a consistent estimator of ψ into (15):

$$\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \hat{\psi}\right) = \boldsymbol{z}'_{st}\widehat{\boldsymbol{\beta}} + \hat{\psi}\boldsymbol{\lambda}'_{st}\boldsymbol{\Sigma}^{-1}(\hat{\psi})(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}), \tag{16}$$

where $\widehat{\boldsymbol{\beta}}$ is given by (12). These EBLUP's are the proposed calibrated values on the log scale.

3.3 Mean square error approximation

To assess the uncertainty of the calibrated values, we adapt the approach of Datta and Lahiri (2000) in approximating the mean square error (MSE) of the $\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \hat{\psi}\right)$ values. It can be shown that

$$\begin{aligned}
\mathrm{MSE}\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \hat{\psi}\right)\right\} &= \mathrm{E}\left[\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \hat{\psi}\right) - \phi_{st}\right\}^2\right] \\
&= \mathrm{E}\left[\left\{\phi_{st}\left(\widetilde{\boldsymbol{\beta}}_{\psi}, \psi\right) - \phi_{st}\right\}^2\right] + \mathrm{E}\left[\left\{\phi_{st}\left(\boldsymbol{\beta}, \psi\right) - \phi_{st}\left(\widetilde{\boldsymbol{\beta}}_{\psi}, \psi\right)\right\}^2\right] \\
&\quad + \mathrm{E}\left[\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \hat{\psi}\right) - \phi_{st}\left(\boldsymbol{\beta}, \psi\right)\right\}^2\right] \\
&= \dot{g}_{1st}(\psi) + \dot{g}_{2st}(\psi) + \dot{g}_{3st}(\psi) + o\left(m^{-1}\right),
\end{aligned} \tag{17}$$

where

$$\dot{g}_{1st}(\psi) = \frac{\psi D_{st}}{\psi + D_{st}},$$

$$\dot{g}_{2st}(\psi) = \left(\frac{\psi(\boldsymbol{z}_{st} - \boldsymbol{x}_{st})' + D_{st}\boldsymbol{z}'_{st}}{\psi + D_{st}} \right) \left[\sum_{u \in \mathcal{A}} (\psi + D_u)^{-1} \boldsymbol{x}_u \boldsymbol{x}'_u \right]^{-1} \times \left(\frac{\psi(\boldsymbol{z}_{st} - \boldsymbol{x}_{st})' + D_{st}\boldsymbol{z}'_{st}}{\psi + D_{st}} \right)',$$

and

$$\dot{g}_{3st}(\psi) = \frac{2D_{st}^2}{(\psi + D_{st})^3} \frac{1}{\sum_{u \in \mathcal{A}} (\psi + D_u)^{-2}}.$$

The terms $\dot{g}_{1st}(\psi)$ and $\dot{g}_{3st}(\psi)$ are identical to the terms $g_{1st}(\psi)$ and $g_{3st}(\psi)$ in §4 of Datta and Lahiri (2000), while $\dot{g}_{2st}(\psi)$ simplifies to $g_{2st}(\psi)$ of that paper in the special case of $\boldsymbol{z}_{st} = \boldsymbol{x}_{st}$. We omit the proofs.

3.4 Mean square error estimation

We now propose an estimator of the MSE approximation in (17). Using arguments like those in §5 of Datta and Lahiri (2000), it can be shown that

$$\begin{aligned} \mathrm{E}\left[\dot{g}_{1st}(\hat{\psi})\right] &\simeq \dot{g}_{1st}(\psi) - \dot{g}_{3st}(\psi) \\ \mathrm{E}\left[\dot{g}_{2st}(\hat{\psi})\right] &\simeq \dot{g}_{2st}(\psi) \\ \mathrm{E}\left[\dot{g}_{3st}(\hat{\psi})\right] &\simeq \dot{g}_{3st}(\psi) \end{aligned}$$

and hence an approximately unbiased estimator of the MSE approximation in (17) is given by

$$\mathrm{mse}\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \hat{\psi}\right)\right\} = \dot{g}_{1st}(\hat{\psi}) + \dot{g}_{2st}(\hat{\psi}) + 2\dot{g}_{3st}(\hat{\psi}). \tag{18}$$

We assess the quality of the asymptotic approximation (17) and its estimator (18) via simulation in §4.1.

3.5 Prediction on the original scale

To compute predictors on the original scale, we back-transform by exponentiating the EBLUP from (16) and adjust for the nonlinearity of the back-transformation using the estimated MSE from (18):

$$\widehat{\exp(\phi_{st})} = \exp\left[\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \hat{\psi}\right) + \frac{1}{2}\mathrm{mse}\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \hat{\psi}\right)\right\}\right], \tag{19}$$

which is an estimator of the best mean square predictor under the normal model, and a standard adjustment even without the normality assumption.

4 Empirical results

4.1 Simulation

In this section, we investigate the performance of our second-order approximation of MSE and the estimated MSE under a setting that mimics the calibration problem of this paper, but with a smaller number of observed time points: 17 states and six years (1985, 1995, 2005, 2010, 2015, and 2016) of six waves each, with telephone effort estimates for all waves, and with mail effort estimates for only the final two years. In this setting, m =(17 states)(6 waves)(6 years)=612. We took the wireless values and US Census population counts from the actual data.

We used as true regression coefficient values the estimates from model (10) fitted to shore data, with intercept, log(population), state indicators, wave indicators, state by log(population) interaction, and state by wave; plus wireless and its interactions with log(population), state indicators, and wave indicators; plus an indicator for presence of a mail survey estimate and the mail indicator's interactions with log(population), state indicators, and wave indicators. We also used $\psi = 0.11$, again from the fit of the model. The simulation model is similar to the final model selected in §4.2 below.

We considered three different patterns for the design variances $\{D_{st}\}$. First, we sampled six actual design variances for each simulated state, arranged the six into a "peaked" seasonal pattern, and replicated this seasonal pattern across all six years to create pattern (b). We considered two additional settings, by multiplying pattern (b) by 0.5 to yield pattern (a), and multiplying pattern (b) by 2.0 to yield pattern (c). The simulated sampling errors $\{e_{st}\}$ in (10) were then generated independently as $\mathcal{N}(0, D_{st})$ under each pattern.

Following Datta et al. (2005), we considered three distributions to simulate the normalized random effects:

- $\{\psi^{-1/2}\nu_{st}\}$ iid $\mathcal{N}(0,1)$;
- $\{\psi^{-1/2}\nu_{st}\}$ iid Laplace$(0, 1/\sqrt{2})$;
- $\{\psi^{-1/2}\nu_{st}\}$ iid centered Exponential(1) (that is, exponential random variables centered to mean zero).

Under each distribution, $\mathrm{E}[\nu_{st}] = 0$ and $\mathrm{Var}(\nu_{st}) = \psi$.

For each combination of sampling variance pattern and random effect distribution, we generated 1000 data sets from model (10). For each simulated data set, we used the `R` package `sae` (Molina and Marhuenda, 2015) to compute $\hat{\psi}$ via REML and $\widehat{\boldsymbol{\beta}}$. We computed the EBLUP's in (16) for the mail targets $\{M_{st}\}$, approximated their MSE's using (17), and estimated their MSE's using (18). We then compared the approximations and the estimates to the true (Monte Carlo) MSE's over the 1000 simulated realizations.

Figure 2 shows plots of the MSE approximation and the estimated MSE versus the true MSE for each of the nine simulation scenarios. Here the gray dots are the MSE approximations and the black circles are the estimated MSE's. The approximations and estimates are nearly overlapping in all cases, indicating that the MSE estimates are essentially unbiased for the MSE approximations. Further, the points are all very close to the (0,1) reference line, indicating that the proposed methodology yields acceptable MSE estimates across a range of settings.

4.2 Calibration of the CHTS and FES estimates

For the data described in §1, we used the `R` package `sae` (Molina and Marhuenda, 2015) to fit a number of models via maximum likelihood for both shore fishing and private boat fishing, and compared the models via their AIC values. The smallest model considered included intercept, log(population), state indicators, wave indicators, state by log(population) interaction, and state by wave interaction. That is, the smallest model includes no differences due to survey methodology and instead drops the terms $\boldsymbol{b}_{st}'\boldsymbol{\mu}$ and

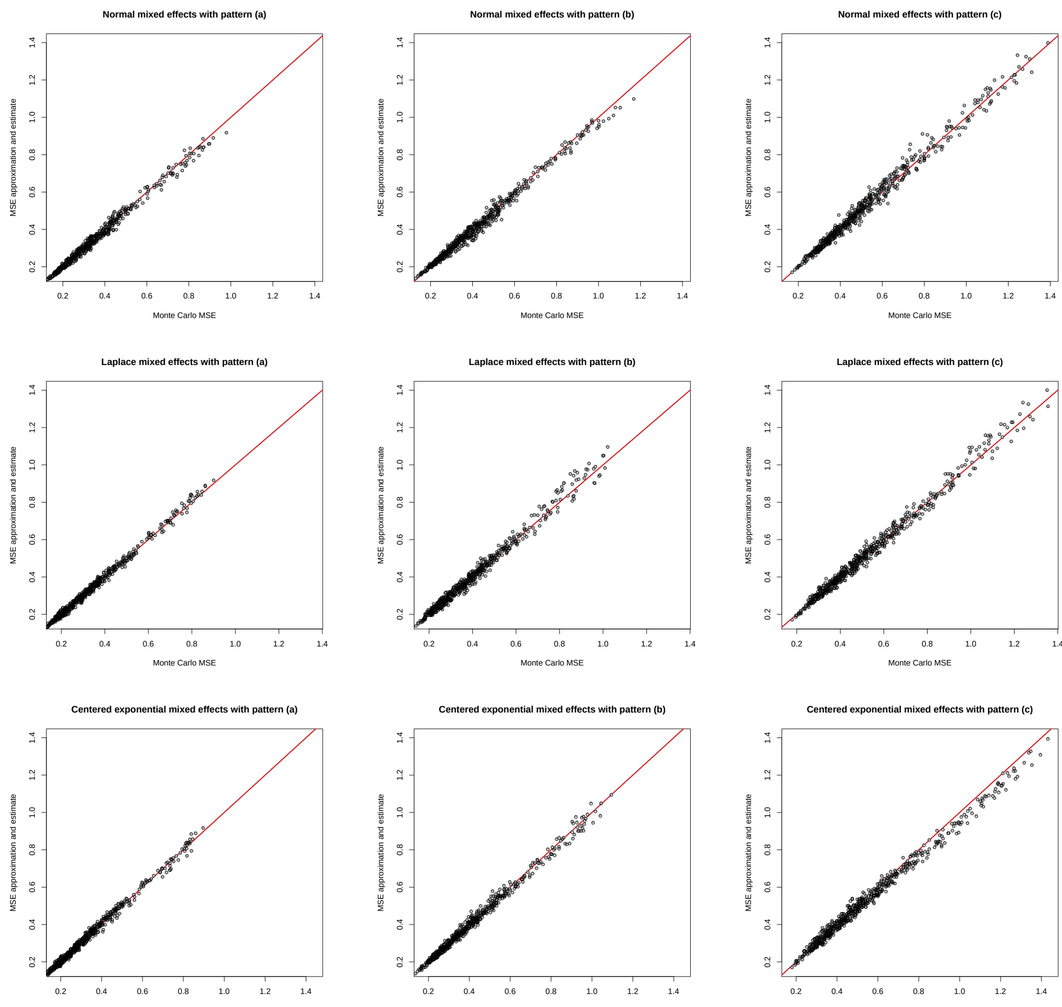


Figure 2: MSE approximation (solid gray dots) and estimated MSE's (open black circles) versus true MSE from Monte Carlo, for random effect distributions normal, Laplace, and centered exponential across the rows, and sampling error patterns (a), (b), and (c) across the columns.

$w_{st}\boldsymbol{c}_{st}'\boldsymbol{\gamma}$ from (1). The largest model considered added wireless and its interactions with log(population), state indicators, wave indicators, and state by log(population), together with an indicator for presence of a mail survey estimate and the mail indicator's interactions with log(population), state indicators, and wave indicators. The omission of the higher order interactions between wireless and the mail indicator is due to parsimony: for the mail indicator in particular, there are only 17 states and 11 waves from which to estimate the parameters $\boldsymbol{\mu}$ in model (1).

Numerous submodels between the smallest and largest were considered; the best four models and additional reference models are given in Table 1 for shore fishing and Table 2 for private boat fishing. The tables are ordered by AIC values, with the best models at the top. The models that ignore some (largest minus all mail, largest minus all wireless) or all (smallest) of the survey mode differences are not competitive with the models that include these factors. The largest model considered is quite competitive, with the best models dropping a small number of interactions from that largest model.

While not the best model for either shore or private boat, the largest model minus the mail by log(population) interaction is third best in both cases. It is operationally convenient to use a common model for both calibrations, and this particular model is further convenient because, when extrapolating back in time, it involves only state by wave level shifts once the effect of wireless has died out. We therefore chose this model as the final model for both modes of fishing, and refitted it using REML to estimate the unknown variance ψ. We then computed EBLUP's of the mail target $\{M_{st}\}$ for all states and waves.

An example for Alabama shore fishing is shown in Figure 3 and an example for Florida private boat fishing is shown in Figure 4. In each figure, we show the effects of successive adjustment, from the telephone log-effort estimates $\{\widehat{T}_{st}\}$, to the estimates $\{\widehat{T}_{st} + \boldsymbol{b}_{st}'\widehat{\boldsymbol{\mu}}\}$ that adjust only for mail methodology effects, to the estimates $\{\widehat{T}_{st} + \boldsymbol{b}_{st}'\widehat{\boldsymbol{\mu}} - w_{st}\boldsymbol{c}_{st}'\widehat{\boldsymbol{\gamma}}\}$ that adjust for both mail and wireless, and finally the EBLUP's themselves. As expected, the effect of wireless is only present in the later years since 2000, and is a relatively modest effect. The EBLUP can be seen as a smoothed version of the estimates adjusted for mail methodology and wireless effects.

Model is largest minus terms below:	log(likelihood)	AIC	df
mail:log(pop) and wireless:wave	-1803.53	3947.06	2798
mail:log(pop), mail:wave, wireless:wave	-1810.49	3950.99	2803
mail:log(pop)	-1801.57	3953.14	2793
nothing (largest)	-1801.23	3954.47	2792
mail:log(pop) and mail:wave	-1808.48	3956.96	2798
mail:log(pop) and mail:state	-1821.50	3961.01	2809
mail interactions	-1828.03	3964.07	2814
wireless interactions	-1942.98	4161.97	2830
all interactions	-1969.05	4170.10	2852
all mail	-1935.15	4176.30	2815
all wireless	-1977.54	4229.09	2831
all mail and all wireless (smallest)	-2109.83	4447.66	2854

Table 1: Maximized log(likelihood), AIC and residual degrees of freedom (df) for various models fitted to effort estimates for shore fishing. See text for description of largest model.

Model is largest minus terms below:	log(likelihood)	AIC	df
mail interactions	-1336.00	2981.99	2816
mail:log(pop) and mail:wave	-1320.07	2982.13	2800
mail:log(pop)	-1315.48	2982.97	2795
mail:log(pop) and mail:state	-1331.70	2983.40	2811
nothing (largest)	-1314.83	2983.66	2794
mail:log(pop) and wireless:wave	-1323.26	2988.52	2800
mail:log(pop), mail:wave, wireless:wave	-1332.19	2996.37	2805
all mail	-1417.45	3142.90	2817
wireless interactions	-1463.00	3204.01	2832
all interactions	-1495.69	3225.37	2854
all wireless	-1548.81	3373.62	2833
all mail and all wireless (smallest)	-1611.74	3453.48	2856

Table 2: Maximized log(likelihood), AIC and residual degrees of freedom (df) for various models fitted to effort estimates for private boat fishing. See text for description of largest model.

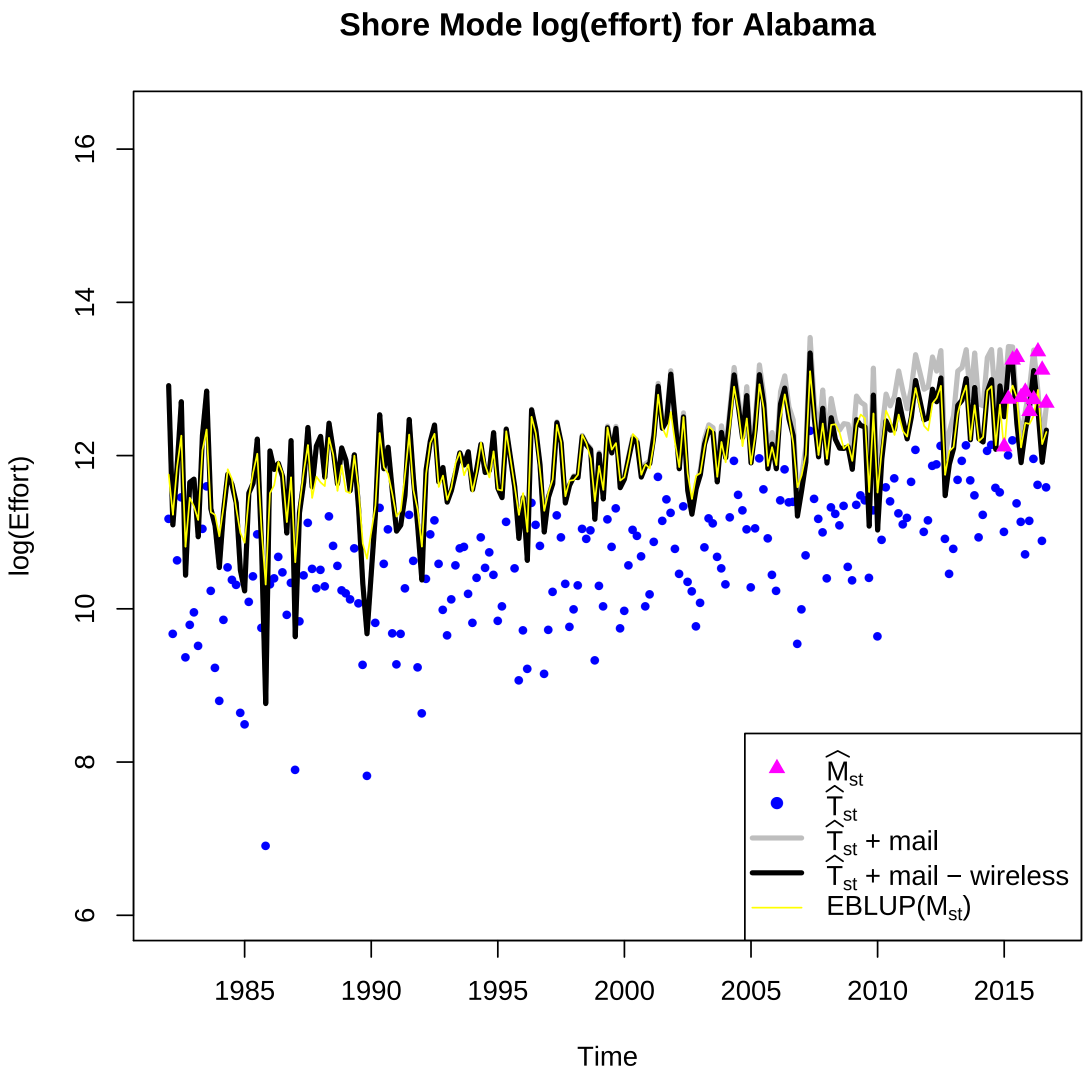


Figure 3: EBLUP's $\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \hat{\psi}\right)\right\}$ (gold curve) of mail targets $\{M_{st}\}$ for shore fishing log-effort in Alabama. Blue dots are telephone log-effort estimates $\{\widehat{T}_{st}\}$ and pink triangles are mail log-effort estimates $\{\widehat{M}_{st}\}$. For comparison to EBLUP's, gray curve is the estimator $\{\widehat{T}_{st} + \boldsymbol{b}'_{st}\widehat{\boldsymbol{\mu}}\}$ that adjusts only for mail methodology effects, and black curve is $\{\widehat{T}_{st} + \boldsymbol{b}'_{st}\widehat{\boldsymbol{\mu}} - w_{st}\boldsymbol{c}'_{st}\widehat{\boldsymbol{\gamma}}\}$ that adjusts for mail and wireless.

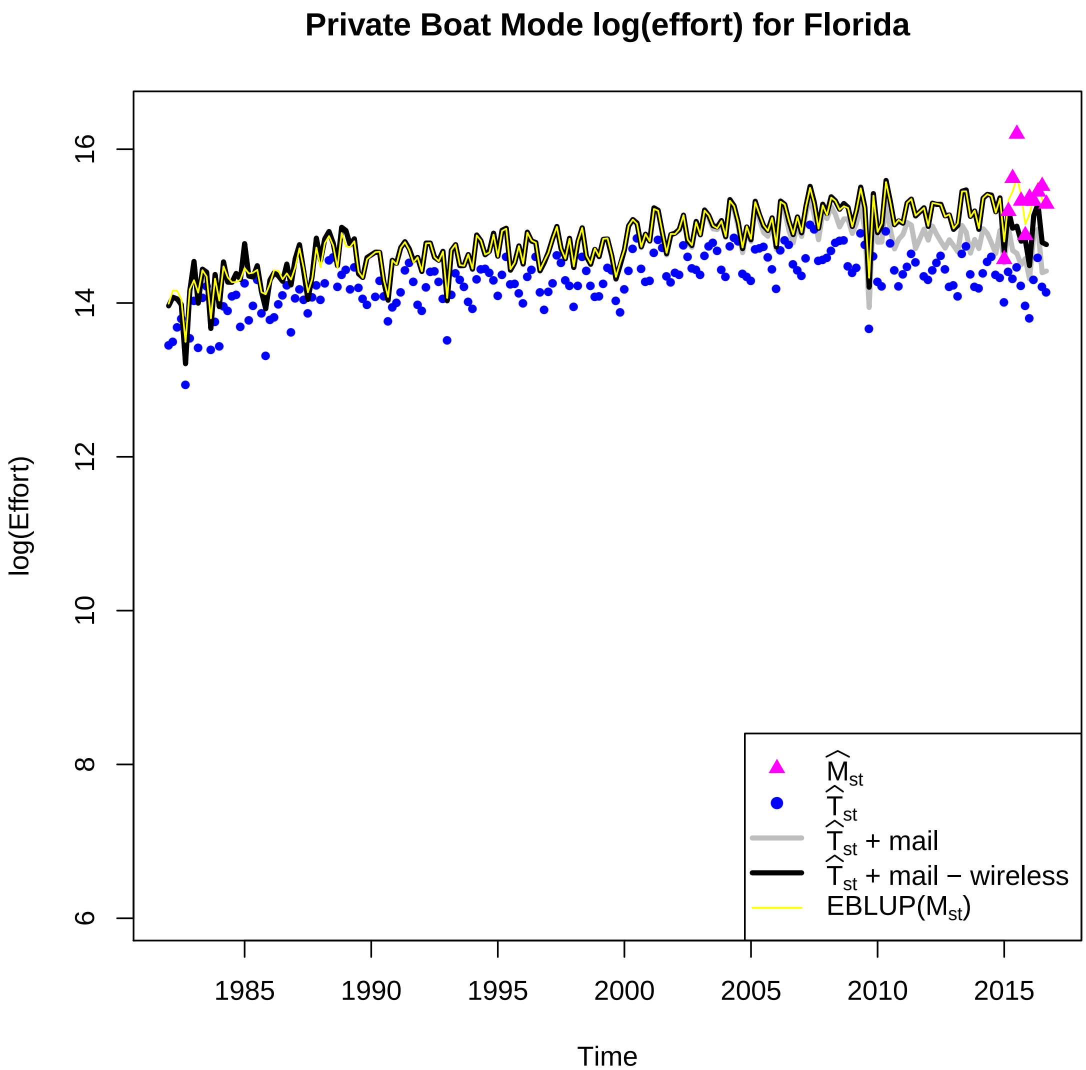


Figure 4: EBLUP's $\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \hat{\psi}\right)\right\}$ (gold curve) of mail targets $\{M_{st}\}$ for private boat fishing in Florida. Blue dots are telephone log-effort estimates $\{\widehat{T}_{st}\}$ and pink triangles are mail log-effort estimates $\{\widehat{M}_{st}\}$. For comparison to EBLUP's, gray curve is the estimator $\{\widehat{T}_{st} + \boldsymbol{b}'_{st}\widehat{\boldsymbol{\mu}}\}$ that adjusts only for mail methodology effects, and black curve is $\{\widehat{T}_{st} + \boldsymbol{b}'_{st}\widehat{\boldsymbol{\mu}} - w_{st}\boldsymbol{c}'_{st}\widehat{\boldsymbol{\gamma}}\}$ that adjusts for mail and wireless.

5 Discussion

The proposed methodology accounts for various sources of variation in the effort series from each survey, including trend, seasonality and irregular terms in the true effort series, together with survey mode effects in the two series. The model assumes that differences in measurement and nonresponse errors between the two surveys would be stable over time, while the changes in coverage error over time due to growth in wireless-only households is explicitly modeled. Further, the methodology accounts for uncertainty due to sampling error, using a novel approach to ensure analytical consistency in mapping design variances estimated on the original scale to design variances estimated on the log scale.

As formulated in this paper, the calibration methodology turns out to follow a standard, well-established procedure: Fay-Herriot small area estimation. This means that the calibrated values turn out to empirical best linear unbiased predictors under a linear mixed model fitted using likelihood-based techniques. The method is flexible enough to provide optimal calibrated values for different problems: predicting mail targets using telephone-only data, or predicting telephone targets using mail-only data, for example.

Uncertainty is quantified via a mean square error approximation that adapts existing methods from the literature. Simulation results show that the mean square error approximation and its estimator are highly accurate for the kinds of sample sizes and sampling errors present in the calibration data. The methodology is readily implemented with standard software.

Acknowledgements.

We thank NOAA Fisheries scientists Rob Andrews and John Foster for assistance with the problem formulation and all aspects of the data compilation. NOAA affiliate Ryan Kitts-Jensen and Westat statistician Mike Brick provided useful discussion.

References

Andrews, R., J. M. Brick, and N. A. Mathiowetz (2014). Development and testing of recreational fishing effort surveys. Technical report.

Blumberg, S. and J. Luke (2013). Wireless substitution: Early release of estimates from the National Health Interview Survey, July–December 2012. Technical report, National Center for Health Statistics.

Curtin, R., S. Presser, and E. Singer (2005). Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly 69*(1), 87.

Datta, G. S. and P. Lahiri (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica 10*, 613–627.

Datta, G. S., J. N. K. Rao, and D. D. Smith (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika 92*(1), 183–196.

Fay, R. E. and R. A. Herriot (1979). Estimation of income from small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association 74*, 269–277.

Molina, I. and Y. Marhuenda (2015). `sae`: An `R` package for small area estimation. *The R Journal 7/1*, 81–98.

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

US Census Bureau (2016). *State Population Totals Datasets: 2010-2016*. `https://www.census.gov/data/datasets/2016/demo/popest/state-total.html`.

Wolter, K. M. (2007). *Introduction to Variance Estimation* (2 ed.). New York: Springer-Verlag Inc.

Summary Report
Marine Recreational Information Program (MRIP)
Fishing Effort Survey (FES) Calibration Review

Calibration Model Review Meeting
June 27-29, 2017
Sheraton Hotel
Silver Spring, MD

December 5, 2017

Final

Panel Members

Paul J. Rago (Chair)
Ali Arab
Robert L. Hicks
Cynthia M. Jones
Jason McNamee
Fredric M. Serchuk
Patrick J. Sullivan

Table of Contents

Executive Summary

A primary objective of the Marine Recreation Information Program (MRIP) is the improvement of the statistical basis of methods for estimating catches of recreationally caught fish in the coastal US. MRIP has implemented a new program for estimating fishing effort that relies on a mail-based survey rather than a historical telephone survey. This report summarizes a technical review of a calibration model to interrelate estimates of recreational shore and private boat fishing effort derived from the Coastal Household Telephone Survey (CHTS) with estimates derived from the new Fishing Effort Survey (FES). The FES is a mail survey that utilizes address-based sampling and a national angler registry. A panel of seven independent scientists met with consultant statisticians and MRIP staff to review a proposed methodology that could express historical estimates of fishing effort in terms of the new FES. A side-by-side experiment of the two methods, conducted in 2015 and 2106, served as the basis for this review.

The proposed methodology builds upon known properties of the CHTS and FES sampling designs, and an extensive time series of historical data. The calibration model relies on standard and highly-regarded methodology known as the Fay-Herriot method for small area estimation. Alternative modeling approaches might have been considered, but the proposed method was reasonable and scientificallydefensible. The authors are commended for introducing several innovations to estimate variances and to achieve analytical consistency. The final estimators have desirable properties and can be implemented with readily available software. The proposed model was considered an elegant approach for dynamic predictions of recreational fishing effort. Particularly notable was the property that allowed for forward and backward estimation by alternate survey modes (i.e., CHTS vs FES). The proposed method preserves design aspects of historical and current surveys and incorporates important differences among states, waves (i.e., two-month calendar periods) and fishing modes. The processes of model identification and variable selection (i.e., consideration of potential predictive covariates) were well done.

The Panel expressed concern on several topics, none of which was considered as sufficient to preclude implementation of the Fay-Herriot model. Comparison of estimates of effort derived from the side-by-side CHTS and FES surveys (2015 and 2016) resulted in large differences (2 to 11-fold). While many hypotheses were considered that might account for these differences, data analyses and the proposed model revealed no single hypothesis (or covariate) was sufficient. Further refinement of the modeling approach, particularly when the results of the 2017 side-by-side experiment are available, is recommended. Refinements include further simulation testing and cross-validation comparisons with the first two years of data. As more information is acquired about the FES there may be additional opportunities to consider alternative models for calibration. Given the importance of such changes for many stock assessments and management decisions, future modifications must be able to demonstrate significant advantages over the proposed small-area estimation model prior to consideration for implementation. The Panel recommended additional efforts to improve communication of these results to scientists, statisticians, fishery managers, and the general public. Each will require varying levels of detail. The Panel also suggests that renewed attention be given to the communication recommendations of two previous NAS reviews of the recreational statistics programs.

1. Introduction

1.1 Background

The Review Panel for the MRIP-FES Calibration Model Review met from June 27 to June 29 to review a statistical model developed by F. Jay Breidt, Teng Liu and Jean D. Opsomer, of Colorado State University. The review committee was composed of three scientists appointed by the Center for Independent Experts (CIE): Robert Hicks, The College of William and Mary, Cynthia Jones, Old Dominion University and Ali Arab, Georgetown University. In addition, representatives from the New England (Patrick Sullivan) and South Atlantic (Fredric Serchuk) Scientific and Statistical Committees, and the Atlantic States Marine Fisheries Commission (Jason McNamee) served on the review panel. The meeting was chaired by Paul Rago as a member of the Mid-Atlantic Fishery Management Council Scientific and Statistical Committee.

The panel reviewed supporting documentation and presentations prepared by NOAA Fisheries' Office of Science and Technology (OST) staff, led by Dave Van Voorhees, and their contractors from the Department of Statistics at Colorado State University. John Foster, Ryan Kitts-Jensen, and Richard Cody acted as rapporteurs, providing valuable daily summaries for the Panel. Other staff and contractors from the OST, notably Karen Pianka, assisted in the efficient handling of documents via a web-based application. Jason Didden of the Mid Atlantic Fishery Management Council provided extensive support for the webinar. Approximately 35 people participated in the open sessions of the meeting. The meeting followed the agenda in Appendix 2 with respect to the sequence but not necessarily the timing of the events. Adjustments were made for differences in the duration of presentations and follow-up questions.

1.2 Review of Activities

About ten days before the meeting the panel was given access to a comprehensive working paper summarizing the proposed statistical model. Prior to the meeting, the chair met with the presenters and MRIP staff via a conference call to discuss the scope of the contributions, presentation format and draft agenda. All supporting documents and presentations were made available to reviewers via a web-based application known as Confluence. In addition, the MRIP staff added a web page to their site that provided members of the public and other managers with access to key papers and presentations. The meetings were broadcast via webinar with the able assistance of Jason Didden of the Mid-Atlantic Fishery Management Council. Mr. Didden also managed all of the in-room computer and audio visual equipment.

The meeting opened on the morning of Tuesday June 27, 2017, with welcoming remarks and comments on the agenda by Van Voorhees and Rago. Participants and audience members introduced themselves. Following introductions, sessions on June 27 were devoted to presentation and initial discussions of five agenda topics. Rob Andrews provided an overview of the pilot study work that led to the development of a new mail survey design (the Fishing Effort Survey, or FES) as a replacement for the legacy telephone survey design (the Coastal Household Telephone Survey, or CHTS). Richard Methot addressed the importance of properly calibrated effort for estimation of catch in stock assessments. Andy Strelcheck addressed the importance of catch information as a basis for fisheries management policies and decisions, such as allocation. Jean Opsomer provided an overview of the challenges of applying calibration methods to historical time series. Jay Breidt led the presentation of the proposed statistical calibration model.

Each presentation was followed by a question and answer period by panel members and as appropriate, by other meeting attendees. Questions from web participants were also addressed at opportune times. A formal public comment period was reserved on each day of the meeting.

The Panel met in closed session at the end of each day to discuss the day's presentations, progress toward answering the agenda, and to make plans for the following day.

Follow-up discussions on the first day presentations were held on Wednesday June 28. The Panel requested additional data and clarification from the presenters, including greater details on the model results. Day two began with an overview of the activities of Day One and an overview of the day's work plan. Most of the Panel's efforts were devoted to questions on the statistical calibration model. Material provided by Jay Breidt and colleagues enhanced the Panel's understanding of the model and its performance. A short presentation by Paul Rago used the results of model predictions to compare results over states and fishing modes (i.e., shore vs private boat).

Day Two also included a formal public comment period and an initial summary of the Panel's findings. This was done to ensure that all participants were aware of the general outcomes of the review. The Panel stressed that this summary was not to be considered a consensus report. Instead it represented a summary of the perspectives of the Panel.

Following the initial presentation of findings, the Panel met in closed session to begin writing the Summary Report. Day Three consisted of a half day meeting for Panelists only. The purpose of the meeting was to summarize the various viewpoints herein with respect to the Terms of Reference.

The Panel completed drafting this Summary Report by correspondence, evaluating each ToR. The Chair compiled and edited the draft Panel Summary Report, which was distributed to the Panel for final review before being submitted to the MRIP. Each Panelist also provided an independent summary of their perspectives and as appropriate, with details on potential improvements to the calibration model and its application. Individual panelist reports for CIE participants were sent to the Center for Independent Experts for initial editing for completeness. Reports of Panelists supported directly by the Agency via contract were sent to the Chair. All reports were made available to MRIP staff for fact checking but were not altered for content.

The Panel agreed that scientific and statistical analyses conducted by the presenters were thorough, statistically sound, and innovative. Specific comments on the details of the analyses are provided below.

2. Review of MRIP FES Calibration Model

2.1 Synopsis of Panel Review

The Panel commented that the proposed methodology builds upon known properties of the existing sampling design, the proposed new method, and extensive time series of historical data. A review of calibration approaches in other disciplines revealed no comparable attempts to adjust a historical times series forward or backward in time in response to new information from a side-by-side comparative surveys. The proposed model was considered to be an elegant approach for dynamic predictions of recreational fishing effort. Particularly notable was the property that allowed for forward and backward estimation by alternate survey modes (i.e., CHTS vs FES). Notably, the proposed method preserves design aspects of historical and current surveys and incorporates important differences among states, waves (i.e., two-month calendar periods) and fishing modes. The Panel acknowledged the extensive exploratory data analyses on model development, alternatives, and testing performed by the MRIP scientific staff and consultants. The processes of model identification and variable selection (i.e., consideration of potential predictive covariates) were well done.

Although the Panel identified several alternative modeling approaches and other candidate covariates that might have been considered, the Panel acknowledged that the proposed method was a reasonable and scientifically defensible estimation approach.

The calibration model relies on standard, well known, and highly regarded methodology. The authors are commended for introducing several innovations to estimate variances and to achieve analytical consistency. The final estimators have desirable properties and can be implemented with readily available software.

The Panel expressed concern on several topics, none of which was considered as sufficient to preclude implementation of the model. Comparison of estimates of effort derived from the side-by-side CHTS and FES surveys (2015 and 2016) resulted in large differences (2 to 11-fold). While many hypotheses were considered that might account for these differences, data analyses and the proposed model revealed no single hypothesis (or covariate) was sufficient.

Model performance was partially assessed by sensitivity analysis of specific alternative hypotheses on the distribution of the "irregular" random effect (an effort effect not accounted for explicitly in the model). However, additional simulation work may be necessary to more thoroughly test overall model performance. As additional information becomes available by the end of the 2017 side-by-side surveys, it is recommended that a series of cross-validation exercises be conducted to compare model results based on the first two years of model results. Other permutations of cross calibration comparisons may be instructive with respect to stability of model parameter estimates and prediction error induced by various data rarefaction methods. As more information is acquired about the FES there may be additional opportunities to consider models for calibration that include alternative causal factors. Given the importance of such changes for many stock assessments and management decisions, future modifications must be able to demonstrate significant advantages over the proposed small-area estimation model prior to consideration for implementation.

The Panel spent considerable time discussing the communication of results. It was recognized that at least three distinct audiences must be addressed: scientists and statisticians, fishery managers, and the general public. Each will require varying levels of detail without compromising the integrity of the model or its underlying principles. A "lay person's" version of the methods would be valuable for communicating results to multiple audiences. Model results, in combination with a similar calibration exercise for the APAIS, have significant downstream impacts for assessments and management. The Panel also suggests that renewed attention be given to the recommendations concerning communications of two previous NAS reviews of the recreational statistics programs.

Despite progress in improving communication with stakeholders, the some members of the Panel, working directly with fishermen, are aware of important misconceptions among the angling communities regarding the transition to the new mail-based survey mode. The new MRIP website is a considerable improvement but direct, pro-active communication and dialogue with fishing groups, perhaps with downloadable podcasts, YouTubes etc. and in-person presentations to the angling community would be valuable.

2.2Evaluation of Terms of Reference

2.2.1 Term of Reference 1

Evaluate the suitability of the proposed model for converting historical estimates of private boat and shore fishing effort produced by the CHTS design to estimates that best represent what would have been produced had the new FES design been used prior to 2017.

- The Panel concurs that this TOR and its subcomponents listed below (1a,1b, 1c, 1d, 1e) were met.

a) Does the proposed model adequately account for differences observed in the estimates produced by the CHTS and FES designs when conducted side-by-side in 2015-2016?

- The results of the side-by-side surveys are central to the development of the proposed model. The model parameterization accounts for these changes but does not provide insight into the underlying mechanisms resulting in differences in estimated angling effort.
- The new mail survey mode has advantages relative to issues of comprehensiveness of angler coverage within households, efficiency of the estimate, a better sampling frame, a more thoughtful consideration of individual angler effort, improved demographic information, better identification of angler residence and enhanced follow-up with respondents to reduce non-response. Collectively these features are thought to yield more reliable metrics of angling effort and serve as a basis for improved understanding in the future as the new survey continues. These advantages are relevant to 2015 and onward but do not necessarily extend back to historical estimates.

b) Is the proposed model robust enough to account for potential differences that would have been observed if the two designs had been conducted side-by-side in years prior to 2015 with regards to time trending biases?

- The Panel had difficulty formulating a response to this TOR as it required conjecture about unidentified underlying causal mechanisms contributing to observed differences and hypothetical comparisons of survey mode responses in the past.
- Insufficient information was provided to inform this decision either before or during the meeting.
- Although the proposed model allows for inclusion of other causal mechanisms, neither the investigators nor the Panel were able to identify covariates that vary over time and meet the criteria necessary for expansion to total angling effort estimates. Furthermore, data collection procedures during the CHTS did not collect information that in retrospect (e. g., demography, gender of angler), might have allowed such inference.

c) How does the approach used in developing the proposed FES/CHTS calibration model compare in terms of strengths or weaknesses with other potential approaches?

- The investigators conducted an extensive analysis of within-model comparisons of reduced model parameterizations using the model selection procedure known as the Akaike Information Criterion. One sub-model included a simple ratio with random effects that had much lower explanatory power. A preliminary analysis was conducted and reviewed by the Panel that corroborated the inappropriateness of the simple ratio estimator.
- Other models exist that could be used, including Bayesian Hierarchical modeling, state-space modeling, and time-varying ratio estimation. The investigators provided the panel with a summary of their experiences with some of these alternatives but the results of these comparisons were not available to the Panel. Given the responses of the investigators, the Panel concurred with the conclusion to focus on the modified Fay-Herriot approach.

d) Does the proposed calibration model help to explain how different factors would have contributed to changes in differences between CHTS and FES results **over time**?

- As noted above a complete set of causal mechanisms resulting in differences between survey estimates remain elusive.
- Raw survey data in the CHTS (rather than aggregated data provided by contractors) could be examined more carefully but it is unknown whether such data exist over a sufficient span of years to support such analyses
  - o As presently configured the model is limited in terms of what can be explored but alternative calibration models may be useful.
  - o Within the existing data, there do not appear to be covariates, other than log(Population) that would explain the major differences seen between survey modes (i.e., CHTS vs FES). The wireless effect

captures a minor component of the contrast. The Panel and Investigators agreed that the wireless effect may be a proxy for a wide range of factors.
  - Demographic information in the CHTS would have been instructive and is essential for proper historical analyses. However, it is uncertain that such data exist over a sufficient span of years to support such analyses.
  - Consideration of spatially differentiated data that has been collected historically at a finer scale (e.g., Census tract) may yet contain information sufficient to illuminate explanatory factors related to this TOR.
- The "Gatekeeper" effect has been proposed as a major influence in the CHTS but a complete understanding remains difficult to identify.

e) Is it reasonable to conclude that revised 1981-2016 private boat and shore fishing effort estimates based on the application of the proposed FES/CHTS calibration model would be more accurate than the estimates that are currently available? Does evidence provided for this determination include an assessment of model uncertainty?

- No conclusions can be reached regarding the accuracy of calibrating self-reported data from one survey mode to the other. However, the Panel noted that bias in the historical CHTS may not be as large as observed in contemporary CHTS samples due to degradation of survey coverage and other temporal trends in other factors such as privacy concerns.
- Gatekeeper effect, recall bias, response rate etc. indicate that the mail survey is preferred to a phone survey, particularly in relation to statistical and operational efficiency. This conclusion was supported by the 2006 and 2017 NRC reports, and also in a separate review conducted by independently selected members of the American Statistical Association's Survey Research Methods Section.

- Response rate per se is not a problem unless differences in fishing activity differ between respondents and non-respondents

2.2.2 Term of Reference 2

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

The following sections highlight the Panel's concerns about the peer review meeting, including preparations before the meeting and follow-up activities. The Panel recognizes the complexity of the revisions of MRIP transition process and the need to satisfy many different audiences. The following recommendations are offered in the context of constructive criticism to improve the quality of future peer-review panels. While there is some redundancy in this section with the Panel's comments in section 2.1, the text below provides additional clarification of issues and more broadly reflects the diversity of the Panelist's opinions. The text below draws heavily from comments provided by the Panelists via correspondence after the meeting. Therefore some sections below may be reflected in part or their entirety in the Panelist's individual reports.

Pre-Meeting Preparations

Four background documents (Section 5 , Working Papers) were provided to Panel members two weeks prior to the meeting, and all additional documents and presentation were made available to the Panel during the meeting via a web-site (i.e., Confluence). The Panel Chair provided each of the reviewers with a proposed meeting Agenda a day prior to the start of the meeting, requesting that any comments and possible changes be provided back to him before the meeting opened. As the proposed Agenda was satisfactory to all of the Panel members, no changes to the Agenda were needed.

Panelists expressed concerns about pre-meeting preparations, noting an inadequate assembly of all the pieces needed to address the terms of reference. Greater overall coordination among presenters would have been desirable to ensure that all the relevant information was covered. Additional background documents would have been useful for the review; for example, the MRIP Handbook should have been provided before to provide more information about the telephone and mail surveys. Comprehensive previous reviews of the MRIP, such as those from the National Academy of Sciences should have been brought to the attention of the Panel, not all of whom had extensive knowledge of the history of MRIP. In this context, basic details about the surveys including similarities and differences in definitions of effort (notably, the definition of angling households), questions on the questionnaires, etc. would have helped the Panel to more effectively conduct the review A valuable adjunct to future technical reviews might be a targeted guide to relevant resources available on the extensive MRIP website.

Proceedings

The review panel proceedings went smoothly. Operationally, the meeting room had sufficient space for the Panel, presenters, and meeting attendees. The sound and projection systems worked well, as did the webinar link. Representatives from the Office of Science and Technology served as Rapporteurs and provided comprehensive summary notes to the Panel.

Discussions during the 2½ day MRIP Calibration Review illuminated various issues related to the results provided in the background documents and the PowerPoint presentations. Many of the concerns involved clarification of the information provided and/or requests for additional data and analyses. Additional data, model outputs and documents were made available to the Panel during the meeting. In all cases, these requests were satisfactorily fulfilled allowing the Panel to gain fuller insight on:

- Sampling designs, strengths, and shortcomings of the telephone (CHTS) and mail (FES) survey methods, including their relative performance and sources of error.

- Development, design, statistical properties, testing, and application of the proposed MRIP FES calibration model. This included consideration of alternative modeling approaches, cross-validation of the modeling framework to examine the stability of model parameter estimates (as well as prediction errors), the sufficiency and explanatory power of the model's covariates, and the possible underlying mechanism(s) affecting the distribution of the "irregular" random effect, which is not explicitly accounted for within the proposed small-area estimation approach.

- Potential impacts of the calibrated recreational fishing effort estimates during 1981-2016 on future stock assessments, and on subsequent fishery management policies and practices.

- Need to effectively communicate the results of the calibration work (as well as the basis and need for continuing only the mail-based survey method in the future) to various constituency groups (i.e., the recreational and commercial fishing communities; scientists; fishery managers; the lay public) so that these groups fully understand and accept the calibration results and their subsequent use in deriving recreational catch estimates for application in stock assessments and in the fishery management process.

The Review Panel acknowledged that the proposed MRIP FES calibration model developed by Breidt *et al.* was a well-suited and statistically-appropriate approach to obtain calibrated estimates of recreational fishing effort (by state and 2-month calendar quarter for shore-based and private boat anglers) during 1982-2016.

<u>Utility of Presentations</u>

The presentations on the implications of revised recreational catch estimates on stock assessments and on management measures and regulatory protocols were instructive, but the Panel would have appreciated more quantitative examples. For example, implications for stock assessment models could have been drawn from the previously completed scoping exercises conducted by the Northeast and Southeast Fisheries Science Centers. Similarly, the Panel noted that detailed simulation exercises would also have been instructive.

The presentation on the Fay-Herriot model was lucid and effective, but the Panel would have appreciated more details on the model components and the model building process. Also, a summary of candidate modeling approaches —and details on the process that led to the proposed model—would have been very useful. Such details, as provided on the second day of the review, were greatly appreciated.

Greater detail would have been appreciated on the survey methodologies in the phone and mail surveys. The simulation exercise was an important start, but further simulation testing beyond those conducted would have lent greater support to the applicability of the Fay-Herriot model to the CHTS vs FES calibration. Further work on simulated data sets is suggested during the third-year comparisons (i.e., when the 2017 telephone and mail survey data are fully available).

Terms of Reference

The presenters did not address the TORs directly, which made it harder for the Panel to assess the relevance of some of the information presented with regard to the TORs. Consequently, the Panel spent a substantial portion of the question/answer periods (and discussion time) on obtaining the requisite information to address the TORs. It was evident during these interactions that the model developers had conducted additional work relevant to the TORs (such as investigation of additional modeling approaches). However, because the developers were unaware of the TORs, neither the primary report nor the presentations specifically addressed the TORs. Follow-up work accomplished by the developers during the meeting and subsequently shared with the Panel gave the Panel confidence that sufficient model scoping had been performed.

The TORs presume that converting CHTS to FES is the appropriate way to standardize the MRIP effort data. However, the statistical work available for the review primarily focused on the mathematical aspects of the calibration and not on which set of estimates reflects a truer representation of fishing effort. Lacking a sufficient statistical justification for standardizing the MRIP data to the FES estimates created problems both during the review and in addressing the TORs.

TOR1e seeks the Panel's opinion concerning the accuracy of effort estimates obtained from the CHTS and the FES. The Panel understands that any survey conducted offsite of the fishery, such as mail or telephone surveys, rely on angler self-reported data which is not subject to verification. Self-reported data is subject to a variety of biases including recall problems which can result in misremembered time and number of trips. Without an external measure of fishing from an onsite survey covering the same population in space and time, angler self-reported data cannot be verified. While the Panel comments on the calibration from CHTS to FES, there is no basis to comment on accuracy of either survey.

Documentation for Meeting

It would have been helpful for the Panel to have been provided (several weeks before the review) additional background documents (available from the MRIP Team and/or the MRIP Website) to enhance a collaborative understanding by Panel members of various aspects of the MRIP program and of recent analyses using MRIP data. For example, the *MRIP Data User Handbook,* and the October 2016 report, '*Possible Effects of Calibration Scenarios on Stock Assessments Planned for the MRIP Fishing Effort Survey*

Transition' would have especially useful for Panel members to have had and read before the actual peer review occurred.

Prior to the presentation and discussion of the Breidt *et al.* report at the Peer Review, this report was difficult to understand for anyone other than a highly-trained statistician. Although a more complete understanding of this report was fostered by distribution of a PowerPoint presentation a week or so before the Review Meeting (and subsequently enhanced at the meeting by direct dialogue and interaction with the authors of the paper who clarified and responded to many issues raised by the Panel), it is recommended that in any future reviews in which a highly technical paper is seminal to the crux of such reviews that efforts be made by the paper authors to present the essence of their work in a manner that facilitates full appreciation and understanding of the import of such work by educated non-specialists. This becomes especially critical when the methods/approach provided in a paper will have significant downstream effects. This matter should be recognized in the future APAIS peer review.

<u>Ancillary Analyses</u>

The Panel appreciated the opportunity to investigate the details of the statistical calibration/prediction model on day 2. The model and assumptions were well thought out, but the Panel needed to better understand model inputs, parameter definitions, and nuances of the Fay-Herriot model. Similarly, the Panel appreciated the opportunity to solicit more information on model development and model selection beyond what was initially available at the meeting. Panelists received model parameter estimates upon request but did not have time at the meeting to explore them fully. Access to more detailed model outputs and the estimation code in R would have been valuable.

Also, apparently, several independent data analyses existed too, separate from the model, and it would have been good to have had a presentation and some discussion on that. Exploratory analyses of the pairwise calibration data was considered useful and should be considered for summarization when the analyses of the 2017 data are conducted.

<u>Communication</u>

Panelists expressed concerns about the need for improved communication at several different levels:

- to the Panel prior to the meeting,
- within the various analytical components,
- to the members of the Transition Team,
- to broader audience of stake holders.

An advantage of the current review was the inclusion of several external independent experts having expertise beyond fisheries science. This helped ensure that the methods were critically evaluated and represented state of the art, but increased the burden during pre-meeting preparations to ensure that all relevant contextual documents were available and fully comprehensible. Concerns were expressed that information essential for the review was not provided at level of detail that the Panel members expected.

The transition from the MRFSS to MRIP has required a massive restructuring of the data collection procedures while maintaining a continuous time series of reliable catch data. Continuity of data has required coordination with governmental, academic, and industry stakeholders. Likewise, the process has involved multiple experiments and survey tests to demonstrate the value of proposed changes and development of advanced calibration approaches. This review constituted one component of this transition. Despite enormous improvements in the MRIP website and availability of raw and processed data at varying degrees of resolution, the Panel recommended greater coordination among the diverse analytical groups. The complexity of the transition requires that technical reviews are both sequential and interdependent. As such the review of any single technical issue (e.g., calibration between CHTS and FES) must rely upon and recognize the conclusions of earlier Panels. In the present review, this Panel relied on the conclusions of the ASA reviewers who noted the superiority of the FES over CHTS. Independent panels of scientists rarely accept prior reviews without questioning. Indeed, this is the nature of science. Hence it essential that each Panel in future reviews be provided with a summary of the full set of previous reviews and their relationship to the current review.

There is a strong need to effectively communicate the results of the calibration work (as well as the basis and need for continuing only the mail-based survey method in the future) to various constituency groups (i.e., the recreational and commercial fishing communities; scientists; fishery managers; the lay public) so that these groups fully understand and accept the calibration results and their subsequent use in deriving recreational catch estimates for application in stock assessments and in the fishery management process. Consideration should be given to a variety of communication approaches including but not limited to public meetings, seminars, podcasts, YouTube, and use of skilled educators.

Finally, it is recommended that an updated report/timetable/chart be prepared to illustrate current progress in meeting the tasks and timelines identified in the FES Transition Plan. This undertaking should also take note of how the recommendations tendered in all previous peer reviews of the MRIP Program (including the 2006 and 2016 NAS Reviews) have been addressed.

Improvements to Future Peer Review Processes

The Panel noted that review process left little time for an intensive review of the data, the model, and the computer code used to develop the results. Such analyses are often part of a stock assessment review (e.g., SAW/SARC https://www.nefsc.noaa.gov/saw/, or SEDAR http://sedarweb.org/). In the spirit of improving future reviews, the Panel suggests consideration of more broadly based working groups based on scientific input within and outside NOAA Fisheries. In stock assessments working groups have a strong technical focus and meet several times prior to the final assessment. Working groups would have the opportunity to examine the proposed methodologies in greater detail, including detailed reviews of the data and methods, and tests with simulated data. Exchanges of code, or reliance on standard packages in stock assessments provide both quality assurance and opportunities for improvements. Moreover, the products of working groups typically assure subsequent reviewers that the products under review are comprehensive and representative of diverse viewpoints. In particular, a working-group process would document the model building process and allay concerns of reviewers who will always wonder why a particular alternative was not considered. Having those prior decisions as a matter of record would enhance the efficiency and quality of the review process.

The Panel recognizes that this recommendation would need to be part of the overall transition from MRFSS to MRIP. Indeed, the current Transition Team process that has regular updates on progress, conversations with stock assessment scientists and various stakeholders, and plans for upcoming tasks, already includes the essential elements of a more focused working group approach. In view of the importance of upcoming technical decisions for stock assessments, managers and harvesters, the Panel strongly urges consideration of this proposal.

3. Bibliography

Background Papers

Many papers and documents on the existing and proposed survey methodology may be found at the following website: http://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/effort-survey-improvements

Background on the MRIP Calibration Model Peer Review may be found at: https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/index.html

The National Academies of Sciences, Engineering, and Medicine. 2016. Review of the Marine Recreational Information Program (MRIP) Washington, DC: The National Academies Press. doi: 10.17226/24640 https://www.nap.edu/catalog/24640/review-of-the-marine-recreational-information-program

National Research Council. 2006. Review of Recreational Fisheries Survey Methods. Committee on the Review of Recreational Fisheries Survey Methods, ISBN: 0-309-66075-0, 202 pages. http://www.nap.edu/catalog/11616.html

Working Papers

Development and Testing of Recreational Fishing Effort Surveys Testing a Mail Survey Design: Final Report. Project Team Members: Rob Andrews, NOAA Fisheries, J. Michael Brick, Westat, Nancy A. Mathiowetz, University of Wisconsin-Milwaukee. July 31, 2014. https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-

Workshop/documents/Report_recommending_FES_to_replace_CHTS--Finalize_Design_of_Fishing_Effort_Surveys.pdf

Marine Recreational Information Program Fishing Effort Survey Transition Progress Report. October 28, 2016. https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/documents/2015_benchmarking_progress_report.pdf

Marine Recreational Information Program Transition Plan for the Fishing Effort Survey Prepared by the Atlantic and Gulf Subgroup of the Marine Recreational Information Program Transition Team May 5, 2015. https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/documents/MRIP_FES_Transition-Plan_FINAL.pdf

A Small Area Estimation Approach for Reconciling Mode Differences in Two Surveys of Recreational Fishing Effort draft: F. Jay Breidt, Teng Liu, Jean D. Opsomer Colorado State University June 10, 2017. https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/documents/DRAFT-Report_of_Calibration_Model.pdf

Presentations

Calibration_Scenarios-20161115.pdf

MRIP FES website link

FES Errors.pptx

Model_Fits.txt

Mode_3_logeffort_poly_fixed.pdf

Mode_7_logeffort_poly_fixed.pdf

The following 4 files are available at https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/pages/additional-materials.html:

EBLUPS.csv

EBLUPS_Variable_Names.csv

Eblup comparisons.docx

MRFSS Fish Hunt Comps.xlsx

Webinar Links

All open sections of the meeting were recorded and available for viewing at the following links.

0 - Intro - Paul Rago

1 - MRIP Fishing Effort Survey - Rob Andrews

2- Catch and Assessments - Rick Methot

3 - Management Implications - Andy Strelcheck

4 - Calibrating Survey Estimates over Time - Jean Opsomer

5 - Calibration from CHTS to FES - Jay Breidt

6 - Initial Calibration Review Discussion - Tuesday Afternoon

7 - Day Two, AM Discussion

8 - Day Two, PM Discussion

9 - Day Two, Initial Findings Summary

4. Appendices

Appendix 1. Terms of Reference for the MRIP FES Calibration Model Review
T*he Review Panel shall assess whether or not the MRIP Working Group has reasonably and satisfactorily completed the following actions.*

1. Evaluate the suitability of the proposed model for converting historical estimates of private boat and shore fishing effort produced by the CHTS design to estimates that best represent what would have been produced had the new FES design been used prior to 2017.

   f) Does the proposed model adequately account for differences observed in the estimates produced by the CHTS and FES designs when conducted side-by-side in 2015-2016?

   g) Is the proposed model robust enough to account for potential differences that would have been observed if the two designs had been conducted side-by-side in years prior to 2015 with regards to time trending biases?

   h) How does the approach used in developing the proposed FES/CHTS calibration model compare in terms of strengths or weaknesses with other potential approaches?

   i) Does the proposed calibration model help to explain how different factors would have contributed to changes in differences between CHTS and FES results over time?

   j) Is it reasonable to conclude that revised 1981-2016 private boat and shore fishing effort estimates based on the application of the proposed FES/CHTS calibration model would be more accurate than the estimates that are currently available? Does evidence provided for this determination include an assessment of model uncertainty?

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

Appendix 2. Draft Review Meeting Agenda

MRIP FES Calibration Review

Sheraton Silver Spring Hotel

Silver Spring, MD

June 27-29, 2017

Day	*Date*	*Time*	*Topic*	*Rapporteur*	*Presenter*
Tuesday	27-Jun	9:00 AM	Welcome and Opening Remarks	TBD	Van Voorhees
		9:20 AM	Introductions		
		9:30 AM	Overview of Meeting	TBD	Rago
		9:45 AM	MRIP Fishing Effort Survey	TBD	Andrews
		10:15 AM	Importance of Calibrated Catch for Stock Assessments	TBD	Methot
		10:45 AM	Break		
		11:00 AM	Importance of Calibrated Catch for Fisheries Management	TBD	Strelcheck
		11:30 AM	Calibrating Survey Estimates over Time	TBD	Opsomer
		12:00 PM	Lunch		
		1:30 PM	A Calibration Methodology for CHTS to FES Transition	TBD	Breidt
		3:30 PM	Break		
		3:45 PM	Public Comment	TBD	
		4:15 PM	Summary of Day 1	TBD	Rago
		4:45 PM	Review Panel Coordination and Writing (closed)		
		6:00 PM	Adjourn		
Wednesday	28-Jun	9:00 AM	Overview of Day 1 and Preview of Day 2	TBD	Rago
		9:10 AM	Follow-up Questions for Presenters	TBD	Various
		10:30 AM	Break		
		10:45 AM	Follow-up Questions for Presenters (cont.)	TBD	Various
		12:00 PM	Lunch		
		1:00 PM	Review Panel Coordination and Writing (closed)		
		2:30 PM	Initial Summary Findings of Review Panel (open)	TBD	Panel
		3:30 PM	Public Comment	TBD	
		4:00 PM	Review Panel Coordination and Writing (closed)		
		6:00 PM	Adjourn		
Thursday	29-Jun	9:00 AM	Review Panel Coordination and Writing (closed)		
		12:30 PM	Adjourn		

Closed sessions allow the panel to discuss and clarify technical issues, and begin initial writing of reports.
Attendance of public, staff and presenters, if at all, is by invitation only and for purposes of clarification.

Appendix 3. Individual Independent Peer Review Report Requirements

Statement of Work

National Oceanic and Atmospheric Administration (NOAA)

National Marine Fisheries Service (NMFS)

External Independent Peer Review

Calibration Model Accounting for a Recreational Fishery Survey Design Change

Background

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

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards.
(http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).

Scope

The Office of Science and Technology requests an independent peer review of a calibration model proposed for use in revising statistics produced by surveys of marine recreational fishing effort on the Atlantic coast and in the Gulf of Mexico. This calibration model is considered by the Marine Recreational Information Program (MRIP) to be very important to adjust historical time series of recreational effort and catch estimates in order to account for biases in past sampling and estimation methods that have become apparent with the development of a new, more statistically sound method. The calibration model is intended to account for past biases in private boat and shore fishing effort estimates that have resulted from the continued use of a legacy random-digit-dial telephone survey design that has degraded over time and will be replaced with the implementation of a new mail survey design (the "Fishing Effort Survey", or FES) in 2018.

Calibration Model for the Fishing Effort Survey

In 2015, MRIP formed a Transition Team to collaboratively plan a transition from a legacy telephone survey design to a new mail survey design for estimating private boat and shore fishing effort by marine recreational anglers. Since 2008, MRIP had conducted six pilot studies to determine the most accurate and efficient survey method for this purpose on the Atlantic and Gulf coasts. The most recent study, conducted in four states in 2012-2013, compared a new mail survey design with the Coastal Household Telephone Survey (CHTS) design that has been used since 1979. MRIP subjected the final report from the pilot project to external peer review in 2014 and certified the new survey design, called the Fishing Effort Survey (FES), in February 2015 as a suitable replacement for the CHTS. The FES is much less susceptible to potential sources of bias than the CHTS because it can reach more anglers, achieve higher response rates, and is less prone to possible recall errors. The pilot project results indicated that FES estimates were substantially higher than CHTS estimates for both private boat fishing and shore fishing.

MRIP recognized the FES should not be implemented immediately as a replacement for the CHTS, and a well thought out transition plan was needed to ensure that the phase-in of the FES is appropriately integrated into ongoing stock assessments and fisheries management actions in a way that minimizes disruptions to these processes, which are based on input from multiple data sources over lengthy time series. The Transition Plan developed by the Transition Team called for side-by-side benchmarking of the FES against the CHTS for three years (2015-2017) with the development and application of a calibration model to enable adjustment of past estimates that account for biases in historical effort and catch statistics after the second year. With this timeline, revised estimates can be incorporated into stock assessments during 2018 using a peer reviewed calibration model, and new Annual Catch Limits (ACLs) can then be set in 2019 for at least some stocks.

Requirements

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

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

Tasks for Reviewers

Pre-review Background Documents

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

Transition Plan for the FES:

https://www.st.nmfs.noaa.gov/Assets/recreational/pdf/MRIP%20FES%20Transition%20Plan%20FINAL.pdf

Report recommending the FES to replace the CHTS: *Finalize Design of Fishing Effort Surveys* (https://www.st.nmfs.noaa.gov/pims/main/public?method=DOWNLOAD_FR_PDF&record_id=1179)

2015 Benchmarking Progress Report:

https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/documents/2015_benchmarking_progress_report.pdf

Report on FES/CHTS Calibration Model:

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

Panel Review Meeting

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

Contract Deliverables - Independent CIE Peer Review Reports

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

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

Other Tasks – Contribution to Summary Report

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

Place of Performance

The place of performance shall be at the reviewers' facilities, and at the NMFS Headquarters in Silver Spring, Maryland.

Period of Performance

The period of performance shall be from the time of award through July 31, 2017. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

Travel

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790).

Restricted or Limited Use of Data
The contractors may be required to sign and adhere to a non-disclosure agreement.

NMFS Project Contact:
Dave Van Voorhees
National Marine Fisheries Service
1315 East West Highway
Silver Spring, MD 20910
dave.van.voorhees@noaa.gov

Annex I: Format and Contents of Independent Peer Review Report

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

2. The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each ToR, in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the ToRs.

   a. Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.

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

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

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

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

3. The report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review

   Appendix 2: A copy of this Statement of Work

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

Annex 2: Terms of Reference for the Peer Review

Calibration Model Accounting for a Recreational Fishery Survey Design Change

1. Evaluate the suitability of the proposed model for converting historical estimates of private boat and shore fishing effort produced by the CHTS design to estimates that best represent what would have been produced had the new FES design been used prior to 2017.
   k) Does the proposed model adequately account for differences observed in the estimates produced by the CHTS and FES designs when conducted side-by-side in 2015-2016?
   l) Is the proposed model robust enough to account for potential differences that would have been observed if the two designs had been conducted side-by-side in years prior to 2015 with regards to time trending biases?

m) How does the approach used in developing the proposed FES/CHTS calibration model compare in terms of strengths or weaknesses with other potential approaches?

n) Does the proposed calibration model help to explain how different factors would have contributed to changes in differences between CHTS and FES results over time?

o) Is it reasonable to conclude that revised 1981-2016 private boat and shore fishing effort estimates based on the application of the proposed FES/CHTS calibration model would be more accurate than the estimates that are currently available? Does evidence provided for this determination include an assessment of model uncertainty?

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

Appendix 4. CIE contract

Statement of Work

National Oceanic and Atmospheric Administration (NOAA)

National Marine Fisheries Service (NMFS)

Center for Independent Experts (CIE) Program

External Independent Peer Review

Calibration Model Accounting for a Recreational Fishery Survey Design Change

Background

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

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards.
(http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).
Further information on the CIE program may be obtained from www.ciereviews.org.

Scope

The Office of Science and Technology requests an independent peer review of a calibration model proposed for use in revising statistics produced by surveys of marine recreational fishing effort on the Atlantic coast and in the Gulf of Mexico. This calibration model is considered by the Marine Recreational Information Program (MRIP) to be very important to adjust historical time series of recreational effort and catch estimates in order to account for biases in past sampling and estimation methods that have become apparent with the development of a new, more statistically sound method. The calibration model is intended to account for past biases in private boat and shore fishing effort estimates that have resulted from the continued use of a legacy random-digit-dial telephone survey design that has degraded over time and will be replaced with the implementation of a new mail survey design (the "Fishing Effort Survey", or FES) in 2018.

Calibration Model for the Fishing Effort Survey

In 2015, MRIP formed a Transition Team to collaboratively plan a transition from a legacy telephone survey design to a new mail survey design for estimating private boat and shore fishing effort by marine recreational anglers. Since 2008, MRIP had conducted six pilot studies to determine the most accurate and efficient survey method for this purpose on the Atlantic and Gulf coasts. The most recent study, conducted in four states in 2012-2013, compared a new mail survey design with the Coastal Household Telephone Survey (CHTS) design that has been used since 1979. MRIP subjected the final report from the pilot project to external peer review in 2014 and certified the new survey design, called the Fishing Effort Survey (FES), in February 2015 as a suitable replacement for the CHTS. The FES is much less susceptible to potential sources of bias than the CHTS because it can reach more anglers, achieve higher response rates, and is less prone to possible recall errors. The pilot project results indicated that FES estimates were substantially higher than CHTS estimates for both private boat fishing and shore fishing.

MRIP recognized the FES should not be implemented immediately as a replacement for the CHTS, and a well thought out transition plan was needed to ensure that the phase-in of the FES is appropriately integrated into ongoing stock assessments and fisheries management actions in a way that minimizes disruptions to these processes, which are based on input from multiple data sources over lengthy time series. The Transition Plan developed by the Transition Team called for side-by-side benchmarking of the FES against the CHTS for three years (2015-2017) with the development and application of a calibration model to enable adjustment of past estimates that account for biases in historical effort and catch statistics after the second year. With this timeline, revised estimates can be incorporated into stock assessments during 2018 using a peer reviewed calibration model, and new Annual Catch Limits (ACLs) can then be set in 2019 for at least some stocks.

Requirements

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

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

Tasks for Reviewers

Pre-review Background Documents

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

Transition Plan for the FES:

https://www.st.nmfs.noaa.gov/Assets/recreational/pdf/MRIP%20FES%20Transition%20Plan%20FINAL.pdf

Report recommending the FES to replace the CHTS: *Finalize Design of Fishing Effort Surveys* (https://www.st.nmfs.noaa.gov/pims/main/public?method=DOWNLOAD_FR_PDF&record_id=1179)

2015 Benchmarking Progress Report:

https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/documents/2015_benchmarking_progress_report.pdf

Report on FES/CHTS Calibration Model:

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

Panel Review Meeting

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

Contract Deliverables - Independent CIE Peer Review Reports

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

Other Tasks – Contribution to Summary Report

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

Foreign National Security Clearance

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

Place of Performance

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

Period of Performance

The period of performance shall be from the time of award through July 31, 2017. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

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

Within two weeks of award	Contractor selects and confirms reviewers
Within four weeks of award	Contractor provides the pre-review documents to the reviewers
June, 2017	each reviewer participates and conducts an independent peer review during the panel review meeting
Within two weeks of panel review meeting	Contractor receives draft reports
Within two weeks of receiving draft reports	Contractor submits final reports to the Government

Applicable Performance Standards

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

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

Travel

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $15,000.

Restricted or Limited Use of Data
The contractors may be required to sign and adhere to a non-disclosure agreement.

NMFS Project Contact:
Dave Van Voorhees
National Marine Fisheries Service
1315 East West Highway
Silver Spring, MD 20910
dave.van.voorhees@noaa.gov

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

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

2. The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each ToR, in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the ToRs.

   a. Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.

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

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

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

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

3. The report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review

   Appendix 2: A copy of this Statement of Work

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

Annex 2: Terms of Reference for the Peer Review

Calibration Model Accounting for a Recreational Fishery Survey Design Change

3. Evaluate the suitability of the proposed model for converting historical estimates of private boat and shore fishing effort produced by the CHTS design to estimates that best represent what would have been produced had the new FES design been used prior to 2017.

    p) Does the proposed model adequately account for differences observed in the estimates produced by the CHTS and FES designs when conducted side-by-side in 2015-2016?

    q) Is the proposed model robust enough to account for potential differences that would have been observed if the two designs had been conducted side-by-side in years prior to 2015 with regards to time trending biases?

    r) How does the approach used in developing the proposed FES/CHTS calibration model compare in terms of strengths or weaknesses with other potential approaches?

    s) Does the proposed calibration model help to explain how different factors would have contributed to changes in differences between CHTS and FES results over time?

    t) Is it reasonable to conclude that revised 1981-2016 private boat and shore fishing effort estimates based on the application of the proposed FES/CHTS calibration model would be more accurate than the estimates that are currently available? Does evidence provided for this determination include an assessment of model uncertainty?

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

Tentative Agenda

Calibration Model Accounting for a Recreational Fishery Survey Design Change

TBD

National Marine Fisheries Service

Office of Science and Technology

1315 East-West Highway

Silver Spring, MD

June, 2017

Point of contact: Front Desk

Appendix 5: CALIBRATION MODEL REVIEW ATTENDEES

MRIP Calibration Model Peer Review Workshop
Sheraton Silver Spring Hotel
Silver Spring, MD
June 27-29, 2017

ATTENDANCE LIST

#	NAME	AFFILIATION
1	Paul Rago	MAFMC SSC
2	Dave Van Voorhees	NOAA Fisheries
3	John Foster	NOAA Fisheries
4	Ali Arab	Georgetown University
5	Rob Hicks	College of William and Mary
6	Cynthia M. Jones	Old Dominion University
7	Richard Cody	NOAA support ECS
8	Teng Liu	Colorado State University
9	Thomas Sminkey	NOAA Fisheries/ST1
10	Steve Turner	NOAA Fisheries SEFSC
11	Andy Strelcheck	NOAA Fisheries - SERO
12	Richard Methot	NOAA Fisheries - HQ
13	Karen Pianka	NOAA Fisheries – ST1
14	Lauren Dolinger Few	NMFS ST1
15	Chris Wright	NMFS - SF
16	Sabrina Lovell	NMFS ST
17	Patrick Lynch	NMFS ST
18	Melissa Karp	NMFS ST
19	Toni Kerns	ASMFC
20	Steve Ander	Gallup
21	Tommy Tran	Gallup
22	Melissa Niles	Fifth Estate/MRIP CET
23	Yong-Woo Lee	NOAA - Fisheries
24	Jay Breidt	Colorado State University
25	Jean Opsomer	Colorado State University
26	Rob Andrews	NOAA Fisheries
27	Ryan Kitts-Jensen	NOAA Fisheries
28	Fred Serchuk	SAFMC SSC
29	Jason McNamee	ASMFC
30	Patrick Sullivan	Cornell/NEFMC
31	Jason Didden	MAFMC
32	Daemian Schreiber	NMFS HQ
33	Laura Diederick	NOAA Fisheries

Peer Review Report for Marine Recreational Information Program (MRIP) Fishing Effort Survey (FES) Calibration Model

Ali Arab

Associate Professor of Statistics

Department of Mathematics and Statistics

Georgetown University

August 2017

Table of Contents

Executive Summary

In order to improve the survey methodology for estimating catch for recreational fishing in the coastal US, the Marine Recreation Information Program (MRIP) has implemented a new program for estimating fishing effort based on a mail-based survey, the Fishing Effort Survey (FES), to replace a historical telephone survey, the Coastal Household Telephone Survey (CHTS). This report provides a technical review of a calibration model for adjusting the historic CHTS estimates using the FES results during the overlapping period. The calibration model was developed and tested using data from side-by-side implementation of the two methods during 2015 and 2016.

The proposed modeling framework has strong theoretical underpinnings and the proposed estimators have desirable properties. The proposed model is equipped with the components to address different sources of variation in the survey data as well as accounting for method-specific effects. The design variance as well as the effort estimates are modeled using predictor information. There are a limited number of potential explanatory variables that are readily available through both surveys. This limits the explanatory and predictive ability of the statistical calibration modeling strategies. Critically, the current model does not provide insight into the underlying mechanisms resulting in differences in estimated effort.

It is recommended that the investigators provide a comprehensive discussion of alternative methods and present a narrative on the reasoning behind selection of the proposed model over the competing alternatives. Although the investigators did not discuss alternative approaches in their report, they informed the Review Panel of the alternative options that they had considered and explored. This list included a reasonable number of options. They provided sufficient discussion on the advantages and disadvantages of some of these approaches and convincingly articulated the reasoning which had led them to choose the proposed method. In particular, the investigators reported on consideration of several popular approaches including time series approaches, and hierarchical Bayesian methods.

It is recommended that the MRIP and the investigators consider efforts to improve several aspects of the current model as well as the presentation and communication of the methodology and results. In particular, efforts should be made to obtain additional potential predictor information to better understand the underlying mechanisms that may explain the differences observed in the effort estimates during the side-by-side experiments. Additional potential predictor information may include state-level or county-level population values (potentially broken down by age groups) and socio-economic factors. Also, comparisons of similarities and dissimilarities among estimates of different states may shed light on area-specific and local drivers of these mechanisms. Additionally, a more comprehensive simulation study of the model to assess the effectiveness and predictive ability of the model is lacking and should be implemented.

1 Introduction

1.1 Background

The Review Panel for the MRIP-FES Calibration Model Review met from June 27 to June 29 in Silver Spring, Maryland to review a statistical model developed by a team of investigators from Colorado State University (F. Jay Breidt, Teng Liu and Jean D. Opsomer). The review committee was composed of six members. Three scientists were appointed by the Center for Independent Experts (CIE): Robert Hicks, The College of William and Mary; Cynthia Jones, Old Dominion University; and Ali Arab, Georgetown University. The other three members on the review panel consisted of representatives from the New England (Patrick Sullivan) and South Atlantic (Fredric Serchuk) Scientific and Statistical Committees, and the Atlantic States Marine Fisheries Commission (Jason McNamee). The meeting was chaired by Paul Rago as a member of the Mid-Atlantic Fishery Management Council Scientific and Statistical Committee.

1.2 Review Activities

The pre-review documents were provided by the NTVI staff on June 19, 2017, about a week before the Panel Review.

Day 1 (Tuesday June 27, 2017): The Panel Review meeting started with welcoming remarks and introductions, followed by presentations on the transition from the telephone survey (CHTS) to the mail survey (FES), the importance of calibration of the CHTS efforts, and the ramifications of the calibrated catch efforts for stock assessment, and fisheries management. The presentations in the afternoon, included presentations by the Colorado State University investigators, Jean Opsomer and Jay Breidt. Opsomer provided an overview of the challenges of calibrating historical time series in general, and the specific challenges for the calibration of the CHTS effort estimates. Breidt presented the proposed calibration model.

The presentations were followed by questions and comments from the Panel, and the audience (present in the room as well as online through the webinar platform).

The Panel met in closed session at the end of Day 1 and discussed the presentations.

Day 2 (Wednesday June 28, 2017): The Panel Review meeting resumed in the morning with a summary discussion of the Panel based on initial reactions and findings. The main focus of the presentations and discussions was on the proposed calibration model. Breidt presented additional material including model results for a limited number of cases and clarified several points raised and requests made by the Panel during Day 1. In particular, Breidt and colleagues provided information on the list of modeling options they had considered and informed the panel of the process which had led them to the proposed model. They also provided additional information and sample results of the calibrated CHTS effort with prediction intervals.

The Panel met in closed session at the end of Day 2 and discussed the presentations.

Day 3 (Thursday June 29, 2017): The Panel met in closed session to discuss the Terms of Reference and draft a summary report. The meeting concluded about mid-day.

2 Review of MRIP FES Calibration Model

The modeling approach is based on well-established classical methodology, and I commend the investigators on their work, especially for making the connection between their initial modeling framework with a well-known model in small area estimation, the Fay-Herriot model (See e.g., Fay and Herriot, 1979; Rao, 2015). The proposed method results in valid analytical forms for the model estimators based on well-established theory.

The main area of improvement in the current modeling framework is to better account for uncertainty of some of the model estimates. In particular, the uncertainty in the design variances is not accounted for in the model. Although I consider this as the main shortcoming of the proposed modeling framework, it is not an unusual consequence of the methodology choice (and in fact, it is a rather common consequence of most classical methods). This may be improved by adapting a Bayesian approach for estimating the model parameters. However, Bayesian approaches have disadvantages too; mainly, the estimation procedures do not rely on analytical results and are based on advanced computational methods.

Below, I list several recommendations to possibly improve the model and its implementation for calibrating the CHTS data.

2.1 Recommendations:

- It is highly recommended that the investigators conduct realistic simulation studies and test the performance of the proposed model (in comparison to other alternative methods). The current simulations, as described by the investigators, are limited to sensitivity analysis for specific assumptions and choices (e.g., sensitivity of the normality assumption for sampling error).

- The model is based on only two years of calibration data (in fact, 11 waves), and although the proposed model structure is based on well-established methodology, it is highly recommended that the calibration is periodically updated based on future data. It is my understanding that the overlapping period between CHTS and FES is scheduled to be three years (two of which data is available for). I highly recommend extending the overlapping period between the two surveys to obtain additional data for the purpose of calibration.
- Given that the model results indicate the wireless effect as the only significant covariate (aside from log of population) with a minor effect size in explaining the differences between the two surveys, I recommend limiting the application of the calibration model to the CHTS data for the period where the wireless phones became relatively prevalent (early 2000's and onwards).
- Also, I recommend considering other potential candidates beyond what has already been considered to serve as predictor information for the model to possibly better explain the differences between the data obtained using the two survey methods. In particular, additional information related to demographics (possibly broken down by age groups) and socio-economic within states may serve as predictor variables.
- Another aspect that does not seem to have been explored is the potential similarities or dissimilarities in trends of CHTS and/or FES data among certain states. This may help better understand the mechanisms underlying these data. To clarify, this recommendation does not necessarily indicate using spatial dependence structure to model the response data, rather the goal is to identify potential common predictor factors specific to certain states through by focusing on similarities (or dissimilarities) between the patterns of survey data in these states.
- Finally, the current description of the proposed model requires familiarity with statistical methodology at a relatively high level. Given that the audience of this product are not statisticians, the methodology should be communicated in a more effective way than the current document prepared by the investigators.

3 Evaluation of Terms of Reference

3.1 Term of Reference 1

1. Evaluate the suitability of the proposed model for converting historical estimates of private boat and shore fishing effort produced by the CHTS design to estimates that best represent what would have been produced had the new FES design been used prior to 2017.

TOR 1 and its subcomponents (a-e) were met.

> *a) Does the proposed model adequately account for differences observed in the estimates produced by the CHTS and FES designs when conducted side-by-side in*

2015-2016?

The general model structure is capable of accounting for the observed differences between the CHTS and FES results during the overlapping period (2015-2016). The model parameterization accounts for different patterns and sources of variability including trend, seasonality (between waves), and unexplained sources (called the 'irregular' effect). Also, the proposed model accounts for the sampling method effect being different between the mail and telephone surveys. Moreover, the design variances are modeled using predictor information. The described parameterization allows for adequately accounting for the differences between the observation from the two survey methods. However, in practice, there are two shortcomings: 1) the period of overlap between the two surveys is short, currently resulting in 11 observations, and thus, the process of learning from data in order to calibrate historic CHTS values is based on limited number of observations; 2) the current model results only identify a few number of predictors as important factors in describing the differences between the two survey results, and these results hardly explain the mechanism underlying these differences.

It should be noted that the described issues are not shortcomings of the proposed model and rather are based on limited availability of data and predictor information.

b) Is the proposed model robust enough to account for potential differences that would have been observed if the two designs had been conducted side-by-side in years prior to 2015 with regards to time trending biases?

The model parametrization, as described previously, contains the required components to account for the differences between the two survey methods. The main shortcoming in this area is due to data availability and inconsistency in collection of auxiliary data (e.g., demographic information about the anglers being surveyed) through the CHTS.

Another important issue is that the investigators were not able to identify the mechanism underlying the differences between the two surveys. The Panel members discussed this issue at length, but were unable to identify an easy solution for this problem. I agree that this is not a simple problem to address but without insight into the underlying mechanisms that explain the differences between the two survey methods, it would be difficult to confidently respond to this ToR. Presumably, if we knew more about the underlying mechanism and had access to additional useful predictor data, the model structure would allow to conduct robust inference.

c) How does the approach used in developing the proposed FES/CHTS calibration model compare in terms of strengths or weaknesses with other potential approaches?

Strengths: The proposed model is developed based on well-established classical methodology and nicely fits into a well-known small area estimation method framework (the Fay-Herriot model). The estimators have desirable properties (e.g., unbiasedness, etc.) and model implementation is straightforward and may be done using available software.

Weaknesses: I consider the disconnect between the uncertainty in estimated design variance and the estimation of effort as the main weakness of the proposed model. In the proposed model, the point estimates for the design variances are used in the model for estimating effort, without accounting for uncertainty in the estimation of design variances. Alternatively, a hierarchical Bayesian approach may be considered to fully account for uncertainty in the design variance estimation.

The investigators described that they had considered and explored additional modeling approaches including a hierarchical Bayesian approach and although they recognized the advantages of some of these methods over their proposed method, they provided convincing arguments in defense of their choice. In particular, the advantages of the proposed method based on the Fay-Herriot model including the nice theoretical properties of the estimators, the availability of analytical forms for the estimators (as oppose to stochastic ones determined using numerical approximations in Bayesian methods), and availability of off-the-shelf software tools outweigh the competing modeling options. In summary, I have no concerns about the scientific credibility and theoretical underpinnings of the proposed method.

d) Does the proposed calibration model help to explain how different factors would have contributed to changes in differences between CHTS and FES results over time?

As previously mentioned, the current model results do not provide a clear understanding of the underlying mechanisms that may describe the differences between the CHTS and FES outcomes. Although the investigators have considered several predictor variables, other than population size (included in the model as the log of population) and a minimal effect of wireless phones, none of these predictor variables showed any statistical significance in explaining the differences between the two surveys. Potentially, availability of auxiliary information about the anglers surveyed through the CHTS (similar to what is available through the FES) would have been helpful to better understand the differences. However, given that these data are lacking for the historical CHTS surveys (pre-2015), it is not clear if much can be done to improve the issue.

Further possibilities that may deem helpful include using population and demographic information at finer scales (e.g., Census tract or county level data). Also, it may be instructive to look at similarities and dissimilarities of data among different geographical locations (e.g., among states) to potentially identify spatially differentiated effects that may help better understand the underlying mechanism of the differences in survey results.

e) Is it reasonable to conclude that revised 1981-2016 private boat and shore fishing effort estimates based on the application of the proposed FES/CHTS calibration model would be more accurate than the estimates that are currently available? Does evidence provided for this determination include an assessment of model uncertainty?

This is a very difficult question to answer as the underlying mechanisms for these surveys are complex and not fully understood. In general, it may be argued that mail surveys are currently more effective than telephone surveys. This is due to a decline in landlines and the rise in prevalence of wireless/mobile phones (which are not used in CHTS) as well as other potential factors. There are other advantages to a mail survey over a telephone survey in this setting including a better recollection of fishing trips, etc. Although some of these arguments hold true for the historic period and thus we may conclude for example that the calibrated historic CHTS values may be more accurate than the observed CHTS values, one may argue that in general, telephone surveys used to be more effective than mail surveys in the past. This is particularly true for the period before wireless phones became popular (and use of landlines started to decline, especially among the younger demographics). In general, there are advantages and disadvantages to both survey methods (For more discussion see e.g., Groves et al. 2001).

The proposed model is capable of accounting for uncertainty in the CHTS calibrated estimates. In particular, prediction intervals may be produced and considered. The investigators did not provide the prediction intervals in the manuscript describing the methodology; however, they provide discussion of the derivation of the estimate variances (i.e., the “MSE”). During the Panel Review meeting, per request from the Panel, the investigators provided sample results which contained prediction intervals. In the future, it would be critical that the produced calibrated CHTS results include prediction intervals, and the importance of accounting for uncertainty in the point estimates should be effectively communicated with the community of users of this product.

3.2 Term of Reference 2

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

On pre-review materials and background documents:

- Additional background documents would have been useful for the review, for example, MRIP Handbook should have been provided before the review meeting in order to provide the reviewers with more detailed background information about the surveys.

- Discussions during the review included several other reports that seemed to be closely related to this review (e.g., the National Academy reports, etc.). However, none of these reports were provided prior to the Panel Review meeting.

- It would have been extremely helpful to have a clearer presentation of the proposed model that would discuss the components of the model in more details. Also, a summary of candidate modeling approaches, and details on the process that led to the proposed model would have been very useful. The investigators provided this summary per request from the Panel. However, it would have been helpful to have the discussion documented and presented to the Panel prior to the Panel Review meeting.

- It would have been extremely helpful to have more information about the surveys prior to the meeting, including similarities and differences in definitions of effort, questions on the questionnaires, etc.

Review panel and presentations:

- I was hoping and expecting to see:
  - more details presented on the survey methodologies used in both surveys,
  - more specific information and simulation regarding impact of the calibration procedure results on stock assessment, and
  - more details on the proposed model beyond the paper that was provided to the reviewers, and information on exploratory data analyses and the process that led to the proposed model (including details on other potential candidate models), and simulation studies based on the proposed model to validate model performance for simulated data sets.
- The presenters did not address the TORs directly, which made it harder for the Panel to assess the relevance of some of the information presented to these TORs. Consequently, the Panel spent substantial portion of questions/answers period (and discussion time) on obtaining answers to address TORs.
- The Panel members and staff were all very knowledgeable and pleasant to work with. Overall, the review process was efficient except for the issues mentioned above. The Panel members worked effectively together and the Chair of the Panel did an extremely well job in making sure the discussions stayed on track.
- In summary, my main concern about the review process and an area that requires attention and improvement for future reviews is communication. The background documents, and the information essential for the review were either not provided or not provided in the level of details that the Panel members expected. This is extremely important, in particular for outside reviewers who may not be familiar with the history of these surveys and past reviews.

Appendix 1: Bibliography of materials provided for review

Fay III, R. E., & Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, *74*(366a), 269-277.

Groves, R. M., Biemer, P. P., Lyberg, L. E., Massey, J. T., Nicholls, W. L., & Waksberg, J. (Eds.). (2001). *Telephone survey methodology* (Vol. 328). John Wiley & Sons.

Rao, J. N. (2015). *Small-Area Estimation*. John Wiley & Sons, Ltd.

Background Papers

Many papers and documents on the existing and proposed survey methodology may be found at the following website: http://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/effort-survey-improvements

Background on the MRIP Calibration Model Peer Review may be found at: https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/index.html

The National Academies of Sciences, Engineering, and Medicine. 2016. Review of the Marine Recreational Information Program (MRIP) Washington, DC: The National Academies Press. doi: 10.17226/24640 https://www.st.nmfs.noaa.gov/confluence/display/FESCALIB?preview=/73074985/73728799/NAS_MRIP_review.pdf

National Research Council. 2006. Review of Recreational Fisheries Survey Methods. Committee on the Review of Recreational Fisheries Survey Methods, ISBN: 0-309-66075-0, 202 pages. http://www.nap.edu/catalog/11616.html

Working Papers

Development and Testing of Recreational Fishing Effort Surveys Testing a Mail Survey Design: Final Report. Project Team Members: Rob Andrews, NOAA Fisheries, J. Michael Brick, Westat, Nancy A. Mathiowetz, University of Wisconsin-Milwaukee. July 31, 2014. https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/documents/Report_recommending_FES_to_replace_CHTS--Finalize_Design_of_Fishing_Effort_Surveys.pdf

Marine Recreational Information Program Fishing Effort Survey Transition Progress Report. October 28, 2016. https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/documents/2015_benchmarking_progress_report.pdf

Marine Recreational Information Program Transition Plan for the Fishing Effort Survey
Prepared by the Atlantic and Gulf Subgroup of the Marine Recreational Information Program Transition Team May 5, 2015
https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/documents/MRIP_FES_Transition-Plan_FINAL.pdf

A Small Area Estimation Approach for Reconciling Mode Differences in Two Surveys
of Recreational Fishing Effort draft: F. Jay Breidt, Teng Liu, Jean D. Opsomer Colorado State University June 10, 2017
https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/documents/DRAFT-Report_of_Calibration_Model.pdf

Presentations

Calibration_Scenarios-20161115.pdf

MRIP FES website link

FESCALIBRATIONNOTESDay2.docx

EBLUPS.csv

EBLUPS_Variable_Names.csv

FESCALIBRATIONNOTESDay1.docx

Eblup comparisons.docx

MRFSS Fish Hunt Comps.xlsx

FES Errors.pptx

Model_Fits.txt

Mode_3_logeffort_poly_fixed.pdf

Mode_7_logeffort_poly_fixed.pdf

Webinar Links

All open sections of the meeting were recorded and available for viewing at the following links.

0 - Intro - Paul Rago

1 - MRIP Fishing Effort Survey - Rob Andrews

2- Catch and Assessments - Rick Methot

3 - Management Implications - Andy Strelcheck

4 - Calibrating Survey Estimates over Time - Jean Opsomer

5 - Calibration from CHTS to FES - Jay Breidt

6 - Initial Calibration Review Discussion - Tuesday Afternoon

7 - Day Two, AM Discussion

8 - Day Two, PM Discussion

9 - Day Two, Initial Findings Summary

Appendix 2: Statement of Work

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

Calibration Model Accounting for a Recreational Fishery Survey Design Change

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

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards. (http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).
Further information on the CIE program may be obtained from www.ciereviews.org.

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

method. The calibration model is intended to account for past biases in private boat and shore fishing effort estimates that have resulted from the continued use of a legacy random-digit-dial telephone survey design that has degraded over time and will be replaced with the implementation of a new mail survey design (the "Fishing Effort Survey", or FES) in 2018.

Calibration Model for the Fishing Effort Survey

In 2015, MRIP formed a Transition Team to collaboratively plan a transition from a legacy telephone survey design to a new mail survey design for estimating private boat and shore fishing effort by marine recreational anglers. Since 2008, MRIP had conducted six pilot studies to determine the most accurate and efficient survey method for this purpose on the Atlantic and Gulf coasts. The most recent study, conducted in four states in 2012-2013, compared a new mail survey design with the Coastal Household Telephone Survey (CHTS) design that has been used since 1979. MRIP subjected the final report from the pilot project to external peer review in 2014 and certified the new survey design, called the Fishing Effort Survey (FES), in February 2015 as a suitable replacement for the CHTS. The FES is much less susceptible to potential sources of bias than the CHTS because it can reach more anglers, achieve higher response rates, and is less prone to possible recall errors. The pilot project results indicated that FES estimates were substantially higher than CHTS estimates for both private boat fishing and shore fishing.

MRIP recognized the FES should not be implemented immediately as a replacement for the CHTS, and a well thought out transition plan was needed to ensure that the phase-in of the FES is appropriately integrated into ongoing stock assessments and fisheries management actions in a way that minimizes disruptions to these processes, which are based on input from multiple data sources over lengthy time series. The Transition Plan developed by the Transition Team called for side-by-side benchmarking of the FES against the CHTS for three years (2015-2017) with the development and application of a calibration model to enable adjustment of past estimates that account for biases in historical effort and catch statistics after the second year. With this timeline, revised estimates can be incorporated into stock assessments during 2018 using a peer reviewed calibration model, and new Annual Catch Limits (ACLs) can then be set in 2019 for at least some stocks.

Requirements

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

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

Tasks for Reviewers

Pre-review Background Documents

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

Transition Plan for the FES:
https://www.st.nmfs.noaa.gov/Assets/recreational/pdf/MRIP%20FES%20Transition%20Plan%20FINAL.pdf

Report recommending the FES to replace the CHTS: *Finalize Design of Fishing Effort Surveys*
(https://www.st.nmfs.noaa.gov/pims/main/public?method=DOWNLOAD_FR_PDF&record_id=1179)

2015 Benchmarking Progress Report:
https://www.st-test.nmfs.noaa.gov/Assets/recreational/pdf/2015_FES_Progress_Report-20161115.pdf

Report on FES/CHTS Calibration Model:
This report will be provided by the contractor (via electronic mail or make available at an FTP site) to the CIE reviewers.

Panel Review Meeting

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

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

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

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

Place of Performance
The place of performance shall be at the contractor's facilities, and at the NMFS Headquarters in Silver Spring, Maryland.

Period of Performance
The period of performance shall be from the time of award through July 31, 2017. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

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

Within two weeks of award	Contractor selects and confirms reviewers
Within four weeks of award	Contractor provides the pre-review documents to the reviewers
June, 2017	each reviewer participates and conducts an independent peer review during the panel review meeting
Within two weeks of panel review meeting	Contractor receives draft reports
Within two weeks of receiving draft reports	Contractor submits final reports to the Government

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

Travel
All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $15,000.

Restricted or Limited Use of Data
The contractors may be required to sign and adhere to a non-disclosure agreement.

NMFS Project Contact:
Dave Van Voorhees
National Marine Fisheries Service
1315 East West Highway
Silver Spring, MD 20910
dave.van.voorhees@noaa.gov

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

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

2. The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each ToR, in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the ToRs.

   a. Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.

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

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

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

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

3. The report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of this Statement of Work
   Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

Calibration Model Accounting for a Recreational Fishery Survey Design Change

1. Evaluate the suitability of the proposed model for converting historical estimates of private boat and shore fishing effort produced by the CHTS design to estimates that best represent what would have been produced had the new FES design been used prior to 2017.
   a) Does the proposed model adequately account for differences observed in the estimates produced by the CHTS and FES designs when conducted side-by-side in 2015-2016?
   b) Is the proposed model robust enough to account for potential differences that would have been observed if the two designs had been conducted side-by-side in years prior to 2015 with regards to time trending biases?
   c) How does the approach used in developing the proposed FES/CHTS calibration model compare in terms of strengths or weaknesses with other potential approaches?
   d) Does the proposed calibration model help to explain how different factors would have contributed to changes in differences between CHTS and FES results over time?
   e) Is it reasonable to conclude that revised 1981-2016 private boat and shore fishing effort estimates based on the application of the proposed FES/CHTS calibration model would be more accurate than the estimates that are currently available? Does evidence provided for this determination include an assessment of model uncertainty?

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

Tentative Agenda

Calibration Model Accounting for a Recreational Fishery Survey Design Change

TBD

National Marine Fisheries Service
Office of Science and Technology
1315 East-West Highway
Silver Spring, MD
June, 2017
Point of contact: Front Desk

Appendix 3: Panel membership

The review committee was composed of six members: three scientists appointed by the Center for Independent Experts (CIE): Robert Hicks, The College of William and Mary, Cynthia Jones, Old Dominion University and Ali Arab, Georgetown University, as well as representatives from the New England (Patrick Sullivan) and South Atlantic (Fredric Serchuk) Scientific and Statistical Committees, and the Atlantic States Marine Fisheries Commission (Jason McNamee) served on the review panel. The meeting was chaired by Paul Rago as a member of the Mid-Atlantic Fishery Management Council Scientific and Statistical Committee.

Independent Peer Review Report on the Calibration Model Accounting for a Recreational Fishery Survey Design Change

Prepared for the Center for Independent Experts (CIE) Program

Robert L. Hicks

Department of Economics and the School of Marine Science
The College of William and Mary
Williamsburg, VA 23187-8795

August 2017

Contents

1 Executive Summary

This document presents my findings on the proposed calibration model for estimating the historical recreational effort one would have estimated had the Fishing Effort Survey (FES) been conducted at some point in the past when only telephone estimates were available from the Coastal Household Telephone Survey (CHTS). The importance of developing a calibration approach that can produce reliable and comparable estimates of recreational effort for long time series (e.g. 1982 - Present) is a key task outlined in the FES transition plan [3]. To that end, data were collected and effort estimated for both the FES and CHTS (during 2015 and 2016) and a new proposed calibration approach uses this data and the past time-series of CHTS data for judging the performance of the calibration model. In this report I find that

1. The proposed model is a reliable and scientifically defensible way to estimate (calibrate) in either FES or CHTS effort units, since
   (a) the approach employs a well-known methodology and provides estimates of model uncertainty that embodies both the prediction and sampling error associated with calibrated estimates.
   (b) the statistical properties of the model are clearly presented and follow from clear and reasonable modeling assumptions.
   (c) the model is well specified for the calibration problem for which it is used.
2. While the calibration model may be intended to predict FES estimates in the past, it can also be used to
   (a) purge the "wireless" effects that have potentially biased CHTS effort estimates during the period 2000 - Present.
   (b) predict what the CHTS would be in some point in the future.

My report also includes some specific recommendations for potentially improving the application of the model and these include:

1. Sensitivity analysis should be performed to investigate the effect of the overlapping mail and telephone specification in the model.
2. Additional covariates should be explored for better capturing the wireless effect in the model
3. The agency should consider revisiting the model once a longer time series of FES data is available so that the FES portion of the model might include time trending covariates.
4. The model results and outputs should be better presented using case studies to show the types of output it can yield (e.g. confidence intervals, effort units rather than log(effort units)) for hindcasting and forecasting.

2 Background

The Marine Recreational Information Program (MRIP) has committed to a full transition from the Coastal Household Telephone Survey (CHTS) to the Fishing Effort Survey (FES) for allowing the estimation of total effort [3] because of likely biases resulting from the random digit dial of coastal household residences sample frame. As pointed out by Andrews et al. [2], there are multiple problems associated with the CHTS that the FES attempts to overcome including

- CHTS undersamples wireless-only households and therefore there are questions about the representativeness of landline households as compared to the total population.
- More efficient sample frame for FES.
- Potential for FES to overcome some of the problems associated with gatekeeper bias.

A further issue that should be pointed out is that the CHTS does not collect sociodemographic information in sufficient detail to enable a re-weighting for possibly overcoming some of these factors. A complete review of the problems with the CHTS and the advantages associated with the FES were the motivation of the change currently ongoing with the MRIP data collection efforts.

Both pilot survey evidence and recent side-by-side sampling show that there can be large and persistent differences resulting from the two sampling methodologies due to a host of recognized factors and the transition plan for moving from CHTS to FES [3] calls for the development of a methodology to calibrate one set of estimates to another (e.g. CHTS to FES, or potentially vice-versa). The differences between Mail and Telephone estimates can be attributed to a range of causes, but the most important ones are arguably

- Mode Effects (phone versus mail)
- A change in the survey instrument
- On-going issues associated with the representativeness of the CHTS sample due to wireless telephone adoption by of U.S. households

A review of the proposed calibration method was organized to analyze the soundness of the statistical approach taken, and to investigate the suitability of the application to the MRIP FES data as outlined in the Terms of Reference (ToR) provided below. It is important to recognize that the review panel was instructed to take the survey methods and estimation methods underlying either the FES and CHTS estimates used in the calibration model as scientifically defensible and therefore, we were tasked to focus only on the calibration methods one might employ *after data is collected and effort is estimated using either FES or CHTS methodologies*.

Three CIE reviewers, three appointed reviewers, and a Chair served on the review panel. The review was conducted during a meeting at the Sheraton Silver Spring, Maryland from June 27th - 29th 2017 and the peer review panel had a conference call for finalizing the Summary Report on July 8, 2017. Each panelist participates in the Panel review meeting and writes their own independent assessment of the approach proposed. While my report is in large measure consistent with the panel's Summary Report, it reflects my own independent findings with respect to the proposed approach.

3 Description of My Role in the Review Activities

Four pre-meeting documents ([3],[2], [11], and [4]) were available and reviewed from June 14, 2017. In addition, the panel was given access to a recorded webinar by F. Jay Breidt on June 23, 2017 for more detail on the statistical method underlying the calibration approach. During the meeting, I participated in the discussion and suggested some exploratory analysis for checking model robustness and model fit. Since the meeting I have performed some exploratory analysis based on the provided model outputs [5], and written a summary of the model and outlined key issues for enhancing my understanding of details, included in Section 5 of this report.

4 Summary of Findings

Below I discuss my findings for each ToR. In some places I reference more detailed discussions contained in my summary of the methodology (Section 5.2).

4.1 Term of Reference 1

Evaluate the suitability of the proposed model for converting historical estimates of private boat and shore fishing effort produced by the CHTS design to estimates that best represent what would have been produced had the new FES design been used prior to 2017.

In my view the proposed model is a reliable and scientifically defensible way to estimate what an FES design estimate would have been had it been conducted at some time since 1982. The approach employs a well-known methodology that is capable of being used to predict either Mail or Telephone effort estimates and provides estimates of error that embodies both the prediction and sampling error associated with calibrated estimates. The proposed calibration method meets ToR 1 and the sub-components (a) - (e).

It is important to note that the model [4] is agnostic with respect to whether CHTS or FES estimates are "best". I believe this is a reasonable position to take given that we are dealing with self-reported data and that for most of the 1980's and 1990's there are strong arguments to be made for Telephone Surveys in general. Notwithstanding the many reasons why more recent CHTS estimates (denoted as $\hat{T}$ hereafter) might be biased downwards, the model allows for projection from Telephone to Mail "units" of effort or vice versa. The proposed approach also allows for wireless effects to be purged from the CHTS estimate to account for the hypothesized downward bias in CHTS estimates since 2000. Given that in the future, only the FES methodology will be used, the model will most likely be used to cast past Telephone estimates into predicted Mail estimates, and it is suited for that. But the model is also equipped to cast future Mail estimates (FES) into predicted Telephone estimates (see discussion in Section 5.2.1). The ability to calibrate in either direction is a strength of the proposed approach particularly if future side-by-side stock assessments or policy analysis is desired using both Mail and Telephone predicted effort.

4.1.1 Term of Reference 1a

Does the proposed model adequately account for differences observed in the estimates produced by the CHTS and FES designs when conducted side-by-side in 2015-2016?

In my opinion, the model accounts for differences in side-by-side Mail and Telephone estimate and based on feedback from the research team, finds that most of the differences are due to an intercept shifter that captures average differences between mail and telephone estimates that are *time invariant* rather than large changes in underlying trends. This intercept shifter would be capturing any systematic difference between the mail and telephone estimate *for each* state and wave,year and might include survey mode effects and/or effects due to differences in the survey instrument itself. While the model "accounts" for the differences, I have seen no evidence that it can explain what is driving the difference, since based on responses by the review team time-invariant mail constants are responsible for most of the differences between mail and telephone.

4.1.2 Term of Reference 1b

Is the proposed model robust enough to account for potential differences that would have been observed if the two designs had been conducted side-by-side in years prior to 2015 with regards to time trending biases?

Since this is a hypothetical comparison we are being asked about, it is difficult to answer. The model is able to adjust for the wireless bias, one of the the primary biases believed to exist with respect to the CHTS since 2000. On average, I would say the model would account for these differences.

The method includes time trends and corrections for changing composition of wireless penetration after 2000 and the bias that might impact telephone effort estimates. Consequently, it is able to predict in two types of Telephone Effort Units: one that purges telephone estimates of effort of potential biases due to the wireless effect (after 2000) and one that does not. The model, therefore, is able to explain how these biases change through time as more wireless-only and wireless-mostly household penetrate study areas, since the wireless covariate is state-specific and varies by year and are interacted with state-level population levels. Consequently, the wireless effect can influence the statistical model either by shifting the average difference between mail and telephone estimates or through time-varying trends. Unobservable factors that impact Telephone and Mail estimates in the same manner and that are not related to model covariates are captured by the model random effect. Any other systematic time-varying differences between mail and telephone estimates not included in the model specification are absorbed in the model error.

While I believe the model as it currently stands is defensible and well developed, I recommend that the model specification [4] for capturing wireless effects should investigate alternative covariates. In Section 5.2.4, I suggest some alternative specifications for the wireless portion of the model for perhaps better capturing the nuances of the wireless effects based on how we believe they are impacting our sample from a random digit dial. My suggestions center on choosing explanatory variables that focus on population for older individuals in coastal counties. Additionally, a more thorough discussion of model results as outlined in Section 5.2.5 would have been beneficial for evaluating this ToR.

4.1.3 Term of Reference 1c

How does the approach used in developing the proposed FES/CHTS calibration model compare in terms of strengths or weaknesses with other potential approaches?

While the study [4] provides no evidence, whether in the form of side-by-side comparisons or simulation experiments for determining this ToR, I am satisfied based on our discussion during the review meeting that the modeling team considered and experimented with a number of alternative approaches including the general linear model, time-series approaches, and Bayesian Heirarchical Models. They settled on this approach after experimentation with the other methods and I can't fault them for not showing the relative performance of the Small Area Estimator compared to these other approaches since they were not fully aware of the Terms of Reference. Their focus was on developing a scientifically defensible calibration methodology with known statistical properties and they have done that. Given the Small Area Estimator approach, the team did perform a number of model selection tests for the choice of final model covariates, and the review panel was given these results.

4.1.4 Term of Reference 1d

Does the proposed calibration model help to explain how different factors would have contributed to changes in differences between CHTS and FES results over time?

Given the short time-period over which Mail survey data and effort estimates exist, it is a very tough ask for the model to identify factors driving differences between the methodologies. As all time varying trends in the model impact either the base telephone portion (telephone estimate purged of wireless) or the telephone + wireless portion of the model any discussion of differences between CHTS and FES over time is being driven by the wireless effect. As an aside, I believe this is a sound modeling decision given the short time-series of Mail estimates. Unfortunately, the review panel was not presented with enough evidence on the magnitude of the wireless effect relative to other model factors to fully evaluate this ToR. I felt the presentation of results in the paper didn't highlight these types of factors enough as I outline in Section 5.2.5.

The current model could (and perhaps should) be re-estimated in the future as more Mail estimates are collected, allowing the possible inclusion of time-varying trends in the mail portion of the model. This would serve two purposes: 1) Allow for time-varying differences between CHTS and FES beyond the wireless effect and 2) provide for a larger sample size and perhaps better specification for identifying the model parameters associated with Mail. These issues are outlined in more detail in Section 5.2.3.

4.1.5 Term of Reference 1e

Is it reasonable to conclude that revised 1981-2016 private boat and shore fishing effort estimates based on the application of the proposed FES/CHTS calibration model would be more accurate than the estimates that are currently available? Does evidence provided for this determination include an assessment of model uncertainty?

I disagree with this conclusion, particularly the statement on accuracy, for several reasons:

- To gauge accuracy, one needs to know the truth. Both sources of data are designed to measure fishing effort and rely on self-reported fishing data. Furthermore, the estimates are derived from different survey instruments and survey modes. The closest we may get to the truth might be to perform a marine fishing census not relying on self-reported data, an enormous undertaking requiring near round the clock monitoring at all possible fishing sites and launch points. As no such census exists, I can't make a judgment about this ToR.
- Even if one knew the truth for gauging accuracy, there isn't strong evidence that the telephone methodology, prior to approximately 2000 and the advent of wireless phones, produced biased estimates. On the contrary, many survey experts advocated the use of telephone surveys as a reliable method for recovering population estimates of behavior during the period 1980-2000. The calibration method proposed here is agnostic as to which method is closer to the truth, and can be used to hindcast mail estimates from telephone-only time periods, or vice-versa.
- As with any prediction, calibrated estimates rely on a model and have uncertainty induced by forecasting as well as sampling error, so perhaps the pre-wireless telephone estimates are in some sense more accurate or are estimated with less uncertainty.

4.2 Term of Reference 2

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

Overall, I found the review process to be a highly effective way to assess the scientific merits of the calibration methodology. Members of the review panel were highly qualified and brought different perspectives to the review that in the end will give the agency a broad yet penetrative look into the proposed calibration method. The deliberative process of the Panel included stimulating discussions and serendipitous feedback among the panelists during question and answer periods. For example, the afternoon session on the second day when the review panel did a deep dive into the statistical details was valuable. The chair of the panel provided invaluable guidance both in making the "trains run on time" and ensuring that diverse viewpoints were heard.

The MRIP staff are knowledgeable and I appreciated their ability to answer questions, and if necessary, get more information in a timely manner. The statisticians in MRIP are impressive and are making sure the agency asks the right questions as data collection methods evolve. Similarly the research team presenting the proposed methodology were also extremely knowledgeable and able to quickly offer clarificatory answers to questions or additional information if needed. Having access to additional information as the review progressed was vital to the review process. Important examples included access to auxiliary model information [5] and a comparison between FES and Fish and Wildlife Marine Fishing Effort Estimates [9].

I feel the review process could be improved. The approach as written in [4] is not helpful beyond the statistical properties of the model. It (or a companion document) needs to focus more on model outputs rather than statistical properties. Because this type of information wasn't included, I had a difficult time addressing some ToR's adequately. A reader should reasonably expect to understand how covariates enter the model and to what degree they impact predictions. In fairness, the Colorado State research team was unaware of the ToR until approximately one week prior to the meetings. Consequently, it was impossible for them to adequately present their approach for getting at the specific concerns highlighted by the ToR's. Finally, technical reviews should include access to code and data. The panel wasn't able to fully engage on the underpinnings of the approach until the second day (after we received some auxiliary information from the research team [5]). Even with the extra information, it would have been beneficial to have access to code and data.

My primary recommendation for future statistical reviews is that they are approached more like a stock assessment review process (as it was described to me by my fellow panelists): reviewers have access to models and data, and can contribute in a give and take process for understanding the method.

5 Calibration of Effort Estimates

This is a summary of the calibration approach [4] along with additional detail for my understanding of the model.

5.1 A Strawman Calibration Model

The calibration approach used in the paper [4] does not mention the "strawman" approach outlined in this section. I include it for a) highlighting issues with more simple approaches that might have been taken and b) showing that the suggested approach overcomes a lot of these problems.

The primary requirement of the calibration approach as I understand it is to allow for the prediction of FES Mail estimates for periods when the mail survey didn't exist (e.g. 1982 - 2015) and in a way that accounts for changing trends that might be systematically driving effort estimates through time. An approach one might take is to focus on the time period where both Mail ($\hat{M}$) and Telephone ($\hat{T}$) estimates exist and estimate a model such as

$$\hat{M} = \mathbf{X}' \quad + \hat{T} \quad + \epsilon \tag{1}$$

where $\mathbf{X}$ is a vector of control variables (including state fixed effects, and state-wave interactions, trend variables, and controls for wireless), and are parameters, and ϵ is the model error which might contain random effects for each state and time period as in the proposed model.

Given an estimate of the model parameters $\hat{}$ and $\hat{}$ one can then predict a mail estimate for state s and year,wave t as

$$\hat{\hat{M}}_{st} = \mathbf{X}'_{st}\,\hat{} + \hat{T}_{st}\,\hat{} \tag{2}$$

Using this simple model, this is the estimated Effort Estimate from a mail survey had it been conducted in year,wave t state s.

This model would provide a direct calibration from past telephone estimates into the prediction of what the mail survey would have yielded had it been undertaken. However, this approach has several shortcomings:

1. There is a very limited set of observations over which both $\hat{T}$ and $\hat{M}$ exist and therefore a reliance on short time periods for identifying time trends.

2. The above approach really only allows a *one-way* method for projecting telephone into mail units.

3. Care would need to be taken to correctly account for the fact that $\hat{T}$ is random and estimated with uncertainty, and how this uncertainty propagates into predictions ($\hat{\hat{M}}_{st}$).

4. If unobserved factors impacting the telephone estimates are also impacting mail estimates, then we have parameter bias due to endogeneity issues since it isn't likely that $E[\hat{T}'\epsilon] = 0$ which is required for unbiasedness.

While the above approach provides a direct mapping between Mail and Telephone estimates and may be a natural way to think about the problem, it does have its shortcomings as outlined above. In contrast, the approach under consideration [4] summarized below avoids these shortcomings and is a way to leverage the full time series of data available from both the CHTS and the FES for calibrating from one "effort unit" into another.

5.2 Summary and Discussion of the Proposed Method

This description of the model largely abstracts from the technical detail provided in the paper outlining the proposed calibration method [4] and focuses on model specification and predictions. From equation (1) in the paper, we have

$$\hat{T}_{st} = \mathbf{a}'_{st}\boldsymbol{\alpha} + \nu_{st} + e^T_{st} \text{ for } t < 2000 \tag{3}$$
$$= \mathbf{a}'_{st}\boldsymbol{\alpha} + w_{st}\mathbf{c}'_{st} \quad + \nu_{st} + e^T_{st} \text{ for } t \quad 2000 \tag{4}$$

where the variables are as described in the paper and the differences in pre and post wireless are modeled beginning for year,waves from 2000 onwards. Similarly, for mail we have

$$\hat{M} = \mathbf{a}'_{st}\boldsymbol{\alpha} + \mathbf{b}'_{st}\boldsymbol{\mu} + \nu_{st} + e^M_{st} \tag{5}$$

Compared to equation (1) from the previous section, we don't have the telephone estimate appearing as an explanatory variable. Instead the paper uses the explanatory variables outlined in Table 1.[1] Note that trends are incorporated for each state and year,wave by interacting population estimates with state fixed effects, by an overall model trend by state. Additionally, the wireless effect has similar trends specified. Consequently if the model needs to predict values in future time periods, it need not be re-estimated since no trend parameters are time-period specific (e.g. a fixed effect by year). Also, since Mail Effort isn't calibrated directly off of the Telephone estimate, the method avoids problem (4) in

[1] This table was developed from the reported parameter estimates from R given to the panel [5]. While it involved some guesswork given variable names to construct the table, I hope it captures the exact model specification in the paper.

the previous section altogether.

Table 1: Model Covariates by Variable Type for a State and Year,Wave Observation

Explanatory Variable	Included in
State,Wave Constant	$\mathbf{a}_{st}$
log(pop)$\times$ State Constant	$\mathbf{a}_{st}$
Wireless Constant (=1 for waves after 1999, else 0)	$\mathbf{c}_{st}$
Wireless Constant $\times$ Wave Constant	$\mathbf{c}_{st}$
Wireless Constant $\times$ State Constant	$\mathbf{c}_{st}$
Wireless Constant $\times$ log(pop)	$\mathbf{c}_{st}$
Wireless Constant $\times$ log(pop) $\times$ State Constant	$\mathbf{c}_{st}$
Mail Constant (=1 if Mail Estimate exists and Mail Obs., else 0)	$\mathbf{b}_{st}$
Mail Constant $\times$ Wave Constant	$\mathbf{b}_{st}$
Mail Constant $\times$ State Constants	$\mathbf{b}_{st}$

5.2.1 Predictions, Hindcasting, and Forecasting

Given model estimates, we have the following model predictions in Table 2.[2]

Table 2: Predictions of log(Effort) Estimates from the Proposed Calibration Model

Type of Prediction	Expression
Telephone	$\mathbf{a}'_{st}\hat{\boldsymbol{\alpha}} + \hat{\nu}_{st}$
Telephone + Wireless	$\mathbf{a}'_{st}\hat{\boldsymbol{\alpha}} + w_{st}\mathbf{c}'_{st}\,\hat{} + \hat{\nu}_{st}$
Mail	$\mathbf{a}'_{st}\hat{\boldsymbol{\alpha}} + \mathbf{b}'_{st}\hat{\boldsymbol{\mu}} + \hat{\nu}_{st}$

Before proceeding with an analysis of some predictions we might make using the model, it is useful thinking about what comprises the differences between some of the expressions in Table 2. First, the differences between Telephone (this is purged of wireless effects) and Telephone + Wireless from Table 1 contains

1. Constants that shift Telephone away from Telephone + Wireless for each time period (i.e., Wireless Constant, Wireless Constant $\times$ Wave Constant, Wireless Constant $\times$ State Constant).

2. Trend variables that allow the difference between Telephone and Telephone + Wireless to vary across time (i.e., Wireless Constant $\times$ log(pop) and Wireless Constant $\times$ log(pop) $\times$ State Constant).

By contrast the difference between Telephone (purged of Wireless) and Mail is solely due to Constants that shift Mail away from Telephone for every time period (Mail Constant $\times$ Wave Constant and Mail Constant $\times$ State Constants). There

[2]These predictions are analogous to what the proposed method refers to as $\phi(.)$ in Section 3.2

are no trend differences between Telephone (purged of Wireless) and Mail in the Model since differences are down to estimated constants and don't include trend effects. Of course differences between Telephone + Wireless and Mail would include the wireless constants, the wireless trend variables, and the mail constants. So it is worth noting that the model implicitly assumes there are no time varying mail effects at play since no mail trend interactions are included. We note this as a technical point rather than as a point of omission in the proposed approach since with very few mail estimates available for estimation, there is no way to really model mail trends.

- **Ratios**

  The difference between a predicted telephone estimate (purged of wireless) and a predicted mail estimate is $\mathbf{b}'_{st}\hat{\boldsymbol{\mu}}$. If one wants to think of the calibration as a ratio, we have for our predictions

  $$\frac{\hat{\hat{M}}_{st}}{\hat{\hat{T}}_{st}} = \frac{\mathbf{a}'_{st}\hat{\boldsymbol{\alpha}} + \mathbf{b}'_{st}\hat{\boldsymbol{\mu}} + \hat{\nu}_{st}}{\mathbf{a}'_{st}\hat{\boldsymbol{\alpha}} + \hat{\nu}_{st}} = 1 + \frac{\mathbf{b}'_{st}\hat{\boldsymbol{\mu}}}{\mathbf{a}'_{st}\hat{\boldsymbol{\alpha}} + \hat{\nu}_{st}}$$

  This ratio would vary by state and year,wave and is itself a random variable.[3] There is a high likelihood that this ratio varies substantially from state to state and wave to wave and this is evidence that a ratio-based simple calibration approach is inferior to the proposed method. Without too much effort, this could be fleshed out to show how the model predictions below outperform the ratio estimator. There may be some value in that since a ratio-based approach is perhaps the first way most people think about calibration (as we heard from the public question).

- **Hindcasting**

  For hindcasting what one would have estimated with a mail survey when one wasn't conducted, we can apply the mail predictor (from above):[4, 5]

  $$\hat{\hat{M}}_{st|t<2015} = \mathbf{a}'_{st}\hat{\boldsymbol{\alpha}} + \mathbf{b}'_{st}\hat{\boldsymbol{\mu}} + \hat{\nu}_{st} \tag{6}$$

  Another useful forecast the model gives us is a re-calibration of historical

[3]Given the model specification, this is the ratio in log units.

[4]It is my understanding that this is what the research team labels as EPLUBM of the preferred model from provided supplementary materials [5].

[5]It is also worth mentioning that one could calibrate directly off of the existing historical telephone estimate ($\hat{T}$). The hindcast of what one would have estimated had a mail survey been done could be calculated as $\hat{T} + \mathbf{b}'_{st}\hat{\boldsymbol{\mu}} - w_{st}\mathbf{c}'_{st}\hat{\boldsymbol{\gamma}}$, but my sense is that the EPLUBM is a better estimate, and comes with a coherent estimate of variance (due to sampling and forecasting error). Figures 3 and 4 in the paper [4] shows the performance relative to the EPLUBM.

telephone estimates (post 2000) purged of the wireless effect

$$\hat{\hat{T}}_{st|t>2000} = \mathbf{a}'_{st}\hat{\boldsymbol{\alpha}} + \hat{\nu}_{st} \tag{7}$$

Both of these estimates are creating a historical time series of data using the model, and are readily calculated given model outputs since all predicted parameters are recovered.

- **Forecasting**

  The model could also be used in a forecasting context to examine what one would have estimated with the telephone survey if it was conducted after 2017. This might be useful in a future stock assessment context, for example, if the analyst wants to compare assessments using both telephone and mail units of recreational effort using the estimated model. In this case, we would use the telephone predictor (e.g., purged of wireless effects) to produce future (from the standpoint of when the calibration statistical model was last run):

  $$\hat{\hat{T}}_{st|t>2017} = \mathbf{a}'_{st}\hat{\boldsymbol{\alpha}} + \hat{\nu}_{st} \tag{8}$$

  In this case, the analyst knows $\hat{\boldsymbol{\alpha}}$, has collected data on $\mathbf{a}$ (including future time periods), but $\hat{\nu}_{st}$ is unknown. For proceeding, one might either

  - Re-estimate the new model and recover new estimates, which would include an estimate for ν_{st} for the future time period, or
  - Perhaps the model as estimated would allow you to back out an estimate for ν_{st} in a future time period, given current parameter estimates. Ideally this should also include a new estimate of variance in that time period as well. Should the method be implemented, more guidance should be given by the research team as to how this should be approached. In the paper [4], equations (14) - (16) could well be covering this but a more thorough explanation of hindcasting versus forecasting would enhance understanding of the approach.

5.2.2 Prediction Uncertainty

For quantification of prediction uncertainty, it is worth noting that:

- Confidence intervals are likely to be large for calibrated values since they embody both sampling and forecasting error, this is especially true for effort measured in levels (rather than logs), and will probably also be large *even for* *states that have high effort levels*. It isn't possible to assess this completely given the current presentation of results.

- Since effort is modeled as log-normal and all modeled units are log(Effort), the confidence intervals of effort units (rather than log effort units) are no longer symmetric about the mean. Any stock assessment or policy analysis that needs to use the effort distribution (rather than only the mean) will need more information from NMFS (and possibly training on how to use that information) than the percent-standard-error approach available now.

5.2.3 Estimation Strategy

The calibration approach uses the well-known Fey-Herriott Small Area Method [8]. The approach has the following advantages:

- Statistical properties are known and understood.
- Can be implemented using existing software packages (e.g. R).
- Allows the mean to contain random effects that, in principle, could be spatially or temporally correlated (although that isn't implemented in the current approach).

While the approach is widely used and accepted in the statistical community, there are some downsides to using the approach for this problem:

- The mean model is estimating separately from the sampling variance model.
- The model as it is currently coded in R (and perhaps other software packages) isn't totally suited for this estimation problem, since given the overlapping data collection for the period 2015-2016, there are *two observations per state and year,wave* whereas the software packages assumes a single observation per state and year,wave. The study team creatively gets around this and I will discuss this in more detail below.
- Since in the calibration context, we have in essence a missing data problem (e.g. no observations of mail estimates until 2015) and there are other methods that could be considered for these types of problems that would have been more of a natural fit (e.g. Bayesian Heirarchical Models). The study team examined this approach and found that it wasn't fruitful.

Defining the set of year,wave time periods for which only telephone estimates are available as T^T, for which only mail estimates are available as T^M, and for which both telephone and mail estimates are available as $T^{T,M}$, based on equation (10) in the paper, construct the design matrix by stacking these time period blocks of observations as

$$\mathbf{x} = \begin{bmatrix} \mathbf{a}_t & \mathbf{0}\mathbf{b}_t & \omega\mathbf{c}_t \\ \mathbf{a}_t & \frac{\mathbf{b}_t}{2} & \frac{\omega\mathbf{c}_t}{2} \\ \mathbf{a}_t & \mathbf{b}_t & \mathbf{0}\omega\mathbf{c}_t \end{bmatrix} \begin{matrix} \text{if } \mathrm{t} \in T^T \\ \text{if } \mathrm{t} \in T^M \\ \text{if } \mathrm{t} \in T^{T,M} \end{matrix} \tag{9}$$

while the dependent variable is

$$\mathbf{y} = \begin{bmatrix} \hat{\mathbf{T}}_t \\ \frac{\hat{\mathbf{T}}_t + \hat{\mathbf{M}}_t}{2} \\ \hat{\mathbf{M}}_t \end{bmatrix} \begin{matrix} \text{if } \mathrm{t} \in T^T \\ \text{if } \mathrm{t} \in T^{T,M} \\ \text{if } \mathrm{t} \in T^M \end{matrix} \tag{10}$$

Given the current state of data collection there are no observations where only the mail survey was collected. Consequently, for estimation purposes *in the current paper*, the data used in estimation looks like this

$$\mathbf{y} = \begin{bmatrix} \hat{\mathbf{T}}_t \\ \frac{\hat{\mathbf{T}}_t + \hat{\mathbf{M}}_t}{2} \end{bmatrix}, \mathbf{x} = \begin{bmatrix} \mathbf{a}_t & \mathbf{0} & \omega\mathbf{c}_t \\ \mathbf{a}_t & \frac{\mathbf{b}_t}{2} & \frac{\omega\mathbf{c}_t}{2} \end{bmatrix} \begin{matrix} \text{if } \mathrm{t} \in T^T \\ \text{if } \mathrm{t} \in T^{T,M} \end{matrix} \tag{11}$$

Without any "Mail Only" time periods, the mail portion of the model is estimated over just 157 state and year,wave observations (for shore mode), while the telephone only part of the model has 2810 observations. All parameters are identified, although it should be pointed out that

- The mail-specific covariates ($\mathbf{b}$) enter the model for year,waves were both the mail and telephone surveys are present and enter as the average. Consequently, the model recovers $\boldsymbol{\mu}$ by fitting an average model over the average mail and telephone survey estimates.

- Since $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ contain similar covariates and all enter the model when mail and telephone estimates exist, there is likely a very high degree of colinearity between the columns of $\mathbf{x}$ for these time periods.

- Due to data constraints, there is no attempt to model trends for the mail portion of the model.

Given that the primary use of the calibration method will be to predict mail estimates in past time periods, I recommend that some sensitivity analysis be performed particularly as it relates to the assumption of averaging mail and telephone estimates for recovering $\boldsymbol{\mu}$. Try estimating a model that drops the overlapping telephone estimates for the period 2015-2016 and run the model over the data:

$$\mathbf{y} = \begin{bmatrix} \hat{\mathbf{T}}_t \\ \hat{\mathbf{M}}_t \end{bmatrix}, \mathbf{x} = \begin{bmatrix} \mathbf{a}_t & \mathbf{0} & \omega \mathbf{c}_t \\ \mathbf{a}_t & \mathbf{b}_t & \mathbf{0} \end{bmatrix} \begin{matrix} \text{if } \text{t} \in T^T \\ \text{if } \text{t} \in T^{T,M} \end{matrix} \tag{12}$$

If large differences are found (in parameters and in predictions) or if mail trend effects are deemed important, then the agency might consider re-visiting specification and estimation of the calibration model once more mail data is collected and, in particular, *including mail-only time periods for estimating the model*. It is important to note that the proposed approach does not strictly require simultaneously collected mail and telephone effort estimates for a given state and year,wave for identification of parameters. In fact, the presence of both estimates has to be creatively dealt with for using existing software. From an efficiency viewpoint it would be advisable to modify the `R SAE` package (or write custom code) to overcome this problem, however custom code has to be maintained by the agency and it is my belief that any efficiency loss associated with this estimation trick is not large enough to warrant a coding extension to this project.

5.2.4 Covariates

Covariates are listed in Table 1. The choice of co-variates included in the model (and experimented with during model development) are defensible from a statistical standpoint and the study team has investigated other covariates but ruled them out using model selection criteria. Covariates are chosen so that forecasting can be done without re-estimating the model, since time trends only enter via the state's population interacted with state fixed effects. This is a reasonable choice given the requirements of the model.

Given the importance of capturing the "wireless effect" and explaining differences between mail and telephone estimates, I was surprised that no efforts were made to try to capture this more directly given what we know about landline only and mostly landline households that tend to consist of older individuals who also tend to fish less. In my view it is advisable to investigate more nuanced variables in the wireless portion of the model ($\mathbf{c}$). For example, data on the total population of coastal counties *and* the total population of older individuals in coastal counties by state should be available from the U.S. Census and could be included in the model. Many Southeastern states have had a large and increasing influx of retirees since 2000 (particularly in coastal areas) and these covariates my help explain cross-state trends that would improve the wireless correction portion of the model.

5.2.5 Results

I found the results section of the paper the most lacking and due to that, the strength of the proposed approach isn't showcased to the degree that it should be. The methodology paper should be expanded to include

- Details on estimated results

  It is difficult to know which covariates are in the model and how "subsets" drive the difference between telephone, telephone with wireless, and the mail portions of the model. The study team should include tables outlining covariates included (with descriptions) and tables of parameter estimates.

- Evidence for each of the 3 types of predictors discussed above

  One of the great strengths of the model is that it can predict into either mail or telephone effort units, and for telephone can predict with or without wireless effects. This isn't clear enough when presenting results, as the focus is only on the Mail estimates (EPLUBM). A nice addition would be to include some calibration case studies to show model capabilities both graphically and in tabular format.

- Details about the impact of the wireless effect

  Given the sometimes large differences between the mail and telephone estimates please provide more evidence about how big the wireless effect is. What is the telephone estimate post 2000 after wireless effects are purged? To what degree does it shrink the difference between Mail and Telephone estimates? A plot like Figure 1 could easily include two plots of $EPLUBT$ one that purges and one including wireless effects. In the figure, eyeballing where the pre-2000 telephone estimator ($\hat{T}$) are on the edge of the 95% confidence interval and after 2000 they fall away, I suspect that an $EPLUBT$ purged of the wireless effect would close some of this gap. That would be evidence the model is working as we expect and provides information that informs us about problems with the telephone survey since 2000.

- Evidence about what is driving the difference between mail and telephone

  This is related to the above point, but it would be useful to quantify what is driving the biggest difference between EPLUBT (wireless purged) and EPLUBM. Given that only the Mail Constant, Wave Constant $\times$ Mail Constant, and State

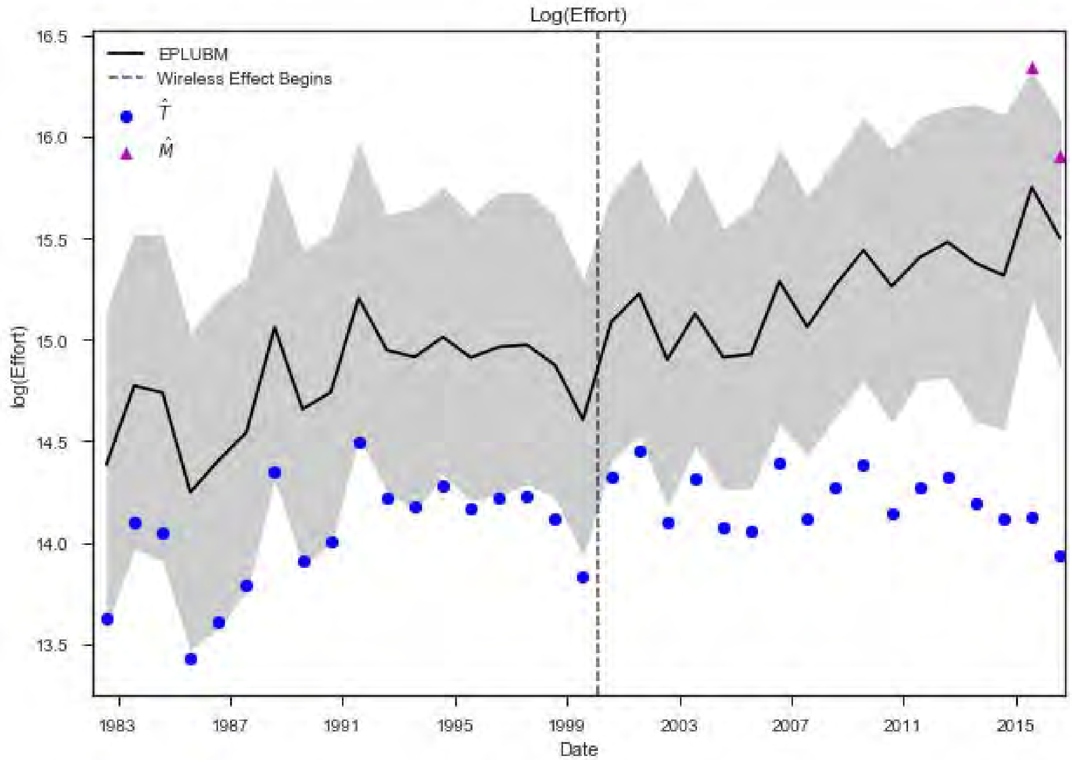


Figure 1: Florida Shore Mode Wave 4

Constant × Mail Constant are in the model, there isn't too much one can do here. One could look at the state and wave constants to see if anything systematic jumps out either spatially or temporally.

- Results in effort rather than log(effort) units

  Model outputs will be used in effort units most of the time. Please provide some figures and/or tables that show model predictions based on effort. Investigate how large prediction confidence intervals are in effort. I suspect that wireless might have relatively more important impact when examined using effort units.

6 Appendix 1: Bibliography

[1] Review of the marine recreational information program. Technical report, National Academy of Science, 2017.

[2] R. Andrews, J. M. Brick, and N. A. Mathiowetz. Development and testing of recreational fishing effort surveys testing a mail survey design: Final report.

[3] Atlantic and Gulf Subgroup of the Marine Recreational Information Program Transition Team, Marine Recreational Information Program, National Marine Fisheries Service. Transition plan for the fishing effort survey.

[4] F. J. Breidt, T. Liu, and J. D. Opsomer. A small area estimation approach for reconciling mode differences in two surveys of recreational fishing effort.

[5] F. J. Breidt, T. Liu, and J. D. Opsomer. Supporting materials. Included the files: EBLUPS.csv, EBLUPS_Variable_Names.csv, Model_Fits.txt, Mode_3_logeffort_poly_fixed.pdf, and Mode_7_logeffort_poly_fixed.pdf.

[6] CIE Review Rapporteur. Notes for day 1.

[7] CIE Review Rapporteur. Notes for day 2.

[8] R. E. Fay III and R. A. Herriot. Estimates of income for small places: an application of james-stein procedures to census data. *Journal of the American Statistical Association*, 74(366a):269–277, 1979.

[9] Marine Recreational Fisheries Program Staff. A comparison between fishing and wildlife's and mrip effort estimates.

[10] Marine Recreational Fisheries Program Staff. Sources of error and mail and chts surveys.

[11] Marine Recreational Information Program, National Marine Fisheries Service. Fishing effort survey transition progress report.

[12] National Marine Fisheries Service. Possible effects of calibration scenarios on stock assessments planned for the mrip fishing effort survey transition.

[13] Personal Correspondence: National Marine Fisheries Service. Effort survey improvements. http://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/effort-survey-improvements.

[14] Presenters at Review. Presentations. Available on Review Website.

[15] P. Rago. Eplub comparisons.

7 Appendix 2: Statement of Work

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

Calibration Model Accounting for a Recreational Fishery Survey Design Change

Background

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

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards.
(http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).
Further information on the CIE program may be obtained from www.ciereviews.org.

Scope

The Office of Science and Technology requests an independent peer review of a calibration model proposed for use in revising statistics produced by surveys of marine recreational fishing effort on the Atlantic coast and in the Gulf of Mexico. This calibration model is considered by the Marine Recreational Information Program (MRIP) to be very important to adjust historical time series of recreational effort and catch estimates in order to account for biases in past sampling and estimation methods that have become apparent with the development of a new, more statistically sound method. The calibration model is intended to account for past biases in private boat and shore fishing effort estimates that have resulted from the continued use of a legacy random-digit-dial telephone survey design that has degraded over time and will be replaced with the implementation of a new mail survey design (the "Fishing Effort Survey", or FES) in 2018.

Calibration Model for the Fishing Effort Survey

In 2015, MRIP formed a Transition Team to collaboratively plan a transition from a legacy telephone survey design to a new mail survey design for estimating private boat and shore fishing effort by marine recreational anglers. Since 2008, MRIP had conducted six pilot studies to determine the most accurate and efficient survey method for this purpose on the Atlantic and Gulf coasts. The most recent study, conducted in four states in 2012-2013, compared a new mail survey design with the Coastal Household Telephone Survey (CHTS) design that has been used since 1979. MRIP subjected the final report from the pilot project to external peer review in 2014 and certified the new survey design, called the Fishing Effort Survey (FES), in February 2015 as a suitable replacement for the CHTS. The FES is much less susceptible to potential sources of bias than the CHTS because it can reach more anglers, achieve higher response rates, and is less prone to possible recall errors. The pilot project results indicated that FES estimates were substantially higher than CHTS estimates for both private boat fishing and shore fishing.

MRIP recognized the FES should not be implemented immediately as a replacement for the CHTS, and a well thought out transition plan was needed to ensure that the phase-in of the FES is appropriately integrated into ongoing stock assessments and fisheries management actions in a way that minimizes disruptions to these processes, which are based on input from multiple data sources over lengthy time series. The Transition Plan developed by the Transition Team called for side-by-side benchmarking of the FES against the CHTS for three years (2015-2017) with the development and application of a calibration model to enable adjustment of past estimates that account for biases in historical effort and catch statistics after the second year. With this timeline, revised estimates can be incorporated into stock assessments during 2018 using a peer reviewed calibration model, and new Annual Catch Limits (ACLs) can then be set in 2019 for at least some stocks.

Requirements

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

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

Tasks for Reviewers

Pre-review Background Documents

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

Transition Plan for the FES:
https://www.st.nmfs.noaa.gov/Assets/recreational/pdf/MRIP%20FES%20Transition%20Plan%20FINAL.pdf

Report recommending the FES to replace the CHTS: *Finalize Design of Fishing Effort Surveys* (https://www.st.nmfs.noaa.gov/pims/main/public?method=DOWNLOAD_FR_PDF&record_id=1179)

2015 Benchmarking Progress Report:
https://www.st-test.nmfs.noaa.gov/Assets/recreational/pdf/2015_FES_Progress_Report-20161115.pdf

Report on FES/CHTS Calibration Model:
This report will be provided by the contractor (via electronic mail or make available at an FTP site) to the CIE reviewers.

Panel Review Meeting

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

Contract Deliverables - Independent CIE Peer Review Reports

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

Other Tasks – Contribution to Summary Report

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

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

Place of Performance
The place of performance shall be at the contractor's facilities, and at the NMFS Headquarters in Silver Spring, Maryland.

Period of Performance
The period of performance shall be from the time of award through July 31, 2017. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

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

Within two weeks of award	Contractor selects and confirms reviewers
Within four weeks of award	Contractor provides the pre-review documents to the reviewers
June, 2017	each reviewer participates and conducts an independent peer review during the panel review meeting
Within two weeks of panel review meeting	Contractor receives draft reports
Within two weeks of receiving draft reports	Contractor submits final reports to the Government

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

Travel
All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $15,000.

Restricted or Limited Use of Data
The contractors may be required to sign and adhere to a non-disclosure agreement.

NMFS Project Contact:
Dave Van Voorhees
National Marine Fisheries Service
1315 East West Highway
Silver Spring, MD 20910
dave.van.voorhees@noaa.gov

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

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

2. The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each ToR, in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the ToRs.

   a. Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.

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

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

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

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

3. The report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of this Statement of Work
   Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

Calibration Model Accounting for a Recreational Fishery Survey Design Change

1. Evaluate the suitability of the proposed model for converting historical estimates of private boat and shore fishing effort produced by the CHTS design to estimates that best represent what would have been produced had the new FES design been used prior to 2017.

   a) Does the proposed model adequately account for differences observed in the estimates produced by the CHTS and FES designs when conducted side-by-side in 2015-2016?

   b) Is the proposed model robust enough to account for potential differences that would have been observed if the two designs had been conducted side-by-side in years prior to 2015 with regards to time trending biases?

   c) How does the approach used in developing the proposed FES/CHTS calibration model compare in terms of strengths or weaknesses with other potential approaches?

   d) Does the proposed calibration model help to explain how different factors would have contributed to changes in differences between CHTS and FES results over time?

   e) Is it reasonable to conclude that revised 1981-2016 private boat and shore fishing effort estimates based on the application of the proposed FES/CHTS calibration model would be more accurate than the estimates that are currently available? Does evidence provided for this determination include an assessment of model uncertainty?

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

Tentative Agenda

Calibration Model Accounting for a Recreational Fishery Survey Design Change

TBD

National Marine Fisheries Service

Office of Science and Technology

1315 East-West Highway

Silver Spring, MD

June, 2017

Point of contact: Front Desk

8 Appendix 3: Panel Membership and List of Attendees

MRIP Calibration Model Peer Review Workshop
Sheraton Silver Spring Hotel
Silver Spring, MD
June 27-29, 2017

ATTENDANCE LIST

	NAME	AFFILIATION
1	Paul Rago	MAFMC SSC
2	Dave Van Voorhees	NOAA Fisheries
3	John Foster	NOAA Fisheries
4	Ali Arab	Georgetown University
5	Rob Hicks	College of William and Mary
6	Cynthia M. Jones	Old Dominion University
7	Richard Cody	NOAA support ECS
8	Teng Liu	Colorado State University
9	Thomas Sminkey	NOAA Fisheries/ST1
10	Steve Turner	NOAA Fisheries SEFSC
11	Andy Strelcheck	NOAA Fisheries - SERO
12	Richard Methot	NOAA Fisheries - HQ
13	Karen Pianka	NOAA Fisheries âĂŞ ST1
14	Lauren Dolinger Few	NMFS ST1
15	Chris Wright	NMFS - SF
16	Sabrina Lovell	NMFS ST
17	Patrick Lynch	NMFS ST
18	Melissa Karp	NMFS ST
19	Toni Kerns	ASMFC
20	Steve Ander	Gallup
21	Tommy Tran	Gallup
22	Melissa Niles	Fifth Estate/MRIP CET
23	Yong-Woo Lee	NOAA - Fisheries
24	Jay Breidt	Colorado State University
25	Jean Opsomer	Colorado State University
26	Rob Andrews	NOAA Fisheries
27	Ryan Kitts-Jensen	NOAA Fisheries
28	Fred Serchuk	SAFMC SSC
29	Jason McNamee	ASMFC
30	Patrick Sullivan	Cornell/NEFMC
31	Jason Didden	MAFMC
32	Daemian Schreiber	NMFS HQ
33	Laura Diederick	NOAA Fisheries

Center for Independent Experts (CIE) Independent Peer Review of the Marine Recreational Information Program (MRIP) Fishing Effort Survey (FES) Calibration Model

Cynthia M. Jones
Director, Center for Quantitative Fisheries Ecology
Old Dominion University

For The Center of Independent Experts (CIE)

August 2017

Table of Contents

Executive Summary

The task of the MRIP Calibration Review Panel was to evaluate the performance of a new calibration model developed by F. Jay Breidt, Teng Liu, and Jean D. Opsomer of Colorado State University that permits conversion of telephone-survey effort to mail-survey effort and vice versa. The review of the MRIP FES Calibration took place at the Sheraton Silver Springs, in Silver Springs, MD on June 27-29, 2017. Dr. Paul Rago chaired the meeting which included three reviewers from the CIE (Ali Arab, Robert Hicks, Cynthia Jones) and three representing the Fisheries Management Councils and ASMFC (Jason McNamee, Fredric Serchuk, Patrick J. Sullivan).

A survey of recreational fishing effort has been conducted through a random-digit dial (RDD) telephone survey of coastal county households (CHTS) since 1981. With the advent of caller ID, portable prefixes and the proliferation of wireless-only households, the response rate has fallen below 10%. NMFS has chosen a mail survey (FES) to replace the CHTS after a three-year period from 2015-2017 with both surveys overlapping. The calibration model has been applied to the first year and one-half that has been completed of that overlapping period.

The proposed calibration model is based on a modification of the Fay-Herriot small area estimation method. The Fay-Herriot method (Fay and Herriot, 1979) is well established in the statistical literature and has known statistical behavior. Drs. Breidt and Opsomer and Mr. Liu modified the variance estimation component of that method to be analytically tractable and readily programmed in widely available software. It is fit as a log-normal model regressed on population size and state-by-wave factors with data from the 17 states along the US Atlantic and Gulf coasts. The differences in the non-sampling errors (e.g. frame coverage differences) were modeled with available covariates such as wireless coverage. The difference in the estimates includes the effect of sampling with different survey methods and an "irrational" factor that includes trends over time that could not be explicitly identified as influential covariates. Although some of the differences in effort estimation could be attributed to the increase in wireless only households, the majority of the difference could not be explained with existing available data. As the next year and one half of data become available, the MRIP team will have an opportunity to cross validate the model and evaluate the stability of model parameters. The Panel report includes recommendations to do so. After much consideration, the Panel concurred that this was an appropriate model for calibration.

Although the Fay-Herriot small-area estimation method is well suited for the CHTS to FES calibration, other approaches exist. The statistical team has examined modifications to their approach. For example, through use of the Akaike Information Criteria (AIC), they were able to determine that a simple time-varying ratio estimate that included error performed poorly compared with the current model. The modelers tested Bayesian approaches, but none were presented at the meeting.

TOR1e requested that the panel comment on the accuracy of the CHTS and the FES, but this is not possible for several reasons. The main reason is that anglers self-report their trip number in surveys that occur off the fishing grounds and there is no external validation of effort by an unbiased observer. Anglers must recall the number of trips that they took within the past two months when asked in the mail or telephone surveys. Many anglers do not keep a diary, although perhaps some keep a calendar, but there is a possibility that these trips are mis-remembered. While there may be little motivation to exaggerate fishing effort, a variety of factors can result in the reported trips differing from the actual number of trips taken and this type of problem is well documented in the survey literature. To measure accuracy one must undertake special surveys that match off site reports with on-site observations and this is best done in small area surveys. Because the effort estimate is combined with CPUE from the on-site angler intercept survey (APAIS) to estimate catch, there is an advantage to the fact that the FES is more efficient, statistically sound, and can potentially have a larger sample size. Larger sample size (more respondents) often results in smaller variance and better characterization of the effort distribution and, thus may result in less uncertainty when combined to produce estimates of catch.

In TOR2, we were asked to comment on the proceedings and issues around them, thus addressing process. I concur with the panel report (Appendix 4).

Having just completed the NAS MRIP Review, and having participated heavily in reviewing the FES and APAIS methodologies, had read much of the literature surrounding the survey methodologies, I was very familiar with the issues underlying the review of the calibration model. However, I noticed that several important reviews, reports, and manuals hadn't been posted for the panel. I and fellow panelists requested these materials on the first day of the meeting and they were promptly made available on the Confluence website. Moreover, the statisticians were not aware of the TORs until shortly prior to the meeting and had less time to prepare their presentations to address the TORs directly. Although they were able to provide us with additional information and presentations by the second day, it would have been better aligned if they had more notice.

During the meeting, I brought up my concerns with communication to the angling public about the calibration model and why the survey method was being changed. I have found that conveying ideas such as a random sample to the lay public challenging even for a trained communicator. These ideas are not simple and the FES is complex. A recent article in the Virginian Pilot by our local outdoor writer complained that NMFS was transitioning to an old-fashioned survey method and why didn't they just use smartphones (Tolliver, 2017)? The difficulty of the task of communicating to the angling public shouldn't be underestimated.

Communication to stock assessment scientists and fishery managers is also vital as the transition to the new survey is completed. The marked difference in effort estimates between the FES and CHTS has ramification of assessment of stock status, how to knit the time-series together, and on the allocation of catch between the commercial and recreational sectors. In some fisheries, the initial impact will be large and possibly disruptive. As time passes and the new survey estimate time series grows longer, problems may diminish. In the meantime, MRIP communication to these two groups will also rely on the difficult task of conveying concepts that underlie survey sampling, an area of statistics not commonly taught even to quantitative scientists.

Background

To develop a survey of recreational fishing, the location of the fishing area and the length of the season must be considered. For the coastal US, marine recreational fishing is extensive in area, covers both public and private access, and can occur year round on a variety of species and gears. One of the appropriate survey types for such a challenging assessment is a *complemented* survey, wherein effort is assessed off site of the fishery and catch-per-unit effort (CPUE) is observed directly on site. Both the Marine Recreational Fishery Statistics Survey (MRFSS) and the MRIP are two types of complemented surveys. MRFSS uses a telephone survey (Coastal Household Telephone Survey, CHTS) to measure effort off site and the Access-Point Angler Intercept Survey (APAIS) to obtain CPUE on site. In contrast, MRIP uses a mail survey, the Fishing Effort Survey (FES) to obtain effort offsite and APAIS for CPUE onsite. The changeover from the CHTS to the FES has resulted in significant differences in estimates of effort that must be reconciled as a new time series of effort is established. The review that I was asked to participate in was to evaluate a model to calibrate effort between the CHTS and FES. Dr. Opsomer noted in his presentation that when other large surveys in the US had change their survey methods, that they didn't try to establish a calibration between the old and new survey methods, so the NMFS MRIP calibration is one of the first of its kind.

Since 1981 the NMFS has monitored recreational fishing effort with the CHTS. The CHTS used random-digit dialing to reach households, using coastal county telephone prefixes. Initially, the CHTS saw high response rates but was inefficient, meaning that many non-angling households were contacted for every angling household that answered. Because the CHTS did not contact non-coastal county anglers, they were captured in the on-site survey component of the survey and the ratio of coastal to non-coastal anglers was used to increase the effort obtained from the CHTS. Several trends have rendered the CHTS less efficient and potentially less reliable over time. Telephone prefixes are now portable, such that a person who first got her telephone number in Kansas may now be living and fishing in Florida. Prefixes can no longer be relied on to indicate a coastal county resident. Moreover, telephone response rates have fallen dramatically with the almost universal use of caller ID. Also, the CHTS relied on land-line telephones and the majority of US households are now wireless only. Wireless-only households have different demographic characteristics than do land-line households, and NMFS can no longer be certain that the CHTS provides unbiased or efficient estimates of effort. NMFS investigated several methods to replace the CHTS and chose a mail survey (FES) that includes a small reward and multiple mailings as is standard practice for such surveys.

The task of the MRIP Calibration Review Panel was to evaluate the performance of a new calibration model developed by F. Jay Breidt, Teng Liu, and Jean D. Opsomer of Colorado State University that permits conversion of telephone-survey effort to mail-survey effort and vice versa. NMFS has undertaken concurrent mail and telephone surveys for 2015-2017 to which the calibration model has been applied. One and one-half years of the concurrent survey evaluation has been completed at the time of this review.

Review Activites:

Review of the MRIP FES Calibration took place at the Sheraton Silver Spring, Silver Spring, MD on June 27-29, 2017.

Prior to the meeting, I reviewed documents that were provided for us on a Confluence web site two weeks before the meeting. For the first two days of the meeting, there was a series of presentations that covered issues related to the two terms of reference and five sub-terms of TOR1. On Wednesday, the reviewers requested further clarification of the presenters on several issues relating to model specification. Meetings included questions from the Panel, the audience and web participants. The Panel began work on the report Thursday. Reviewers contributed equally to the discussions. On Friday July 7, Dr. Rago conducted a conference call to further discuss TOR 2. Upon my return home, I re-read the documents, reviewed the presentations and rapporteurs' notes, and obtained several other references to help me clarify my understanding of the calibration model. These are listed in the references section of this document. I participated via email in further edits of the Panel report prior to its submission.

A very detailed review of activities is included in the Panel Review (Appendix 4).

Summary of findings for each TOR wherein weaknesses and strengths are described, with conclusions and recommendations in accordance with terms of reference:

Calibration Model Accounting for a Recreational Fishery Survey Design Change

TOR1. Evaluate the suitability of the proposed model for converting historical estimates of private boat and shore fishing effort produced by the CHTS design to estimates that best represent what would have been produced had the new FES design been used prior to 2017.

The Panel concurred that is TOR was met.

1a) Does the proposed model adequately account for differences observed in the estimates produced by the CHTS and FES designs when conducted side-by-side in 2015-2016?

I concur with the Panel's statement under TOR 1a and agree with the statements included in the Panel Review Report (Appendix 4).

It is concerning that there is a 4 to 11 fold difference in estimated trips between the CHTS and the FES and this begs an explanation.

The National Academy of Sciences (2017) and the American Statistical Association have both reviewed the FES design and agree the methodology is statistically sound. The sampling frames differ between the CHTS and the FES. The CHTS uses coastal county prefixes with random digit dialing (RDD) to contact potential angling households, while the FES uses a list of addresses of coastal state residents overlain probabilistically with the list of residences of anglers holding state licenses. The FES also gives higher selection probability to the coastal county addresses (Thereby permitting potential comparisons between the CHTS and FES strata albeit with different sampling frames). The FES is a more efficient survey because of how the angler lists are used to increase inclusion probabilities of angling households. Moreover, anglers will answer a survey differently based on the mode of contact, mail or telephone (Dillman 2014). With RDD, the angler has no prior warning that they will be asked about their fishing trips and they may also be influenced by the survey agent asking the questions. They can ask the agent for clarifications, but may not have a calendar nearby to prompt their recall on the number of trips that they took in the past two months. However, depending on when the call is received there is a chance that not all anglers in the household would be home. With the FES, the angler has time to review their calendar (if they use one) or to think about the trips that they took, and all anglers in the household have time to answer the survey. However, if the respondents have a question not included on the FAQ sheet sent with the survey, then they may mis-interpret a question. In both cases, the answers are self-reported by the angler with no external verification as to trip number or location.

Some of the differences that might occur between the surveys have been explored as predictive covariates to the model, but none were influential except, to a small degree, the increase in wireless telephone coverage over time beginning in 2000. Initially, telephone response rates were high, but with the increasing proliferation of wireless-only households and caller ID, telephone response rates have plummeted. Thus, land-line households may represent a different demographic from the target population of marine anglers that the survey seeks to contact. I am not aware if there has been a study of the demography of the anglers responding to the CHTS or the FES that might help to uncover the differences in trips reported. Please note that response bias and response rates are two different issues. Just because response rate is low does not mean that the anglers contacted differ from those not answering. A non-response survey is necessary to discover bias. However, if the CHTS is not covering the full target population and if the demographics of those who respond have different fishing characteristics, then there is cause for concern that bias might exist. Without further investigation, one is left to conjecture with no proof.

Nonetheless, the FES rests on a statistically sound sampling design with known sampling inclusion probabilities, and is far more efficient than the telephone survey at reaching an angling household. Because the response rate has been higher for mail surveys, sample size can also be larger with potential concomitant decrease in variance –thereby lessening uncertainty. Additionally, with greater sample size, the underlying distribution of number of trips per household can be better characterized.

1b) Is the proposed model robust enough to account for potential differences that would have been observed if the two designs had been conducted side-by-side in years prior to 2015 with regards to time trending biases?

I concur with the Panel's statement under TOR 1b and agree with the statements included in the Panel Review Report (Appendix 4).

Although there are studies in other fields that have tried to uncover differences between survey modes (How the survey is delivered), without actual side-by-side assessments an answer is pure conjecture. One has to assume that any trends, for example in demographic types of recreation, have been influential on participation in recreational angling and in addition, that such trends would be consistent. Although NMFS conducted a short pilot study in North Carolina for 2012-2013 on the mail survey design, there are simply no data upon which to form a conclusion. To date, none of the possible factors that are hypothesized to cause differences in effort estimates between the CHTS and the FES has been shown to account for the differences seen in trips reported.

After returning from the Panel meeting, I have been wondering if the MRIP team have any data to explore the role of "gatekeeper" in the telephone survey. The gatekeeper is the person who answers the phone. I have been wondering whether such persons answered for themselves only, which could account for the difference. I don't know whether there are data to compare trips reported based on number of anglers in a household, or even if that has been done already. However, one could also hypothesize a difference if the demographic has been changing in the CHTS to older people who don't fish as often – hence the full target population is not being reached. Again, without data, all of this is pure conjecture.

1c) How does the approach used in developing the proposed FES/CHTS calibration model compare in terms of strengths or weaknesses with other potential approaches?

I concur with the Panel's statement under TOR 1c and agree with the statements included in the Panel Review Report (Appendix 4).

The advantage to the current calibrations model is the use of a modified Fay-Herriot small-area approach which is widely respected by statisticians (Datta et al., 2005, among others). The statisticians who developed the calibration model are skilled in this approach; the model has well-defined statistical properties, and can be used to evaluate potential factors that might explain differences in the number of reported trips. The calibration team has also derived a new way of formulating the variance estimators for the model that now allows for the use of off-the-shelf software. Having readily available, tested software saves time and lowers costs of producing estimates of effort and variance for either forward or back projecting units of effort in FES or CHTS equivalents.

The Panel also discussed other types of models that could be used for calibration. Even though this was not the task assigned to us in this review, the use of other models would have value. Dr. Sullivan suggested that the team look into the use of a Bayesian approach. That had been attempted by the Calibration Team with less than good success, but may be better implemented by different software and modeling approaches. The value of other models is that they may validate the difference seen in the two surveys or may be better able to retrieve explanatory variables that drive the differences. I would endorse this approach but think that the differences are more probably a result of problems in telephone coverage of the full target population, having better access to all household anglers through a mail survey, and a fundamental difference in how people respond to mail and telephone surveys. Hence, I don't think there is an easy answer to understanding the effort differences.

1d) Does the proposed calibration model help to explain how different factors would have contributed to changes in differences between CHTS and FES results over time?

I concur with the Panel's statement under TOR 1d and agree with the statements included in the Panel Review Report (Appendix 4).

The calibration model developed by Breidt, Teng and Opsomer permits the inclusion of covariates that can be used to uncover factors that account for differences in the effort estimates from the FES and CHTS. To date, there is no single factor that thoroughly accounts for the changes in the number of trips provided by the telephone survey. Trends in non-responses for telephone have not been explicitly modeled by factors other than the increase in wireless coverage that began in 2000. Even so, this factor accounts only for five percent of the modeled differences between the FES and CHTS projected back through time. It is important to note that only one year and one-half of three years of the side-by-side testing has been completed at this time. The model includes an "irrational" factor that the models have been unable to attribute to a known factor despite extensive efforts to uncover the reason for the different estimates.

The calibration model is detailed to the state and wave level, and even with such a short side-by-side survey has fit the data well, in part because of the small-area estimators that underlie the model. It will be important to test the stability of the model parameters as the next half of the data is included. The Panel has suggested that the model be cross validated with that new data, and I concur that will be an important test of the model. The model will not be used on the survey data until the three-year period of data collection in completed, and this will give the statisticians time to fine tune the model.

1e) Is it reasonable to conclude that revised 1981-2016 private boat and shore fishing effort estimates based on the application of the proposed FES/CHTS calibration model would be more accurate than the estimates that are currently available? Does evidence provided for this determination include an assessment of model uncertainty?

I concur with the Panel's statement under TOR 1e and agree with the statements included in the Panel Review Report (Appendix 4).

I was rather surprised by the wording of this TOR subcomponent. It seeks the panel to evaluate accuracy of the estimates, when in fact that is not possible. It led me to think that there is confusion about the type of data that are provided by offsite surveys such as the CHTS or FES. Anglers self-report their trip numbers in these surveys and there is no external validation of effort. The anglers' trips are not counted while they are fishing or when they complete their trip on site, but rather they must recall the number of trips that they took within the past two months. Many anglers do not keep a diary, perhaps some keep a calendar, but there is a possibility that these trips are mis-remembered. While there may be little motivation to exaggerate fishing effort, a variety of factors can result in the reported trips differing from the actual number of trips taken and this type of problem is well documented in the survey literature. To determine accuracy, a validation study would need to be devised that paired an onsite validation with the offsite survey. For such a large scale survey effort, this would be difficult and very expensive.

The calibration model does provide an estimate of uncertainty even though it doesn't explain the differences in the estimates. I believe that this is the best approach at this time with the data available.

Because the effort estimate is combined with CPUE from the APAIS to estimate catch, there is an advantage to the fact that the FES is more efficient, statistically sound, and can potentially have a larger sample size. A larger sample size (more respondents) often results in smaller variance and better characterization of the effort distribution and, thus may result in less uncertainty when combined to produce estimates of catch.

TOR2. Briefly describe the panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.

I concur with the Panel's statement under TOR 2 and agree with the statements included in the Panel Review Report (Appendix 4). The Panel took this TOR very seriously, we provided a detailed response to the TOR, and I will not repeat what we presented in the report.

Having just completed the NAS MRIP Review, and having participated heavily in reviewing the FES and APAIS methodologies, I was very familiar with the issues underlying the review of the calibration model. Even so, I wished that more material had been available prior to the meeting to inform me and fellow panelists of the previous reviews and workshops that address the issue for this panel review. Moreover, the statisticians were not aware of the TORs until shortly prior to the meeting and had less time to prepare their presentations to address the TORs directly. The statisticians on this project are among the best in the world and they were able to provide us with much information in a short period of time. However, we did not see detailed information on their initial explorations into model choice that would have led to a more productive meeting. They explained that they had tried other models that weren't as good as the Fay-Herriot approach and on the second day, they provided results of an Akaike Information Criteria test of different model configurations including the simple ratio estimator with error. Because there is a serious issue that will potentially affect allocation between fishing sectors given the new estimates, it was important that we had as much information as possible. The Panelists and statisticians understood the importance of this issue and did extra work to fill in gaps that were a consequence of this. For example, I went over the ASA evaluation that I hadn't seen previously, and amended my reading with other statistical papers on the Fay-Herriot approach.

I commend the presenters, panelists, and coordinators with a very professionally run meeting. Panelists were fully engaged, and the presenters very responsive to our questions, provided responses within 24 hours. The Confluence website was easy to access and made my work much easier than other CIE websites I have used. The conference room was well equipped and located conveniently. It was easy to see the presentations and hear the discussions. Dr. Rago did an outstanding job as Panel chairperson.

During the meeting, I brought up my concerns with communication of the calibration model and why the survey method was being changed, especially to the angling public. In my experience over 30 years with recreational angling surveys, I know that the estimates are only as good as the data and that the quality of the self-reported data especially will rest on the angler's belief in the legitimacy of the survey itself. I have found that conveying ideas such as a random sample to the lay public is challenging, even to a trained communicator. These ideas are not simple and the FES is complex. A recent article in the Virginian Pilot by our local outdoor writer complained that NMFS was transitioning to an old-fashioned survey method, and asked why didn't they just use smartphones (Tolliver, 2017)? I expect that the MRIP team will find challenges in conveying to the average angler that the mail survey is superior because of its probability basis compared with a volunteer smartphone survey that has unknown inclusion probabilities and sampling frame. I was contacted after the meeting by Gordon Colson who provided me with additional information on the MRIP communication approach. Nonetheless, the difficulty of the task of communicating to the angling public shouldn't be underestimated.

Communication to stock assessment scientists and fishery managers is also vital as they transition exclusively to the FES. The marked difference in effort estimates between the FES and CHTS has ramifications on assessments of stock status, on how to knit the time-series together, and on the allocation of catch between the commercial and recreational sectors. In some fisheries, the initial impact will be large and possibly disruptive. The MRIP communication to these two groups will also rely on the difficult task of conveying concepts that underlie survey sampling, an area of statistics not commonly taught even to quantitative scientists.

Appendix 1: Bibliography of materials provided for review

Transition Plan for the FES:

https://www.st.nmfs.noaa.gov/Assets/recreational/pdf/MRIP%20FES%20Transition%20Plan%20FINAL.pdf

Report recommending the FES to replace the CHTS: *Finalize Design of Fishing Effort Surveys* (https://www.st.nmfs.noaa.gov/pims/main/public?method=DOWNLOAD_FR_PDF&record_id=1179)

2015 Benchmarking Progress Report:

https://www.st-test.nmfs.noaa.gov/Assets/recreational/pdf/2015_FES_Progress_Report-20161115.pdf

Report on FES/CHTS Calibration Model:

BACKGROUND INFORMATION

(1) Presentations at the review

- Introduction – Paul Rago
- MRIP Fishing Effort Survey – Rob Andrews
- Importance of calibrated catch for fishery stock assessments – Richard Methot
- Importance of Calibrated Catch for Fisheries Management – Andy Strelcheck
- Calibrating survey estimates over time – Jean Opsomer
- A Calibration Methodology for CHTS to FES
- Transition – Jay Breidt
- Day One Review – Paul Rago
- Follow Up on Comments for " A Calibration Methodology for CHTS to FES" – Jay Breidt

(2) Other Papers that I Read

Datta, G.S., Rao, J.N.K., and Smith, D.D. 2005. On measuring the variability of small area estimators under a basic area level model. Biometriks 92-1: 183-196.

Dillman, D.A., Smyth, J.D. and Christian, L.M. 2014. Internet, Phone, Mail, and Mixed-Mode Surveys: a tailored design method. 4th Edition, Wiley.

Fay III, R.E. and Herriot, R.A. 1979. Journal of the American Statistical Association, Vol. 74, No. 366 (Jun., 1979), pp. 269-277.

NAS. 2017. Review of the Marine Recreational Information Program (MRIP). National Academy Press. Washington, D.C.

Tolliver, J. 2017. How many fish are really in the ocean? Some congressmen think federal fisheries can do a better job of finding out. Virginian Pilot, April 25, 2017. https://pilotonline.com/news/local/environment/how-many-fish-are-really-in-the-ocean-some-congressmen/article_dfc2f052-dab8-590c-829a-5d510dd8e983.html.

Appendix 2: A copy of this Statement of Work

Statement of Work

National Oceanic and Atmospheric Administration (NOAA)

National Marine Fisheries Service (NMFS)

Center for Independent Experts (CIE) Program

External Independent Peer Review

Calibration Model Accounting for a Recreational Fishery Survey Design Change

Background

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

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards. (http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).

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

Scope

The Office of Science and Technology requests an independent peer review of a calibration model proposed for use in revising statistics produced by surveys of marine recreational fishing effort on the Atlantic coast and in the Gulf of Mexico. This calibration model is considered by the Marine Recreational Information Program (MRIP) to be very important to adjust historical time series of recreational effort and catch estimates in order to account for biases in past sampling and estimation methods that have become apparent with the development of a new, more statistically sound method. The calibration model is intended to account for past biases in private boat and shore fishing effort estimates that have resulted from the continued use of a legacy random-digit-dial telephone survey design that has degraded over time and will be replaced with the implementation of a new mail survey design (the “Fishing Effort Survey”, or FES) in 2018.

Calibration Model for the Fishing Effort Survey

In 2015, MRIP formed a Transition Team to collaboratively plan a transition from a legacy telephone survey design to a new mail survey design for estimating private boat and shore fishing effort by marine recreational anglers. Since 2008, MRIP had conducted six pilot studies to determine the most accurate and efficient survey method for this purpose on the Atlantic and Gulf coasts. The most recent study, conducted in four states in 2012-2013, compared a new mail survey design with the Coastal Household Telephone Survey (CHTS) design that has been used since 1979. MRIP subjected the final report from the pilot project to external peer review in 2014 and certified the new survey design, called the Fishing Effort Survey (FES), in February 2015 as a suitable replacement for the CHTS. The FES is much less susceptible to potential sources of bias than the CHTS because it can reach more anglers, achieve higher response rates, and is less prone to possible recall errors. The pilot project results indicated that FES estimates were substantially higher than CHTS estimates for both private boat fishing and shore fishing.

MRIP recognized the FES should not be implemented immediately as a replacement for the CHTS, and a well thought out transition plan was needed to ensure that the phase-in of the FES is appropriately integrated into ongoing stock assessments and fisheries management actions in a way that minimizes disruptions to these processes, which are based on input from multiple data sources over lengthy time series. The Transition Plan developed by the Transition Team called for side-by-side benchmarking of the FES against the CHTS for three years (2015-2017) with the development and application of a calibration model to enable adjustment of past estimates that account for biases in historical effort and catch statistics after the second year. With this timeline, revised estimates can be incorporated into stock assessments during 2018 using a peer reviewed calibration model, and new Annual Catch Limits (ACLs) can then be set in 2019 for at least some stocks.

Requirements

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

NMFS will provide a Chair who has experience with U.S. fisheries stock assessments and their application to fisheries management. The Chair would ensure that reviewers understand the importance of maintaining a comparable time series of marine recreational fisheries catch statistics for use in stock assessments and their application to fisheries management. The Chair will not be selected by the contractor and will be responsible for facilitating the meeting,

developing and finalizing a summary report and working with the CIE reviewers to make sure that the ToRs are addressed in their independent reviews.

Tasks for Reviewers

Pre-review Background Documents

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

Transition Plan for the FES:

https://www.st.nmfs.noaa.gov/Assets/recreational/pdf/MRIP%20FES%20Transition%20Plan%20FINAL.pdf

Report recommending the FES to replace the CHTS: *Finalize Design of Fishing Effort Surveys* (https://www.st.nmfs.noaa.gov/pims/main/public?method=DOWNLOAD_FR_PDF&record_id=1179)

2015 Benchmarking Progress Report:

https://www.st-test.nmfs.noaa.gov/Assets/recreational/pdf/2015_FES_Progress_Report-20161115.pdf

Report on FES/CHTS Calibration Model:

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

Panel Review Meeting

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

Contract Deliverables - Independent CIE Peer Review Reports

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

Other Tasks – Contribution to Summary Report

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

the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Foreign National Security Clearance

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

Place of Performance

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

Period of Performance

The period of performance shall be from the time of award through July 31, 2017. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables in accordance with the following schedule. Within two weeks of award	Contractor selects and confirms reviewers
Within four weeks of award	Contractor provides the pre-review documents to the reviewers
June, 2017	each reviewer participates and conducts an independent peer review during the panel review meeting
Within two weeks of panel review meeting	Contractor receives draft reports
Within two weeks of receiving draft reports	Contractor submits final reports to the Government

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

MRIP Calibration Model Peer Review Workshop

Sheraton Silver Spring Hotel

Silver Spring, MD

June 27-29, 2017

ATTENDANCE LIST

#	NAME	AFFILIATION
1	Paul Rago	MAFMC SSC
2	Dave Van Voorhees	NOAA Fisheries
3	John Foster	NOAA Fisheries
4	Ali Arab	Georgetown University
5	Rob Hicks	College of William and Mary
6	Cynthia M. Jones	Old Dominion University
7	Richard Cody	NOAA support ECS
8	Teng Liu	Colorado State University
9	Thomas Sminkey	NOAA Fisheries/ST1
10	Steve Turner	NOAA Fisheries SEFSC
11	Andy Strelcheck	NOAA Fisheries - SERO
12	Richard Methot	NOAA Fisheries - HQ
13	Karen Pianka	NOAA Fisheries – ST1
14	Lauren Dolinger Few	NMFS ST1
15	Chris Wright	NMFS - SF
16	Sabrina Lovell	NMFS ST
17	Patrick Lynch	NMFS ST
18	Melissa Karp	NMFS ST
19	Toni Kerns	ASMFC
20	Steve Ander	Gallup
21	Tommy Tran	Gallup
22	Melissa Niles	Fifth Estate/MRIP CET
23	Yong-Woo Lee	NOAA - Fisheries
24	Jay Breidt	Colorado State University
25	Jean Opsomer	Colorado State University
26	Rob Andrews	NOAA Fisheries
27	Ryan Kitts-Jensen	NOAA Fisheries
28	Fred Serchuk	SAFMC SSC
29	Jason McNamee	ASMFC
30	Patrick Sullivan	Cornell/NEFMC
31	Jason Didden	MAFMC
32	Daemian Schreiber	NMFS HQ
33	Laura Diederick	NOAA Fisheries

1

Appendix 4. Amended Panel Report to include text body only

Summary Report
Marine Recreational Information Program (MRIP)
Fishing Effort Survey (FES) Calibration Review

Calibration Model Review Meeting
June 27-29, 2017
Sheraton Hotel
Silver Spring, MD

July 14, 2017

Draft #4

Panel Members

Paul Rago (Chair)
Ali Arab
Robert Hicks
Cynthia Jones
Jason McNamee
Fredric Serchuk
Patrick J. Sullivan

2

Table of Contents

Executive Summary

A primary objective of the Marine Recreation Information Program (MRIP) is the improvement of the statistical basis of methods for estimating catches of recreationally caught fish in the coastal US. MRIP has implemented a new program for estimating fishing effort that relies on a mail-based survey rather than a historical telephone survey. This report summarizes a technical review of a calibration model to interrelate estimates of recreational fishing effort derived from the Coastal Household Telephone Survey (CHTS) with the Fishing Effort Survey (FES). The FES uses a mail survey and national angler registry. A panel of seven independent scientists met with consultant statisticians and MRIP staff to review a proposed methodology that could express historical estimates of fishing effort in terms of the new FES. A side-by-side experiment of the two methods, conducted in 2015 and 2016, served as the basis for this review.

The proposed methodology builds upon known properties of the CHTS and FES sampling designs, and an extensive time series of historical data. The calibration model relies on standard and highly-regarded methodology known as the Fay-Herriot method for small area estimation. Alternative modeling approaches might have been considered, but the proposed method was reasonable and scientifically-defensible. The authors are commended for introducing several innovations to estimate variances and to achieve analytical consistency. The final estimators have desirable properties and can be implemented with readily available software. The proposed model was considered an elegant approach for dynamic predictions of recreational fishing effort. Particularly notable was the property that allowed for forward and backward estimation by alternate survey modes (i.e., CHTS vs FES). The proposed method preserves design aspects of historical and current surveys and incorporates important differences among states, waves (i.e., two-month calendar periods) and fishing modes. The processes of model identification and variable selection (i.e., consideration of potential predictive covariates) were well done.

The Panel expressed concern on several topics, none of which was considered as sufficient to preclude implementation of the Fay-Herriot model. Comparison of estimates of effort derived from the side-by-side CHTS and FES surveys (2015 and 2016) resulted in large differences (2 to 11-fold). While many hypotheses were considered that might account for these differences, data analyses and the proposed model revealed no single hypothesis (or covariate) was sufficient. Further refinement of the modeling approach, particularly when the results of the 2017 side-by-side experiment are available, is recommended. Refinements include further simulation testing and cross-validation comparisons with the first two years of data. As more information is acquired about the FES there may be additional opportunities to consider alternative models for calibration. Given the importance of such changes for many stock assessments and management decisions, future modifications must be able to demonstrate significant advantages over the proposed small-area estimation model prior to consideration for implementation. The Panel recommended additional efforts to improve communication of these results to scientists, statisticians, fishery managers, and the general public. Each will require varying levels of detail. The Panel also suggests that renewed attention be given to the recommendations of two previous NAS reviews of the recreational statistics programs.

1. Introduction

1.1 Background

The Review Panel for the MRIP-FES Calibration Model Review met from June 27 to June 29 to review a statistical model developed by F. Jay Breidt, Teng Liu and Jean D. Opsomer, of Colorado State University. The review committee was composed of three scientists appointed by the Center for Independent Experts (CIE): Robert Hicks, The College of William and Mary; Cynthia Jones, Old Dominion University; and Ali Arab, Georgetown University. In addition, representatives from the New England (Patrick Sullivan) and South Atlantic (Fredric Serchuk) Scientific and Statistical Committees, and the Atlantic States Marine Fisheries Commission (Jason McNamee) served on the review panel. The meeting was chaired by Paul Rago as a member of the Mid-Atlantic Fishery Management Council Scientific and Statistical Committee.

The panel reviewed supporting documentation and presentations prepared by MRIP staff, led by Dave Van Voorhees, and their contractors from the Department of Statistics at Colorado State University. John Foster, Ryan Kitts-Jensen, and Richard Cody of MRIP acted as rapporteurs. Other staff from the Office of the Science and Technology, notably Karen Pianka, assisted in the handling of documents via a web-based application. Jason Didden of the Mid-Atlantic Fishery Management Council provided support for the webinar. Approximately 35 people participated in the open sessions of the meeting. The meeting followed the agenda in Appendix 2 with respect to the sequence but not necessarily the timing of the events. Adjustments were made for differences in the duration of presentations and follow-up questions.

1.2 Review of Activities

About ten days before the meeting the panel was given access to a comprehensive working paper summarizing the proposed statistical model. Prior the meeting, the chair met with the presenters and MRIP staff via a conference call to discuss the scope of the contributions, presentation format and draft agenda. All supporting documents and presentations were made available to reviewers via a web-based application known as Confluence. In addition, the MRIP staff added a web page to their site that provided members of the public and other managers with access to key papers and presentations. The meetings were broadcast via webinar with able assistance of Jason Didden of the Mid-Atlantic Fishery Management Council. Mr. Didden also managed all of the in-room computer and audio visual equipment.

The meeting opened on the morning of Tuesday June 27, 2017, with welcoming remarks and comments on the agenda by Van Voorhees and Rago. Participants and audience members introduced themselves. Following introductions, sessions on June 27 were devoted to presentation and initial discussions of five agenda topics. Robert Andrews provided an overview of the transition from the fishing effort surveys based on a Coastal Household Telephone Survey (CHTS) to the Fishing Effort Survey (FES), based on a mail survey. Richard Methot addressed the importance of properly calibrated effort for estimation of catch in stock assessments. Andy Strelcheck addressed the importance of catch information as a basis for fisheries management policies and decisions, such as allocation. Jean Opsomer provided an overview of the challenges of applying calibration methods to historical time series. Jay Breidt led the presentation of the proposed statistical calibration model.

Each presentation was followed by a question and answer period by panel members and as appropriate, by other meeting attendees. Questions from web participants were also addressed at opportune times. A formal public comment period was reserved on each day of the meeting.

The Panel met in closed session at the end of each day to discuss the day's presentations, progress toward answering the agenda, and to make plans for the following day.

Follow-up discussions on the first day presentations were held on Wednesday June 28. The Panel requested additional data and clarification from the presenters, including greater details on the model results. Day two began with an overview of the activities of Day One and an overview of the day's work plan. Most of the Panel's efforts were devoted to questions on the statistical calibration model. Material provided by Jay Breidt and colleagues enhanced the Panel's understanding of the model and its performance. A short presentation by Paul Rago used the results of model predictions to compare results over states and fishing modes (i.e., shore vs private boat).

Day Two also included a formal public comment period and an initial summary of the Panel's findings. This was done to ensure that all participants were aware of the general outcomes of the review. The Panel stressed that this summary was not to be considered a consensus report. Instead it represented a summary of the perspectives of the Panel.

Following the initial presentation of findings, the Panel met in closed session to begin writing the Summary Report. Day Three consisted of a half day meeting for Panelists only. The purpose of the meeting was to summarize the various viewpoints herein with respect to the Terms of Reference.

The Panel completed drafting this Summary Report by correspondence, evaluating each ToR. The Chair compiled and edited the draft Panel Summary Report, which was distributed to the Panel for final review before being submitted to the MRIP. Each Panelist also provided an independent summary of their perspectives and as appropriate, with details on potential improvements to the calibration model and its application. Individual panelist reports for CIE participants were sent to the Center for Independent Experts for initial editing for completeness. Reports of Panelists supported directly by the Agency via contract were sent to the Chair. All reports were made available to MRIP staff for fact checking but were not altered for content.

The Panel agreed that scientific and statistical analyses conducted by the presenters were thorough, statistically sound, and innovative. Specific comments on the details of the analyses are provided below.

2. Review of MRIP FES Calibration Model

2.1 Synopsis of Panel Review

The Panel commented that the proposed methodology builds upon known properties of the existing sampling design, the proposed new method, and extensive time series of historical data. A review of calibration approaches in other disciplines revealed no comparable attempts to adjust a historical times series forward or backward in time in response to new information from a side-by-side comparative surveys. The proposed model was considered to be an elegant approach for dynamic predictions of recreational fishing effort. Particularly notable was the property that allowed for forward and backward estimation by alternate survey modes (i.e., CHTS vs FES). Notably, the proposed method preserves design aspects of historical and current surveys and incorporates important differences among states, waves (i.e., two-month calendar periods) and fishing modes. The Panel acknowledged the extensive exploratory data analyses on model development, alternatives, and testing performed by the MRIP scientific staff and consultants. The processes of model identification and variable selection (i.e., consideration of potential predictive covariates) were well done.

Although the Panel identified several alternative modeling approaches and other candidate covariates that might have been considered, the Panel acknowledged that the proposed method was a reasonable and scientifically defensible estimation approach.

The calibration model relies on standard, well known, and highly regarded methodology. The authors are commended for introducing several innovations to estimate variances and to achieve analytical consistency. The final estimators have desirable properties and can be implemented with readily available software.

The Panel expressed concern on several topics, none of which was considered as sufficient to preclude implementation of the model. Comparison of estimates of effort derived from the side-by-side CHTS and FES surveys (2015 and 2016) resulted in large differences (2 to 11-fold). While many hypotheses were considered that might account for these differences, data analyses and the proposed model revealed no single hypothesis (or covariate) was sufficient.

Model performance was partially assessed by sensitivity analysis of specific alternative hypotheses on the distribution of the “irregular” random effect (an effort effect not accounted for explicitly in the model). However, additional simulation work may be necessary to more thoroughly test overall model performance. As additional information becomes available by the end of the 2017 side-by-side surveys, it is recommended that a series of cross-validation exercises be conducted to compare model results based on the first two years of model results. Other permutations of cross calibration comparisons may be instructive with respect to stability of model parameter estimates and prediction error induced by various data rarefaction methods. As more information is acquired about the FES there may be additional opportunities to consider alternative models for calibration. Given the importance of such changes for many stock assessments and management decisions, future modifications must be able to demonstrate significant advantages over the proposed small-area estimation model prior to consideration for implementation.

The Panel spent considerable time discussing the communication of results. It was recognized that at least three distinct audiences must be addressed: scientists and statisticians, fishery managers, and the general public. Each will require varying levels of detail without compromising the integrity of the model or its underlying principles. A “lay person’s” version of the methods would be valuable for communicating results to multiple audiences. Model results, in combination with a similar calibration exercise for the APAIS, have significant downstream impacts for assessments and management. The Panel also suggests that renewed attention be given to the recommendations concerning communications of two previous NAS reviews of the recreational statistics programs.

Despite progress in improving communication with stakeholders, the Panel is aware of important misconceptions among the angling communities regarding the transition to the new mail-based survey mode. The new MRIP website is a considerable improvement but direct, pro-active communication and dialogue with fishing groups, perhaps with downloadable podcasts, YouTubes etc. and in-person presentations to the angling community would be valuable.

2.2 Evaluation of Terms of Reference

2.2.1 Term of Reference 1

Evaluate the suitability of the proposed model for converting historical estimates of private boat and shore fishing effort produced by the CHTS design to estimates that best represent what would have been produced had the new FES design been used prior to 2017.

- The Panel concurs that this TOR and its subcomponents listed below (1a,1b, 1c, 1d, 1e) were met.

a) Does the proposed model adequately account for differences observed in the estimates produced by the CHTS and FES designs when conducted side-by-side in 2015-2016?

- The results of the side-by-side surveys are central to the development of the proposed model. The model parameterization accounts for these changes but does not provide insight into the underlying mechanisms resulting in differences in estimated angling effort.
- The new mail survey mode has advantages relative to issues of comprehensiveness of angler coverage within households, efficiency of the estimate, a better sampling frame, a more thoughtful consideration of individual angler effort, improved demographic information, better identification of fishing location, and enhanced follow-up with respondents to reduce non-response. Collectively these features are thought to yield more reliable metrics of angling effort and serve as a basis for improved understanding in the future as the new survey continues. These advantages are relevant to 2015 and onward but do not necessarily extend back to historical estimates.

b) Is the proposed model robust enough to account for potential differences that would have been observed if the two designs had been conducted side-by-side in years prior to 2015 with regards to time trending biases?

- The Panel had difficulty formulating a response to this TOR as it required conjecture about unidentified underlying causal mechanisms contributing to observed differences and hypothetical comparisons of survey mode responses in the past.

- Insufficient information was provided to inform this decision either before or during the meeting. Potential approaches were discussed but could not be implemented in the time available.
- Although the proposed model allows for inclusion of other causal mechanisms, neither the investigators nor the Panel were able to identify covariates that vary over time and meet the criteria necessary for expansion to total angling effort estimates. Furthermore, data collection procedures during the CHTS did not collect information that in retrospect (e. g., demography, gender), might have allowed such inference.

c) How does the approach used in developing the proposed FES/CHTS calibration model compare in terms of strengths or weaknesses with other potential approaches?

- The investigators conducted an extensive analysis of within-model comparisons of reduced model parameterizations using the model selection procedure known as the Akaike Information Criterion. One sub-model included a simple ratio with random effects that had much lower explanatory power. A preliminary analysis was conducted and reviewed by the Panel that corroborated the inappropriateness of the simple ratio estimator.
- Other models exist that could be used, including Bayesian Hierarchical modeling, state-space modeling, and time-varying ratio estimation. The investigators provided the panel with a summary of their experiences with some of these alternatives but the results of these comparisons were not available to the Panel. Given the responses of the investigators, the Panel concurred with the conclusion to focus on the modified Fay-Herriot approach.

d) Does the proposed calibration model help to explain how different factors would have contributed to changes in differences between CHTS and FES results **over time**?

- As noted above the causal mechanisms resulting in differences between survey estimates remain elusive.
- Raw survey data in the CHTS could be examined more carefully but it is unknown whether such data exist over a sufficient span of years to support such analyses
    - As presently configured the model is limited in terms of what can be explored but alternatives may be useful.
    - Within the existing data, there do not appear to be covariates, other than log(Population) that would explain the major differences seen between survey modes. The wireless effect captures a minor component of the contrast. The Panel and Investigators agreed that the wireless effect may be a proxy for a wide range of factors.
    - Demographic information in the CHTS would have been instructive and is essential for proper historical analyses. However, it is uncertain that such data exist over a sufficient span of years to support such analyses.

    - Consideration of spatially differentiated data that has been collected historically at a finer scale (e.g., Census tract) may yet contain information sufficient to illuminate explanatory factors related to this TOR.
- The “Gate keeper” effect has been documented as a major influence in the CHTS but a complete understanding remains difficult to identify.

e) Is it reasonable to conclude that revised 1981-2016 private boat and shore fishing effort estimates based on the application of the proposed FES/CHTS calibration model would be more accurate than the estimates that are currently available? Does evidence provided for this determination include an assessment of model uncertainty?

- No conclusions can be reached regarding the accuracy of calibrating self-reported data from one survey mode to the other. However, the Panel noted that bias in the historical CHTS may not be as large as observed in contemporary CHTS samples due to degradation of survey coverage and other factors.
- Gatekeeper, recall bias, response rate etc. indicate that the mail survey is preferred to a phone, particularly in relation to statistical and operational efficiency. This conclusion was supported by the 2006 and 2017 NRC reports, and also in a separate review conducted by the ASA.

- Response rate per se is not a problem unless differences in fishing activity differ between respondents and non-respondents

2.2.2 Term of Reference 2

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

The following sections highlight the Panel’s concerns about the peer review meeting, including preparations before the meeting and follow-up activities. The Panel recognizes the complexity of the revisions of MRIP transition process and the need to satisfy many different audiences. The following recommendations are offered in the context of constructive criticism to improve the quality of future peer-review panels. While there is some redundancy in this section with the Panel’s comments in section 2.1, the text below provides additional clarification of issues and more broadly reflects the diversity of the Panelist’s opinions. he text below draws heavily from comments provided by the Panelists via correspondence after the meeting. Therefore some sections below may be reflected in part or their entirety in the Panelist’s individual reports.

<u>Pre-Meeting Preparations</u>

Four background documents (Section 5 , Working Papers) were provided to Panel members two weeks prior to the meeting, and all additional documents and presentation were made available to the Panel during the meeting via a web-site (i.e., Confluence). The Panel Chair provided each of the reviewers with a proposed meeting Agenda a day prior to the start of the meeting, requesting that any comments and possible changes be provided back to him before the meeting opened. As the proposed Agenda was satisfactory to all of the Panel members, no changes to the Agenda were needed.

Panelists expressed concerns about pre-meeting preparations, noting an inadequate assembly of all the pieces needed to address the terms of reference. Greater overall coordination among presenters would have been desirable to ensure that all the relevant information was covered. Additional background documents would have been useful for the review; for example, the MRIP Handbook should have been provided before to provide more information about the telephone and mail surveys. Comprehensive previous reviews of the MRIP, such as those from the National Academy of Sciences should have been brought to the attention of the Panel, not all of whom had extensive knowledge of the history of MRIP. In this context, basic details about the surveys including similarities and differences in definitions of effort (notably, the definition of angling households), questions on the questionnaires, etc. would have helped the Panel to more effectively conduct the review.

<u>Proceedings</u>

The review panel proceedings went smoothly. Operationally, the meeting room had sufficient space for the Panel, presenters, and meeting attendees. The sound and projection systems worked well, as did the webinar link. Representatives from the Office of Science and Technology served as Rapporteurs and provided comprehensive summary notes to the Panel.

Discussions during the 2½ day MRIP Calibration Review illuminated various issues related to the results provided in the background documents and the PowerPoint presentations. Many of the concerns involved clarification of the information provided and/or requests for additional data and analyses. Additional data, model outputs and documents were made available to the Panel during the meeting. In all cases, these requests were satisfactorily fulfilled allowing the Panel to gain fuller insight on:

- Sampling designs, strengths, and shortcomings of the telephone (CHTS) and mail (FES) survey methods, including their relative performance and sources of error.

- Development, design, statistical properties, testing, and application of the proposed MRIP FES calibration model. This included consideration of alternative modeling approaches, cross-validation of the modeling framework to examine the stability of model parameter estimates (as well as prediction errors), the sufficiency and explanatory power of the model's covariates, and the possible underlying mechanism(s) affecting the distribution of the "irregular" random effect, which is not explicitly accounted for within the proposed small-area estimation approach.

- Potential impacts of the calibrated recreational fishing effort estimates during 1981-2016 on future stock assessments, and on subsequent fishery management policies and practices.

- Need to effectively communicate the results of the calibration work (as well as the basis and need for continuing only the mail-based survey method in the future) to various constituency groups (i.e., the recreational and commercial fishing communities; scientists; fishery managers; the lay public) so that these groups fully understand and accept the calibration results and their subsequent use in deriving recreational catch estimates for application in stock assessments and in the fishery management process.

The Review Panel acknowledged that the proposed MRIP FES calibration model developed by Breidt *et al.* was a well-suited and statistically-appropriate approach to obtain calibrated estimates of recreational fishing effort (by state and 2-month calendar quarter for shore-based and private boat anglers) during 1982-2016.

Utility of Presentations

The presentations on the implications of revised recreational catch estimates on stock assessments and on management measures and regulatory protocols were instructive, but the Panel would have appreciated more quantitative examples. For example, implications for stock assessment models could have been drawn from the previously completed scoping exercises conducted by the Northeast and Southeast Fisheries Science Centers. Similarly, the Panel noted that detailed simulation exercises would also have been instructive.

The presentation on the Fay-Herriot model was lucid and effective, but the Panel would have appreciated more details on the model components and the model building process. Also, a summary of candidate modeling approaches —and details on the process that led to the proposed model—would have been very useful. Such details, as provided on the second day of the review, were greatly appreciated.

Greater detail would have been appreciated on the survey methodologies in the phone and mail surveys. The simulation exercise was an important start, but further simulation testing beyond those conducted would have lent greater support to the applicability of the Fay-Herriot model to the CHTS vs FES calibration. Further work on simulated data sets is suggested during the third-year comparisons (i.e., when the 2017 telephone and mail survey data are fully available).

Terms of Reference

The presenters did not address the TORs directly, which made it harder for the Panel to assess the relevance of some of the information presented with regard to the TORs. Consequently, the Panel spent a substantial portion of the question/answer periods (and discussion time) on obtaining the requisite information to address the TORs. It was evident during these interactions that the model developers had conducted additional work relevant to the TORs (such as investigation of additional modeling approaches). However, because the developers were unaware of the TORs, neither the primary report nor the presentations specifically addressed the TORs. Follow-up work accomplished by the developers during the meeting and subsequently shared with the Panel gave the Panel confidence that sufficient model scoping had been performed.

The TORs presume that converting CHTS to FES is the appropriate way to standardize the MRIP effort data. However, the statistical work available for the review primarily focused on the mathematical aspects of the calibration and not on which set of estimates reflects a truer representation of fishing effort. Lacking a sufficient statistical justification for standardizing the MRIP data to the FES estimates created problems both during the review and in addressing the TORs.

TOR1e seeks the Panel's opinion concerning the accuracy of effort estimates obtained from the CHTS and the FES. The Panel understands that any survey conducted offsite of the fishery, such as mail or telephone surveys, rely on angler self-reported data which is not subject to verification. Self-reported data is subject to a variety of biases including recall problems which can result in misremembered time and number of trips. Without an external measure of fishing from an onsite survey covering the same population in space and time, angler self-reported data cannot be verified. While the Panel comments on the calibration from CHTS to FES, there is no basis to comment on accuracy of either survey.

Documentation for Meeting

It would have been helpful for the Panel to have been provided (several weeks before the review) additional background documents (available from the MRIP Team and/or the MRIP Website) to enhance a collaborative understanding by Panel members of various aspects of the MRIP program and of recent analyses using MRIP data. For example, the *MRIP Data User Handbook,* and the October 2016 report, '*Possible Effects of Calibration Scenarios on Stock Assessments Planned for the MRIP Fishing Effort Survey Transition'* would have especially useful for Panel members to have had and read before the actual peer review occurred

Prior to the presentation and discussion of the Breidt *et al.* report at the Peer Review, this report was difficult to understand for anyone other than a highly-trained statistician.

Although a more complete understanding of this report was fostered by distribution of a PowerPoint presentation a week or so before the Review Meeting (and subsequently enhanced at the meeting by direct dialogue and interaction with the authors of the paper who clarified and responded to many issues raised by the Panel), it is recommended that in any future reviews in which a highly technical paper is seminal to the crux of such reviews that efforts be made by the paper authors to present the essence of their work in a manner that facilitates full appreciation and understanding of the import of such work by educated non-specialists. This becomes especially critical when the methods/approach provided in a paper will have significant downstream effects. This matter should be recognized in the future APAIS peer review.

Ancillary Analyses

The Panel appreciated the opportunity to investigate the details of the statistical calibration/prediction model on day 2. The model and assumptions were well thought out, but the Panel needed to better understand model inputs, parameter definitions, and nuances of the Fay-Herriot model. Similarly, the Panel appreciated the opportunity to solicit more information on model development and model selection beyond what was initially available at the meeting. Panelists received model parameter estimates upon request but did not have time at the meeting to explore them fully. Access to more detailed model outputs and the estimation code in R would have been valuable.

Also, apparently, several independent data analyses existed too, separate from the model, and it would have been good to have had a presentation and some discussion on that. Exploratory analyses of the pairwise calibration data was considered useful and should be considered for summarization when the analyses of the 2017 data are conducted.

Communication

Panelists expressed concerns about the need for improved communication at several different levels:

- to the Panel prior to the meeting,
- within the various analytical components,
- to the members of the Transition Team,
- to broader audience of stake holders.

An advantage of the current review was the inclusion of several external independent experts having expertise beyond fisheries science. This helped ensure that the methods were critically evaluated and represented state of the art, but increased the burden during pre-meeting preparations to ensure that all relevant contextual documents were available and fully comprehensible. Concerns were expressed that information essential for the review was not provided at level of detail that the Panel members expected.

The transition from the MRFSS to MRIP has required a massive restructuring of the data collection procedures while maintaining a continuous time series of reliable catch data.

Continuity of data has required coordination with governmental, academic, and industry stakeholders. Likewise, the process has involved multiple experiments and survey tests to demonstrate the value of proposed changes and development of advanced calibration approaches. This review constituted one component of this transition. Despite enormous improvements in the MRIP website and availability of raw and processed data at varying degrees of resolution, the Panel recommended greater coordination among the diverse analytical groups. The complexity of the transition requires that technical reviews are both sequential and interdependent. As such the review of any single technical issue (e.g., calibration between CHTS and FES) must rely upon and recognize the conclusions of earlier Panels. In the present review, this Panel relied on the conclusions of the ASA reviewers who noted the superiority of the FES over CHTS. Independent panels of scientists rarely accept prior reviews without questioning. Indeed, this is the nature of science. Hence it essential that each Panel in future reviews be provided with a summary of the full set of previous reviews and their relationship to the current review.

There is a strong need to effectively communicate the results of the calibration work (as well as the basis and need for continuing only the mail-based survey method in the future) to various constituency groups (i.e., the recreational and commercial fishing communities; scientists; fishery managers; the lay public) so that these groups fully understand and accept the calibration results and their subsequent use in deriving recreational catch estimates for application in stock assessments and in the fishery management process. Consideration should be given to a variety of communication approaches including but not limited to public meetings, seminars, podcasts, YouTube, and use of skilled educators.

Finally, it is recommended that an updated report/timetable/chart be prepared to illustrate current progress in meeting the tasks and timelines identified in the FES Transition Plan. This undertaking should also take note of how the recommendations tendered in all previous peer reviews of the MRIP Program (including the 2006 and 2016 NAS Reviews) have been addressed.

Improvements to Future Peer Review Processes

The Panel noted that review process left little time for an intensive review of the data, the model, and the computer code used to develop the results. Such analyses are often part of a stock assessment review (e.g., SAW/SARC https://www.nefsc.noaa.gov/saw/, or SEDAR http://sedarweb.org/). In the spirit of improving future reviews, the Panel suggests consideration of more broadly based working groups based on scientific input within and outside NOAA Fisheries. In stock assessments working groups have a strong technical focus and meet several times prior to the final assessment. Working groups would have the opportunity to examine the proposed methodologies in greater detail, included detailed reviews of the data and methods, and tests with simulated data. Exchanges of code, or reliance on standard packages in stock assessments provide both quality assurance and opportunities for improvements. Moreover, the products of working groups typically assure subsequent reviewers that the products under review are comprehensive and representative of diverse viewpoints. In particular, a working-group process would document the model building process and allay concerns of reviewers who will always wonder why a particular alternative was not considered. Having those prior decisions as a matter of record would enhance the efficiency and quality of the review process.

The Panel recognizes that this recommendation would need to be part of the overall transition from MRFSS to MRIP. Indeed, the current Transition Team process that has regular updates on progress, conversations with stock assessment scientists and various stakeholders, and plans for upcoming tasks, already includes the essential elements of a more focused working group approach. In view of the importance of upcoming technical decisions for stock assessments, managers and harvesters, the Panel strongly urges consideration of this proposal.

Individual Reviewer Report
Marine Recreational Information Program (MRIP) Fishing Effort Survey (FES) Calibration Review

Calibration Model Review Meeting
June 27-29, 2017
Sheraton Hotel
Silver Spring, MD

July 16, 2017

Panel Member Review from

Jason McNamee

Table of Contents

1. Executive Summary

This report summarizes the technical review from one of seven independent scientists of a calibration model to interrelate estimates of recreational fishing effort derived from the Coastal Household Telephone Survey (CHTS) with the Fishing Effort Survey (FES). A side-by-side experiment of the two methods, conducted in 2015 and 2106, served as the basis for this review.

The proposed modeling methodology uses a time series of historical recreational effort data and a set of explanatory covariates to convert the effort metric from one currency to another. This can be done in either direction, meaning FES can be converted to CHTS and vice versa. This is an attribute of this selected approach. Alternative modeling approaches were investigated by the researchers, but were not presented formally to the review panel. Despite this, the proposed method was deemed reasonable and scientifically-defensible and the authors are commended for their work on the Fay-Herriot model for this calibration application. An attribute of the approach the researchers used is that the model is implemented in R statistical software, making the model code accessible to other researchers for additional testing and future development. The proposed model is considered an elegant approach for dynamic predictions of recreational fishing effort, allowing for forward and backward estimation in different currencies of effort (i.e., can be calculated in CHTS or FES effort metrics). Differences among states and seasonal changes in effort (as represented by two-month periods referred to as waves) are accounted for in the model parameters, a very important aspect to the future use of this approach to account for recreational effort changes through time.

There were concerns on several topics, but as noted in the summary report, none of the concerns prohibit implementation of the Fay-Herriot model for the MRIP calibration. No single hypothesis (or covariate) was sufficient to explain the differences between the CHTS and FES estimates and this will make the explanation to the public difficult. This difficulty in outreach should not be underestimated by the MRIP program. When the results of the 2017 side-by-side experiment are available, it is recommended that some additional work be conducted and documented including simulation testing beyond that already done for the irregular term in the model. This testing will better answer some of the terms of reference that were not well addressed during the current workshop. Additionally, there may be an opportunity during this update to better document alternative models that are tested for the calibration exercise, allowing the researchers to better support why the Fay-Herriot method was deemed a superior method to other options available. Further refinement of some of the important covariates will be a worthwhile effort when the 2017 side-by-side data becomes available, namely, the population covariate can be filtered to better represent the population of interest (i.e. coastal communities) rather than the broad population growth of the entire state. Finally, while recognizing that resources are limited, future side-by-side comparative survey experiments should be considered to test how the model parameter estimates are holding up over time.

2. Introduction

2.1 Background

For the sake of completeness, section 2 of this individual report is reproduced from the review panel summary report. The Review Panel for the MRIP-FES Calibration Model Review met from June 27 to June 29 to review a statistical model developed by F. Jay Breidt, Teng Liu and Jean D. Opsomer, of Colorado State University. The review committee was composed of three scientists appointed by the Center for Independent Experts (CIE): Robert Hicks, The College of William and Mary, Cynthia Jones, Old Dominion University and Ali Arab, Georgetown University. In addition, representatives from the New England (Patrick Sullivan) and South Atlantic (Fredric Serchuk) Scientific and Statistical Committees, and the Atlantic States Marine Fisheries Commission (Jason McNamee) served on the review panel. The meeting was chaired by Paul Rago as a member of the Mid-Atlantic Fishery Management Council Scientific and Statistical Committee.

The panel reviewed supporting documentation and presentations prepared by MRIP staff, led by Dave Van Voorhees, and their contractors from the Department of Statistics at Colorado State University. John Foster, Ryan Kitts-Jensen, and Richard Cody of MRIP acted as rapporteurs, providing valuable daily summaries for the Panel. Other staff from the Office of the Science and Technology, notably Karen Pianka, assisted in the efficient handling of documents via a web-based application. Jason Didden of the Mid Atlantic Fishery Management Council provided extensive support for the webinar. Approximately 35 people participated in the open sessions of the meeting. The meeting followed the agenda in Appendix 2 with respect to the sequence but not necessarily the timing of the events. Adjustments were made for differences in the duration of presentations and follow-up questions.

2.2 Review of Activities

About ten days before the meeting the panel was given access to a comprehensive working paper summarizing the proposed statistical model. Prior the meeting, the chair met with the presenters and Marine Recreational Information Program (MRIP) staff via a conference call to discuss the scope of the contributions, presentation format and draft agenda. All supporting documents and presentations were made available to reviewers via a web-based application known as Confluence. In addition, the MRIP staff added a web page to their site that provided members of the public and other managers with access to key papers and presentations. The meetings were broadcast via webinar with the able assistance of Jason Didden of the Mid-Atlantic Fishery Management Council. Mr. Didden also managed all of the in-room computer and audio visual equipment.

The meeting opened on the morning of Tuesday June 27, 2017, with welcoming remarks and comments on the agenda by Van Voorhees and Rago. Participants and audience members introduced themselves. Following introductions, sessions on June 27 were devoted to presentation and initial discussions of five agenda topics. Robert Andrews provided an overview of the transition from the fishing effort surveys based on a Coastal Household Telephone Survey (CHTS) to the Fishing Effort Survey (FES), based on a mail survey. Richard Methot addressed the importance of properly calibrated effort for estimation of catch in stock assessments. Andy Strelcheck addressed the importance of catch information as a basis for fisheries management policies and decisions, such as allocation. Jean Opsomer provided an overview of the challenges of applying calibration methods to historical time series. Jay Breidt led the presentation of the proposed statistical calibration model.

Each presentation was followed by a question and answer period by panel members and as appropriate, by other meeting attendees. Questions from web participants were also addressed at opportune times. A formal public comment period was reserved on each day of the meeting.

The Panel met in closed session at the end of each day to discuss the day's presentations, progress toward answering the agenda, and to make plans for the following day.

Follow-up discussions on the first day presentations were held on Wednesday June 28. The Panel requested additional data and clarification from the presenters, including greater details on the model results. Day two began with an overview of the activities of Day One and an overview of the day's work plan. Most of the Panel's efforts were devoted to questions on the statistical calibration model. Material provided by Jay Breidt and colleagues enhanced the Panel's understanding of the model and its performance. A short presentation by Paul Rago used the results of model predictions to compare results over states and fishing modes (i.e., shore vs private boat).

Day Two also included a formal public comment period and an initial summary of the Panel's findings. This was done to ensure that all participants were aware of the general outcomes of the review. The Panel stressed that this summary was not to be considered a consensus report. Instead it represented a summary of the perspectives of the Panel.

Following the initial presentation of findings, the Panel met in closed session to begin writing the Summary Report. Day Three consisted of a half day meeting for Panelists only. The purpose of the meeting was to summarize the various viewpoints herein with respect to the Terms of Reference.

The Panel completed drafting this Summary Report by correspondence, evaluating each TOR. The Chair compiled and edited the draft Panel Summary Report, which was distributed to the Panel for final review before being submitted to the MRIP. Each Panelist also provided an independent summary of their perspectives and as appropriate, with details on potential improvements to the calibration model and its application. Individual panelist reports for CIE participants were sent to the Center for Independent Experts for initial editing for completeness. Reports of Panelists supported directly by the Agency via contract were sent to the Chair. All reports were made available to MRIP staff for fact checking but were not altered for content.

The Panel agreed that scientific and statistical analyses conducted by the presenters were thorough, statistically sound, and innovative. Specific comments on the details of the analyses are provided below.

3. Review of MRIP FES Calibration Model

3.1 Synopsis of Individual Panel Member Review

As noted in the review panel summary report, the proposed methodology builds upon known properties of the existing sampling design and the extensive time series of historical data on important potential covariates that could impact effort information. The presentation given during the review on the synthesis of other attempts at calibrating survey information in other disciplines revealed no comparable attempts to adjust a historical times series backward in time in response to new information from a side-by-side comparison. Having no additional knowledge of projects conducted to calibrate surveys in this manner, the premise that this was a unique investigation was accepted, and this illustrated that the research conducted to calibrate the effort information being produced by the two survey approaches was not as simple as retrofitting some previously tested approach to the MRIP effort estimation information.

The proposed model was considered to be a well-designed approach for dynamic predictions of recreational fishing effort. It was also agreed that the property allowing for forward and backward estimation by alternate survey modes (i.e., CHTS vs FES) was an attribute of this approach. Because of the ability to switch the “currency” of the estimate between CHTS and FES, additional comparisons can be made in the future to test how well the model is able to estimate past CHTS data given new FES data, which would allow for additional judgement as to how well the model performs through time as conditions potentially change. It would be beneficial to conduct future side by side comparisons to provide new data with which to test how well the model continues to perform in to the future, but it is understood that resources are limited.

The lack of information presented on alternative modeling approaches and other candidate covariates that might have been considered was an item of note. The proposed method was a reasonable and scientifically defensible estimation approach, but it was difficult to judge whether this approach was truly superior to other potential approaches that could have been used. For instance, one of the hypotheses of why the CHTS has become unreliable is that there is a change in behavior of anglers with regard to the use of caller ID and switching to cell phones from landline telephone systems. This effect could be a time trending effect, and there are state space modeling approaches that can estimate time trending effects (Newman et al 2014) , and there are also Bayesian hierarchical techniques (Gelman et al 2013) that can function in this same way to better account for and quantify process errors that may occur within modeling frameworks. It appeared that at least some of these types of approaches were investigated by the researchers, however this information came out during discussion so was not formally presented to the reviewers nor included in any of the pre-meeting materials, making it difficult for the reviewers to judge for themselves the logic of modeling approach used by the researchers.

The final selected calibration model chosen by the researchers is a well-founded and appropriate choice, and an additional attribute is that the researchers implemented the model using R statistical software (R core team 2016), which is free and readily available. This will allow future running and future development of the model. It would have been useful and appropriate to have had the source code provided by the researchers to the reviewers as this would have allowed for a more mechanistic understanding of the model which was somewhat difficult to fully grasp from the working paper provided on the model alone.

In accordance with the summary report from the review panel, the concerns expressed above aren't considered sufficient to preclude implementation of the model. Echoing one important concern, however, the result of the calibration increases effort by a large margin. This will have major implications on the outcome of stock assessment information, and as importantly, this result will impact many facets of management such as proportion of harvest across fishing modes (i.e. party and charter boat mode effort is not impacted by this calibration while private boat and shore angling modes are increased) and may have impacts to allocations of important recreational species amongst states. Given the magnitude and importance of the changes of the calibration results to our fisheries processes, it will be important to better define what the causative factors are for this change so that this information can be communicated out to the fisheries community at large. Without this systematic understanding of what caused the changes between the two different effort survey methodologies, it will be difficult for constituents to buy in to the information being produced by the model.

3.2 Evaluation of Terms of Reference

3.2.1 Term of Reference 1

Evaluate the suitability of the proposed model for converting historical estimates of private boat and shore fishing effort produced by the CHTS design to estimates that best represent what would have been produced had the new FES design been used prior to 2017.

- The Individual Panel Member concurs that this TOR and its subcomponents listed below (1a,1b, 1c, 1d, 1e) were met.

a) Does the proposed model adequately account for differences observed in the estimates produced by the CHTS and FES designs when conducted side-by-side in 2015-2016?

While in agreement that the model is suitable for understanding differences between the survey methodologies, similar concerns to those expressed in the summary report remain. The model converts CHTS to FES effort metrics, allowing for a retrospective recalibration of the effort levels back in time, which is critical to being able to better assess fish stocks with high recreational participation. However, the model and the investigation in to the data failed to determine any one or set of covariates that would account for why the results between the two survey estimates of effort are so different from each other in a mechanistic way. This is not a fault of the researchers, many data sources and potential covariates were investigated during model development to test various hypotheses on why the effort calculations differed between the two survey types, which was an attribute of the project, but this point is brought up to highlight the need to continue to investigate the underlying data and to seek out new data sources that may better explain in a mechanistic way why the changes occurred due simply to a change in survey method, and why the changes are so large.

b) Is the proposed model robust enough to account for potential differences that would have been observed if the two designs had been conducted side-by-side in years prior to 2015 with regards to time trending biases?

In agreement with the summary report, the main covariate investigated to test the hypothesis of a time varying trend in the performance of the CHTS was a significant effect in the model (namely, the increase in wireless-only households), however the practical effect of that covariate did not appear to be strong enough to create the differences seen in the output by itself. This finding will make it difficult for the fishing community to understand why the effect of the model is so large. Further investigation in to additional explanatory covariates should continue and their impact on the model should be tested through time. Not only is this important for the edification of the fishing community, but if and when there is a better causal mechanism identified for the changes in effort estimation, there will be more confidence that the model is not misspecified and will continue to produce reliable effort calibration calculations forward in time.

With regard to how robust the model is, the researchers focused on one area of sensitivity testing, and that had to do with the error distribution assumption around the “irregular” terms. This was a strength of the research, and the researchers proved that their model was robust to different assumptions with regard to this error distribution. This strategy could have been extended to other areas of the model, and a more comprehensive simulation testing could have been done to test the models performance to different biases in underlying data. A fuller simulation testing procedure would have more comprehensively met this term of reference, but the simulation testing that did occur was appreciated and gave confidence in the model performance to this specific assumption.

Issues with not identifying the main causal mechanism notwithstanding, the model does appear to produce output consistent with the underlying hypothesis that the CHTS information has degraded through time, and the output when converting from CHTS to FES information shows the magnitude of the differences between the two surveys decreasing when applied to the historical time series. This is gives some confidence that the model as specified is picking up and accounting for the signal in the data.

c) How does the approach used in developing the proposed FES/CHTS calibration model compare in terms of strengths or weaknesses with other potential approaches?

This was an area of weakness found during the review. It was apparent that the researchers did rigorous internal model testing to find the best fitting model given the data that they investigated, which was documented during the presentation and was covered in the working paper. What was not apparent was how the researchers ended up at their preferred approach, the Fay-Herriot model. During the review the researchers did mention that they tested alternative modeling approaches including some of the approaches mentioned by the review panel in the summary report, however this was not documented in the working paper nor was it a highlight of the presentation given by the researchers. The researchers did verbally explain to the reviewers that this approach vetting did occur, however, given that this was a direct TOR for the review workshop, it would have been preferable to have had more information on this part of the research project.

It would still be worthwhile to produce some information on the approach vetting that occurred during this project in an effort to document and support the Fay-Herriot procedure for this use. Beyond the additional support for the CHTS to FES calibration, a better documentation of the approach vetting procedure will prove valuable for the other calibration efforts that the MRIP will be undergoing in the near future, such as the calibration of the new Access Point Angler Intercept Survey (APAIS) procedures to the old intercept methodology.

As a side note, it was noted that the researchers were not provided the TORs that the reviewers were working under until the week prior to the review workshop, which may have led to a number of the concerns expressed by the reviewers. For future calibration work undertaken by MRIP, an effort should be made to get the review TORs to the researchers so that they may highlight these pieces of information, which will make the review workshops run smoother and allow for easier evaluation of the research projects relative to the given TORs.

d) Does the proposed calibration model help to explain how different factors would have contributed to changes in differences between CHTS and FES results **over time**?

The calibration model certainly helps to explain the differences found between the two survey methods through time. The identification of the underlying causal mechanism remains to be better defined as mentioned previously, however the existing set of covariates chosen for the model seems to account for the differences between the two survey methods, and also seems to account for the fact that these effects change through time as evidenced in Figures 3 and 4 from the Breidt et al working paper (Appendix 1).

Some of the data that was used could be better defined. Specifically, the population covariate used was a broad population metric, but filtering this metric to the population considered to be in close proximity to the coast might be a better way to investigate the population effect in the model. Different trends in population changes in coastal areas relative to the overall population of a state may be informative and could provide a better statistical fit of the model to the data.

Despite these comments, the model does show how the data sources in the model effect the output over time. This was further highlighted by work produced by Review Workshop Chairman Paul Rago during the workshop, showing how trends in the data changed depending on the years investigated.

e) Is it reasonable to conclude that revised 1981-2016 private boat and shore fishing effort estimates based on the application of the proposed FES/CHTS calibration model would be more accurate than the estimates that are currently available? Does evidence provided for this determination include an assessment of model uncertainty?

As noted in the summary report, there was no information provided with regard to evaluating accuracy, nor would this be possible in the context of the information available as this whole project centers around determining differences in self-reported data. Without doing a study specific to investigate the accuracy of a self-reporting program, which would be very different from the research done for the calibration workshop, this information could not be produced by the researchers nor evaluated by the reviewers.

The only possibility that could have been investigated would have been simulation testing of the model with regard to known hypothetical data. The researchers could have produced datasets with specific know biases, and then investigated how the model performed relative to those biases. This would have produced information on the robustness of the model to various forms of bias, however not on “accuracy” in the technical sense of the term.

3.2.2 Term of Reference 2

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

In accordance with the review workshop summary report, the following are reviewer specific comments following the same section format used in the summary report. Some of the following is duplicative with those comments in the summary report.

<u>Pre-Meeting Preparations</u>

Background documents were provided to review panel members prior to the meeting, but additional documents and presentations were only made available during the meeting after it was realized additional information was needed to better evaluate the TORs for the workshop.

Coordination between the researchers and the MRIP with regard to the TORs would have created better flow in the workshop and less on the fly information would have been needed if the TORs had been available to the researchers with an understanding that the review panel was going to be evaluating their work relative to those TORs.

Additional background documents would have been useful for the review as well, in particular existing information of the previous comprehensive reviews of the MRIP, such as the one from the National Academy of Sciences (NAS). In this context, basic details about the surveys including similarities and differences in definitions of effort, questions in the new FES survey, etc. would have helped the reviewers to more effectively conduct the review. On the positive side, the review panel was fortunate to have had two of the participants from this previous NAS review on the panel to help with the understanding of these previous determinations.

<u>Proceedings</u>

In accordance with the review panel summary report, the meeting and proceedings went well. The researchers did an excellent job producing information during the workshop to help the reviewers with their task of evaluating the calibration model, the concerns noted above notwithstanding. Additionally, the workshop chairman did an exemplary job of keeping the researchers and reviewers on track to complete the review in the time allotted.

Given the effectiveness of the proceedings and the ability of the researchers to produce needed information during the workshop, it is believed that the proposed MRIP FES calibration model developed by Breidt *et al.* is a well-suited and statistically-appropriate approach to obtain calibrated estimates of recreational fishing effort (by state and 2-month calendar quarter for shore-based and private boat anglers) during 1982-2016.

Utility of Presentations

The presentations on the implications of revised recreational catch estimates on stock assessments, management measures, and regulatory protocols were helpful and helped put the workshop in to context, but additional presentations, would have been very informative for more specific context of the impacts of the calibration exercise. As an example, there are previously completed stock assessment exercises conducted by the Northeast and Southeast Fisheries Science Centers that could have been presented to show what the effect of the new estimates are relative to previously assessed population information.

Similarly, as mentioned above, more comprehensive simulation exercises would have been useful in the evaluation of the TORs, and so could have been presented in addition to the specific model information that was presented.

The presentation on the Fay-Herriot model was well done and helped with the interpretation of the working paper, but more details on the model components and the model building process would have been appreciated. Also, a summary of other candidate modeling approaches that were vetted would have been useful. Such details, as provided on the second day of the review, were greatly appreciated and helped the reviewers complete their evaluation of the TORs. Further work on simulated data sets is suggested for the final year comparisons.

Terms of Reference

The presenters did not address the TORs directly, which made it hard for the reviewers to assess the relevance of some of the information presented with regard to the TORs. Consequently, the reviewers spent a substantial portion of the discussion periods on obtaining the requisite information to address the TORs, some of which were not able to be addressed fully due to the constraint of time. Follow-up work accomplished by the researchers during the meeting gave the reviewers confidence that sufficient model scoping had been performed, though more information on this topic should be aggregated for the benefit of future review workshops on the various MRIP transitions in progress.

TOR 1e sought information concerning the accuracy of effort estimates obtained from the CHTS and the FES. Self-reported data is subject to a variety of biases that result from forgotten aspects of fishing trip. Without an external measure of fishing from an onsite survey covering the same population in space and time, angler self-reported data cannot be verified or tested for accuracy. While the review panel commented on the calibration from CHTS to FES, there was no basis to comment on accuracy of either survey to meet that TOR.

Documentation for Meeting

The technical report on the Breidt et al. calibration modeling approach was difficult to understand. The researchers did a great job of enhancing understanding during the meeting, including an informative exchange on Day 2 of the workshop between the reviewers and the researchers, and this helped inform evaluation of the TORs on the model by clarifying what the modeling approach was actually doing with regard to the data examined. This should be better appreciated in the future APAIS peer review to allow that workshop to proceed in a more efficient fashion.

Ancillary Analyses

The presentation and documentation of the model and assumptions were well thought out, but the reviewers would have appreciated more information on the model inputs, parameter definitions, and nuances of the Fay-Herriot model. Panelists received model parameter estimates upon request but did not have time at the meeting to explore them fully. Access to more detailed model outputs and the estimation code in R would have been valuable.

Additionally, several independent data analyses existed, separate from the model, which came out during the workshop. It would have been helpful to have had a presentation and some discussion on these alternate approaches. Exploratory analyses of the pairwise calibration data was considered useful and should be considered for summarization when the analyses of the 2017 data are conducted.

Communication

There was a lot of discussion on the communication of the MRIP transition process to the public and other stakeholder groups, of which this calibration model is one element. While this was not a direct TOR for the review workshop, these points were believed to be important for the MRIP to consider. A detailed outline of the importance of the communication of the calibration model, and the MRIP transition process in general, is given in the review panel's summary report and is not reproduced here, but this reviewer will emphasize the importance of heeding those comments as the MRIP transition proceeds.

4. Bibliography

Background Papers

Background on the MRIP Calibration Model Peer Review may be found at:
https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/index.html

The National Academies of Sciences, Engineering, and Medicine. 2016. Review of the Marine Recreational Information Program (MRIP) Washington, DC: The National Academies Press. doi: 10.17226/24640 https://www.st.nmfs.noaa.gov/confluence/display/FESCALIB?preview=/73074985/73728799/NAS_MRIP_review.pdf

National Research Council. 2006. Review of Recreational Fisheries Survey Methods. Committee on the Review of Recreational Fisheries Survey Methods, ISBN: 0-309-66075-0, 202 pages. http://www.nap.edu/catalog/11616.html

Working Papers

Development and Testing of Recreational Fishing Effort Surveys Testing a Mail Survey Design: Final Report. Project Team Members: Rob Andrews, NOAA Fisheries, J. Michael Brick, Westat, Nancy A. Mathiowetz, University of Wisconsin-Milwaukee. July 31, 2014. https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/documents/Report_recommending_FES_to_replace_CHTS--Finalize_Design_of_Fishing_Effort_Surveys.pdf

Marine Recreational Information Program Fishing Effort Survey Transition Progress Report. October 28, 2016. https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/documents/2015_benchmarking_progress_report.pdf

Marine Recreational Information Program Transition Plan for the Fishing Effort Survey Prepared by the Atlantic and Gulf Subgroup of the Marine Recreational Information Program Transition Team May 5, 2015 https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/documents/MRIP_FES_Transition-Plan_FINAL.pdf

A Small Area Estimation Approach for Reconciling Mode Differences in Two Surveys of Recreational Fishing Effort draft: F. Jay Breidt, Teng Liu, Jean D. Opsomer Colorado State University June 10, 2017 https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/documents/DRAFT-Report_of_Calibration_Model.pdf

Additional References

Gelman, A, Carlin, JB, Stern, HS, Dunson, DB, Vehtari, A, Rubin, DB. 2013. Bayesian Data Analysis (3rd edition). Chapman and Hall, CRC. 667 p.

Newman, KB, Buckland, ST, Morgan, BT, King, R, Borchers, DL, Cole, DJ, Besbeas, P, Gimenez, O, Thomas, L. 2014. Modelling Population Dynamics: Model formulation, fitting and assessment using state-space methods. Springer. 215 p.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing,
Vienna, Austria. URL https://www.R-project.org/.

5. Appendices

Appendix 1. Figures referred to in this review report

From Breidt et al working paper:

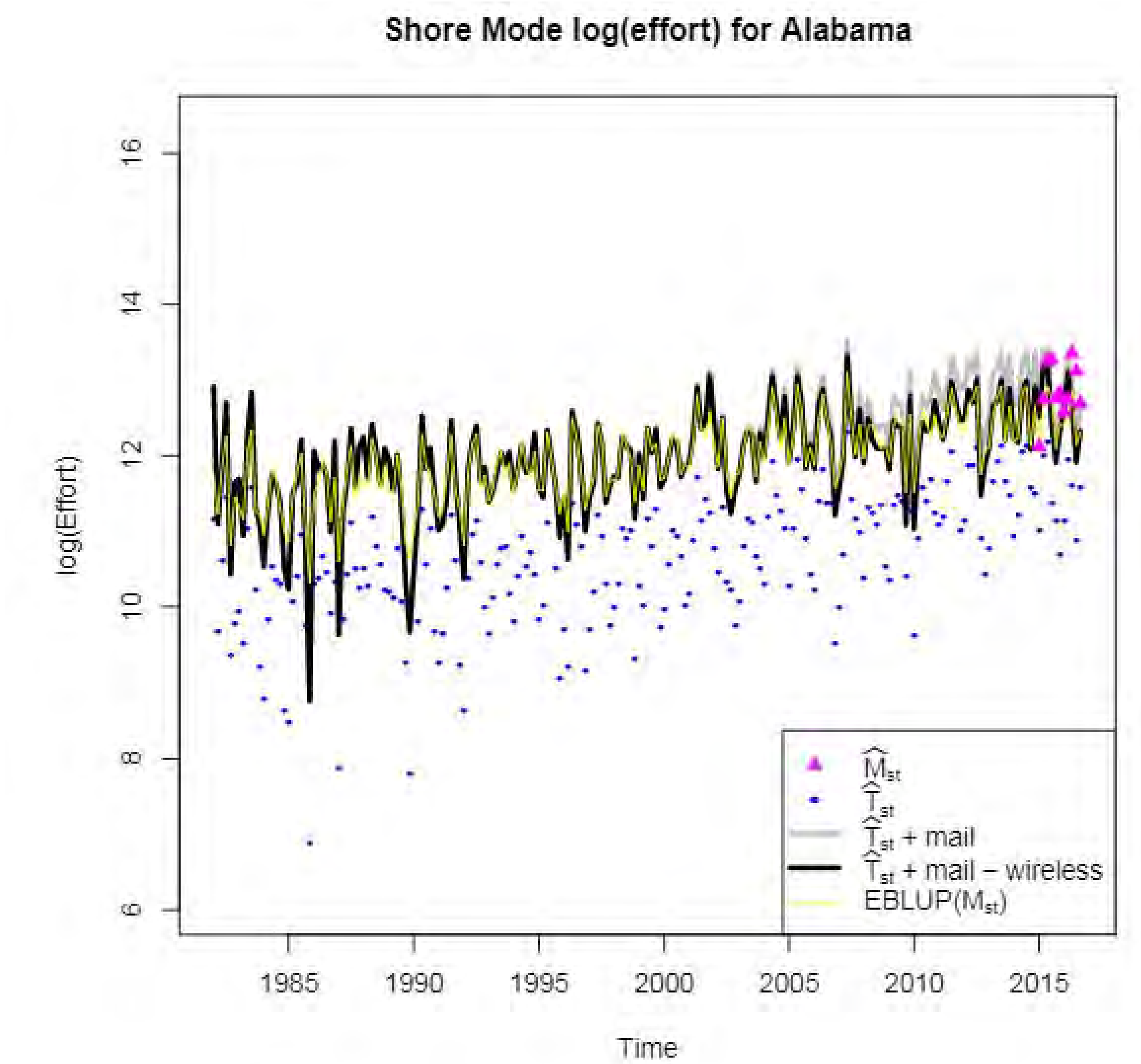


Figure 3: EBLUP's $\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \hat{\psi}\right)\right\}$ (gold curve) of mail targets $\{M_{st}\}$ for shore fishing log-effort in Alabama. Blue dots are telephone log-effort estimates $\{\widehat{T}_{st}\}$ and pink triangles are mail log-effort estimates $\{\widehat{M}_{st}\}$. For comparison to EBLUP's, gray curve is the estimator $\{\widehat{T}_{st} + \boldsymbol{b}'_{st}\widehat{\boldsymbol{\mu}}\}$ that adjusts only for mail methodology effects, and black curve is $\{\widehat{T}_{st} + \boldsymbol{b}'_{st}\widehat{\boldsymbol{\mu}} - w_{st}\boldsymbol{c}'_{st}\widehat{\gamma}\}$ that adjusts for mail and wireless.

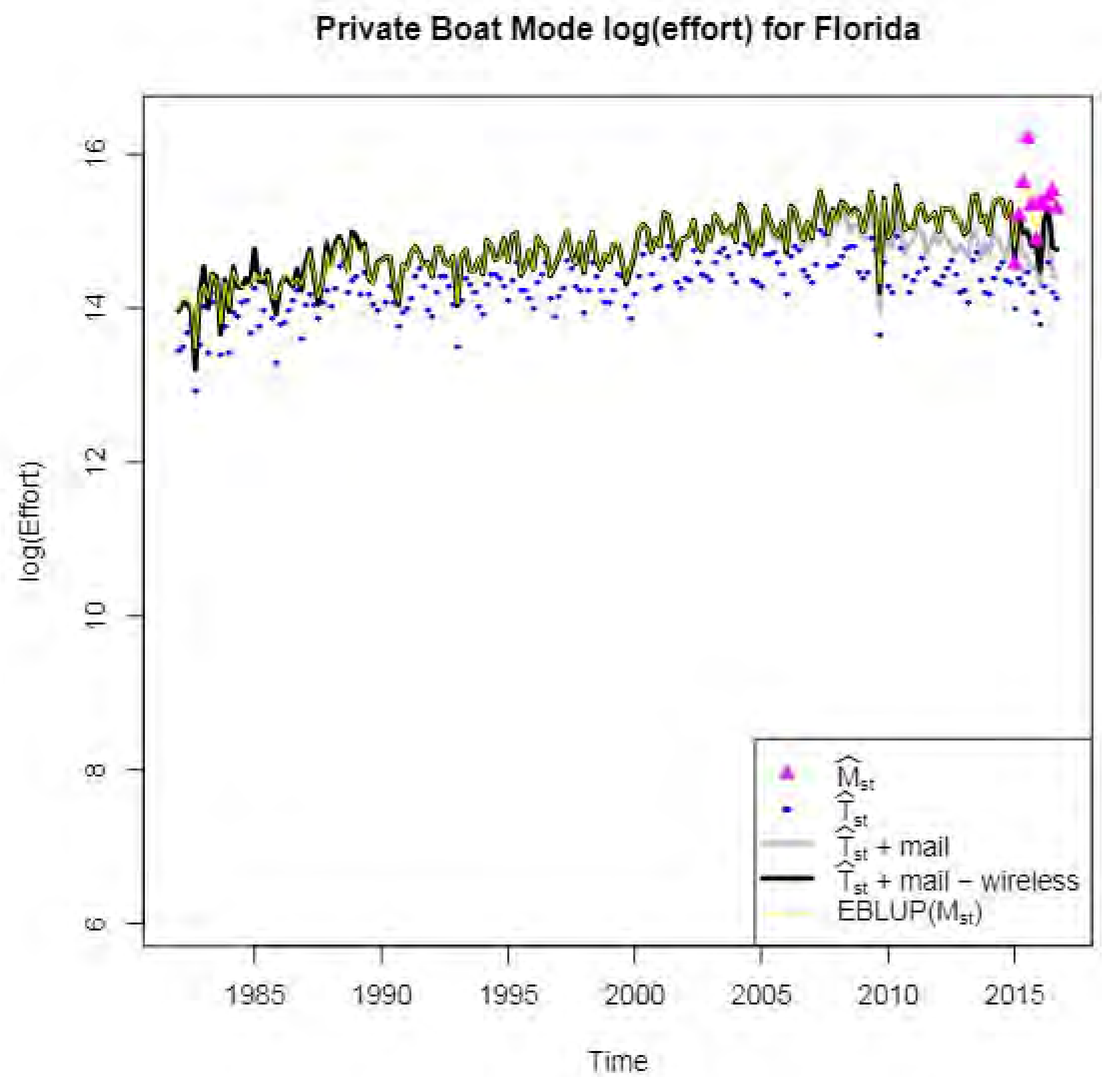


Figure 4: EBLUP's $\left\{\phi_{st}\left(\widehat{\beta},\hat{\psi}\right)\right\}$ (gold curve) of mail targets $\{M_{st}\}$ for private boat fishing in Florida. Blue dots are telephone log-effort estimates $\{\widehat{T}_{st}\}$ and pink triangles are mail log-effort estimates $\{\widehat{M}_{st}\}$. For comparison to EBLUP's, gray curve is the estimator $\{\widehat{T}_{st}+\boldsymbol{b}'_{st}\widehat{\boldsymbol{\mu}}\}$ that adjusts only for mail methodology effects, and black curve is $\{\widehat{T}_{st}+\boldsymbol{b}'_{st}\widehat{\boldsymbol{\mu}}-w_{st}\boldsymbol{c}'_{st}\widehat{\boldsymbol{\gamma}}\}$ that adjusts for mail and wireless.

`

Independent External Peer Review Report

Marine Recreational Information Program (MRIP) Fishing Effort Survey (FES) Calibration

Calibration Model Accounting for a Recreational Survey Design Change

Sheraton Hotel, Silver Spring, MD

June 27-29, 2017

Reviewer Report to NMFS MRIP

Prepared By:

Fredric M. Serchuk

July 2017

Executive Summary

a) This report is an independent peer review of the *Calibration Model Accounting for a Recreational Fishery Survey Design Change* presented at the MRIP Fishing Effort Survey (FES) Calibration Model Review meeting held 27-29 June 2017 at the Sheraton Hotel in Silver Spring, Maryland.

b) About two weeks prior to the review meeting, the Peer Review Panel—comprising six independent reviewers—was provided with the Terms of Reference (ToRs) for the Peer Review, as well as with four pre-review background documents. One of these documents was a working paper entitled *A Small Area Estimation Approach for Reconciling Mode Differences in Two Surveys of Recreational Fishing Effort* (by F. Jay Breidt, Teng Liu, and Jean D. Opsomer, Colorado State University, June 10, 2017). This paper provided a description of the proposed model and statistical approach developed to calibrate the time series of recreational fishing effort estimates derived from the Coastal Telephone Survey (CHTS) during 1982-2016 with the effort estimates derived from the mail-based Fishing Effort Survey (FES) available in 2015 and 2016. A comparison of the CHTS and FES effort estimates from the contemporaneous 2015 and 2016 surveys (which will also continue in 2017) revealed large differences, with the mail survey estimates very much higher (2 to 11-fold) than the telephone estimates.

c) Three presentations were given to the Panel on the first day of the review meeting to provide additional background information on (1) the MRIP fishing effort survey; (2) the importance of calibrated catch for stock assessments; and (3) the importance of calibrated catch for fisheries management. Two other presentations were also given: one of these focused on the general issue of calibrating survey estimates over time, while the second provided an in-depth explanation of the development, design structure, analytical methodologies, estimators, and testing/performance of the proposed fishing effort calibration model (*i.e.,* the Breidt *et al*. model).

d) The second day of the review primarily involved follow-up discussions and dialogue with the calibration modelers to gain a fuller understanding by the Panel of the calibration model, particularly regarding variable selection and model parameterization. Several additional analyses were performed by the modelers and provided to the Panel in response to specific questions and concerns by the reviewers.

e) The calibration model is a statistically valid approach to obtain calibrated estimates of recreational fishing effort during 1982-2016, even though the casual mechanism(s) for the differences between the CHTS and FES effort estimates remain unknown. The model uses standard and highly respected methodologies (e.g., the Fay-Harriot small area estimation procedure) and can be implemented with off-the-shelf software. Although many other modeling approaches could have considered (and indeed a few of these were evaluated by the developers), the Breidt *et al.* model is certainly an appropriate and scientifically credible statistical approach for calibrating CHTS/FES effort data.

f) An additional year of contemporaneous data telephone and mail survey effort data will be available at the end of 2017. It is highly recommended that a series of cross-validation analyses be conducted to evaluate the calibration modeling results based on the first, second, and third years of data to ensure that the modelling framework—and the model parameter estimates and predictions errors—are stable. As but one approach, the current model (based on the 2015 and 2016 surveys), should be used to predict the 2017 FES effort given the actual 2017 CHTS effort estimate (and/or vice-versa) – and then compare this to the actual effort obtained from the FES survey. Because the calibration procedure should work equally well whether converting from CHTS to FES or FES to CHTS, this exercise should be illuminating.

g) It is important to effectively communicate the calibrated effort results and their impacts (as well as to clearly describe the model used in the calibration) to a variety of user and stakeholder groups as the calibrated data will have significant downstream effects on future stock assessments and on various fishery management programs and activities. A variety of pro-active communication approaches should be used to dispel any misconceptions that may currently exist regarding the legitimacy of the calibration and the transition to the FES system.

h) Finally, it is recommended that an updated report/timetable/chart be prepared illustrating current progress in meeting the tasks and timelines identified in the FES Transition Plan. This undertaking should take note of how the recommendations tendered in the current peer review, as well as those in all previous peer reviews of the MRIP Program (including the 2006 and 2016 NAS Reviews), have been addressed.

Background

This document reports on an independent peer review of a calibration model proposed for use in revising statistics produced by surveys of marine recreational fishing effort on the Atlantic coast and in the Gulf of Mexico. This calibration model is considered by the Marine Recreational Information Program (MRIP) to be very important to adjust historical time series of recreational effort and catch estimates to account for biases in past sampling and estimation methods that have become apparent with the development of a new, more statistically sound method. The calibration model is intended to account for past biases in private boat and shore fishing effort estimates that have resulted from the continued use of a random-digit-dial telephone survey design (known as the "Coastal Household Telephone Survey" [CHTS]) that has degraded over time and will be replaced with the implementation of a new mail survey design (the "Fishing Effort Survey", or FES) in 2018. During 2015-2017, a side-by-side benchmarking of the FES against the CHTS has been occurring to facilitate the development and application of a calibration model "to enable adjustment of past estimates that account for biases in historical effort and catch statistics after the second year."

The purpose of MRIP—FEC Calibration Model Review held during 27-29 June 2017 was to provide an independent peer review of a statistical model for calibrating CHTS and FES effort estimates so that a single time series of effort (from 1981 onward) could be used in the future. The statistical model developed by F. Jay Breidt, Teng Liu, and Jean D. Opsomer (all from Colorado State University) was the subject of the Peer Review. The model was described in a working paper entitled *A Small Area Estimation Approach for Reconciling Mode Differences in Two Surveys of Recreational Fishing Effort* provided to the peer reviewers about two weeks before the meeting.

The Review Panel meeting was chaired by Paul Rago (a member of the Mid-Atlantic Fishery Management Council Scientific and Statistical Committee) and the Panel included six other scientists: Robert Hicks, Cynthia Jones, and Ali Arab (all appointed by the Center for Independent Experts [CIE]), and Patrick Sullivan, Fredric Serchuk, and Jason McNamee (selected, respectively, as representatives from the New England and South Atlantic Fishery Management Council Scientific and Statistical Committees, and from the Atlantic States Marine Fisheries Commission.

Four background documents were provided to members of the Review Panel approximately two weeks prior to the meeting. These included the Breidt *et al.* working paper, the MRIP Transition Plan for the Fishing Effort Survey, a MRIP Fishing Effort Survey Transition Progress Report (dated October 28, 2016), and a report by Rob Andrews, J. Michael Brick, and Nancy A. Mathiowetz entitled *Development and Testing of Recreational Fishing Effort Surveys, Testing a Mail Survey Design* Final Report (dated July 31, 2014). Panel members were also given electronic access to a PowerPoint presentation on the Breidt *et al.* calibration model about a week prior to the review meeting.

The reviewer's Statement of Work is provided in Annex 1, the Terms of Reference (ToRs) for the Peer Review in Annex 2, a Bibliography listing Background and Working Papers for the Peer Review (as well as the Presentations and Hyperlinks provided at the Peer Review) is found at Annex 3, attendees at the Peer Review meeting are listed in Annex 4, and the draft Agenda for the Peer Review meeting is provided in Annex 5.

Review Activities

This reviewer independently read all documents provided in preparation of the review, participated actively in the review meeting (and in the Panel closed sessions at the end of each day and on the last day of the meeting), identified key issues and concerns during the review, contributed to the drafting and editing of the summary report (at the closed session held on the last day of the meeting, by email correspondence several days after the meeting, and during a Panel teleconference held on Friday, 7 July), and authored this review report. As well, this reviewer interacted with the Panel Chair (in person and via email) prior to the review seeking clarification of several of the ToRs and discussing several aspects of the Breidt *et al.* working paper.

The Peer Meeting and Peer Review Process

The Peer Review meeting encompassed 2½ days from 9 am, 27 June 2017 to 1:30 pm, 29 June 2017. The meeting opened with welcoming comments by Dave Van Vorhees (NMFS MRIP) who provided background on the Agency's planned transition from the telephone survey approach (CHTS) to obtaining estimates of marine recreational fishing effort to a mail survey (FES) for obtaining such estimates. He stated that a 3-year benchmarking process was underway (2015-2017) in which the two surveys are being conducted contemporaneously to provide the requisite data to facilitate the development and application of a calibration model to generate a single historical series of fishing effort (from 1981 onwards) that would be expressed in FES equivalents. The FES mail survey has greater coverage and higher response rates than the CHTS and is considered to represent a major improvement over the CHTS (see the 2016 review of the MRIP program conducted by the National Academy of Sciences). The FES is also much less susceptible to potential sources of bias than the CHTS. Initial examination of the data from the side-by-side 2015 CHTS and FES surveys indicate that the FES overall response rate was about 5X higher than CHTS, and that the overall FES effort estimate was 4.7X larger than the CHTS estimate. Hence, the FES is thought to be a more much efficient and inclusive survey approach than the CHTS, and is believed to produce more accurate information.

The *MRIP Transition Plan for the Fishing Effort Survey* (May 2015) calls for the development and evaluation of *"one or more calibration models . . . for possible use in correcting past catch statistics. Alternative models should be considered and one should be selected and defended as the most appropriate validated by external peer review."*

The Peer Review Panel was accordingly tasked (see ToR 1 for the Peer Review) *to evaluate the proposed* [Breidt *et al*. calibration] *model for converting historical estimates of private boat and shore fishing effort produced by the CHTS design to estimates that best represent what would have been produced had the new FES design been used prior to 2017.*

Following up on the introductory remarks by Van Vorhees, the Review Panel Chair, Paul Rago, also welcomed participants and meeting attendees (both those who were physically present and those who joined the meeting via a webinar) and requested that everyone introduce themselves. The draft meeting agenda was then reviewed by the Panel Chair and adopted by the Panel without change. The Chair encouraged lively and friendly debate among meeting participants and attendees, and then briefly reviewed the TORs and several administrative details relating to the responsibilities of the Panel members.

The remainder of the first day of the meeting was devoted to five PowerPoint presentations with Panel discussions following each of these. Rob Andrews (NOAA Fisheries, MRIP) provided an overview of the MRIP CHTS and FES surveys. He noted a number of significant shortcoming with the CHTS (e.g., susceptibility to non-sampling errors, including non-coverage of cell-phone only households, declining response rates, and inaccurate reporting of fishing activity) and indicated that the CHTS was inefficient for sampling recreational anglers. He briefly described the development and sampling design of the FES and highlighted that the FES had been tested in 2012 in four states before being implemented in 2015. The DES is much less susceptible to non-sampling error than the CHTS and has resulted in greater coverage, higher response rates, and given sufficient time for anglers to consider their responses before mailing back their questionnaires. The use of license lists to screen and stratify the address-based sampling has significantly increased survey efficiency and helped target the sampling to fishing households.

The next two presentations focused on the implications of calibrated catches in subsequent science and management activities. Rick Methot (NOAA Fisheries Senior Scientist for Stock Assessments) presented information on the importance of calibrated catch for fishery stock assessments noting that changes in catch streams can significantly impact stock assessment results with respect to stock abundance and exploitation rates, and also affect biological reference points. Andy Strelcheck (NOAA Fisheries, Deputy Regional Administrator, Southeast Region) then gave a presentation (Importance of Calibrated Catch for Fisheries Management) on how MRIP data are used by fishery managers (a) in setting quotas and annual catch limit, and in quota/catch monitoring; (b) in setting sector allocations; and (c) in evaluating regulatory policies. He also noted that the MRIP data are used in a variety of biological and economic models and analyses. Any changes to the baseline catches presently used in the above activities (i.e., effected through the MRIP calibrations) will affect many user and stakeholder groups (some more than others) and therefore have significant economic and social impacts. This situation will likely be exacerbated because not all stocks with recreational fisheries will be re-assessed immediately after the calibrated MRIP data become available. Hence, some stocks will be assessed, managed, and monitored using pre-calibration data, while others will use calibrated data. As well, the calibrated data may cause shifts in existing allocations among sectors and user groups. In the years ahead before fully transition to FES, successfully addressing these issues will be a major challenge for fishery managers.

The last two presentations on day 1 of the Peer Review meeting were by Jean Opsomer, Colorado State University (Calibrating Survey Estimates Over Time) and by Jay Breidt, Colorado State University (A Calibration Methodology for CHTS to FES Transition). In his presentation, Jean provided background information on the characteristics of "good" surveys (e.g., sample populations according to a prescribed statistical sampling design; have probability-weighted estimators, and allow for design-based inference; have methodologies that minimize sampling error; and are implemented following formal, documented protocols). Surveys that rely on voluntary participation and self-reported information (such as the CHTS and FES) typically result in non-response rates, and are subject to recall and reporting errors. If these attributes change over time, interpretability and estimator consistency of the survey results can become problematic. This seems to be the case for the CHTS as nonresponse rates have continued to decrease, landline-only telephone samples are no longer representative, coastal-county sampling has known coverage problems, and the CHTS does not take advantage of fishing license databases. So changing to FES makes sense but calibration presents challenges in that any calibration model will have uncheckable assumptions and unquantified uncertainty associated with the extrapolation effect. Moreover, no factor or covariate has yet been identified that can explain the large difference between the effort estimates obtained during booth the 2015 and 2016 CHTS and FES surveys. Nonetheless, the proposed calibration approach developed by Breidt *et al.* "is firmly grounded in established statistical principles and methodologies [and] allows for quantification of design and model uncertainty."

The presentation by Jay Breidt (*A Calibration Methodology for CHTS to FES Transition*) described the methodological approach used in developing and testing the proposed calibration model to allow the construction of a new, consistent time series of recreational fishing effort estimates. The calibration issue was approached statistically by identifying sources of uncertainty, applying best analytical practices, making all assumptions explicit, and evaluating the sensitivity of the model with regard to failure to meet model assumptions.

The data used for the calibration work were the side-by-side CHTS and FES effort estimates obtained during 2015 and 2016 (by state and 2-month period) and the historical times series of CHTS effort estimates of shore and private boat fishing (1982-2016) available by state and 2-month period. The calibration model assumed that both the telephone and mail estimates target a common underlying time series of true effort, but that each survey estimate is affected by both sampling and non-sampling errors. This true effort is described by a classical time series model comprising trend, seasonal and irregular components. Although the sampling error properties (and the design variances) of the CHTS and FES are well known based on the statistical designs of these surveys, the non-sampling errors (called the "Irregular Effect') cannot be isolated from the true effort series. However, because of the side-by-side results from the two surveys, the difference in the non-sampling errors can be estimated and then modeled with covariates to allow extrapolation backward (or forward) in time. The proposed calibration approach combines the two sets of efforts estimates using a well-known mixed model called the Fay-Harriot model. The model was run accounting for temporal dynamics through regression on population size and state-by-2 month period seasonal factors, and also accounting for changing coverage properties in the CHTS due to expanded wireless telephone usage from the 1990s onward (as the CHTS only used landline telephones in sampling the recreational anglers). A desirable attribute of the model is that it can be run using readily available software.

Several novel innovations were incorporated within the model to estimate variances and to ensure analytical consistency. A large number of exploratory analyses (including simulations and sensitivity analyses) were conducted during model development to assess model structure and performance, to select appropriate covariates, and to evaluate alternative hypotheses regarding the distribution of the "Irregular Effect".

Although the Review Panel posed many questions for the modelers about various aspects of the calibration model and its development and performance (which led to a second presentation by Jay Breidt on the second day of the meeting in which all of these issues were addressed), all Panel members were in agreement that the calibration model is a statistically valid and innovative approach to obtain calibrated estimates of recreational fishing effort during 1982-2016, although the casual mechanism(s) for the differences between the CHTS and FES effort estimates remain unknown.

During the morning of Day 2 of the meeting, Jay Breidt (as noted above) gave his follow-up presentation (*Followup on Comments for "A Calibration Methodology for CHTS to FES Transition"*) to the Panel that responded to the various technical concerns and questions raised by panel numbers the previous day. As well, analyses and figures requested by Panel members were provided and explained. A lengthy and wide-ranging discussion ensued on both the model configuration and performance, as well as on a variety of issues related to the CHTS and FES surveys themselves (particularly as related to a lack of external validation of the self-reported data obtained in both surveys and what the "wireless effect" is really aliasing). Given that the 2017 side-by-side surveys results will become available at the end of this year, the Panel recommended that a series of cross-validation exercises be conducted to ascertain whether the model and its predictive performance remain stable after the addition of the third (and final) year of contemporaneous CHTS-FES data.

The afternoon of Day 2—and all of the morning and the early part of the afternoon of Day 3, were spent by the Panel in closed session in crafting portions of the Summary Report and in exchanging views regarding individual responses to the ToRs.

Evaluation of the Terms of Reference

1. **Evaluate the suitability of the proposed model for converting historical estimates of private boat and shore fishing effort produced by the CHTS design to estimates that best represent what would have been produced had the new FES design been used prior to 2017.**

   This TOR—and its subsections (1a, 1b, 1c, 1d, and 1e)—were satisfactorily met. The proposed calibration model developed by Breidt *et al.* is a statistically valid approach to obtain calibrated estimates of recreational fishing effort during 1982-2016, even though the casual mechanism(s) for the differences between the CHTS and FES effort estimates remain unknown. The model uses standard and highly respected methodologies (e.g., the Fay-Harriot small area estimation procedure) and can be implemented with off-the-shelf software. Although many other modeling approaches could have considered (and indeed a few of these were evaluated by the developers), the Breidt *et al.* model is certainly an appropriate and scientifically credible statistical approach for calibrating the CHTS/FES effort data

   a) **Does the proposed model adequately account for differences observed in the estimates produced by the CHTS and FES designs when conducted side-by-side in 2015-2016?**

      The proposed modeling approach uses the effort estimates obtained from the 2015-2016 concurrent CHTS and FES surveys as the foundation for developing and parameterizing the calibration model, and for estimating the difference in the non-sampling errors associated each of the two survey modes so that this difference can be modeled with covariates to allow extrapolation backward in time. The modeling approach preserves the design features of the surveys (among states, 2-month sampling periods, fishing mode [private boat fishing and shore fishing]). The proposed model is an appropriate and scientifically credible statistical approach for calibrating the CHTS/FES effort data series.

   b) **Is the proposed model robust enough to account for potential differences that would have been observed if the two designs had been conducted side-by-side in years prior to 2015 with regards to time trending biases?**

      It is difficult to assess whether the proposed model is robust enough to account for potential differences in trend biases that would have been observed between the CHTS and FES had these surveys been concurrently conducted prior to 2015. There are simply no data available to evaluate this hypothesis. Some insights regarding the robustness of the calibration model may be gleaned from cross-validation exercises comparing model results based on using only the 2015-2016 side-by-side survey data vs the full three years (2015-2017) of side-by-side survey estimates. As well, estimating either one of the 2017 effort estimates based on applying the model crafted using the 2015-2016 data and the other 2017 estimate would be informative regarding model stability.

Lastly, the CHTS did not collect ancillary data on the demography (e.g., age, sex, etc.) of the survey respondents that could inform inferences concerning possible time trending biases.

c) How does the approach used in developing the proposed FES/CHTS calibration model compare in terms of strengths or weaknesses with other potential approaches?

The approach used in developing the proposed model was statistically well-founded and pursued in a systematic and comprehensive manner taking explicit account of the CHTS/FES methodologies, sources of variability and uncertainty, sensitivity of model assumptions, and the explanatory power of various covariates. The Fay-Harriot approach used in the model well is a highly regarded, well-established statistical methodology that easily allows for incorporation of covariates, and leads to empirical best linear unbiased predictors of either CHTS or FES effort. Performance of the model was tested through various simulations. Overall, the proposed calibration approach is an appropriate and scientifically credible statistical approach for calibrating the CHTS/FES effort data. Although no model is perfect—and while other potential modeling approaches could have been more thoroughly pursued (and a few of these approaches actually were considered during the model development phase)—the Breidt et al. calibration approach is aptly suited for modeling and for calibrating the existing time series of recreational effort estimates.

d) Does the proposed calibration model help to explain how different factors would have contributed to changes in differences between CHTS and FES results over time?

Although a number of factors have been identified as contributing to differences between the CHTS and FES estimates in terms of survey error (i.e., the FES survey design is less susceptible to error than the CHTS resulting from nonresponse and non-coverage issues in the CHTS; responses in the FES are likely to be more accurate than in the CHTS because the CHTS required respondents to answer on-the-spot during the phone call rather than having a sufficient time period as in the FES to more thoroughly consider their responses often using the help of memory aids such as datebooks, conversations with family members, etc.; a number of biases have been identified in the CHTS related to (a) underreporting of fishing effort due to a 'gatekeeper effect' (which person in the household actually answered the telephone), (b) non-coverage of wireless-only households whose members are more likely to fish than those in land-line households; and (c) insufficient sample size to detect fishing activity in some sampling strata during low-activity time waves), none of these singularly explains the temporal differences in the CHTS and FES results. The major covariate in the calibration model is population size. Although, the "wireless effect" covariate in the calibration model is statistically significant, it only accounts for a minor component of the difference between the CHTS and FES results.

As noted by Jay Breidt, there is no estimated regression coefficient in the model that is the "smoking gun" accounting for the differences the two survey estimates over time, and hence the causal mechanism(s) resulting in the large disparities in the survey estimates remain elusive.

e) Is it reasonable to conclude that revised 1981-2016 private boat and shore fishing effort estimates based on the application of the proposed FES/CHTS calibration model would be more accurate than the estimates that are currently available? Does evidence provided for this determination include an assessment of model uncertainty?

Because both the CHTS and FES effort estimates are based on self-reported information that has never been externally validated, the accuracy of any of the estimates cannot be ascertained. There are known shortcomings and biases in the CHTS estimates (see comments in subcomponent [d] above) because of design and coverage issues that are not present with the FES estimates. The FES is clearly the superior approach for obtaining estimates of private boat and shore fishing, and calibrating the 1981-2016 effort estimates to FES equivalents is sensible if only the FES approach will be used in the future.

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

The review panel proceedings went smoothly. Operationally, the meeting room had sufficient space for the Panel, presenters, and meeting attendees. The sound and projection systems worked well, as did the webinar link. Four background documents were provided to Panel members two weeks prior to the meeting, and all additional documents and presentation were made available to the Panel during the meeting via a web-site (i.e., Confluence).

The Panel Chair provided each of the reviewers with a proposed meeting Agenda a day prior to the start of the meeting, requesting that any comments and possible changes be provided back to him before the meeting opened. As the proposed Agenda was satisfactory to all the Panel members, no changes to the Agenda were needed.

Discussions during the 2½ day MRIP Calibration Review illuminated various issues related to the results provided in the background documents and the PowerPoint presentations. Many of the concerns involved clarification of the information provided and/or requests for additional data and analyses. In all cases, these requests were satisfactorily fulfilled allowing the Panel to gain fuller insight on:

1. The sampling designs, strengths, and shortcomings of the telephone (CHTS) and mail (FES) survey methods, including their relative performance and sources of error.

2. The development, design, statistical properties, testing, and application of the proposed MRIP FES calibration model. This included consideration of alternative modeling approaches, cross-validation of the modeling framework to examine the stability of model parameter estimate (as well as prediction errors), the sufficiency and explanatory power of the model's covariates, and the possible underlying mechanism(s) affecting the distribution of the "Irregular" random effect, which is not explicitly accounted for within the proposed small-area estimation approach.
3. The potential impacts of the calibrated recreational fishing effort estimates during 1981-2016 on future stock assessments, and on subsequent fishery management policies and practices.
4. The need to effectively communicate the results of the calibration work (as well as the basis and need for continuing only the mail-based survey method in the future) to various constituency groups (i.e., the recreational and commercial fishing communities; scientists; fishery managers; the lay public) so that these groups fully understand and accept the calibration results and their subsequent use in deriving recreational catch estimates for application in stock assessments and in the fishery management process.

The Review Panel acknowledged that proposed MRIP FES calibration model developed by Breidt *et al.* was a well-suited and statistically-appropriate approach to obtain calibrated estimates of recreational fishing effort (by state and 2-month calendar quarter for shore-based and private boat anglers) during 1982-2016.

Although the Peer Review process worked very well and the Panel concluded that all of the TORs for the Review were met, I believe that there are few areas in which the process could have worked even better. These include:

1. It would have been helpful for the Panel to have been provided (several weeks before the review) additional background documents (available from the MRIP Team and/or the MRIP Website) to enhance a collaborative understanding by Panel members of (a) various aspects of the MRIP program and (2) of recent analyses using MRIP data. For example, the *MRIP Data User Handbook,* and the October 2016 report, '*Possible Effects of Calibration Scenarios on Stock Assessments Planned for the MRIP Fishing Effort Survey Transition'* would have especially useful for Panel members to have had and read before the actual peer review occurred.

2. Prior to the presentation and discussion of the Breidt *et al.* report at the Peer Review, this report was difficult to understand for anyone other than a highly trained statistician. Although a more complete understanding of this report was fostered by distribution of a PowerPoint presentation a week or so before the Review Meeting (and subsequently enhanced at the meeting by direct dialogue and interaction with the authors of the paper who clarified and responded to many issues raised by the Panel), it is recommended that in any future reviews in which a highly technical paper is seminal to the crux of such reviews that efforts be made by the paper authors to present the essence of their work in a manner that facilitates full appreciation and understanding of the import of such work by educated non-specialists. This becomes especially critical when the methods/approach provided in a paper will have significant downstream effects. This matter should be recognized in the future APAIS peer review.

3. In its comments on the various subcomponents of TOR 1 (1a, 1b, 1c, 1d, 1e), the Review Panel highlighted a number of issues related to additional work and analyses that might be undertaken to provide additional insight into the performance and robustness of the proposed CHTS/FES calibration model and the efficacy of the effort collection survey methodologies. It is recommended that the MRIP Team (in collaboration where necessary with Breidt *et al.*) develop a protocol to facilitate the timely accomplishment of the highlighted additional work.

4. Finally, it is recommended that an updated report/timetable/chart be prepared illustrating current progress in meeting the tasks and timelines identified in the FES Transition Plan. This undertaking should also take note of how the recommendations tendered in all previous peer reviews of the MRIP Program (including the 2006 and 2016 NAS Reviews) have been addressed.

Annex 1. Statement of Work

Statement of Work

National Oceanic and Atmospheric Administration (NOAA)

National Marine Fisheries Service (NMFS)

External Independent Peer Review

Calibration Model Accounting for a Recreational Fishery Survey Design Change

Background

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

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards.
(http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).

Scope

The Office of Science and Technology requests an independent peer review of a calibration model proposed for use in revising statistics produced by surveys of marine recreational fishing effort on the Atlantic coast and in the Gulf of Mexico. This calibration model is considered by the Marine Recreational Information Program (MRIP) to be very important to adjust historical time series of recreational effort and catch estimates in order to account for biases in past sampling and estimation methods that have become apparent with the development of a new, more statistically sound method. The calibration model is intended to account for past biases in private boat and shore fishing effort estimates that have resulted from the continued use of a legacy random-digit-dial telephone survey design that has degraded over time and will be replaced with the implementation of a new mail survey design (the "Fishing Effort Survey", or FES) in 2018.

Calibration Model for the Fishing Effort Survey

In 2015, MRIP formed a Transition Team to collaboratively plan a transition from a legacy telephone survey design to a new mail survey design for estimating private boat and shore fishing effort by marine recreational anglers. Since 2008, MRIP had conducted six pilot studies to determine the most accurate and efficient survey method for this purpose on the Atlantic and Gulf coasts. The most recent study, conducted in four states in 2012-2013, compared a new mail survey design with the Coastal Household Telephone Survey (CHTS) design that has been used since 1979. MRIP subjected the final report from the pilot project to external peer review in 2014 and certified the new survey design, called the Fishing Effort Survey (FES), in February 2015 as a suitable replacement for the CHTS. The FES is much less susceptible to potential sources of bias than the CHTS because it can reach more anglers, achieve higher response rates, and is less prone to possible recall errors. The pilot project results indicated that FES estimates were substantially higher than CHTS estimates for both private boat fishing and shore fishing.

MRIP recognized the FES should not be implemented immediately as a replacement for the CHTS, and a well thought out transition plan was needed to ensure that the phase-in of the FES is appropriately integrated into ongoing stock assessments and fisheries management actions in a way that minimizes disruptions to these processes, which are based on input from multiple data sources over lengthy time series. The Transition Plan developed by the Transition Team called for side-by-side benchmarking of the FES against the CHTS for three years (2015-2017) with the development and application of a calibration model to enable adjustment of past estimates that account for biases in historical effort and catch statistics after the second year. With this timeline, revised estimates can be incorporated into stock assessments during 2018 using a peer reviewed calibration model, and new Annual Catch Limits (ACLs) can then be set in 2019 for at least some stocks.

Requirements

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

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

Tasks for Reviewers

Pre-review Background Documents

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

Transition Plan for the FES:

https://www.st.nmfs.noaa.gov/Assets/recreational/pdf/MRIP%20FES%20Transition%20Plan%20FINAL.pdf

Report recommending the FES to replace the CHTS: *Finalize Design of Fishing Effort Surveys* (https://www.st.nmfs.noaa.gov/pims/main/public?method=DOWNLOAD_FR_PDF&record_id=1179)

2015 Benchmarking Progress Report:

https://www.st-test.nmfs.noaa.gov/Assets/recreational/pdf/2015_FES_Progress_Report-20161115.pdf

Report on FES/CHTS Calibration Model:

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

Panel Review Meeting

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

Contract Deliverables - Independent CIE Peer Review Reports

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

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

Other Tasks – Contribution to Summary Report

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

Place of Performance

The place of performance shall be at the reviewers' facilities, and at the NMFS Headquarters in Silver Spring, Maryland.

Period of Performance

The period of performance shall be from the time of award through July 31, 2017. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

Travel

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790).

Restricted or Limited Use of Data
The contractors may be required to sign and adhere to a non-disclosure agreement.

NMFS Project Contact:
Dave Van Voorhees
National Marine Fisheries Service
1315 East West Highway
Silver Spring, MD 20910
dave.van.voorhees@noaa.gov

Annex 2. Terms of Reference

Terms of Reference for the Peer Review

Calibration Model Accounting for a Recreational Fishery Survey Design Change

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

1. Evaluate the suitability of the proposed model for converting historical estimates of private boat and shore fishing effort produced by the CHTS design to estimates that best represent what would have been produced had the new FES design been used prior to 2017.

   a) Does the proposed model adequately account for differences observed in the estimates produced by the CHTS and FES designs when conducted side-by-side in 2015-2016?

   b) Is the proposed model robust enough to account for potential differences that would have been observed if the two designs had been conducted side-by-side in years prior to 2015 with regards to time trending biases?

   c) How does the approach used in developing the proposed FES/CHTS calibration model compare in terms of strengths or weaknesses with other potential approaches?

   d) Does the proposed calibration model help to explain how different factors would have contributed to changes in differences between CHTS and FES results over time?

   e) Is it reasonable to conclude that revised 1981-2016 private boat and shore fishing effort estimates based on the application of the proposed FES/CHTS calibration model would be more accurate than the estimates that are currently available? Does evidence provided for this determination include an assessment of model uncertainty?

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

Annex 3. Bibliography of Documents and Presentations

Background Papers

Many papers and documents on the existing and proposed survey methodology may be found at the following website:

http://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/effort-survey-improvements

Background on the MRIP Calibration Model Peer Review may be found at:

https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/index.html

The National Academies of Sciences, Engineering, and Medicine. 2016. Review of the Marine Recreational Information Program (MRIP) Washington, DC: The National Academies Press. doi: 10.17226/24640

https://www.st.nmfs.noaa.gov/confluence/display/FESCALIB?preview=/73074985/73728799/NAS_MRIP_review.pdf

National Research Council. 2006. Review of Recreational Fisheries Survey Methods. Committee on the Review of Recreational Fisheries Survey Methods, ISBN: 0-309-66075-0, 202 pages. http://www.nap.edu/catalog/11616.html

Working Papers

Development and Testing of Recreational Fishing Effort Surveys Testing a Mail Survey Design: Final Report. Project Team Members: Rob Andrews, NOAA Fisheries, J. Michael Brick, Westat, Nancy A. Mathiowetz, University of Wisconsin-Milwaukee. July 31, 2014.

https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES Workshop/documents/Report_recommending_FES_to_replace_CHTS--Finalize_Design_of_Fishing_Effort_Surveys.pdf

Marine Recreational Information Program Fishing Effort Survey Transition Progress Report. October 28, 2016.

https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FESWorkshop/documents/2015_benchmarking_progress_report.pdf

Marine Recreational Information Program. Transition Plan for the Fishing Effort Survey. Prepared by the Atlantic and Gulf Subgroup of the Marine Recreational Information Program Transition Team, May 5, 2015

https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/documents/MRIP_FES_Transition-Plan_FINAL.pdf

A Small Area Estimation Approach for Reconciling Mode Differences in Two Surveys of Recreational Fishing Effort draft: F. Jay Breidt, Teng Liu, Jean D. Opsomer. Colorado State University June 10, 2017

https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/documents/DRAFT-Report_of_Calibration_Model.pdf

Presentations

Calibration_Scenarios-20161115.pdf

MRIP FES website link

FESCALIBRATIONNOTESDay2.docx

EBLUPS.csv

EBLUPS_Variable_Names.csv

FESCALIBRATIONNOTESDay1.docx

Eblup comparisons.docx

MRFSS Fish Hunt Comps.xlsx

FES Errors.pptx

Model_Fits.txt

Mode_3_logeffort_poly_fixed.pdf

Mode_7_logeffort_poly_fixed.pdf

Webinar Links

All open sections of the meeting were recorded and available for viewing at the following links.

0 - Intro - Paul Rago

1 - MRIP Fishing Effort Survey - Rob Andrews

2- Catch and Assessments - Rick Methot

3 - Management Implications - Andy Strelcheck

4 - Calibrating Survey Estimates over Time - Jean Opsomer

5 - Calibration from CHTS to FES - Jay Breidt

6 - Initial Calibration Review Discussion - Tuesday Afternoon

7 - Day Two, AM Discussion

8 - Day Two, PM Discussion

9 - Day Two, Initial Findings Summary

Annex 4. Attendees at the Peer Review Meeting

MRIP Calibration Model Peer Review Workshop
Sheraton Silver Spring Hotel
Silver Spring, MD
June 27-29, 2017

ATTENDANCE LIST

#	NAME	AFFILIATION
1	Paul Rago	MAFMC SSC
2	Dave Van Voorhees	NOAA Fisheries
3	John Foster	NOAA Fisheries
4	Ali Arab	Georgetown University
5	Rob Hicks	College of William and Mary
6	Cynthia M. Jones	Old Dominion University
7	Richard Cody	NOAA support ECS
8	Teng Liu	Colorado State University
9	Thomas Sminkey	NOAA Fisheries/ST1
10	Steve Turner	NOAA Fisheries SEFSC
11	Andy Strelcheck	NOAA Fisheries - SERO
12	Richard Methot	NOAA Fisheries - HQ
13	Karen Pianka	NOAA Fisheries – ST1
14	Lauren Dolinger Few	NMFS ST1
15	Chris Wright	NMFS - SF
16	Sabrina Lovell	NMFS ST
17	Patrick Lynch	NMFS ST
18	Melissa Karp	NMFS ST
19	Toni Kerns	ASMFC
20	Steve Ander	Gallup
21	Tommy Tran	Gallup
22	Melissa Niles	Fifth Estate/MRIP CET
23	Yong-Woo Lee	NOAA - Fisheries
24	Jay Breidt	Colorado State University
25	Jean Opsomer	Colorado State University
26	Rob Andrews	NOAA Fisheries
27	Ryan Kitts-Jensen	NOAA Fisheries
28	Fred Serchuk	SAFMC SSC
29	Jason McNamee	ASMFC
30	Patrick Sullivan	Cornell/NEFMC
31	Jason Didden	MAFMC
32	Daemian Schreiber	NMFS HQ
33	Laura Diederick	NOAA Fisheries

Annex 5. MRIP Calibration Model Peer Review Draft Meeting Agenda

Day	*Date*	*Time*	*Topic*	*Rapporteur*	*Presenter*
Tuesday	27-Jun	9:00 AM	Welcome and Opening Remarks	TBD	Van Voorhees
		9:20 AM	Introductions		
		9:30 AM	Overview of Meeting	TBD	Rago
		9:45 AM	MRIP Fishing Effort Survey	TBD	Andrews
		10:15 AM	Importance of Calibrated Catch for Stock Assessments	TBD	Methot
		10:45 AM	Break		
		11:00 AM	Importance of Calibrated Catch for Fisheries Management	TBD	Strelcheck
		11:30 AM	Calibrating Survey Estimates over Time	TBD	Opsomer
		12:00 PM	Lunch		
		1:30 PM	A Calibration Methodology for CHTS to FES Transition	TBD	Breidt
		3:30 PM	Break		
		3:45 PM	Public Comment	TBD	
		4:15 PM	Summary of Day 1	TBD	Rago
		4:45 PM	Review Panel Coordination and Writing (closed)		
		6:00 PM	Adjourn		
Wednesday	28-Jun	9:00 AM	Overview of Day 1 and Preview of Day 2	TBD	Rago
		9:10 AM	Follow-up Questions for Presenters	TBD	Various
		10:30 AM	Break		
		10:45 AM	Follow-up Questions for Presenters (cont.)	TBD	Various
		12:00 PM	Lunch		
		1:00 PM	Review Panel Coordination and Writing (closed)		
		2:30 PM	Initial Summary Findings of Review Panel (open)	TBD	Panel
		3:30 PM	Public Comment	TBD	
		4:00 PM	Review Panel Coordination and Writing (closed)		
		6:00 PM	Adjourn		
Thursday	29-Jun	9:00 AM	Review Panel Coordination and Writing (closed)		
		12:30 PM	Adjourn		

Closed sessions allow the panel to discuss and clarify technical issues, and begin initial writing of reports.
Attendance of public, staff and presenters, if at all, is by invitation only and for purposes of clarification.

Marine Recreational Information Program (MRIP)
Fishing Effort Survey (FES) Calibration Review

Calibration Model Review Meeting
June 27-29, 2017
Sheraton Hotel
Silver Spring, MD

By
Patrick J. Sullivan
Cornell University
NEFMC SSC

July 18, 2017

Contents

Executive Summary

A sophisticated statistical model for providing temporally consistent estimates of fishing effort, based on data gathered from two different survey sampling modes (CHTS telephone survey vs. FES mail survey), was presented to the MRIP Calibration Review Panel during a meeting that took place in Silver Spring, MD, on June 27-29, 2017. The proposed statistical model does not estimate a single calibration factor, in that it does not provide a single constant multiplier that can be applied to an entire time series in order to put everything into the same units. Instead, the method defines a statistical relationship between the two survey modes and predicts fishing effort based on the type of survey information available (taken from one mode, the other, or both) while including other factors such as the state and seasonal wave in which the survey took place, population size and the degree of cell phone coverage. The model proposed by Breidt, Liu and Opsomer (2017) is an elegant and state-of-the-art statistical procedure that appears to me to be a valid method for providing a consistent time series of fishing effort estimates. However, explaining how the model works to scientists, managers and stakeholders will prove challenging. Furthermore, the sizable differences in fishing effort estimated under the two survey sampling modes indicates to me that a good introduction and explanation of the overall statistical application will be sought after. The proposed model does not itself identify which fishing effort estimates, those derived by telephone or those derived by mail, are more representative of actual fishing effort. However, the model can be used to derive fishing effort estimates in the context of either the telephone survey or the mail survey. Previous reviews confirm what was foreseen by the 2006 NAS review, namely that, with a better sampling frame, greater coverage and more up-to-date statistical methods, a statistical procedure such as the current mail survey method would result in an estimator with greater precision. But, it must be pointed out that one cannot necessarily draw the conclusion from this alone that the FES mode of estimation is the more accurate of the two (precision represented by the variance is different than accuracy as represented by lack of bias). The time period during which both survey methods were simultaneously applied is short (3 years), which is not much time for identifying all the factors critical to understanding this system given that so many of the components are changing. The move towards implementing the new fishing effort calculations would benefit greatly from further analysis into the causes of the differences between fishing effort estimates from the two survey modes. It was indicated at the review meeting that some data exploration had been done to examine this issue, but no single factor could conclusively be said to be the cause of the difference. The Testing Report by Andrews et al. (2014) would seem to indicate that the FES method is both more precise (more efficient statistically) and more accurate. I would encourage the MRIP team to develop additional inroads to resolving this concern about causes by examining further how the different components (e.g. coverage, population demographic differences, cell-phone response rates) incrementally contribute to the differences in estimates and how this affects the quality of the estimates. Elucidating more fully and clearly the reasons for the differences will aid in the acceptance of the new survey mode and effort estimation methods as well as provide insight on how best to interpret and use the data at hand.

Background

The Review Panel for the MRIP-FES Calibration Model Review met from June 27 to June 29 to review a statistical model developed by F. Jay Breidt, Teng Liu and Jean D. Opsomer, of Colorado State University. The review committee was composed of three scientists appointed by the Center for Independent Experts (CIE): Robert Hicks, The College of William and Mary, Cynthia Jones, Old Dominion University and Ali Arab, Georgetown University. In addition, representatives from the New England (Patrick Sullivan) and South Atlantic (Fredric Serchuk) Scientific and Statistical Committees, and the Atlantic States Marine Fisheries Commission (Jason McNamee) served on the review panel. The meeting was chaired by Paul Rago as a member of the Mid-Atlantic Fishery Management Council Scientific and Statistical Committee.

The panel reviewed supporting documentation and presentations prepared by MRIP staff, led by Dave Van Voorhees, and their contractors from the Department of Statistics at Colorado State University. John Foster, Ryan Kitts-Jensen, and Richard Cody of MRIP acted as rapporteurs. Other staff from the Office of the Science and Technology, notably Karen Pianka, assisted in the handling of documents via a web-based application. Jason Didden of the Mid Atlantic Fishery Management Council provided support for the webinar. Approximately 35 people participated in the open sessions of the meeting.

Terms of Reference for the Peer Review

1. Evaluate the suitability of the proposed model for converting historical estimates of private boat and shore fishing effort produced by the CHTS design to estimates that best represent what would have been produced had the new FES design been used prior to 2017.
    a. Does the proposed model adequately account for differences observed in the estimates produced by the CHTS and FES designs when conducted side-by-side in 2015-2016?
    The model can be used to characterize private boat and shore fishing effort either entirely in terms of CHTS or entirely in terms of FES. The Terms of Reference question about "accounting for differences" is difficult to address. The method does not provide a global calibration factor that can easily be applied as a multiplier, but instead uses a model to predict fishing effort from the two modes of survey estimates while incorporating other auxiliary information. The model itself cannot provide an explanation for the difference, nor should it be expected to. And, because auxiliary information beyond the information contained in side-by-side estimates is being used, side-by-side estimates cannot be compared directly in any kind of global sense using this model as currently constructed. Still, some simpler statistical analyses that compare "side-by-side" estimates on a pairwise basis have been done outside of this modeling context and might be used to facilitate greater understanding and interpretation of the data outside of and in conjunction with the model. We were not provided any side-by-side comparative statistical analyses for this review.

b. Is the proposed model robust enough to account for potential differences that would have been observed if the two designs had been conducted side-by-side in years prior to 2015 with regards to time trending biases?

In theory, yes, provided the assumptions of the model hold over the entire time series. Unfortunately, we have not observed the behavior of the system throughout its operation historically and so may be missing some important components that would better capture and explain biases. Further work should be done in this area. Possible directions would be independent validation of effort metrics as well as gathering historic information where available (e.g. demographic changes, population attitudes towards fishing as a leisure activity, historical coverage) that might shed greater light on calibration differences.

c. How does the approach used in developing the proposed FES/CHTS calibration model compared in terms of strengths or weaknesses with other potential approaches?

Because the MRIP team and collaborators were not provided with the Terms of Reference beforehand the Panel had to inquire about what other approaches were explored during the meeting. Methods such as Bayesian hierarchical modeling, state-space modeling, time-varying ratio estimation and expanded versions of the proposed Fay-Herriot approach were all raised for consideration by the Panel, but the CSU contractors indicated that these and other approaches were explored with the research focus converging to the current version of the model. Had the CSU scientists known of the Terms of Reference they might have been able to provide a more comprehensive report on what models they had explored and why the current one was selected. That said, the model reviewed, in its current form, is a reasonable means for estimating fishing effort over the time series where the survey modes have changed.

d. Does the proposed calibration model help to explain how different factors would have contributed to changes in differences between CHTS and FES results over time?

The fishing effort estimation model accounts for differences by state and wave, population change, and degree of cell-phone coverage. While it also accounts for differences due to survey mode, it cannot be used to explain these differences. It is recommended that further research be put into quantifying the cumulative influence various factors contribute to current and past differences.

e. Is it reasonable to conclude that revised 1981-2016 private boat and shore fishing effort estimates based on the application of the proposed FES/CHTS calibration model would be more accurate than the estimates that are currently available? Does evidence provided for this determination include an assessment of model uncertainty?

Here I repeat what was stated in the Panel Summary report as that succinctly characterizes the issue of accuracy as raise in this Terms of Reference, which is really outside the scope of this review as structured by the information provided to the reviewers and the statistical methods available for review.

- No conclusions can be reached regarding the accuracy of calibrating self-reported data from one survey mode to the other. However, the Panel noted that bias in the historical CHTS may not be as large as observed in contemporary CHTS samples due degradation of survey coverage and other factors.

- Gatekeeper, recall bias, response rate etc. indicate that the mail survey is preferred to a phone, particularly in relation to statistical and operational efficiency. This conclusion was supported by the 2006 and 2017 NRC reports, and also in a separate review conducted by the ASA.

- Response rate per se is not a problem unless differences in fishing activity differ between respondents and non-respondents

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

One challenging problem that became apparent during the meeting, was that the presenters did not have the Terms of Reference prior to their preparation for the meeting. The Panel had to spend extra time with the presenters in order to get the information needed to achieve the Terms of Reference.

Several of the presentations did not provide enough informative depth relevant to their particular topic. It would have helped with the review to have had that knowledge. Greater coordination and communication between collaborators on this project would have benefited the quality of the information coming into the review, but would also have aided the MRIP overall.

I greatly appreciated the web space provided for the documents and that the documents as well as data were posted shortly after being requested. The staff support for this was great.

The documentation initially provided prior to the meeting was rather sparse, but the availability of the documents improved as the meeting progressed. It would have been beneficial, had it been possible to obtain records like the NAS reports and the MRIP user handbook prior to the meeting. Likewise, reports on model selection, model development and the auxiliary statistical analyses conducted outside the context of the model to enumerate and assess causal factors would have been good to have had available in advance, but certainly the overall process of implementing MRIP itself would benefit still from having such documents available.

The Terms of Reference presumed that converting CHTS to FES is the appropriate direction to go. Yet, the statistical work under review primarily focused on the mathematical aspects of the calibration and not on which set of estimates reflected a truer representation of fishing effort. Not recognizing this assumption in the preparation for this meeting created major challenges for the review and in addressing the Terms of Reference.

More information could have been provided on stock assessment modeling responses to data updates for this review. This could have been used to highlight which assumptions of the model were likely to have the greatest downstream influence on products such as population estimates and allocation.

I appreciated that we spent an hour or more on the second day going through the details of the statistical calibration/prediction model. The model and assumptions were well thought out, but the committee needed to better understand model inputs, parameter definitions and nuances of the Fay-Herriot estimator. Given the terms of reference, we needed to solicit more information on model development and model selection than was initially available at the meeting. Furthermore it appears that separate from the model several independent data analyses exist. It would have been good to have had a presentation and some discussion on those. This would also have been relevant to addressing the Terms of Reference. I welcomed MRIP Review Panel Chair Paul Rago's workup of the pairwise calibration data. Something like that should have been provided with an associated report prior to the meeting presumably by someone from the Fisheries Statistics staff. We received model parameter estimates upon request, however, we did not have time at the meeting to explore them fully. Now that I have time to look at them, I am not sure the entire set of estimates is provided in the output. Making the model code and estimates available will assist with future interpretation and potential acceptance of the estimation method.

In general, I thought the meeting was well organized, and run by Chair Paul Rago, as well as all the staff named in the Panel Summary Report, but for some reason pre-meeting preparation was poorly executed in terms of thoughtful assembly of all the pieces needed to address the Terms of Reference. Some overall coordination among presenters would have helped as well to have made sure that all the relevant information was covered. But what is even more disconcerting is that it appears that the different subgroups, i.e. data gatherers, CSU statistics folks, and end users such as modelers and managers, have not had much opportunity to communicate with each other. At least I saw very little evidence of this despite hearing all about the transition considerations. This, I find, worrisome. In the end, MRIP will be more than the sum of its parts. I'm convinced here, as when I led the earlier MRFSS review (NAS 2006), that the synthesis and communication of information must make or break the implementation of the program.

Appendix 1: Bibliography

Background Papers

Many papers and documents on the existing and proposed survey methodology may be found at the following website: http://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/effort-survey-improvements

Background on the MRIP Calibration Model Peer Review may be found at: https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/index.html

The National Academies of Sciences, Engineering, and Medicine. 2016. Review of the Marine Recreational Information Program (MRIP) Washington, DC: The National Academies Press. doi: 10.17226/24640 https://www.st.nmfs.noaa.gov/confluence/display/FESCALIB?preview=/73074985/73728799/NAS_MRIP_review.pdf

National Research Council. 2006. Review of Recreational Fisheries Survey Methods. Committee on the Review of Recreational Fisheries Survey Methods, ISBN: 0-309-66075-0, 202 pages. http://www.nap.edu/catalog/11616.html

Working Papers

Development and Testing of Recreational Fishing Effort Surveys Testing a Mail Survey Design: Final Report. Project Team Members: Rob Andrews, NOAA Fisheries, J. Michael Brick, Westat, Nancy A. Mathiowetz, University of Wisconsin-Milwaukee. July 31, 2014. https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/documents/Report_recommending_FES_to_replace_CHTS--Finalize_Design_of_Fishing_Effort_Surveys.pdf

Marine Recreational Information Program Fishing Effort Survey Transition Progress Report. October 28, 2016. https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/documents/2015_benchmarking_progress_report.pdf

Marine Recreational Information Program Transition Plan for the Fishing Effort Survey Prepared by the Atlantic and Gulf Subgroup of the Marine Recreational Information Program Transition Team May 5, 2015 https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/documents/MRIP_FES_Transition-Plan_FINAL.pdf

A Small Area Estimation Approach for Reconciling Mode Differences in Two Surveys of Recreational Fishing Effort draft: F. Jay Breidt, Teng Liu, Jean D. Opsomer Colorado State University June 10, 2017 https://www.st.nmfs.noaa.gov/recreational-fisheries/MRIP/FES-Workshop/documents/DRAFT-Report_of_Calibration_Model.pdf

Presentations

Calibration_Scenarios-20161115.pdf
MRIP FES website link
FESCALIBRATIONNOTESDay2.docx
EBLUPS.csv
EBLUPS_Variable_Names.csv
FESCALIBRATIONNOTESDay1.docx
Eblup comparisons.docx
MRFSS Fish Hunt Comps.xlsx
FES Errors.pptx
Model_Fits.txt
Mode_3_logeffort_poly_fixed.pdf
Mode_7_logeffort_poly_fixed.pdf

Webinar Links

All open sections of the meeting were recorded and available for viewing at the following links.

0 - Intro - Paul Rago
1 - MRIP Fishing Effort Survey - Rob Andrews
2- Catch and Assessments - Rick Methot
3 - Management Implications - Andy Strelcheck
4 - Calibrating Survey Estimates over Time - Jean Opsomer
5 - Calibration from CHTS to FES - Jay Breidt
6 - Initial Calibration Review Discussion - Tuesday Afternoon
7 - Day Two, AM Discussion
8 - Day Two, PM Discussion
9 - Day Two, Initial Findings Summary

Appendix 2: Statement of Work

Statement of Work
National Oceanic and Atmospheric Administration (NOAA)
National Marine Fisheries Service (NMFS)
External Independent Peer Review

Calibration Model Accounting for a Recreational Fishery Survey Design Change

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

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards.
(http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).

Scope
The Office of Science and Technology requests an independent peer review of a calibration model proposed for use in revising statistics produced by surveys of marine recreational fishing effort on the Atlantic coast and in the Gulf of Mexico. This calibration model is considered by the Marine Recreational Information Program (MRIP) to be very important to adjust historical time series of recreational effort and catch estimates in order to account for biases in past sampling and estimation methods that have become apparent with the development of a new, more statistically sound method. The calibration model is intended to account for past biases in private boat and shore fishing effort estimates that have resulted from the continued use of a legacy random-digit-dial telephone survey design that has degraded over time and will be replaced with the implementation of a new mail survey design (the "Fishing Effort Survey", or FES) in 2018.

Calibration Model for the Fishing Effort Survey
In 2015, MRIP formed a Transition Team to collaboratively plan a transition from a legacy telephone survey design to a new mail survey design for estimating private boat and shore fishing effort by marine recreational anglers. Since 2008, MRIP had conducted six pilot studies to determine the most accurate and efficient survey method for this purpose on the Atlantic and Gulf coasts. The most recent study, conducted in four states in 2012-2013, compared a new mail survey design with the Coastal Household Telephone Survey (CHTS) design that has been used since 1979. MRIP subjected the final report from the pilot project to external peer review in 2014 and certified the new survey design, called the Fishing Effort Survey (FES), in February 2015 as a suitable replacement for the CHTS. The FES is much less susceptible to potential sources of bias than the CHTS because it can reach more anglers, achieve higher response rates, and is less prone to possible recall errors. The pilot project results indicated that FES estimates were substantially higher than CHTS estimates for both private boat fishing and shore fishing.

MRIP recognized the FES should not be implemented immediately as a replacement for the CHTS, and a well thought out transition plan was needed to ensure that the phase-in of the FES is appropriately integrated into ongoing stock assessments and fisheries management actions in a way that minimizes disruptions to these processes, which are based on input from multiple data sources over lengthy time series. The Transition Plan developed by the Transition Team called for side-by-side benchmarking of the FES against the CHTS for three years (2015-2017) with the development and application of a calibration model to enable adjustment of past estimates that account for biases in historical effort and catch statistics after the second year. With this timeline, revised estimates can be incorporated into stock assessments during 2018 using a peer reviewed calibration model, and new Annual Catch Limits (ACLs) can then be set in 2019 for at least some stocks.

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

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

Tasks for Reviewers

Pre-review Background Documents

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

Transition Plan for the FES:
https://www.st.nmfs.noaa.gov/Assets/recreational/pdf/MRIP%20FES%20Transition%20Plan%20FINAL.pdf

Report recommending the FES to replace the CHTS: *Finalize Design of Fishing Effort Surveys* (https://www.st.nmfs.noaa.gov/pims/main/public?method=DOWNLOAD_FR_PDF&record_id=1179)

2015 Benchmarking Progress Report:
https://www.st-test.nmfs.noaa.gov/Assets/recreational/pdf/2015_FES_Progress_Report-20161115.pdf

Report on FES/CHTS Calibration Model:
This report will be provided by ECS (via electronic mail or make available at an FTP site) to the reviewers.

Panel Review Meeting

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

Contract Deliverables - Independent CIE Peer Review Reports

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

Other Tasks – Contribution to Summary Report

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

Place of Performance
The place of performance shall be at the reviewers' facilities, and at the NMFS Headquarters in Silver Spring, Maryland.

Period of Performance
The period of performance shall be from the time of award through July 31, 2017. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

Travel
All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790).

Restricted or Limited Use of Data
The contractors may be required to sign and adhere to a non-disclosure agreement.

NMFS Project Contact:
Dave Van Voorhees
National Marine Fisheries Service
1315 East West Highway
Silver Spring, MD 20910
dave.van.voorhees@noaa.gov

Annex I: Format and Contents of Independent Peer Review Report

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

2. The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each ToR, in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the ToRs.

   a. Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.

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

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

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

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

3. The report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of this Statement of Work
   Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

Calibration Model Accounting for a Recreational Fishery Survey Design Change

1. Evaluate the suitability of the proposed model for converting historical estimates of private boat and shore fishing effort produced by the CHTS design to estimates that best represent what would have been produced had the new FES design been used prior to 2017.

   a) Does the proposed model adequately account for differences observed in the estimates produced by the CHTS and FES designs when conducted side-by-side in 2015-2016?

   b) Is the proposed model robust enough to account for potential differences that would have been observed if the two designs had been conducted side-by-side in years prior to 2015 with regards to time trending biases?

   c) How does the approach used in developing the proposed FES/CHTS calibration model compare in terms of strengths or weaknesses with other potential approaches?

   d) Does the proposed calibration model help to explain how different factors would have contributed to changes in differences between CHTS and FES results over time?

   e) Is it reasonable to conclude that revised 1981-2016 private boat and shore fishing effort estimates based on the application of the proposed FES/CHTS calibration model would be more accurate than the estimates that are currently available? Does evidence provided for this determination include an assessment of model uncertainty?

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

Appendix 3: Calibration Model Review Attendees List

MRIP Calibration Model Peer Review Workshop
Sheraton Silver Spring Hotel
Silver Spring, MD
June 27-29, 2017

ATTENDANCE LIST

#	NAME	AFFILIATION
1	Paul Rago	MAFMC SSC
2	Dave Van Voorhees	NOAA Fisheries
3	John Foster	NOAA Fisheries
4	Ali Arab	Georgetown University
5	Rob Hicks	College of William and Mary
6	Cynthia M. Jones	Old Dominion University
7	Richard Cody	NOAA support ECS
8	Teng Liu	Colorado State University
9	Thomas Sminkey	NOAA Fisheries/ST1
10	Steve Turner	NOAA Fisheries SEFSC
11	Andy Strelcheck	NOAA Fisheries - SERO
12	Richard Methot	NOAA Fisheries - HQ
13	Karen Pianka	NOAA Fisheries – ST1
14	Lauren Dolinger Few	NMFS ST1
15	Chris Wright	NMFS - SF
16	Sabrina Lovell	NMFS ST
17	Patrick Lynch	NMFS ST
18	Melissa Karp	NMFS ST
19	Toni Kerns	ASMFC
20	Steve Ander	Gallup
21	Tommy Tran	Gallup
22	Melissa Niles	Fifth Estate/MRIP CET
23	Yong-Woo Lee	NOAA - Fisheries
24	Jay Breidt	Colorado State University
25	Jean Opsomer	Colorado State University
26	Rob Andrews	NOAA Fisheries
27	Ryan Kitts-Jensen	NOAA Fisheries
28	Fred Serchuk	SAFMC SSC
29	Jason McNamee	ASMFC
30	Patrick Sullivan	Cornell/NEFMC
31	Jason Didden	MAFMC
32	Daemian Schreiber	NMFS HQ
33	Laura Diederick	NOAA Fisheries

Preliminary Response to Recommendations Provided by Peer Reviewers of the FES/CHTS Calibration Model Proposed by MRIP

Recommendations for the Calibration Model

The Marine Recreational Information Program (MRIP) Team developed a protocol for additional work and analyses aimed at evaluating the performance and robustness of the peer reviewed Coastal Household Telephone Survey (CHTS)/Fishing Effort Survey (FES) calibration model when the third year of benchmarking data became available in mid-April of 2018. This protocol was vetted by the MRIP Transition Team's Atlantic and Gulf Subgroup (Transition Team Subgroup) to ensure open communication with all partners. The protocol includes the following:

1. The MRIP Team will re-evaluate the possible effects of different covariates upon inclusion of the third year of side-by-side FES and CHTS data into the calibration model. In addition, the Team will look at the possible significant effects of additional covariates and make sure to consider those suggested by the reviewers.
2. Upon inclusion of the third year of benchmarking data, the MRIP Team will conduct further analyses to evaluate the performance of the calibration model and the relative stability of its statistical outputs. These analyses will be based on model development with and without the third year of data.
3. The MRIP Team will revisit the potential suitability of alternative modeling approaches upon inclusion of the third year of benchmarking data and will document the advantages and disadvantages of considered alternatives relative to the preferred approach in the final report describing the calibration model.

One reviewer recommended extending the benchmarking period for the FES and the CHTS beyond three years. The MRIP Team understands the potential benefits of extending the benchmarking period, but NOAA Fisheries decided not to continue the CHTS beyond 2017. We did not feel we could justify continuing to fund and conduct the CHTS as a survey of fishing effort given its apparent reporting errors and its continuously declining coverage and response rates.

Recommendations for the Calibration Model Report

The MRIP Team will revise the report on the calibration model after inclusion of the third year of benchmarking data and the planned conduct of further analyses to evaluate its performance. At that time, more information will be provided on vetting alternative modeling approaches, the details of estimated results, and the effects of potential explanatory covariates. The final report on the model will be completed and available to the public in July 2018.

Recommendations for Communications

MRIP understands the importance of developing appropriate proactive communication approaches to explain the rationale for transitioning from the CHTS to the new FES, developing a calibration model for converting past CHTS estimates into FES equivalents, and using the calibrated effort and catch statistics in future stock assessments and fisheries management. MRIP also recognizes it will be important to share what we have learned from our ongoing research about the possible causes of the large differences between FES and CHTS estimates of private boat and shore fishing effort, as well as why we have more confidence in the FES estimates. The MRIP Communications and Education Team (CET) developed a strategic communications plan aimed at a wide variety of audiences with different levels of statistical expertise. The CET has been vetting that strategy with the Subgroup and working collaboratively with them to effectively execute it.

Through engagement and two-way dialogue, the MRIP Team and members of the Transition Team Subgroup have been educating and informing internal and external partners on the transition process through updates presented at council and interstate commission meetings, as well as other fishery management and scientific forums. The Team has also been providing information through the MRIP website and NOAA Fisheries newscasts. These efforts will continue. Also through engagement and two-way dialogue, the MRIP Team will educate and inform stakeholders, including Congress, anglers, and eNGOs to secure support of the FES and its effects on fisheries management.

Recommendations for Future Peer Reviews

The MRIP Team incorporated many of the reviewers' recommendations for improving future peer reviews in its planning for the March 2018 workshop to peer review the proposed Access Point Angler Intercept Survey design-change calibration model. In particular, The Team took the following actions:

1. We shared the Terms of Reference (ToR) collaboratively developed by the members of the MRIP Team and Transition Team Subgroup with all presenters and peer reviewers at least one month prior to the planned workshop.
2. We asked the authors of the report on the proposed calibration model to specifically address the ToR in their report.
3. We asked all presenters who provided background information and/or potential impacts of the planned calibration to address the ToR in their workshop presentations.
4. Prior to the workshop, we convened a meeting of the collaborators involved in the development of the calibration model, the authors of the calibration model report, and all of the invited presenters to coordinate how they would address the ToR at the workshop.
5. We provided the reviewers with access to all pertinent background material three weeks prior to the workshop. Pertinent materials included reports on APAIS pilot studies, the new weighted estimation method for the APAIS, and the new sampling design. In addition, we provided access to all previous peer reviews of the new APAIS methods, including what was provided in the 2017 National Academies review of MRIP.

6. We asked the authors of the report on the proposed statistical approach to complete it at least two weeks prior to the workshop, so we could provide it to the reviewers at that time. In the report, we asked the authors to explain how the models proposed in the 2014 Southeast Data Assessment and Review (SEDAR)/MRIP workshop were evaluated and provide the rationale for selecting the proposed method as the best to account for any changes in APAIS estimates caused by the change to an improved sampling design.
7. We asked the authors to provide a webinar explaining the proposed approach to the members of the Transition Team Subgroup two weeks prior to the workshop, and we made a recording of that webinar available to the peer reviewers prior to the workshop.
8. We asked the authors of the report on the proposed statistical approach to take into account varying levels of statistical expertise among the reviewers of the report to be sure that their description of the technical approach is easily understood by both statisticians and non-statisticians.

One reviewer recommended approaching future statistical reviews more like a stock assessment review process with reviewers having access to models and data, so they can contribute in a give and take process for understanding the method. The MRIP Team recognizes that this recommended approach would be useful for at least some future statistical reviews but decided not to use this approach for the peer review of the APAIS design-change calibration model in March 2018. This was largely because a collaborative process was used in 2014 to propose and begin evaluation of three alternative approaches for the APAIS calibration in the MRIP/SEDAR calibration workshop. The March peer review assessed MRIP’s final evaluation of those approaches along with its justification for a new preferred method to account for the APAIS design change.

APAIS data calibration methodology report

John Foster
National Marine Fisheries Service

F. Jay Breidt
Colorado State University

Jean D. Opsomer
Colorado State University

March 11, 2018

1 Background

In 2013, new design and estimation procedures were implemented for the Access Point Angler Intercept Survey (APAIS) conducted by the National Marine Fisheries Service (NMFS). The new procedures were introduced as part of on-going efforts to improve the statistical validity and reliability of the recreational marine fisheries estimates produced by NMFS, which followed the recommendations of a National Academies of Sciences panel review (Sullivan et al. 2006). The most important design changes include improved protocols for interview assignments in terms of interview sites and times of day, and changes to the randomization of assignments so that they better covered the target population, again in terms of sites and times of day. Associated with those design changes were changes to the estimation methods, which are now fully weighted to reflect the unequal probability sampling design. APAIS data collected since March (wave 2) 2013 follow the new design and estimation procedures.

APAIS data have been collected since 1981, and NMFS staff clearly recognize the importance of preserving the integrity of the time series of catch estimates despite these design and estimation changes. Because of this, an adjustment procedure was developed to create "pseudo-weights" for APAIS data collected between January 2004 and March 2013. These weights were constructed based on a combination of site pressures and empirical site visit frequencies, and on estimated expected fractions of trips that took place during the time the interviewer was on site relative to the daily total number of trips. Weighting the observed trips on a given site-day assignment by the inverse of this estimated fraction was meant to correct for differential representation of sampled trips within site-days. The fractions of trips were predicted by a small area estimation model fitted to data from the

Coastal Household Telephone Survey (CHTS), see Hernandez-Stumpfhauser et al. (2016). The combination of modeled site-day selection probabilities and within-day probabilities resulted in weights that better reflect the population of trips during the period 2004-2013, with respect to its overall size and distribution across states, waves and modes.

However, an implicit assumption underlying the validity of this approach is that the trips occurring during the time period the interviewer is on site are representative of those that take place during the full day. This is satisfied if either the time on site is randomly selected within the day, or the trip characteristics are not related to the time of day. The first condition was definitely not satisfied, because the large majority of site visits were made at what was considered the busiest time of the day and were also subject to a degree of interviewer discretion. The second condition appears not to be satisfied either, according to analysis reported in a technical report (see MRIP Staff 2014). Hence, there is a need to supplement the weighting procedure that is based solely on fraction of daily trips within selected site-days by a procedure that accounts for differences in trip characteristics between those that were observed during the site visit intervals and those outside of it.

There is also a desire to adjust the time series for the period prior to 2004. For that earlier period, not only are the selection probabilities within site-days unknown as above, but information allowing the construction of site-day visit probabilities is incomplete or missing, with the required design information becoming progressively more limited going back in time. Further complicating matters, the sampling procedures, including site selection and sampling intensity, underwent changes during that period, and documentation for these changes is no longer available. Hence, separate procedures are needed to calibrate estimates prior to 2004.

Correcting time series of survey data following changes in design, data collection and/or estimation methods is a challenging statistical issue. The "gold standard" approach involves conducting side-by-side measurements under the old and new methods, fitting a suitable calibration model relating estimates under both methods, and developing and applying adjustment factors based on the model results. This approach is currently being implemented by NMFS to calibrate the trip estimates obtained under CHTS and its replacement survey, the Fishing Effort Survey (FES). See NMFS Staff (2015) for more details on the CHTS and FES surveys and the transition between them.

While explicit statistical calibration would in principle be attractive for the APAIS time series as well, there are a number of reasons why that is not possible. First and most critically, there is no overlap period between the old and new designs, so that the data needed for fitting a calibration model are not available. Second, unlike CHTS and FES, which primarily involve estimating the total number of trips for a given region and time period, APAIS is used to produce numerous different estimates, covering a wide range of trip characteristics and detailed catch by species, location and type. Hence, even if an overlap period were available, it is not clear that statistical calibration would be feasible, since multiple models would likely be required for the different types of variables.

For these reasons, the proposed APAIS calibration will rely on a weight adjustment approach, which is conceptually similar to the pseudo-weight approach described above for the 2004-2013 data. By adjusting weights rather than modeling the estimates themselves, the data collected prior to 2013 are preserved but their weights are suitably modified so that the distribution of trips better reflects the actual population distribution. By incorporating the calibration adjustments into the survey weights, the historical data can continue to be made available as survey public-use (micro) datasets, greatly facilitating their acceptance among the current data users.

2 Adjustment approach for 2004-2013 data

We first consider the adjustment of the weights for the period 2004-2013 (wave 1). Because there is no overlap period between the old and new designs and the CHTS contains only limited information on trip characteristics, no direct comparison distribution is available on which to calibrate. Instead, calibration will be performed using the trip distribution for the period 2013 (wave 2)-2016 as the target distribution. This is reasonable if the mix of trip characteristics has remained constant over time, at least over the periods being considered. The validity of this assumption cannot directly be assessed, because differences in observed trip characteristics before and after 2013 can be explained by both the design and estimation changes as well as by actual changes in the fishery. However, we will modify the proposed method in situations in which we observed a significant "drift" in important trip characteristics over time, see Section 3 below. For now, assume that it is reasonable to work under the assumption that differences in trip characteristic distributions between the periods 2004-2013 and 2013-2016 are likely primarily due to the design and estimation method changes. Hence, the weight adjustment method will calibrate the weights for trips in 2004-2013 (wave 1) to the weight distribution for 2013 (wave 2)-2016.

The key decision in the proposed method is which trip characteristics to adjust for. Following the analysis results shown in MRIP Staff (2014), the following trip-level variables were identified as both important trip characteristics and ones for which the distribution in the data collected prior to 2013 deviated from those under the new methods:

- state and sub-state region (if applicable)
- year and wave

- mode
- area fished
- coastal/non-coastal household
- for-hire boat frame membership.

The values for each of these variables defines categories of trips. For instance, there are 4 modes (shore, private boat, headboat, charter) and each trip belongs to one of those modes. Taken in combination, the values for these variables define large numbers of trip domains. Under the new design and estimation approach, by summing up the weights of the trips corresponding to a given set of values for these variables, we obtain an estimate of the number of trips of that type.

As an example to explain the adjustment procedure, let $U_{D,2014}$ present the domain of all trips that occurred, say, in a particular substate region in Florida during wave 2 on a private boat by a coastal household in state waters, in 2014. The true total number of such trips that took place is equal to $N_{D,2014}$. It is unknown but it can be estimated based on APAIS intercepted trips under the new design and estimation methods, by $\widehat{N}_{D,2014} = \sum_s w_i I_{\{i \in U_{D,2014}\}} = \sum_{s_{D,2014}} w_i$. This can be repeated for any combination of values of the classification variables. However, while statistically valid, these estimates are likely to be quite variable for some of these domains because they contain only small numbers of observed trips.

Likewise, we can compute estimates for the same domains for years prior to the design change, e.g. 2012: $\widehat{N}_{D,2012} = \sum_{s_{D,2012}} w_i$. This estimate might not be valid, however, because of the recognized shortcomings of the design and estimation methods in effect at that time. If $N_{D,2012}$ were known, we might therefore decide to adjust the weights so that they sum up to $N_{D,2012}$. This is readily accomplished by replacing all w_i for $i \in s_{D,2012}$ by

$$w_i^* = \frac{N_{D,2012}}{\sum_{j \in s_{D,2012}} w_j} w_i. \tag{1}$$

The weights w_i^* in sample domain $s_{D,2012}$ now sum to the new control total $N_{D,2012}$, and can be applied to any variable y_i collected in the survey. This type of calibration to known control totals is commonly applied in surveys, to improve the precision of estimators.

Since we do not know $N_{D,2012}$, implementing this ratio-type adjustment requires that it be replaced by a sample-based quantity. As noted above, we propose to use estimates based on the data collected under the new design since 2013. In order to reduce the variability of the control total estimates and also because individual years are not meaningful targets (i.e. we are not interested in adjusting 2012 weights to match the 2014 totals, but rather, adjust pre-2013 years to post-2013 years), both the control targets and the adjustment ratios are averaged across years. Hence, the unfeasible adjustment in (1) is replaced by

$$w_i^* = \frac{\widehat{N}_{D,\text{new}}}{\widehat{N}_{D,\text{old}}}\, w_i, \tag{2}$$

where $\widehat{N}_{D,\text{new}}$ is the average of $\widehat{N}_{D,2013}$, $\widehat{N}_{D,2014}$, $\widehat{N}_{D,2015}$ and $\widehat{N}_{D,2016}$ (with the first of these omitted if the domain is in wave 1) and $\widehat{N}_{D,\text{old}}$ is the average of $\widehat{N}_{D,2004}, \ldots, \widehat{N}_{D,2012}$ (and $\widehat{N}_{D,2013}$, only if the domain is in wave 1). Unlike the (unfeasible) adjusted weights in (1), the weights w_i^* do not sum to a control total for a particular year. Instead, they correct for the overall under- or over-representation of trips in domain U_D under the old design and estimation methods relative to the new methods implemented since 2013, which is expected to lead to improved estimates for variables of interest that are related to the domain that is being adjusted.

While averaging the adjustment ratios across years as in (2) reduces their variability, the fine definition of the domains (as intersections of numerous control variables) is still expected to lead to unreliable adjustments in many domains. Therefore, the full ratio adjustment in (2) is replaced by a *raking ratio* adjustment, originally proposed in Deming and Stephan (1940) and widely used in survey calibration. The motivation for this procedure is that instead of adjusting at the finest domain level, adjustments are made iteratively on a set of a coarser domains. These coarser domains are determined by a subset of the variables mentioned above. For each of them, it is possible to compute the averages of the annual estimates as described for $\widehat{N}_{D,\text{new}}$ above. We denote the ones we use in our adjustment procedure as follows:

- AF (state, wave, mode and area fished): $\widehat{N}_{D,\text{new,AF}}$
- HS (state, wave, mode and coastal/non-coastal household status): $\widehat{N}_{D,\text{new,HS}}$
- FH (state, wave, mode and for-hire boat frame status): $\widehat{N}_{D,\text{new,FH}}$
- RE (state, wave, mode and substate region): $\widehat{N}_{D,\text{new,RE}}$

While not explicit in this notation, for each of these domains, the averages are for each of the categories of these variables. So for instance, $\widehat{N}_{D,\text{new,AF}}$ are averages of estimates for each state-wave-mode-area fished combination, and so on for the other domain definitions above.

The raking ratio algorithm, also sometimes called iteratively proportional fitting, then proceeds as follows:

1. Initialize: set $t = 0$, set the adjusted weights $w_i^{(t)}$ equal to the initial weights w_i for the period 2004–2013 (wave 1), and compute the $\widehat{N}_{D,\text{new,AF}}$, $\widehat{N}_{D,\text{new,HS}}$, $\widehat{N}_{D,\text{new,FH}}$ and $\widehat{N}_{D,\text{new,RE}}$.

2. Let $\widehat{N}^{(t)}_{D,\text{old},\text{AF}}$ be the averages of the estimated AF domain totals for the period 2004–2012 (include 2013 for wave 1) using weights $w_i^{(t)}$, compute the ratios $R^{(t)}_{\text{AF}} = \widehat{N}_{D,\text{new},\text{AF}}/\widehat{N}^{(t)}_{D,\text{old},\text{AF}}$, and set $w^{(t)}_{i,\text{AF}} = R^{(t)}_{\text{AF}} w_i^{(t)}$.

3. Starting from the weights $w^{(t)}_{i,\text{AF}}$, do the same as in 2 for the HS domains, resulting in ratios $R^{(t)}_{\text{HS}}$ and weights $w^{(t)}_{i,\text{HS}}$.

4. Starting from the weights $w^{(t)}_{i,\text{HS}}$, do the same as in 2 for the FH domains, resulting in ratios $R^{(t)}_{\text{FH}}$ and weights $w^{(t)}_{i,\text{FH}}$.

5. Starting from the weights $w^{(t)}_{i,\text{FH}}$, do the same as in 2 for the RE domains, resulting in ratios $R^{(t)}_{\text{RE}}$ and weights $w^{(t)}_{i,\text{RE}}$.

6. Set $w_i^{(t+1)} = w^{(t)}_{i,\text{RE}}$.

7. Repeat steps 2–6 until convergence, which is evaluated by measuring the change in the ratios $R^{(t)}_{\text{AF}}, R^{(t)}_{\text{HS}}, R^{(t)}_{\text{FH}}, R^{(t)}_{\text{RE}}$ for different t. Set the final adjusted weights w_i^* equal to the iterated weights $w_i^{(t)}$.

This raking ratio procedure ensures that the weights w_i^* are adjusted to match each of the "marginal" raking variables (AF, HS, FH, RE), but not the fine domains defined by the combinations of these raking variables. This prevents adjusting to overly small domains, with associated overfitting and weight instability issues.

3 Modification for temporal changes in fishery characteristics

We now return to the assumption of constant trip characteristics over time. As noted, the raking procedure described in Section 2 is based on the assumption that the estimated trip distribution since 2013 is a reasonable target for the trip distributions prior to 2013. However, if the trip characteristics in the fishery have changed over that time period, observed differences between the pre-2013 and post-2013 periods are likely to be due to a combination of the design-estimation changes and actual fishery changes. Raking as in Section 2 in this situation will result in a weight adjustment that is too large, because it will remove both the design-induced change and the actual fishery change. We therefore implemented a two-step procedure to decrease the risk of over-adjusting the weights, described in this section.

Consider a single set of control domains first, say AF above. Prior to raking, for a given state, mode and area fished, we create a dataset containing the estimated domain totals for each year and wave combination between 2004 and 2013 (wave 1), resulting in a time series of 145 data points. There are multiple such time series, for each combination of state, move and area fished. We perform a simple linear regression of the totals against a time index for each time series, and test whether the slope is significantly different from zero at the 97.5% confidence level. If the null hypothesis cannot be rejected for a given time series, we maintain the raking adjustment described in Section 2 for the AF domains. If the null hypothesis is rejected for a time series, step 2 in the raking algorithm is modified for that particular control domain, so that only the years 2010–2013 (wave 1) are used in the computation of $\widehat{N}^{(t)}_{D,\text{old,AF}}$. Hence, the AF ratio adjustment $R^{(t)}_{\text{AF}} = \widehat{N}_{D,\text{new,AF}}/\widehat{N}^{(t)}_{D,\text{old,AF}}$ is only based on the most recent years instead of the full time period, in those domains for which a significant time trend is detected.

The same testing and modifications are applied to the remaining three control domains (HS, FH, RE). The full adjustment procedure that accounts for temporal trends therefore consists of the linear regression tests followed by shortening of the time period used for computing the ratio adjustments in any of the control domains for which a non-zero slope is detected, following by the raking algorithm.

4 Adjustments for prior periods: 1993-2003

Weight adjustments for data collected prior to 2004 were performed following the computation of the adjusted weights for 2004-2013 (wave 1). The major difficulty for the earlier periods was that unlike for 2004-2012, it is not possible to construct meaningful initial sample weights for the APAIS data. As such, the weight adjustment method described in sections 2 and 3 could not be applied directly and needed to be extended to address effects of the 2013 APAIS design change as well as any effects associated with initial weighting of the 2004-2012 APAIS data. Using 1 as the initial base weight for intercepted angler-trips was not adequate as the sample sizes, in terms of sampled site-days, were known to vary considerably over time. Unfortunately, the exact sample sizes were unavailable for these earlier years.

It was decided to divide this period in two pieces overall. This provides a hedge against incorrect time trend adjustments masking actual changes in the fishery, as well as unaccounted-for changes in design. This is similar to the argument in Section 3, but was applied globally prior to any further adjustments. Hence, we performed the adjustments for 1993-2003 and 1981-1992 separately.

Considering first 1993-2003, we investigated two approaches for creating initial weights. In a first approach, these weights were calculated by using the MRFSS effort estimates as counts of angler-trips and dividing this by the number of intercepted angler-trips. This calculation was performed in cells defined by state, year, wave, mode, area fished and sub-state region. However, while these initial weights account for the overall magnitude of the fishing effort in a cell, they completely miss relative changes in the number and distribution of site-day assignments that occurred during this period. This lead to stability issues in the development of final weights.

Hence, a second approach was developed using counts of site-days with intercepts to account for changes in site-day assignments. For this approach, counts of site-days with intercepts were tallied in cells defined by state, year, wave and mode. While the exact sampling design was unknown, these counts are a useful proxy for it, in the sense that changes in the number of site-days in these cell over time very likely correspond to changes in the underlying sampling design.

In order to incorporate the design changes, the maximum count was identified within each unique combinations of state, mode, and wave across years. Initial weights at the angler-trip level in a state-year-wave-mode cell were calculated as the count of site-days with intercepts in that cell, divided by the maximum count for that state-wave-mode combination. Hence, for cells corresponding to the year with the maximum count, the angler-trip weight is set equal 1, and for any other cell, the weight is greater than 1.

Under this approach, the initial weights will not be correct for the total number of trips, since they only account for relative changes in the design over time. This is justified by the fact that the overall "scale" of the weights, accounting for the volume of angler-trips, is not of interest in APAIS estimation, in which only rates are estimated.

Starting from these initial weights, a raking algorithm was again implemented to create updated weights. As a further adjustment for unobserved design effects, several raking control domains were added to those used for the 2004-2013 period:

- KOD (state, wave, mode and kind-of-day)
- MG (state, wave, mode and month groups)
- AC (state, wave, mode and site activity class).

The first of these corresponds to the usual weekday-weekend/holiday classification of angler-trips, but the other two require further explanation. For the MG domains, raking was attempted using individual month cells, but there were cases that would not converge. Months were therefore grouped into three classes: (1) January, March, October, December; (2) May, June, July, August; and (3) February, April, September, November. Class 1 represents the traditionally lower activity month during transition periods (month 1 in waves 1 and 2, month 2 in waves 5 and 6). Class 2 represents the peak activity period when sample sizes are generally similar or equally allocated among months within waves 3 and 4. Class 3 represents traditionally higher activity month during transition periods (month 2 in waves 1 and 2, month 1 in waves 5 and 6). For the AC domains, sites are divided into two groups, high activity and low activity, based on annual counts of intercepts by fishing mode. Sites with counts above the annual mean within cells defined by state, mode, year and sub-state region were classified as high; sites at or below the mean were classified as low.

The raking algorithm described in Section 2 was applied including these additional control variables, with the adjusted estimates for period 2004-2013 (wave 1) as the "new" estimates and those obtained with the initial weights described above for the period 1993-2003 as the "old" estimates. The linear regression testing for trend described in Section 3 was also performed, but with the modification that it was applied for both the new and the old periods. For any domains where a trend was detected in the old period, the adjustment ratio was computed on the years 2001-2003 instead of on the full period. Similarly, for domains where a trend was detected in the new period, the adjustment ratio was computed using 2004-2006 instead of the full period.

5 Adjustments for prior periods: 1981-1992

The adjustment procedure for 1981-1992 follows the same procedure as that for 1993-2003. The initial weights are again created based on relative counts of site-day assignments, and the raking procedure uses the additional control domains described in Section 4. "New" estimates are those obtained with the adjusted weights for 1993-2003 and "old" estimates are those for 1981-1992. Significant trends resulted in shortening of the period used for the raking ratios to 1990-1992 for the old period and to 1993-1995 for the new period.

References

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

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

MRIP Staff (2014, August 20). A descriptive analysis of the Access Point Angler Intercept Survey 2013 design change. Internal report, National Marine Fisheries Service.

NMFS Staff (2015, May 5). Transition plan for the Fishing Effort Survey. Internal report, National Marine Fisheries Service.

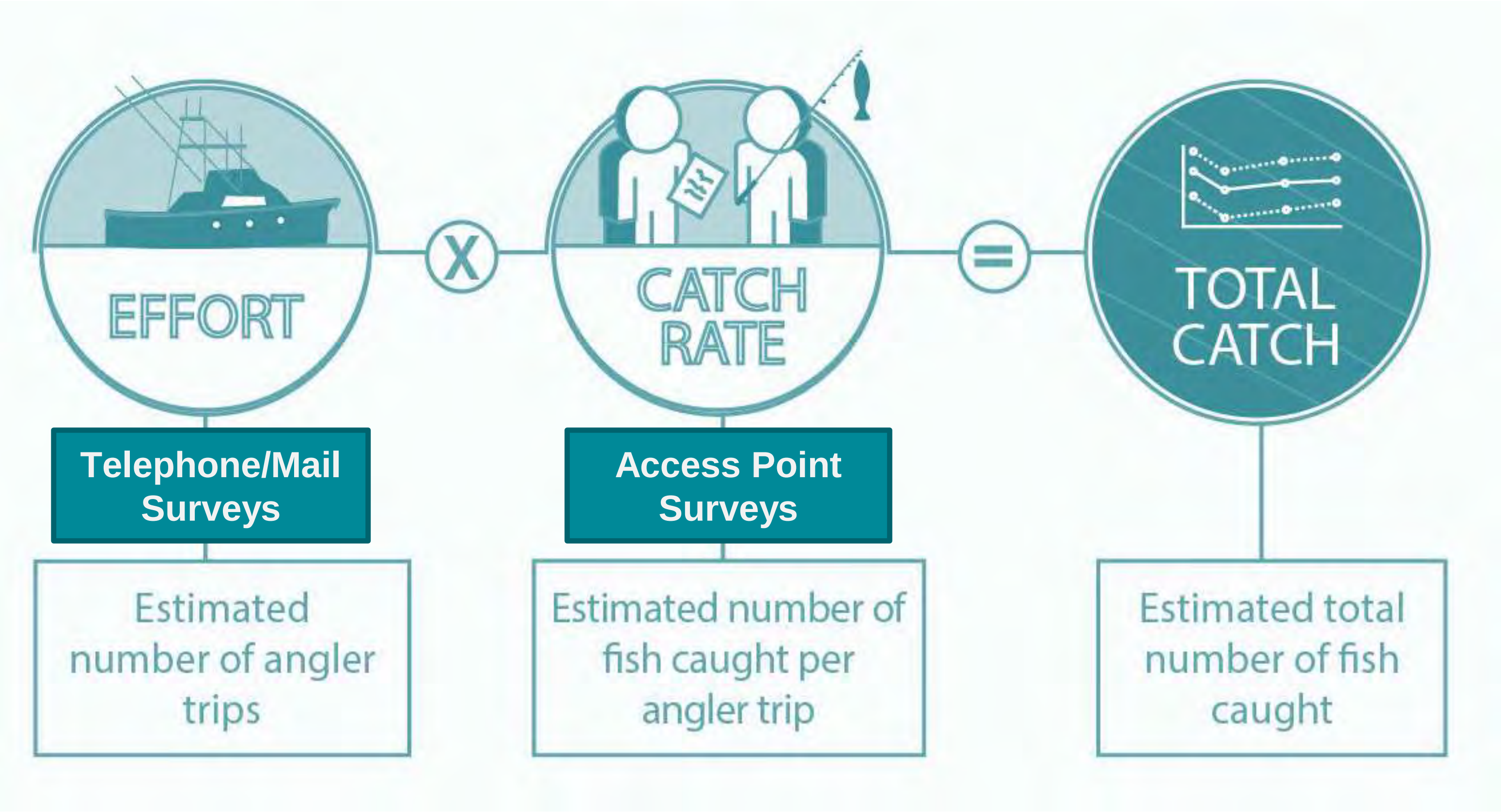
Sullivan, P. J., F. J. Breidt, R. B. Ditton, B. A. Knuth, B. M. Leaman, V. M. O'Connell, G. R. Parsons, K. H. Pollock, S. J. Smith, and S. L. Stokes (2006). *Review of Recreational Fisheries Survey Methods*. Washington, DC: National Academies Press.

Marine Recreational Information Program

Transition to Improved Survey Designs

John Foster
Kelly Denit

Estimating Total Recreational Catch



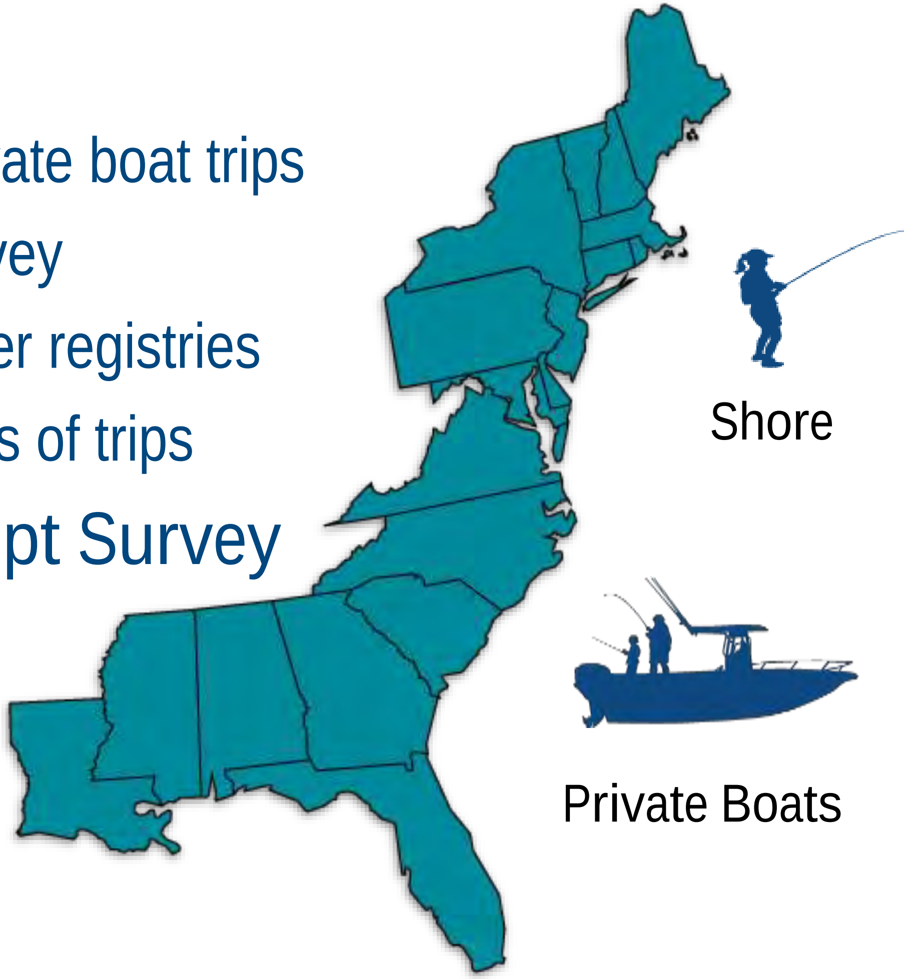
Improved Survey Designs

Fishing Effort Survey

- New mail survey of shore & private boat trips
- Replaces legacy telephone survey
- Uses USPS database and angler registries
- Higher, more accurate estimates of trips

Access Point Angler Intercept Survey

- New design for sampling catch
- Better time-of-day coverage
- More statistically sound



Fishing Effort Survey

We’re reaching more anglers.

Surveys get into the right hands.

Much higher response rate.

Improved questionnaire.

More complete answers.

Extensive Testing and Peer Review

“The [Fishing Effort Survey] methodologies, including the address-based sampling survey design, are major improvements.”

“The current methods used in the APAIS … are a vast improvement over the previous sampling and estimation procedures and reflect state of the art methods in survey sampling.”

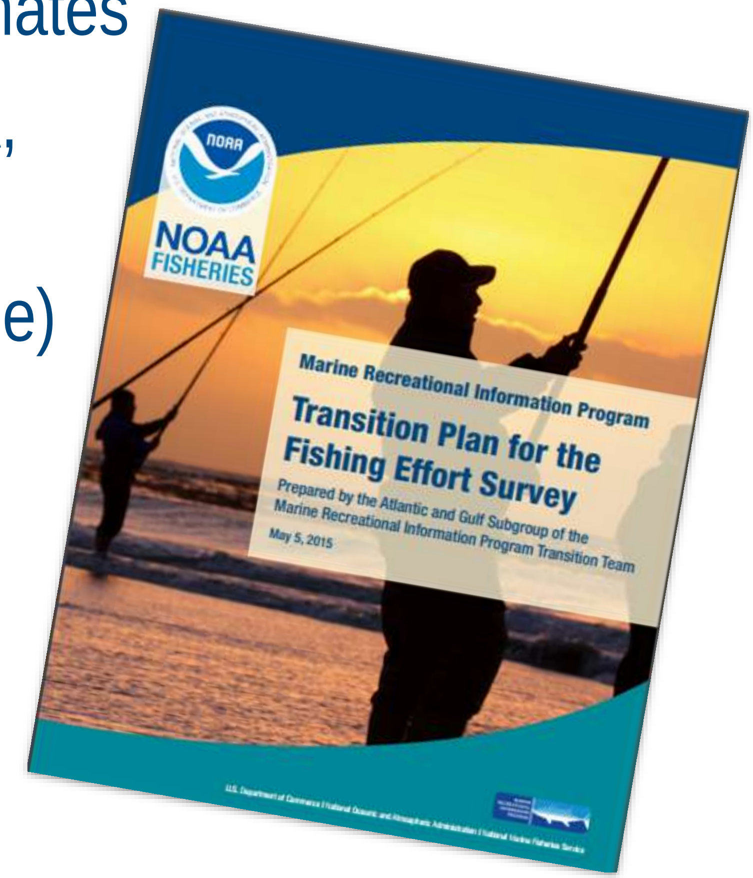
The National Academies of SCIENCES ENGINEERING MEDICINE

Transition Plan

Improving designs can change estimates

Transition Plan developed by NOAA, states, councils, and commissions.

- Benchmarking (mail vs. phone)
- Calibration (old vs. new)
- Convert “old” estimates into “new” currency
    - Stock assessments
    - Management decisions



MRIP Calibration Models

Fishing Effort Survey

- Fay-Herriot small area estimation model.
- Developed with independent consultants at Colorado State University.

Access Point Angler Intercept Survey

- Sample weight adjustment method.
- Developed with independent consultants at Colorado State University.

MRIP Calibration Model Peer Reviews

Fishing Effort Survey

- June 27-29, 2017 workshop in Silver Spring, MD.
- Reviewers unanimously endorsed the proposed FES/CHTS calibration model.

Access Point Angler Intercept Survey

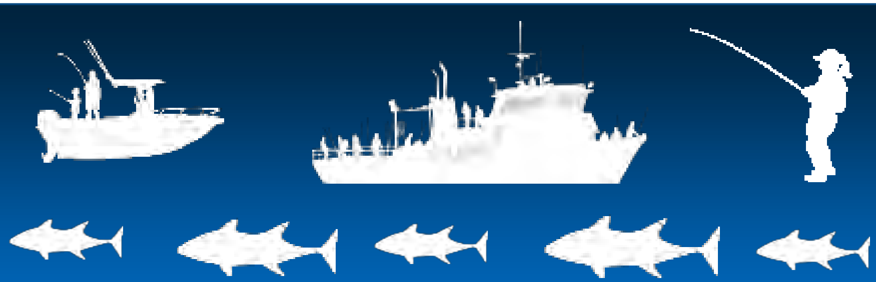
- March 20-22, 2018 workshop in Silver Spring, MD.
- Initial reviews positive, anticipating full review reports in coming weeks.

2018 Annual Catch Limits

2018 ACLs set using CHTS estimates

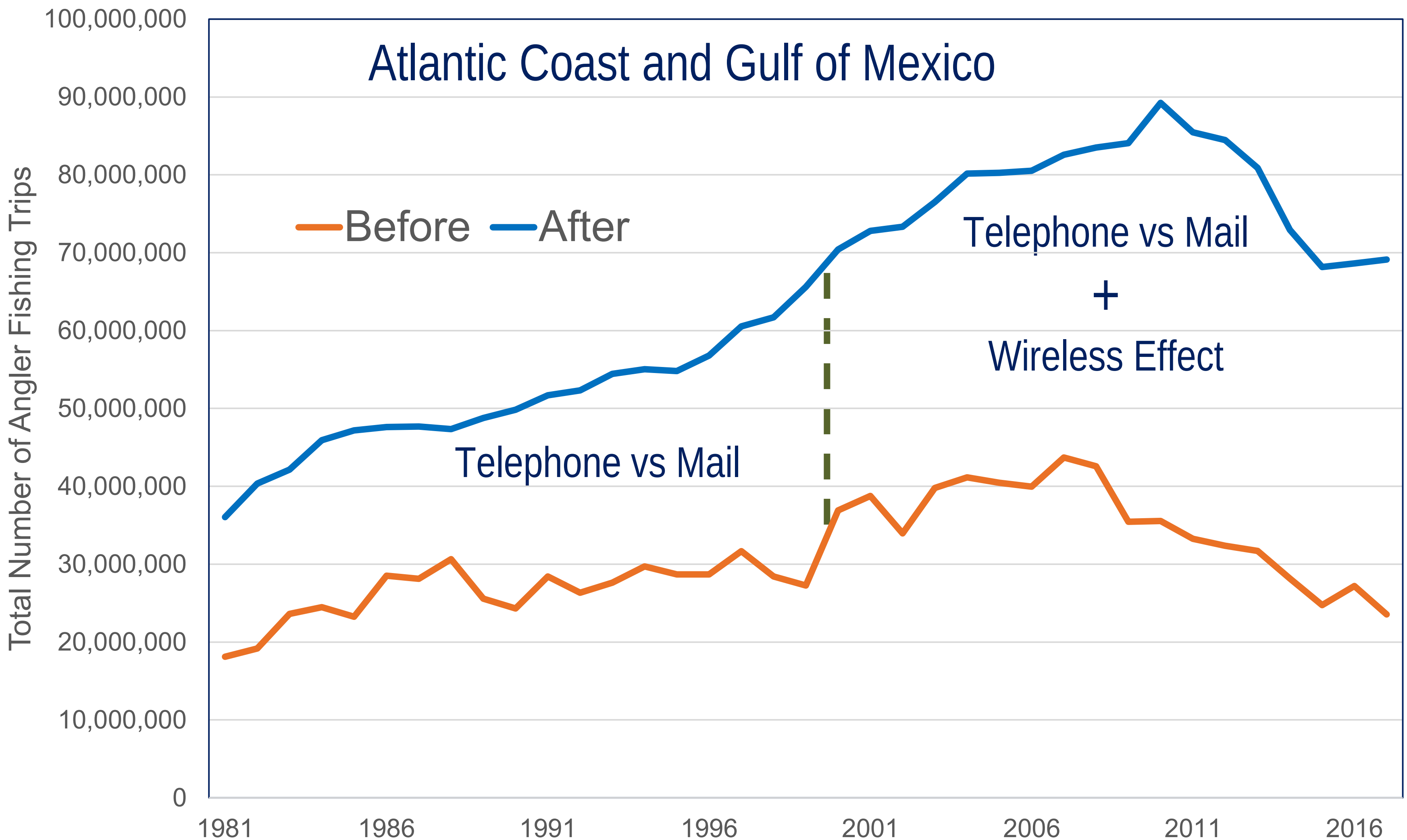
FES ⟷ CHTS

Calibration model converts both directions

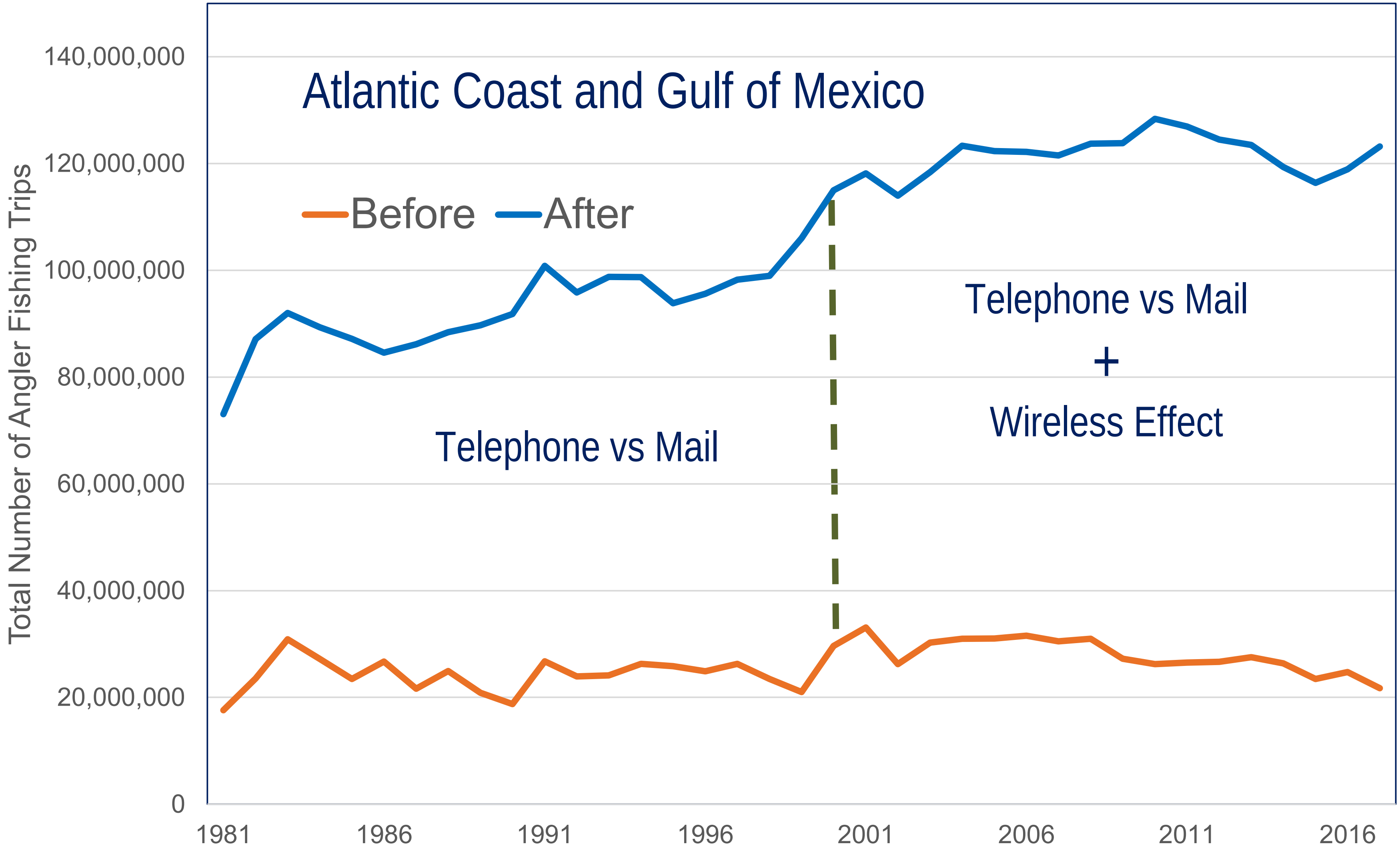


Year-**end estimates in same “currency”**

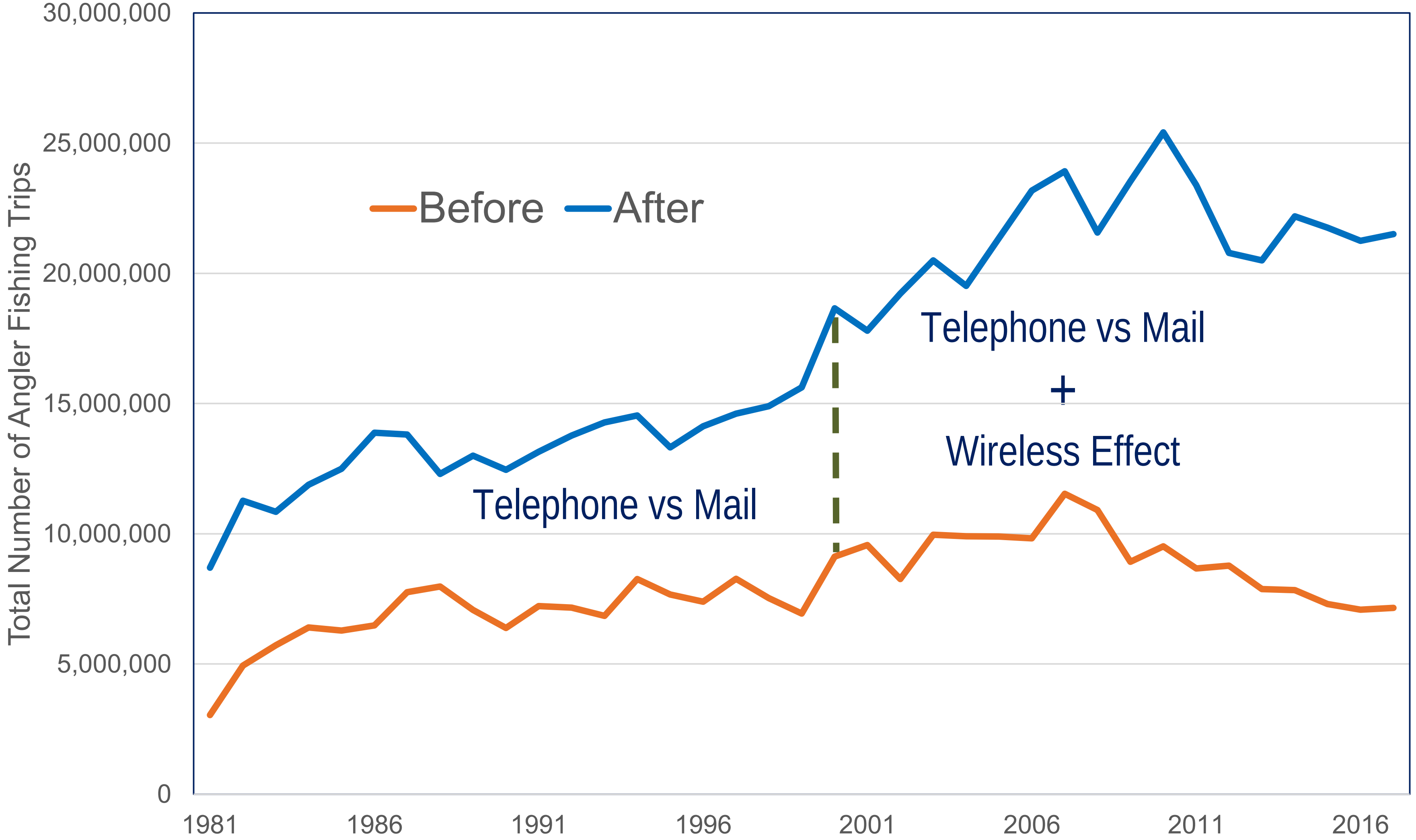
Private Boat Effort



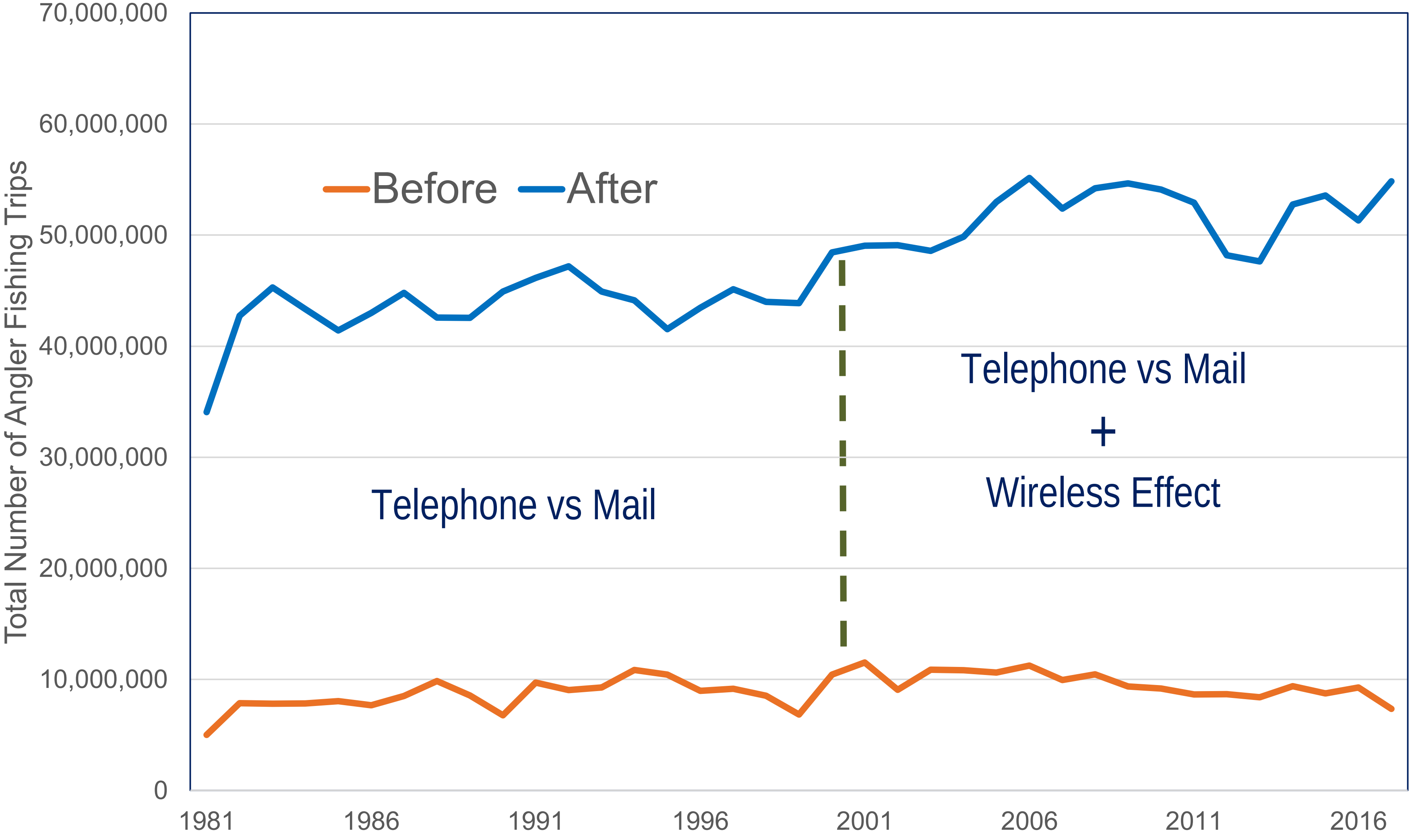
Shore Effort



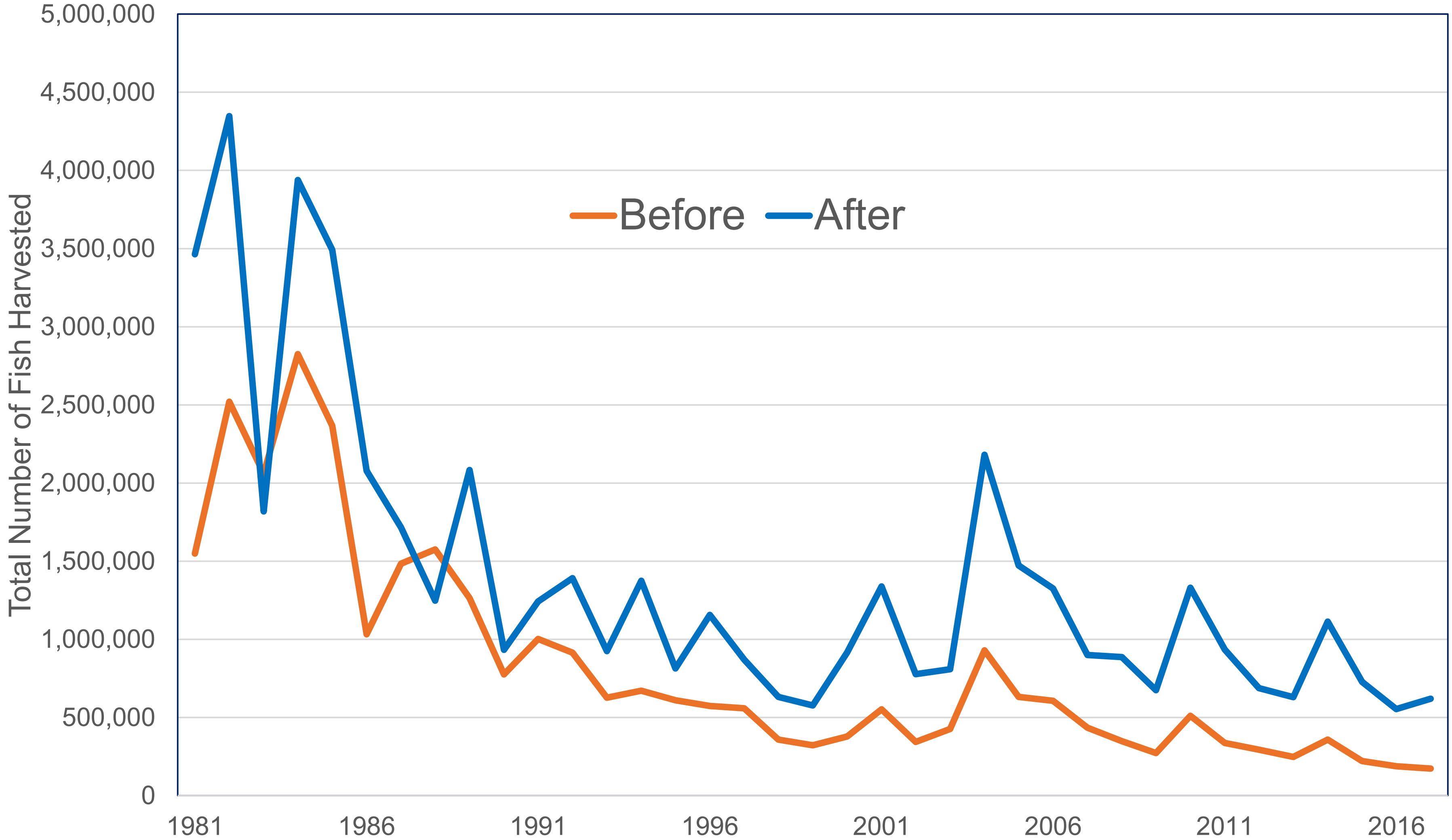
South Atlantic Private Boat Effort



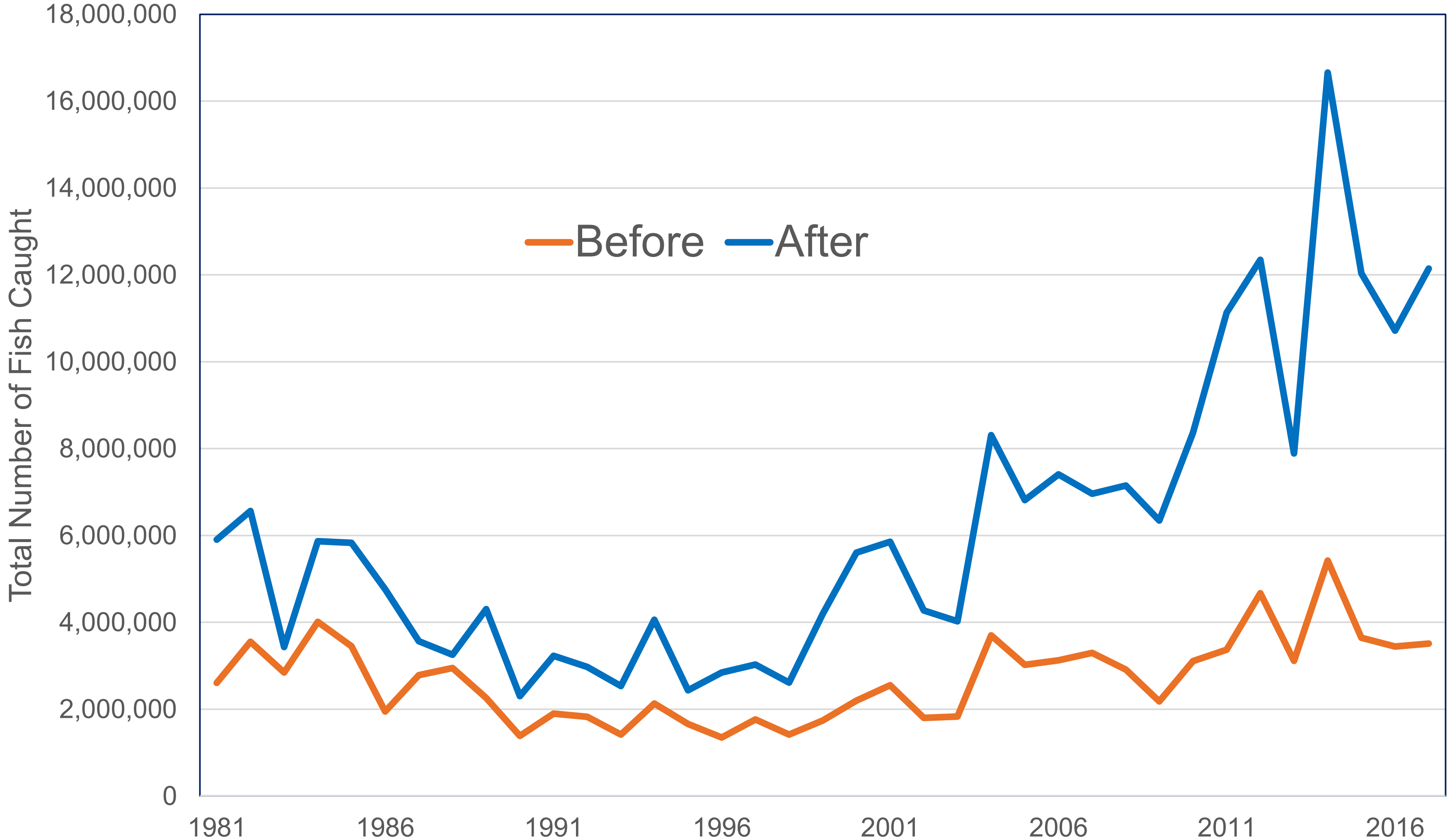
South Atlantic Shore Effort



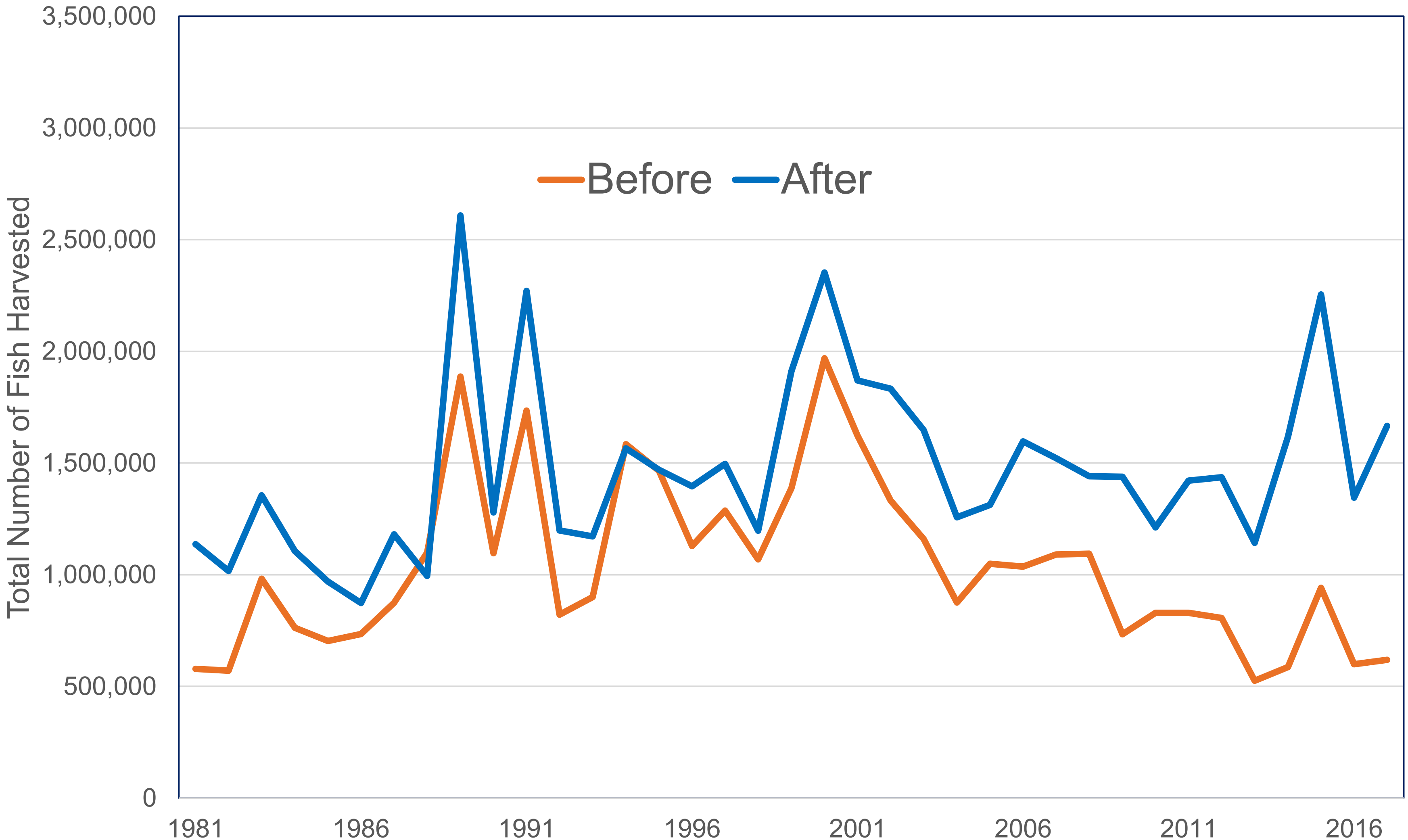
Black Sea Bass Harvest



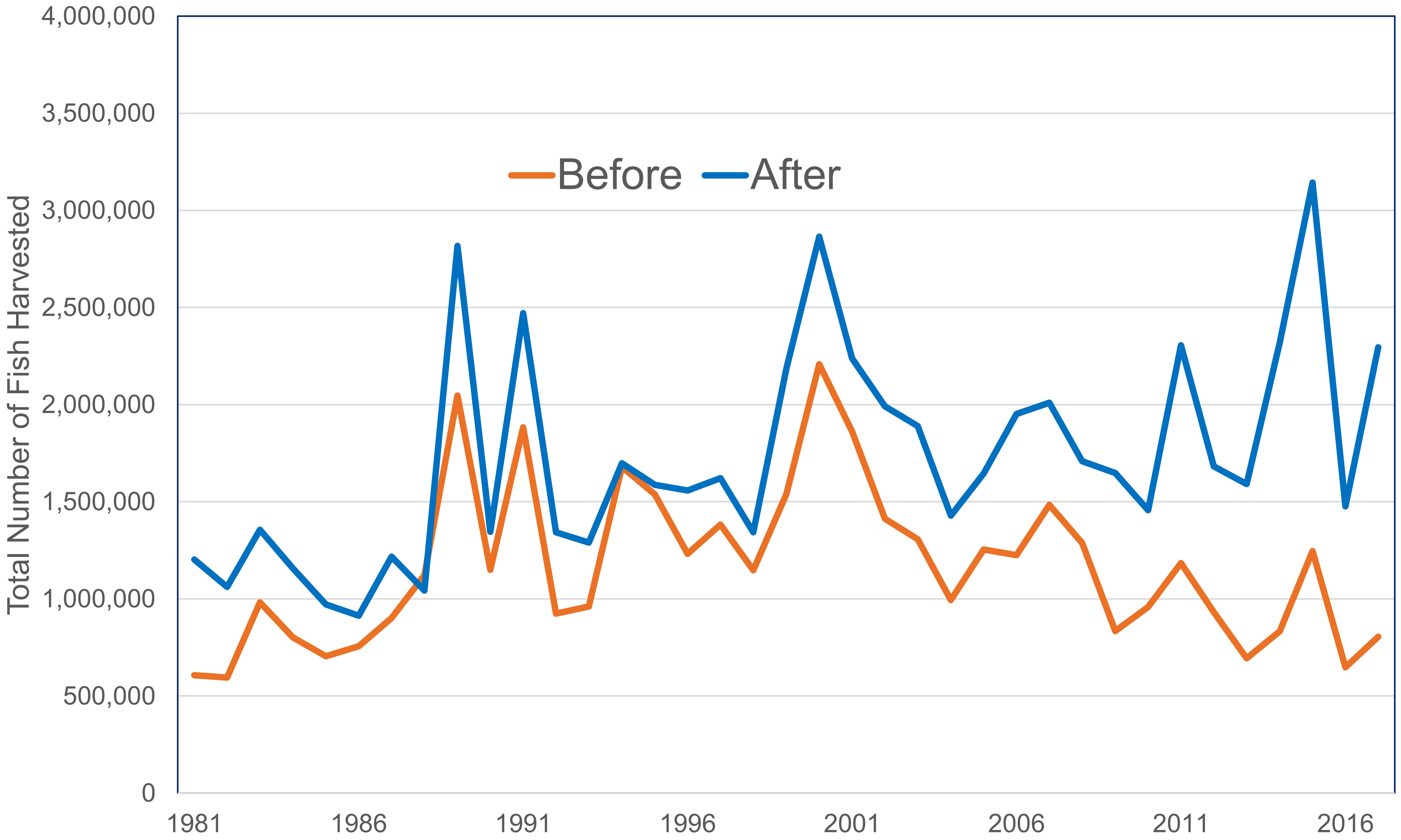
Black Sea Bass Total Catch



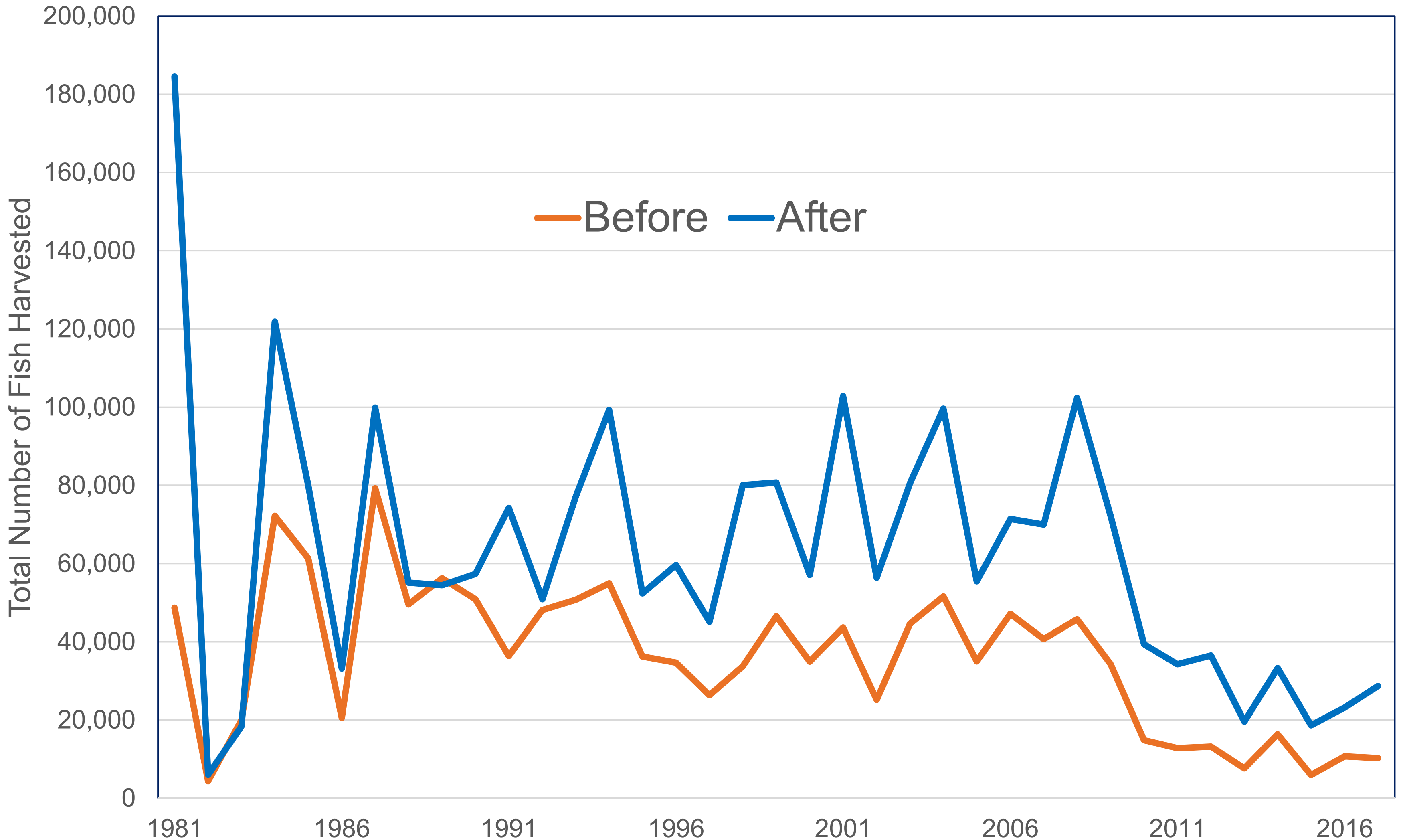
Dolphin Harvest



Dolphin Total Catch

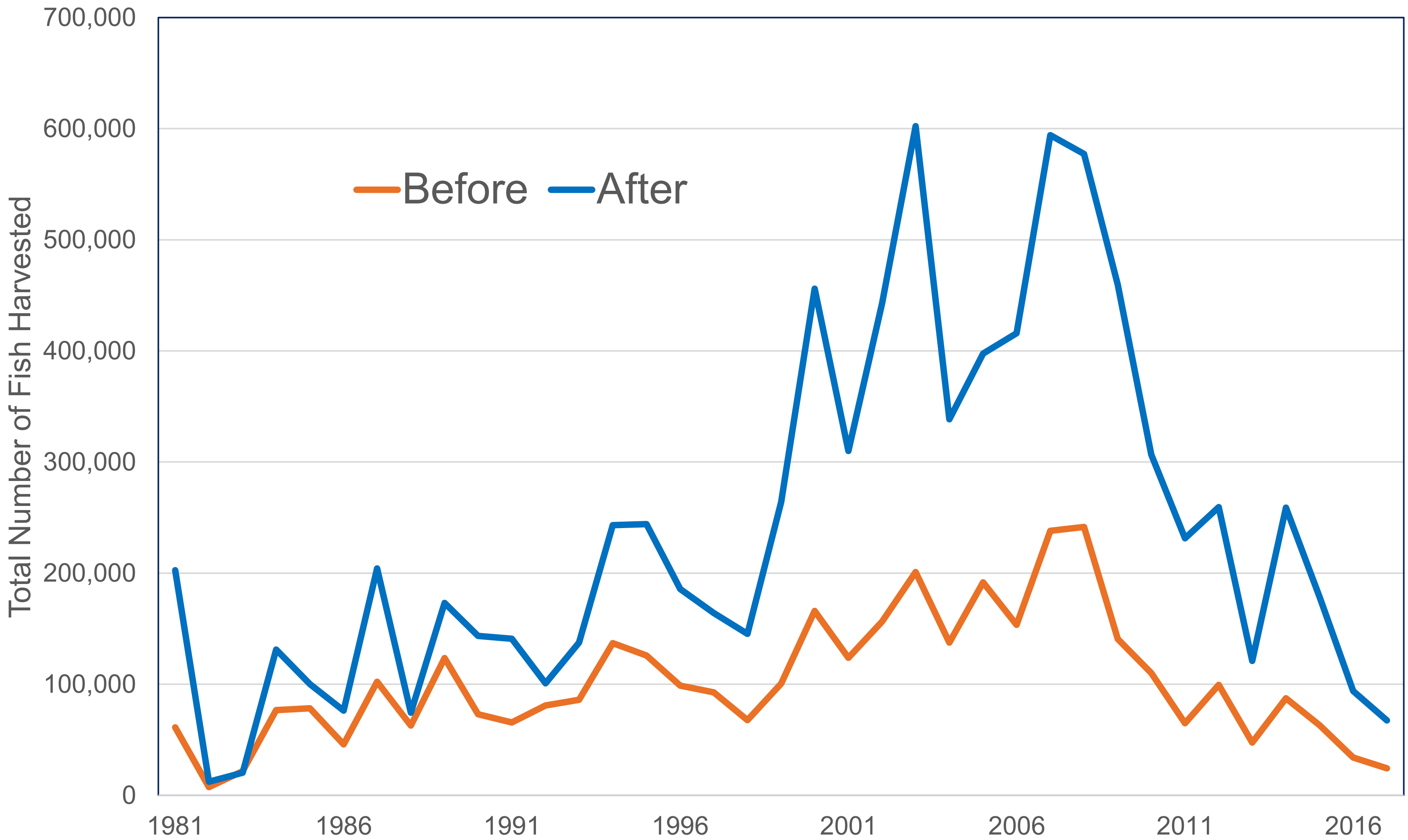


Gag Harvest

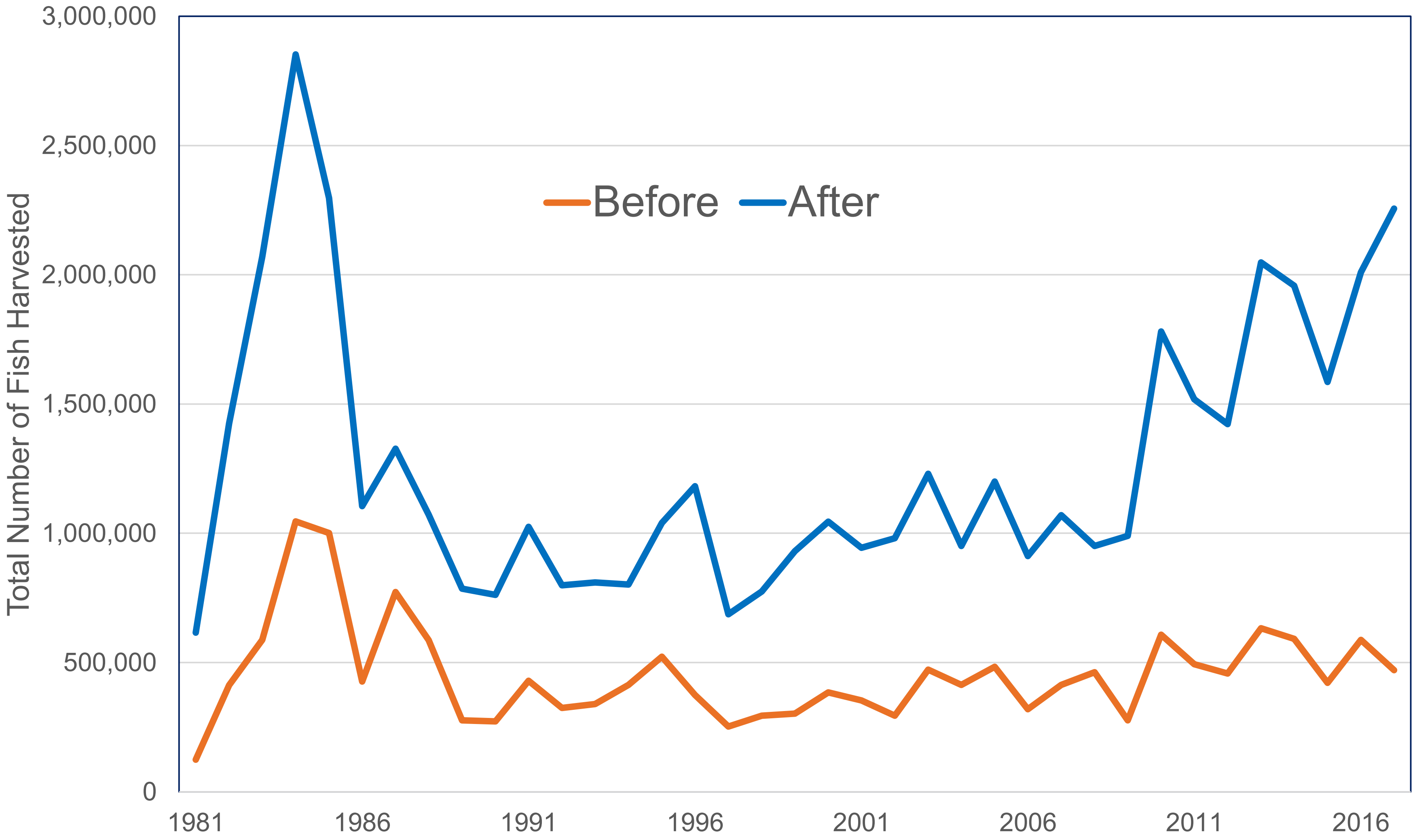


Total Number of Fish Harvested
200,000
180,000
160,000
140,000
120,000
100,000
80,000
60,000
40,000
20,000
0
1981
1986
1991
1996
2001
2006
2011
2016
Before
After

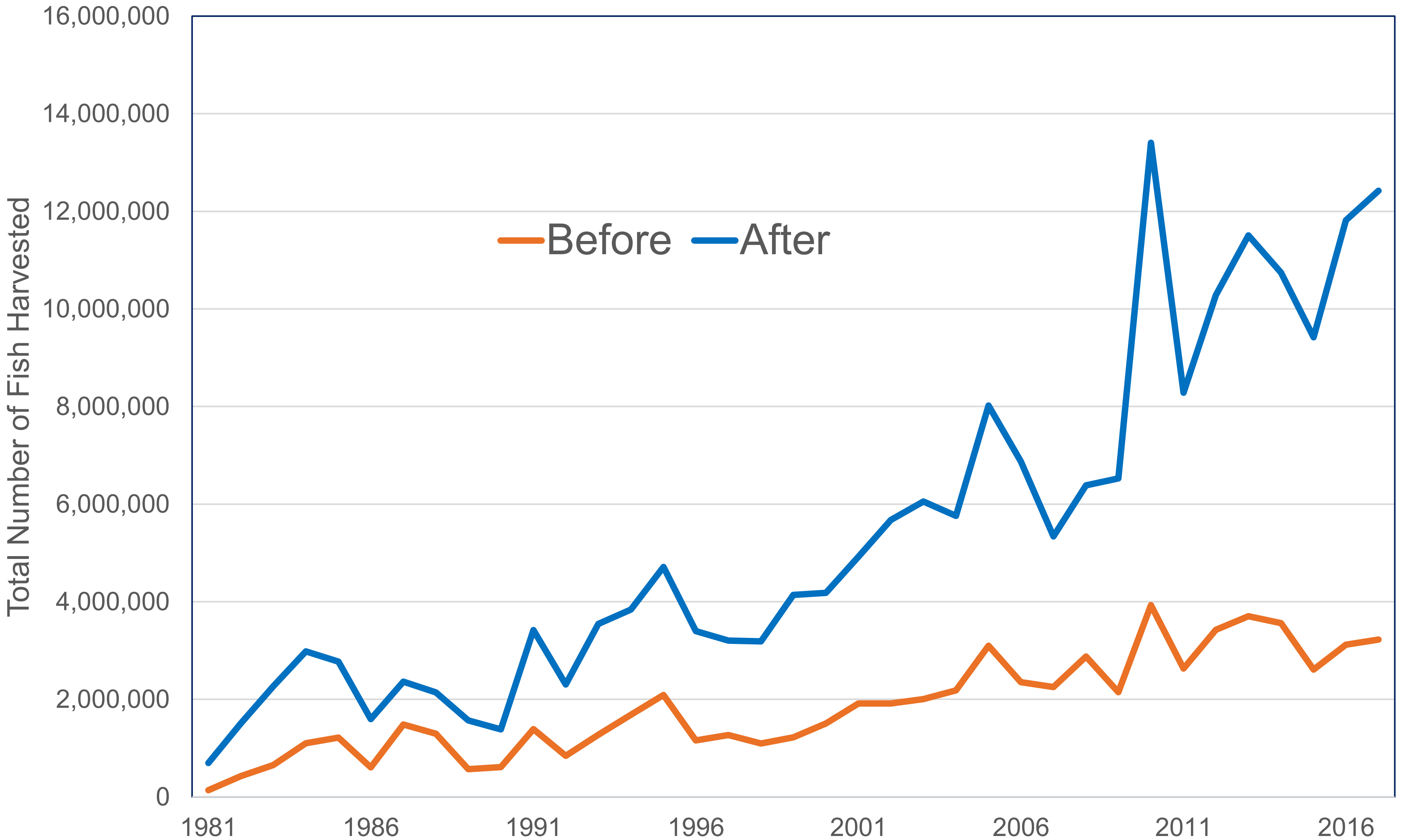
Gag Total Catch



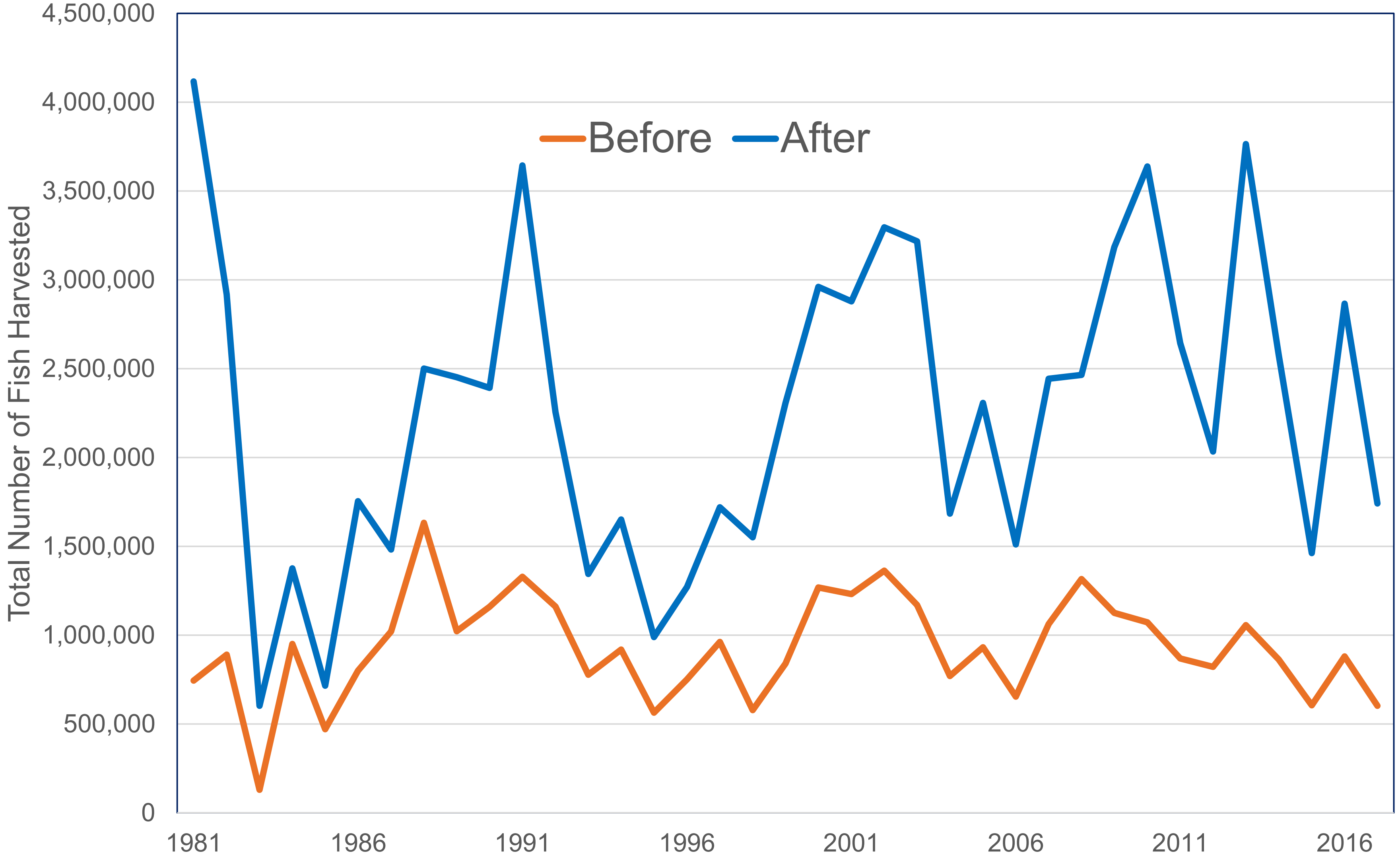
Red Drum Harvest



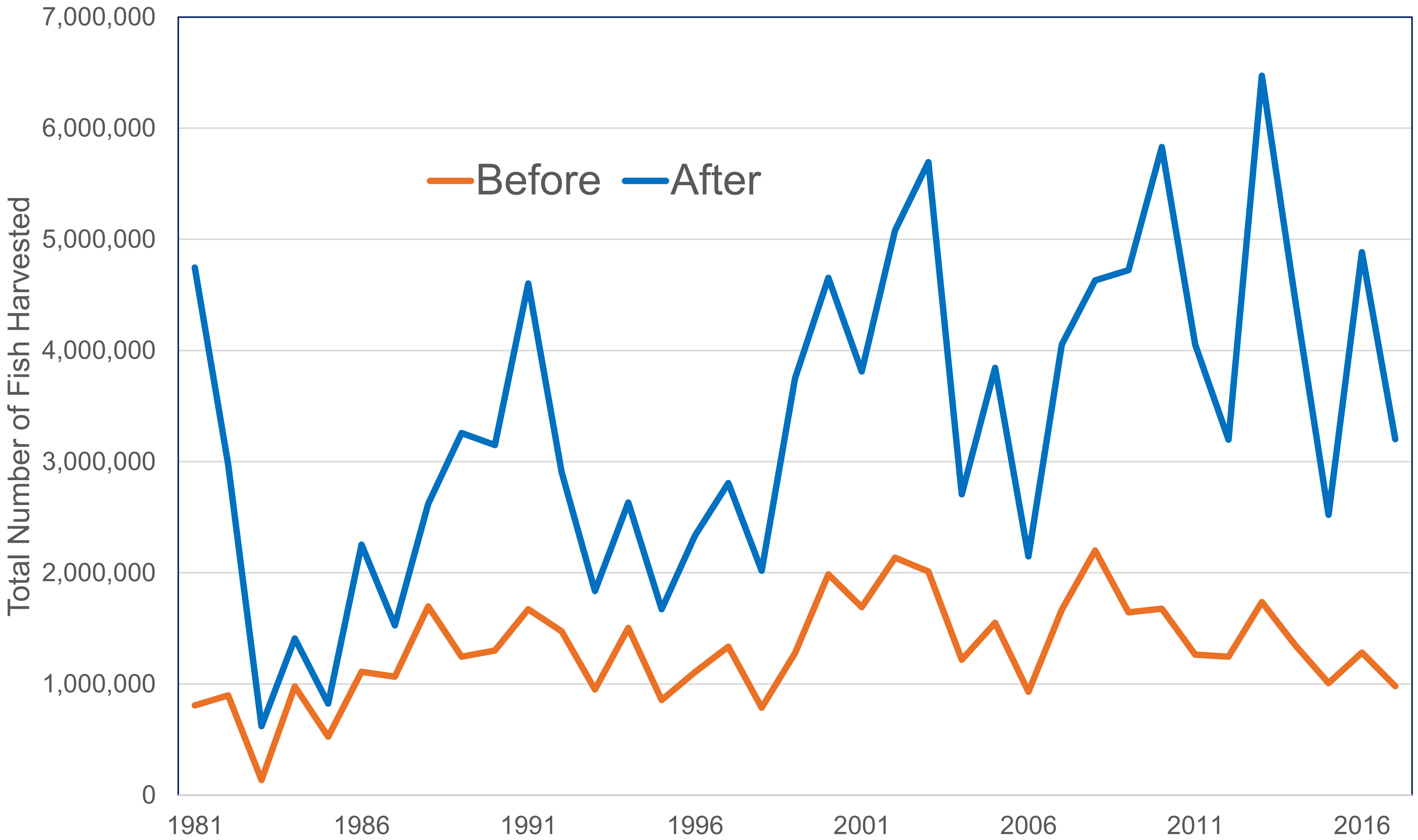
Red Drum Total Catch



Spanish Mackerel Harvest



Spanish Mackerel Total Catch



Impacts on Recreational Fishing

Stock Assessment Schedule

Species	Region	Type	Timing
Striped Bass	Atlantic Coast	Benchmark	Fall 2018
Summer Flounder	Mid-Atlantic Coast	Benchmark	Fall 2018
Red Snapper	Gulf of Mexico	Data Update	Fall 2018
Atlantic Cod	Gulf of Maine	Operational	Early 2019
Black Sea Bass	Mid-Atlantic Coast	Operational	Early 2019
Bluefish	Atlantic Coast	Operational	Early 2019
Scup	Atlantic Coast	Operational	Early 2019
Spot	Atlantic Coast	Operational	Early 2019
Red Porgy	Southern Atlantic Coast	Full Update	May 2019
Greater Amberjack	Southern Atlantic Coast	Full Update	May 2019
Red Grouper	Gulf of Mexico	Full Update	July 2019
Gray Triggerfish	Gulf of Mexico	Full Update	Aug. 2019
Yellowtail Snapper	Southern Atlantic Coast/Gulf of Mexico	Benchmark	Nov. 2019
Red Snapper	South Atlantic	Full Update	2020

Key Takeaways

- We see a substantial increase in effort, especially in the shore mode.
- Those stocks that have high proportion of catch from shore are more heavily impacted.
- The changes in effort are generally larger in more recent years, mostly driven by the “wireless” effect.
- 2018 catch will be back-calibrated to ensure ACLs and catch are in the same “currency.”
- Stock assessments incorporating this new data will be used to determine stock status and ACLs.

What’s Next

2018

- Revised total catch estimates now available for use in planned stock assessments.
- 2018 ACLs and catch estimates will be in the same **“currency.”**

2019

- Preliminary management changes may be made for re-assessed stocks.
- Calibrated statistics incorporated into additional stock assessments.

2020

- Based on new stock assessments, management changes could occur for a number of species.

Upcoming Presentations

August 2

- Atlantic States Marine Fisheries Commission Assessment Science Committee and Management and Science Committee

August 14

- Mid-Atlantic Fishery Management Council

September 24

- New York State Department of Environmental Conservation, Division of Marine Resources, Marine Resources Advisory Council

Scheduling Underway

New England Fishery Management Council

September 25-27

- New England Fishery Management Council

October 2-4

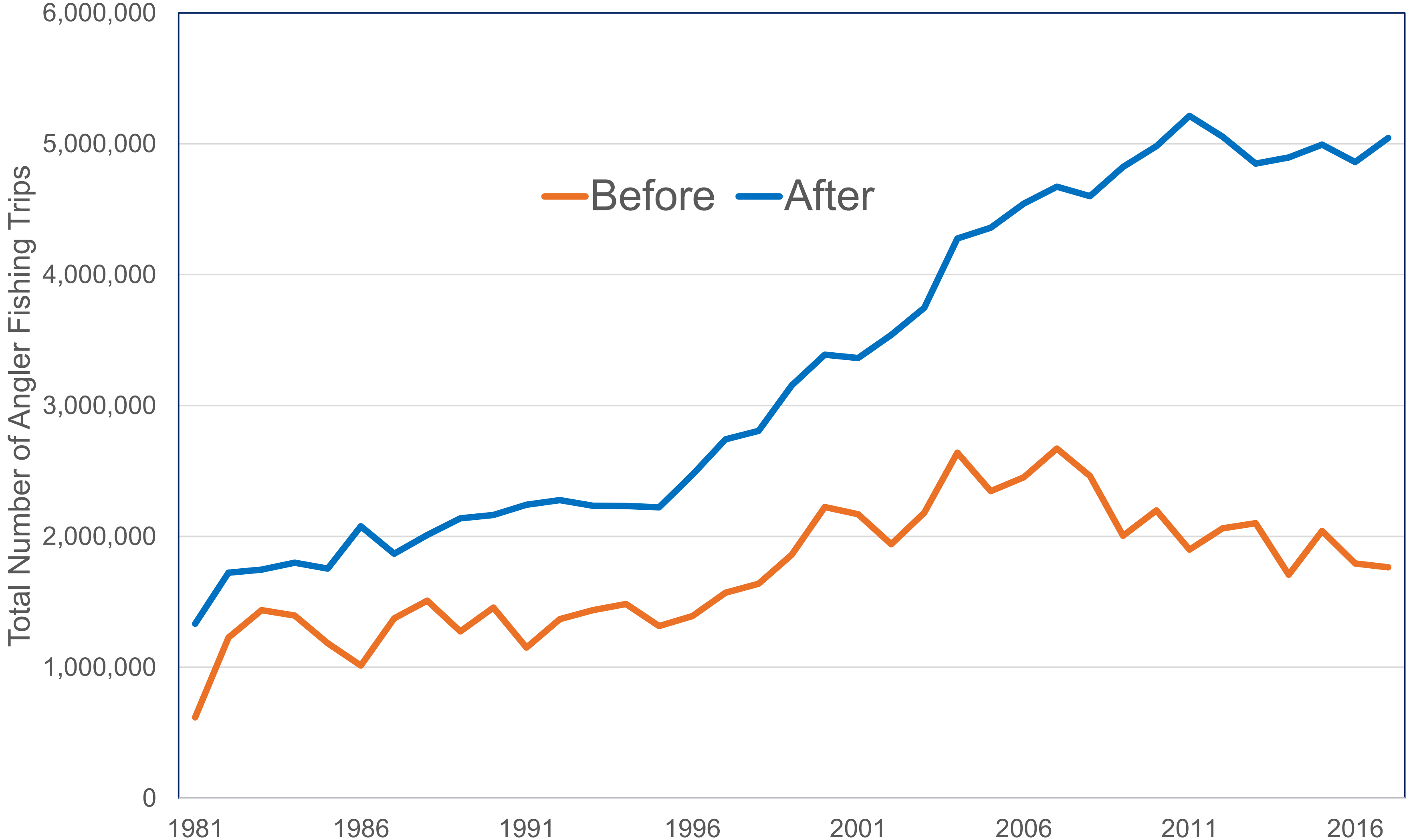
- Gulf of Mexico Fishery Management Council Scientific & Statistical Committee

October 22-25

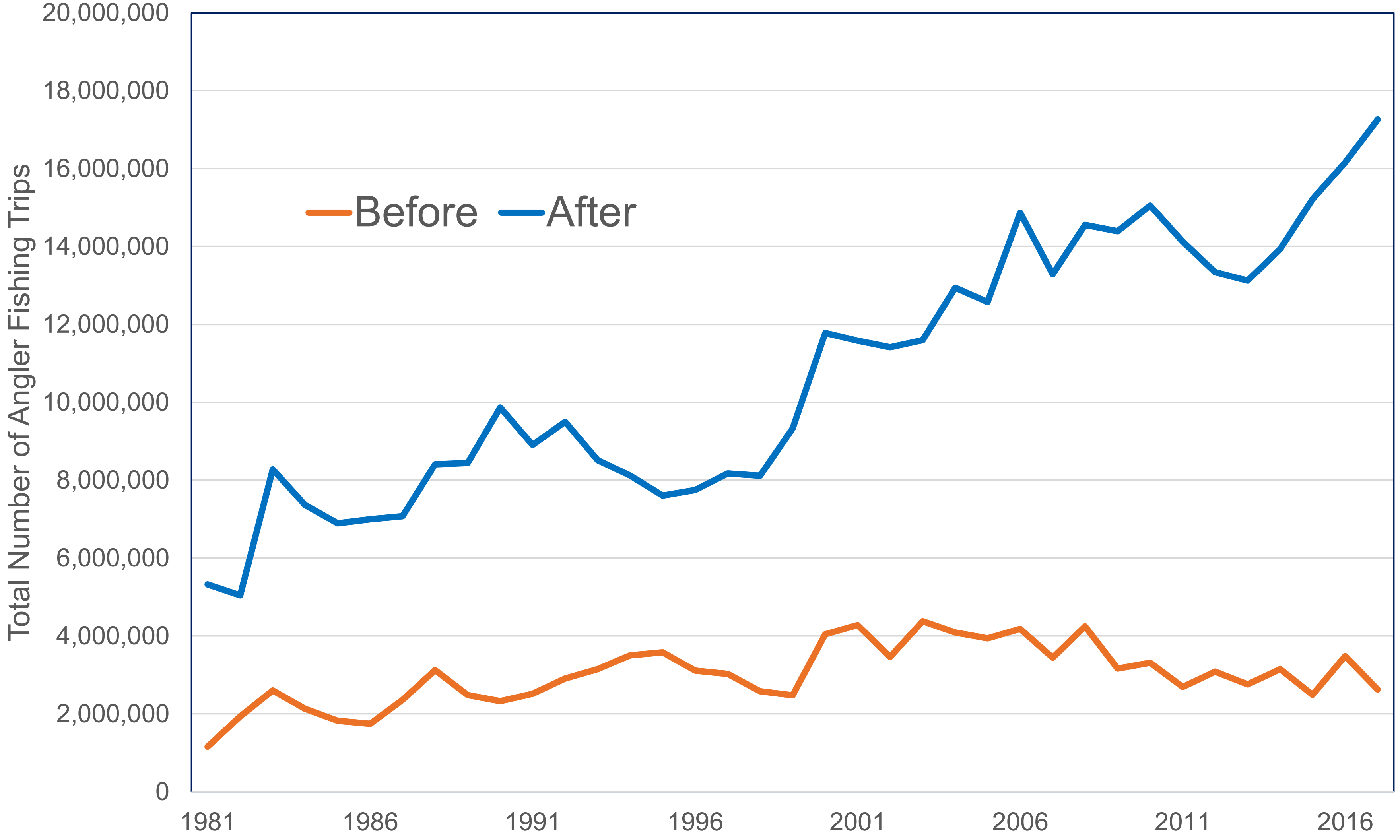
- Gulf of Mexico Fishery Management Council

Questions?

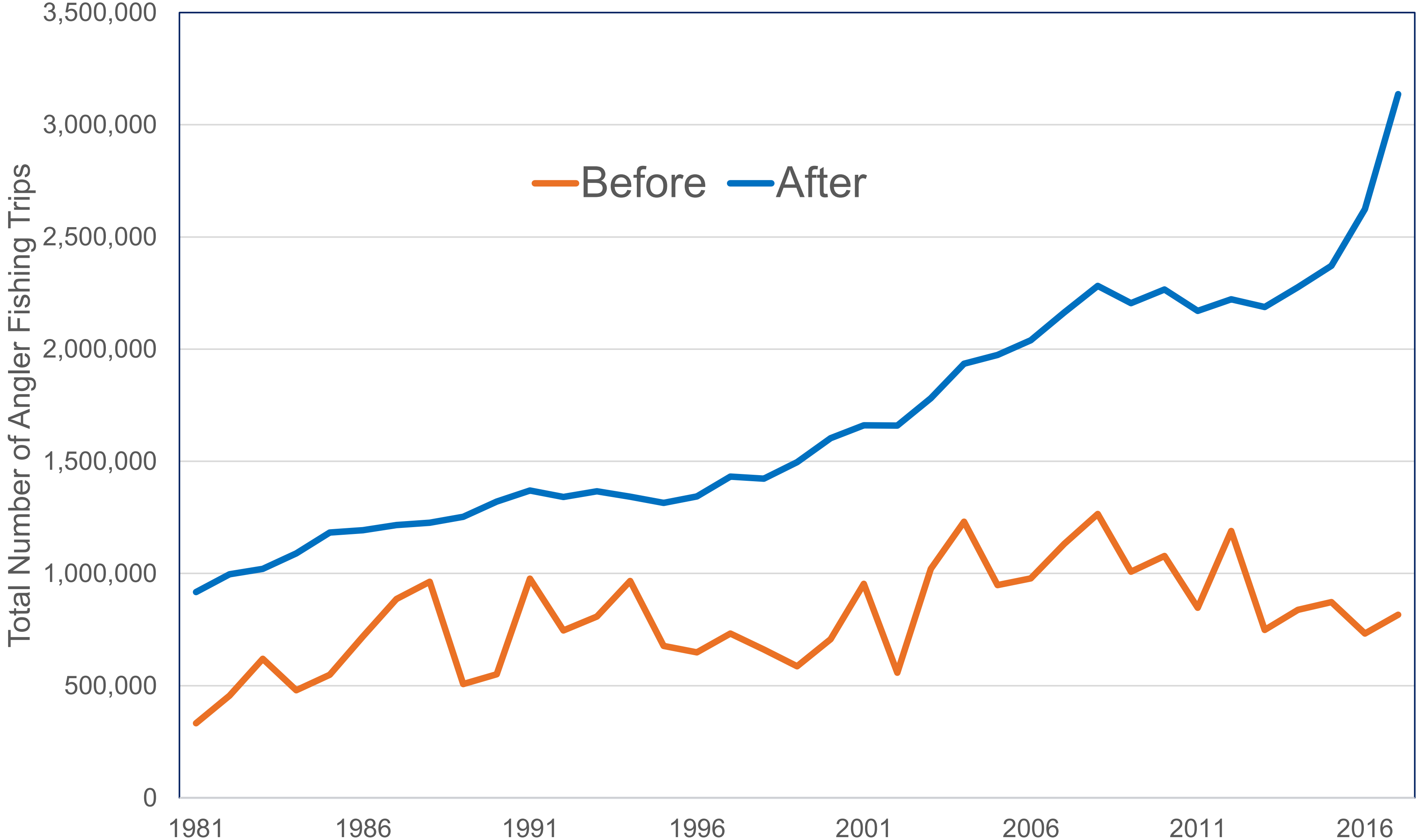
North Carolina Private Boat Effort



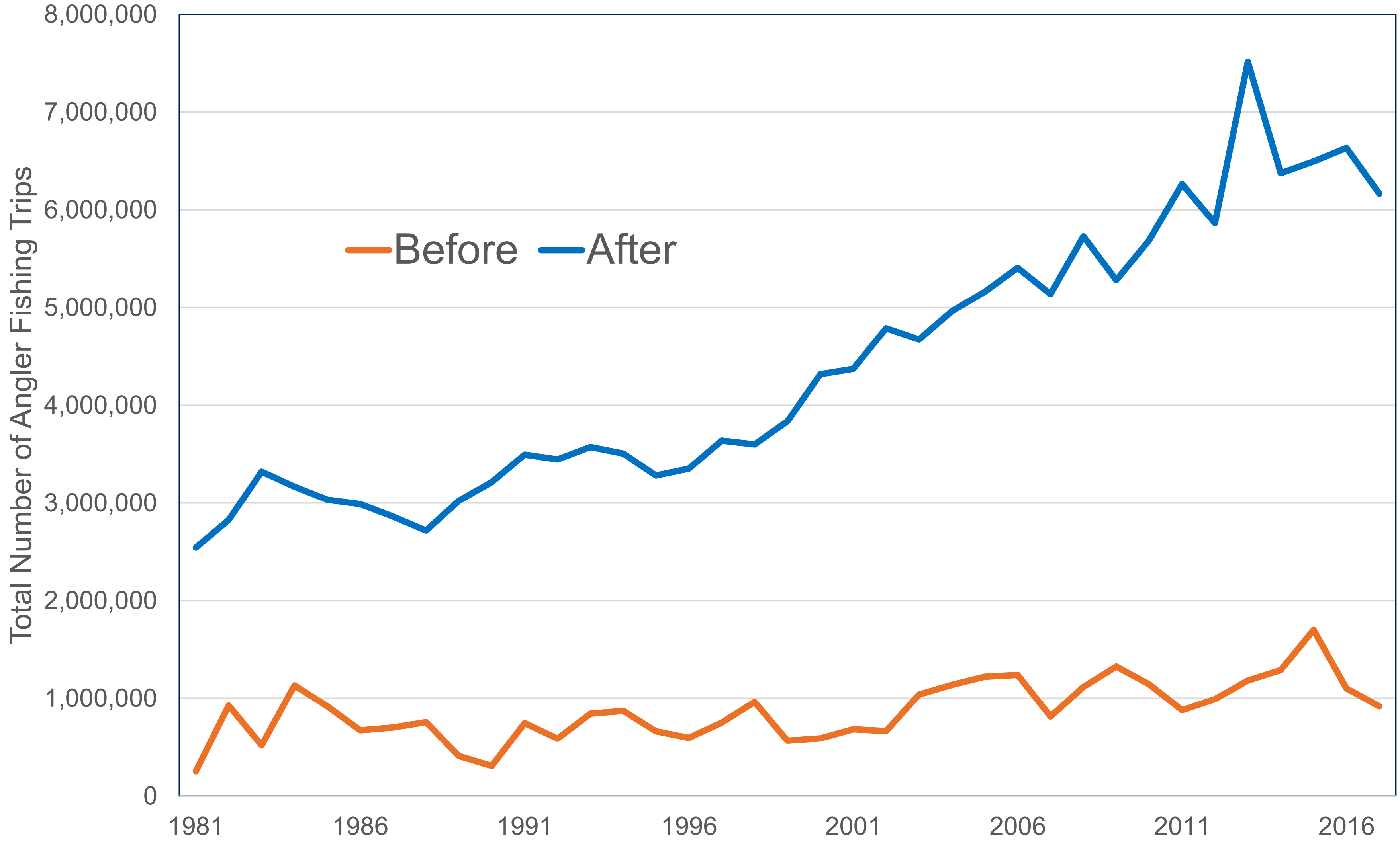
North Carolina Shore Effort



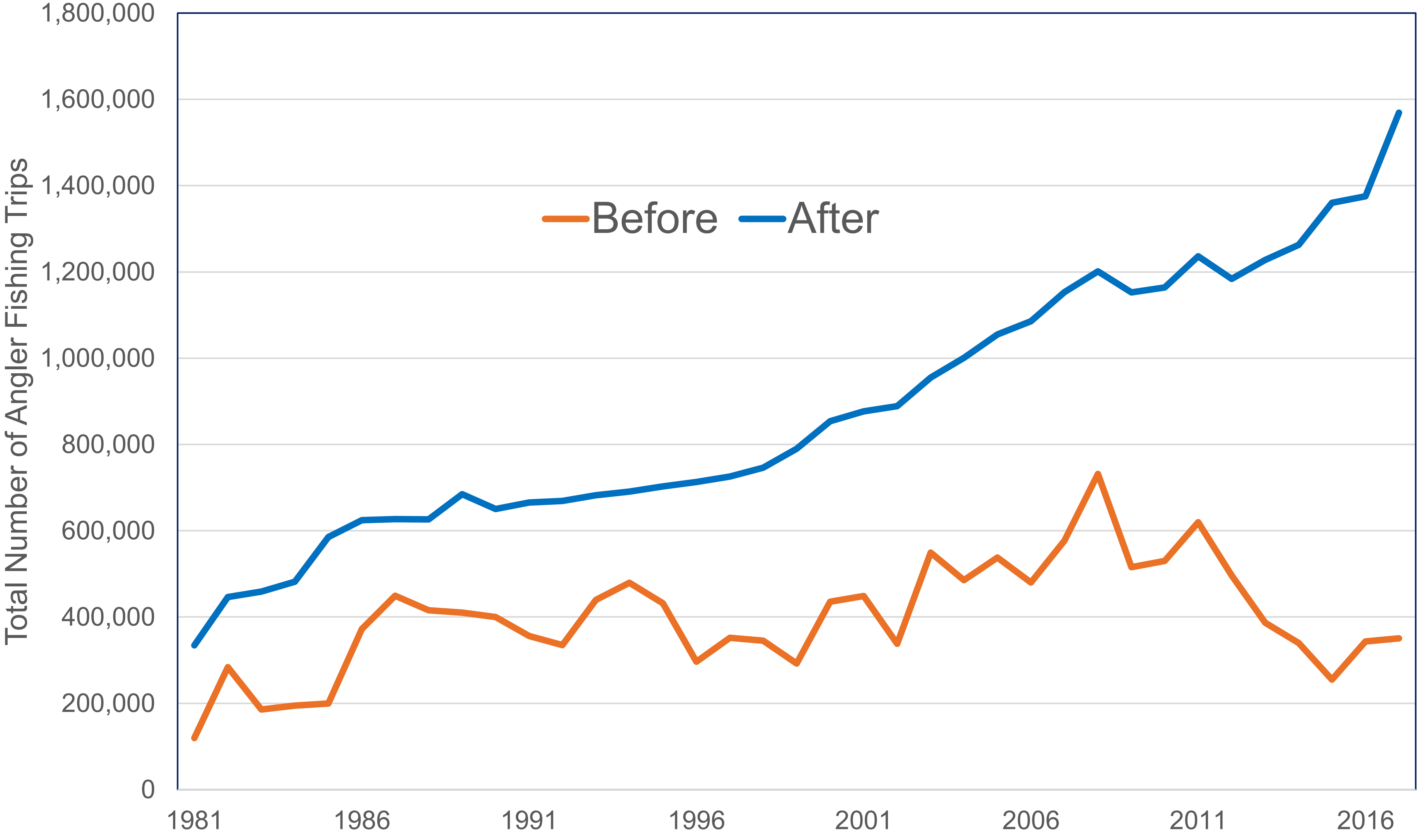
South Carolina Private Boat Effort



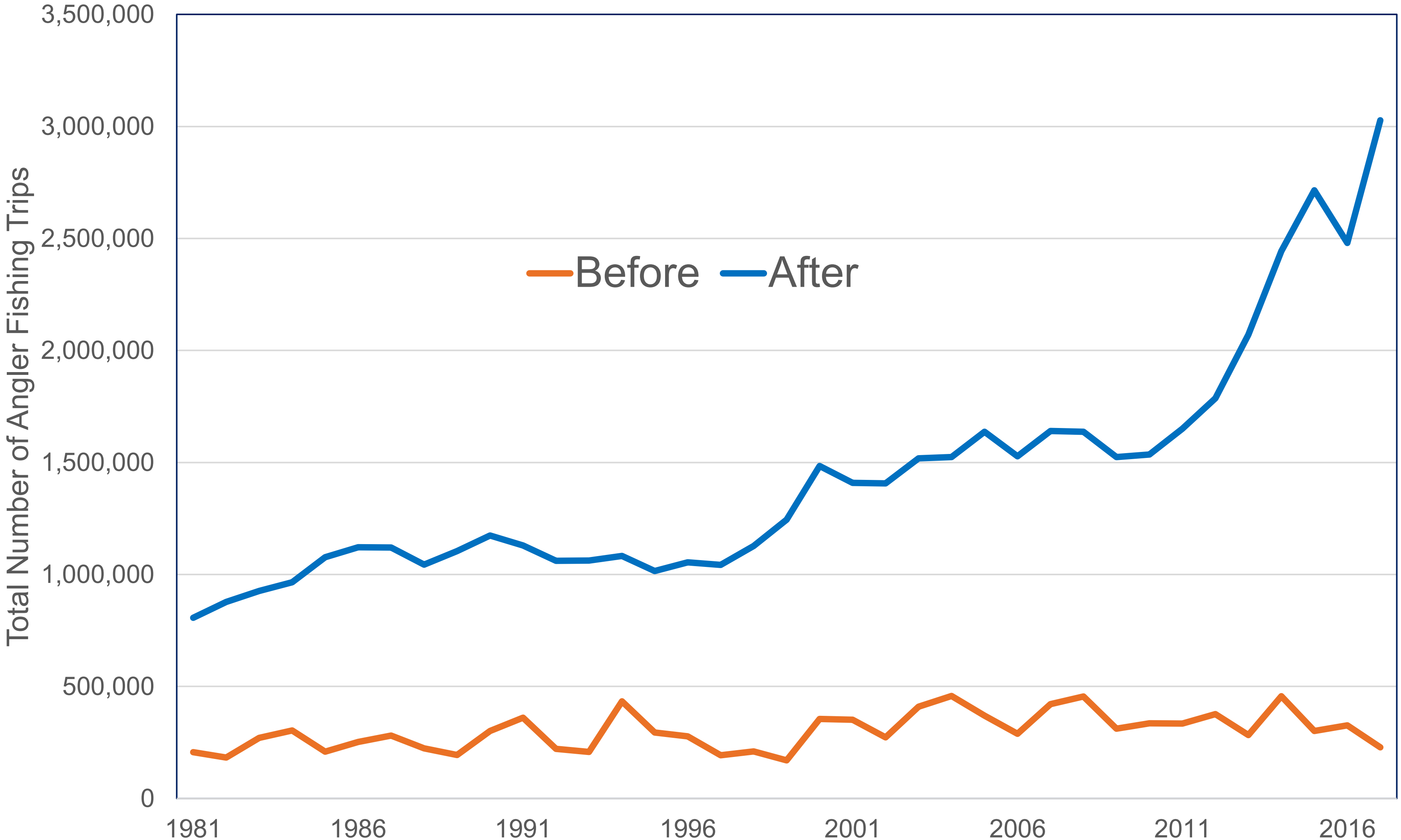
South Carolina Shore Effort



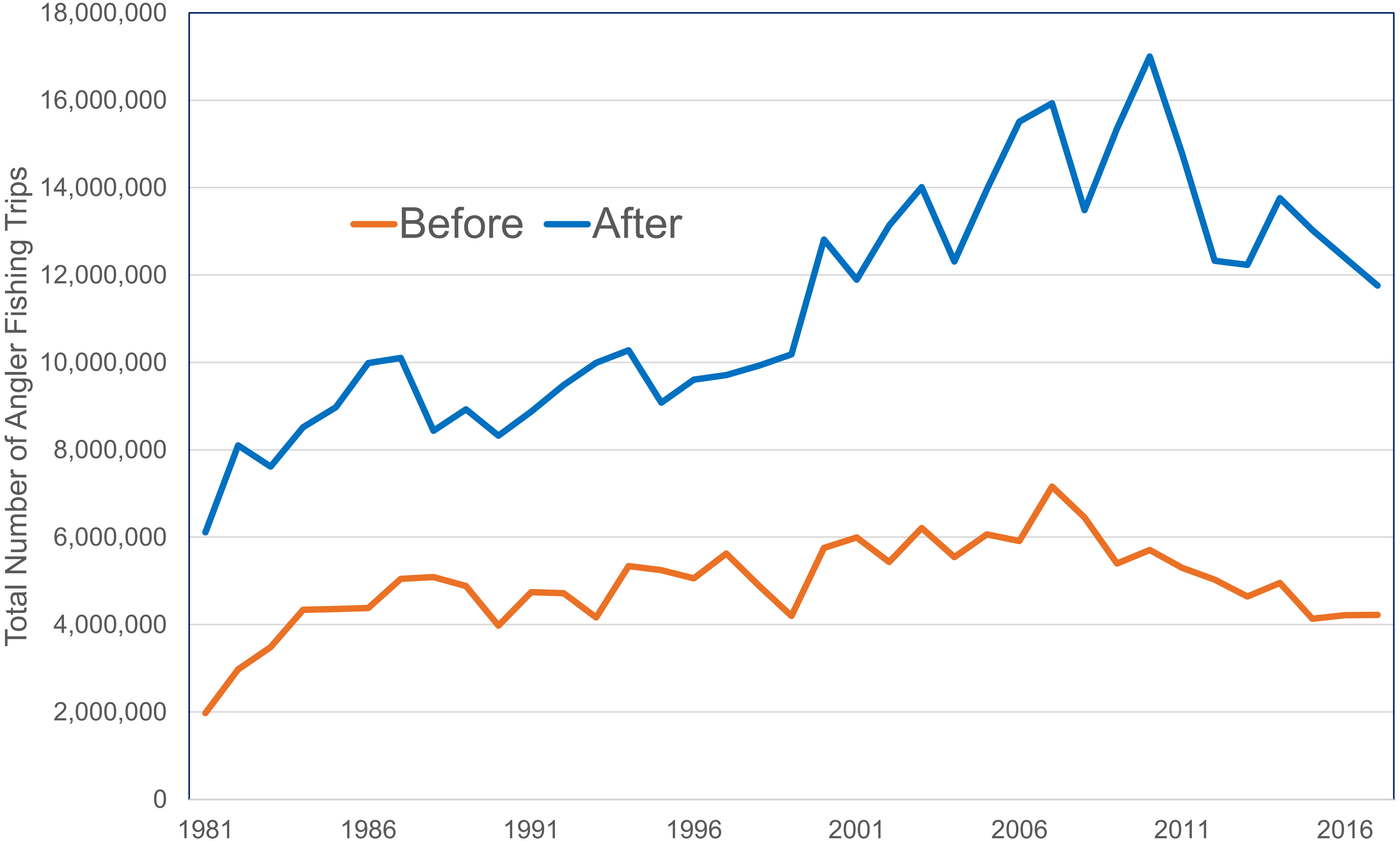
Georgia Private Boat Effort



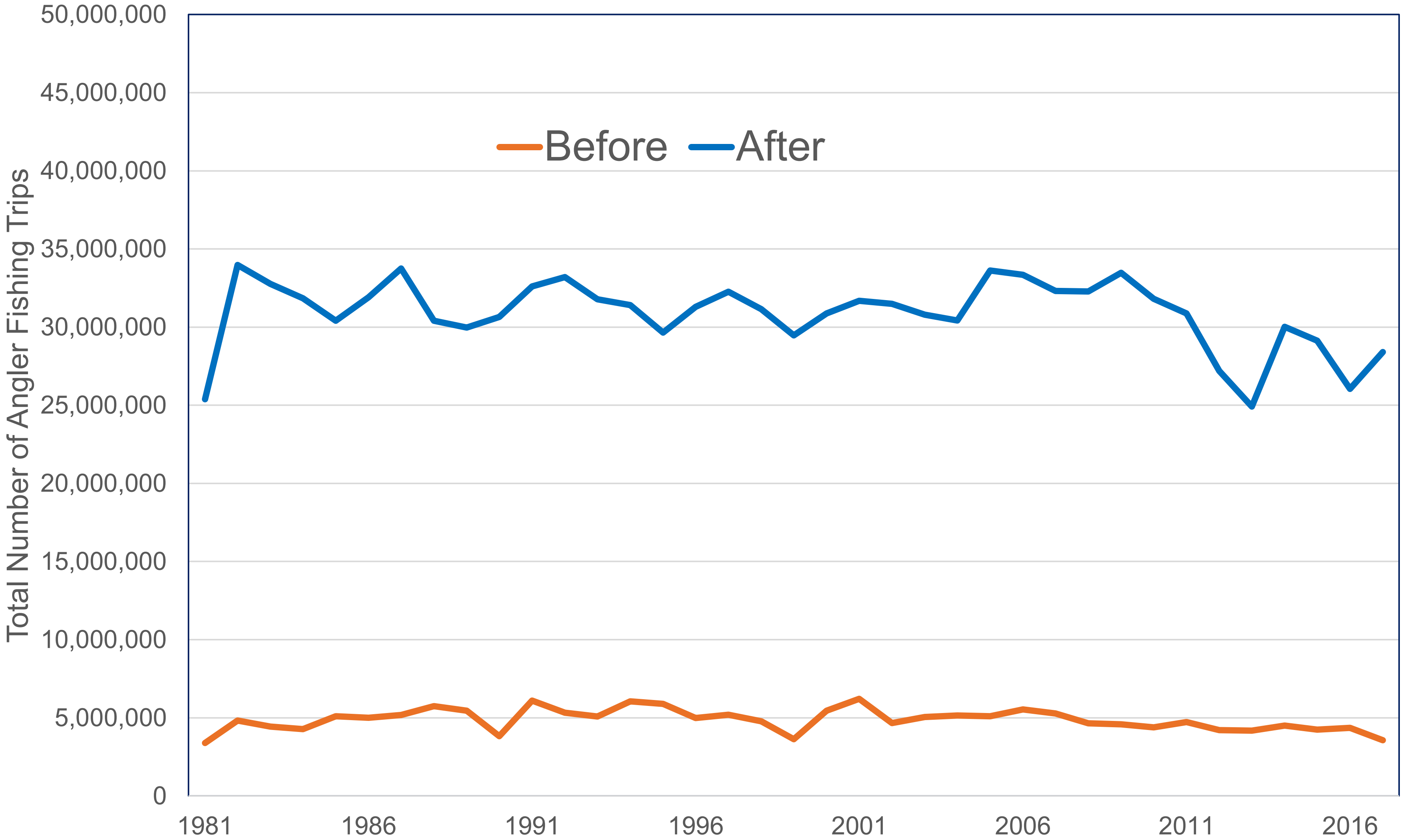
Georgia Shore Effort



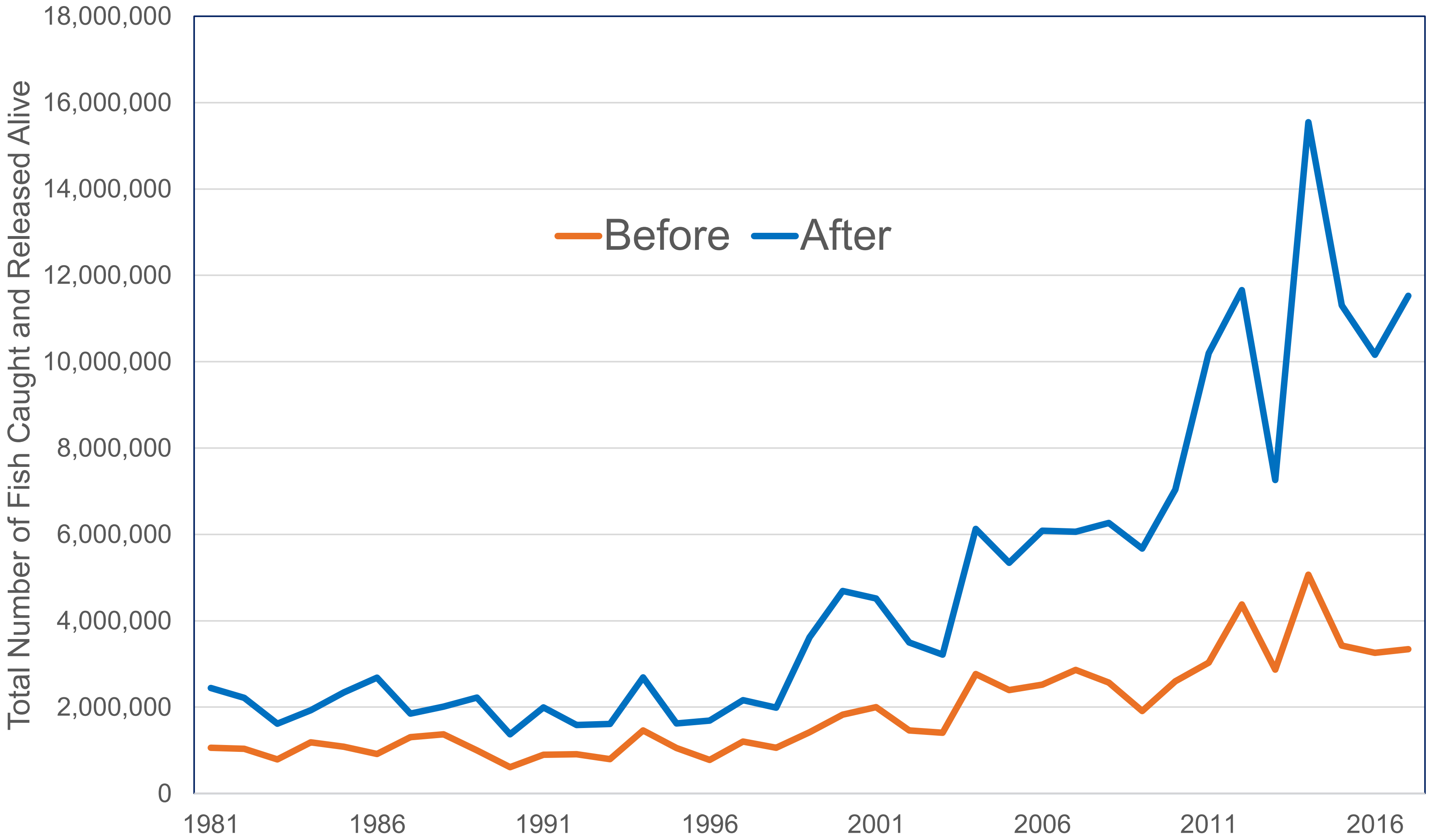
East Florida Private Boat Effort



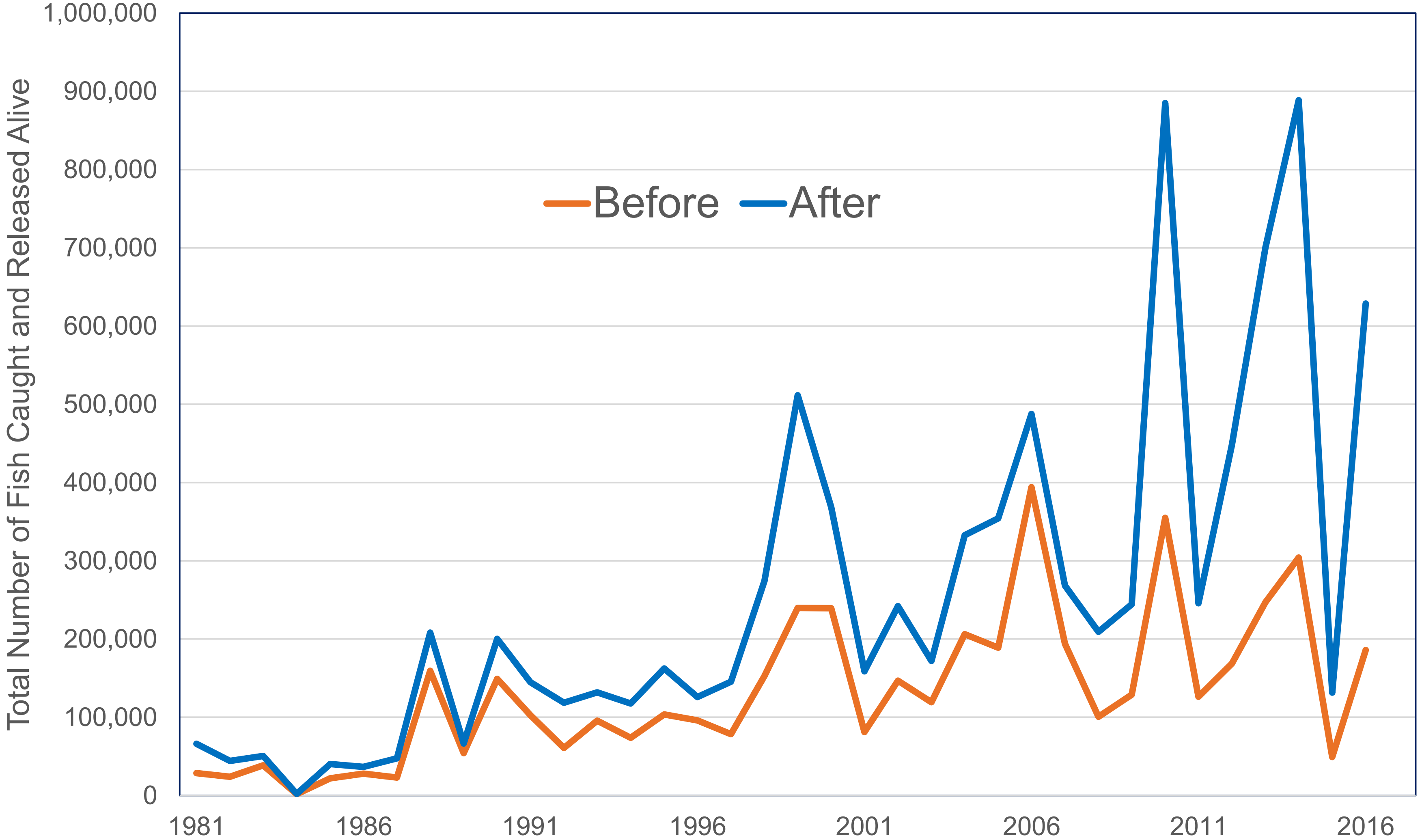
East Florida Shore Effort



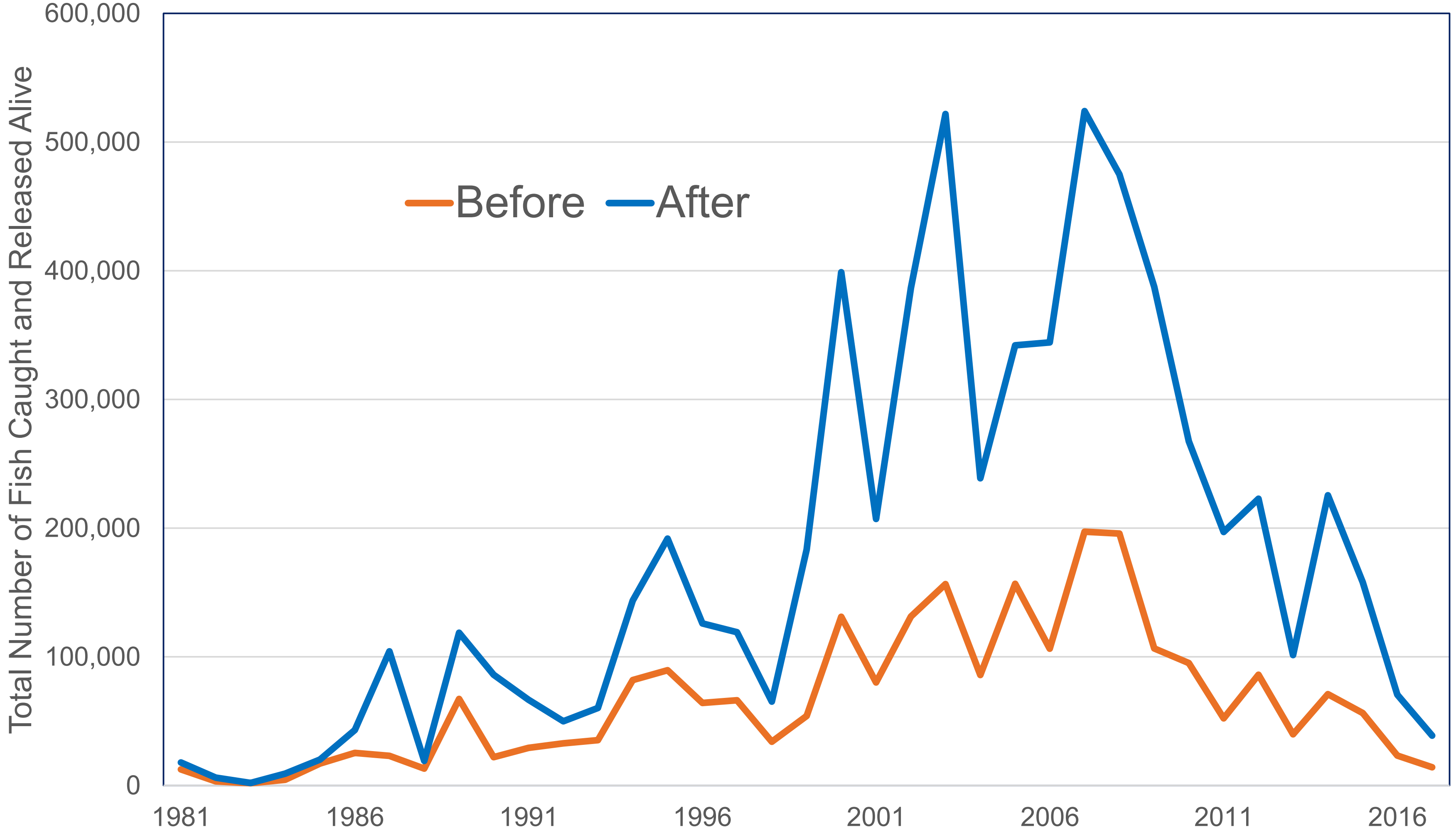
Black Sea Bass Released Catch



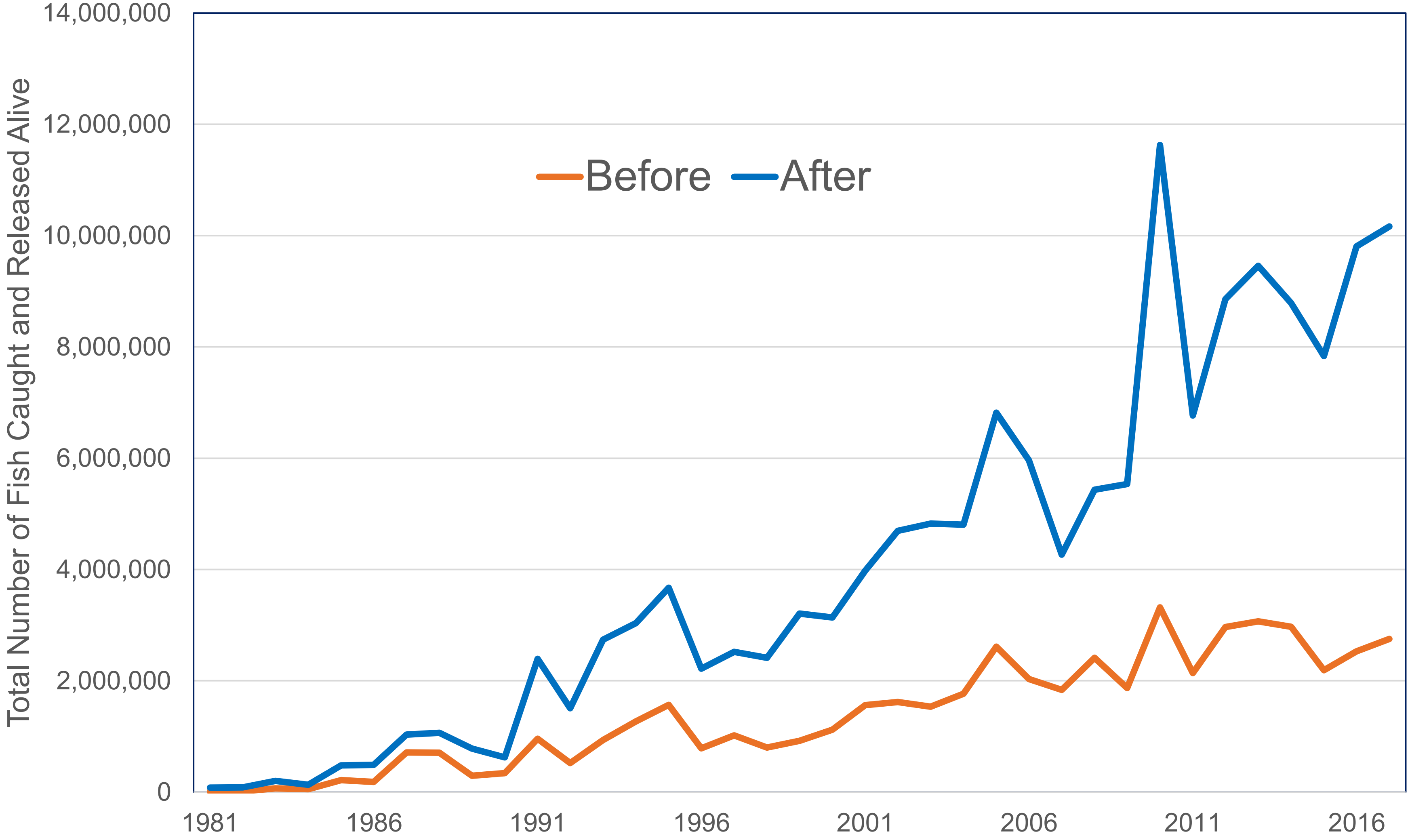
Dolphin Released Catch



Gag Released Catch



Red Drum Released Catch



Spanish Mackerel Released Catch