Documentation and Information

FES Calibration and Review

The collection of documents contained in this pdf are only a part of the resources available on the FES calibration and review workshop website. That website also includes recorded webinar presentations given during the review workshop for the calibration models. That information can be found at the address below.

https://www.fisheries.noaa.gov/event/fishing-effort-survey-calibration-model-peer-review

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DRAFT: PLEASE DO NOT CITE OR DISTRIBUTE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,

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

where the sampling errors $\{e^T_{st}\}$ and $\{e^M_{st}\}$ 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

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

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 $\{e^T_{st}\}$ are independent $\mathcal{N}(0, \sigma^2_{Tst})$ random variables, with known design variances σ^2_{Tst};
- the sampling errors $\{e^M_{st}\}$ are independent $\mathcal{N}(0, \sigma^2_{Mst})$ random variables, with known design variances σ^2_{Mst};
- 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;
- $\{e^T_{st}\}$, $\{e^M_{st}\}$ 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 $\{e^T_{st}\}$ and $\{e^M_{st}\}$ 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

$$\boldsymbol{x}'_{st} = \begin{cases} \boldsymbol{x}'_{Tst}, & \text{if no mail estimate is available;} \\ \boldsymbol{x}'_{Mst}, & \text{if no telephone estimate is available;} \\ (\boldsymbol{x}_{Tst} + \boldsymbol{x}_{Mst})'/2, & \text{otherwise.} \end{cases}$$

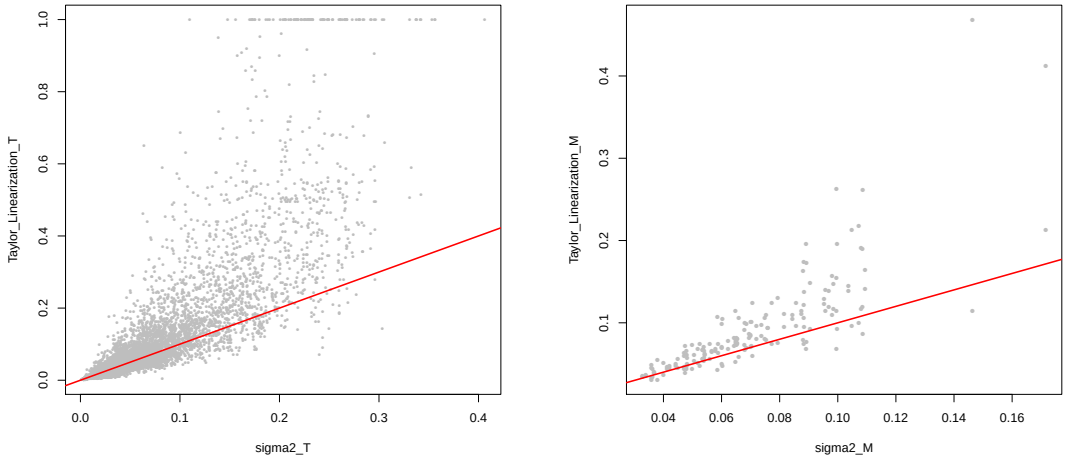


Figure 1: Estimated design variances for log-effort via Taylor linearization versus solution of the quartic polynomial equations (8) for telephone (left panel) and (9) for mail (right panel).

Then it is convenient to write

$$
\begin{aligned}
Y_{st} &= \begin{cases} \widehat{T}_{st}, & \text{if no mail estimate is available;} \\ \widehat{M}_{st}, & \text{if no telephone estimate is available;} \\ \left(\widehat{T}_{st} + \widehat{M}_{st}\right)/2, & \text{otherwise;} \end{cases} \\
&= \begin{cases} \boldsymbol{x}'_{Tst}\boldsymbol{\beta} + \nu_{st} + e^T_{st}, & \text{if no mail estimate is available;} \\ \boldsymbol{x}'_{Mst}\boldsymbol{\beta} + \nu_{st} + e^M_{st}, & \text{if no telephone estimate is available;} \\ (\boldsymbol{x}_{Tst} + \boldsymbol{x}_{Mst})'\boldsymbol{\beta}/2 + \nu_{st} + (e^T_{st} + e^M_{st})/2, & \text{otherwise;} \end{cases} \\
&= \boldsymbol{x}'_{st}\boldsymbol{\beta} + \nu_{st} + e_{st}. \qquad (10)
\end{aligned}
$$

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

$$\boldsymbol{\Sigma}(\psi) = \text{Var}\,(\boldsymbol{Y}) = \text{diag}\{\psi + D_{st}\}_{(s,t)\in\mathcal{A}}.$$

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

$$\boldsymbol{z}_{st} = [\boldsymbol{a}'_{st}, \boldsymbol{0}', \boldsymbol{0}']$$

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):

$$\widehat{\exp(\phi_{st})} = \exp\left[\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \hat{\psi}\right) + \frac{1}{2}\text{mse}\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \hat{\psi}\right)\right\}\right], \tag{19}$$

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

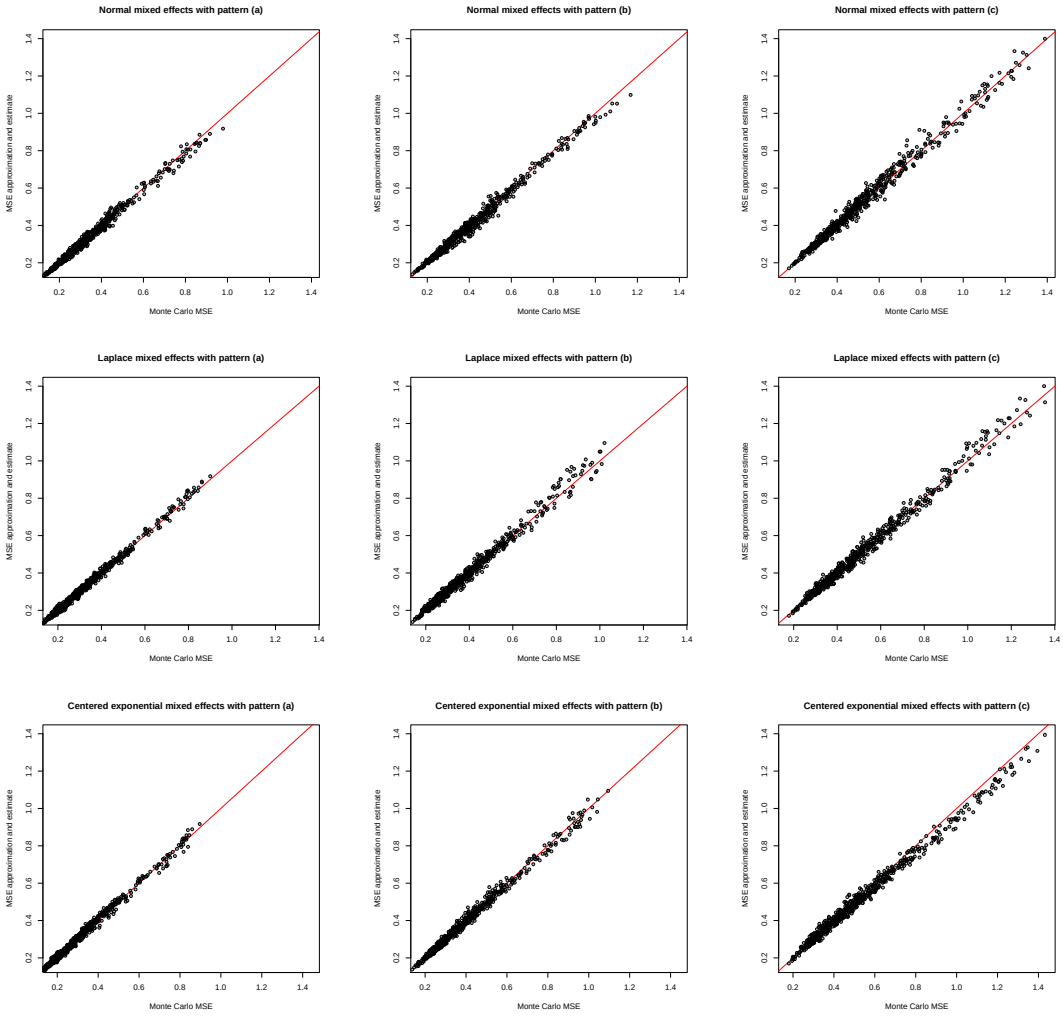


Figure 2: MSE approximation (solid gray dots) and estimated MSE's (open black circles) versus true MSE from Monte Carlo, for random effect distributions normal, Laplace, and centered exponential across the rows, and sampling error patterns (a), (b), and (c) across the columns.

$w_{st}\boldsymbol{c}'_{st}\boldsymbol{\gamma}$ 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.

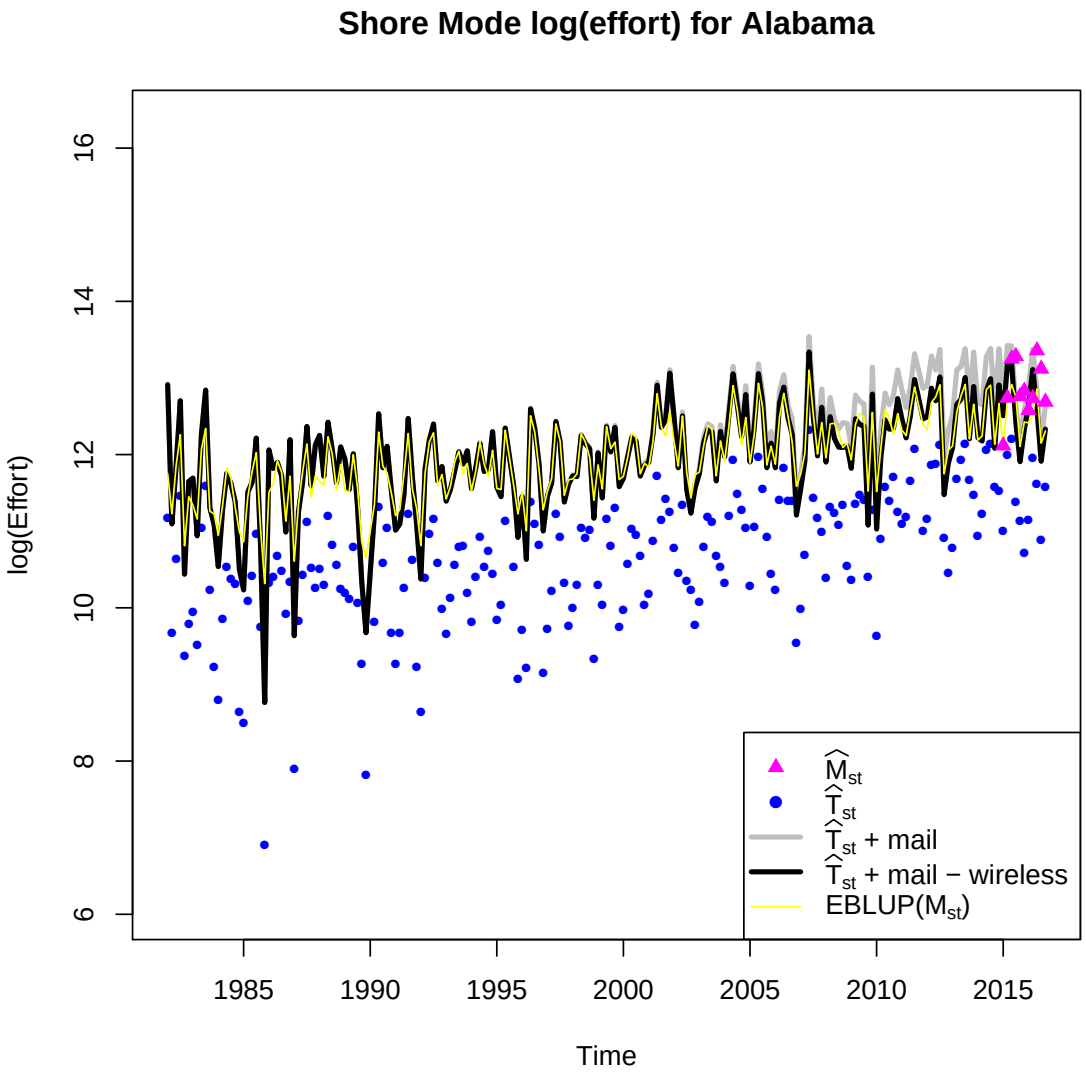


Figure 3: EBLUP's $\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \hat{\psi}\right)\right\}$ (gold curve) of mail targets $\{M_{st}\}$ for shore fishing log-effort in Alabama. Blue dots are telephone log-effort estimates $\{\widehat{T}_{st}\}$ and pink triangles are mail log-effort estimates $\{\widehat{M}_{st}\}$. 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.

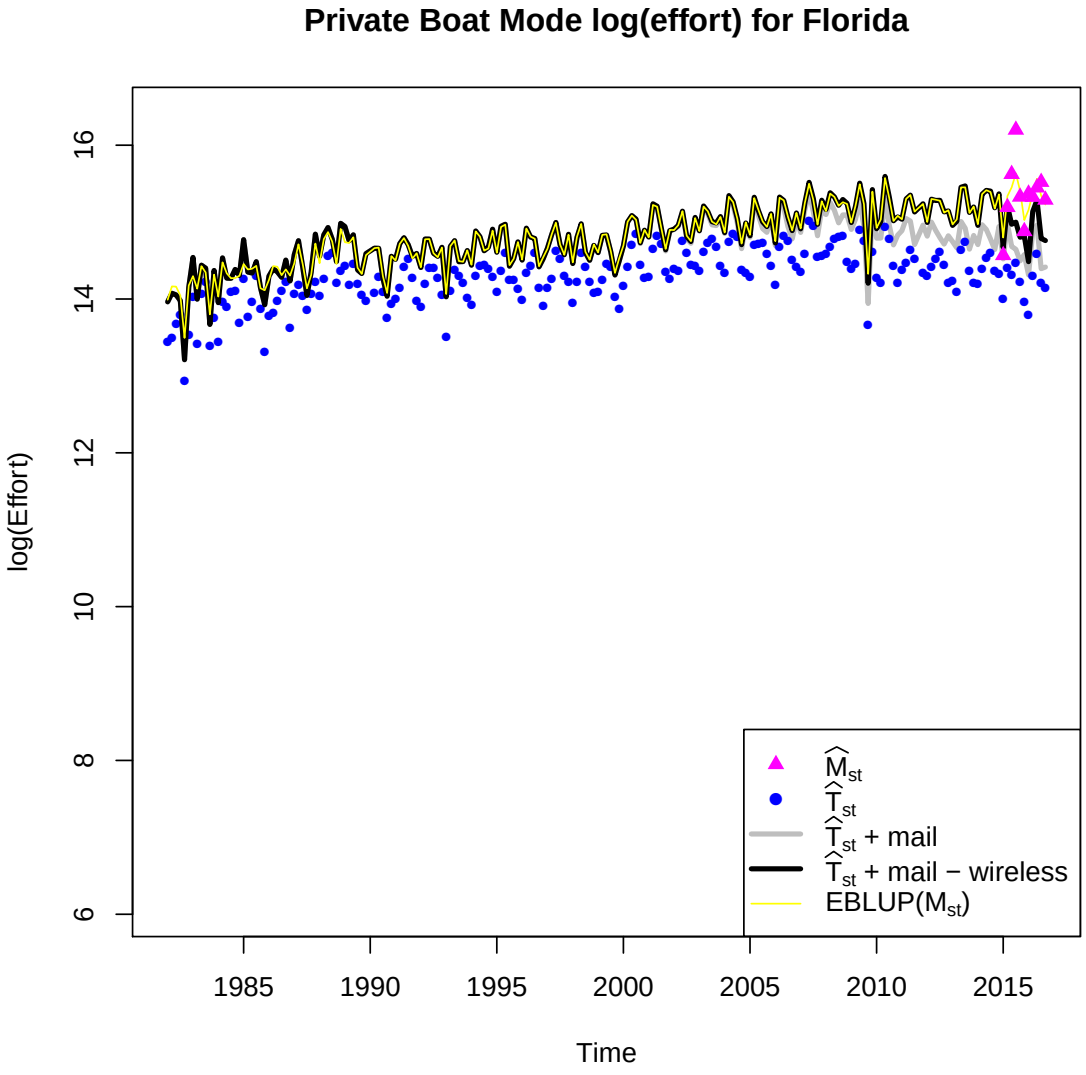


Figure 4: EBLUP's $\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \hat{\psi}\right)\right\}$ (gold curve) of mail targets $\{M_{st}\}$ for private boat fishing in Florida. Blue dots are telephone log-effort estimates $\{\widehat{T}_{st}\}$ and pink triangles are mail log-effort estimates $\{\widehat{M}_{st}\}$. 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
    - As presently configured the model is limited in terms of what can be explored but alternative calibration models 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 (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.

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,

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 ˆ and ˆ 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

$$\hat{T}_{st} = \mathbf{a}'_{st}\boldsymbol{\alpha} + \nu_{st} + e^T_{st} \text{ for } t < 2000 \tag{3}$$

$$= \mathbf{a}'_{st}\boldsymbol{\alpha} + w_{st}\mathbf{c}'_{st} + \nu_{st} + e^T_{st} \text{ for } t \quad 2000 \tag{4}$$

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

$$\mathbf{x} = \begin{bmatrix} \mathbf{a}_t & \mathbf{0}\mathbf{b}_t & \omega\mathbf{c}_t \\ \mathbf{a}_t & \frac{\mathbf{b}_t}{2} & \frac{\omega\mathbf{c}_t}{2} \\ \mathbf{a}_t & \mathbf{b}_t & \mathbf{0}\omega\mathbf{c}_t \end{bmatrix} \begin{array}{l} \text{if t} \in T^T \\ \text{if t} \in T^M \\ \text{if t} \in T^{T,M} \end{array} \tag{9}$$

while the dependent variable is

$$\mathbf{y} = \begin{bmatrix} \hat{\mathbf{T}}_t \\ \frac{\hat{\mathbf{T}}_t + \hat{\mathbf{M}}_t}{2} \\ \hat{\mathbf{M}}_t \end{bmatrix} \begin{array}{l} \text{if t} \in T^T \\ \text{if t} \in T^{T,M} \\ \text{if t} \in T^M \end{array} \tag{10}$$

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

$$\mathbf{y} = \begin{bmatrix} \hat{\mathbf{T}}_t \\ \frac{\hat{\mathbf{T}}_t + \hat{\mathbf{M}}_t}{2} \end{bmatrix}, \mathbf{x} = \begin{bmatrix} \mathbf{a}_t & \mathbf{0} & \omega\mathbf{c}_t \\ \mathbf{a}_t & \frac{\mathbf{b}_t}{2} & \frac{\omega\mathbf{c}_t}{2} \end{bmatrix} \begin{array}{l} \text{if t} \in T^T \\ \text{if t} \in T^{T,M} \end{array} \tag{11}$$

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:

$$\mathbf{y} = \begin{bmatrix} \hat{\mathbf{T}}_t \\ \hat{\mathbf{M}}_t \end{bmatrix}, \mathbf{x} = \begin{bmatrix} \mathbf{a}_t & \mathbf{0} & \omega\mathbf{c}_t \\ \mathbf{a}_t & \mathbf{b}_t & \mathbf{0} \end{bmatrix} \begin{array}{l} \text{if t} \in T^T \\ \text{if t} \in T^{T,M} \end{array} \tag{12}$$

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

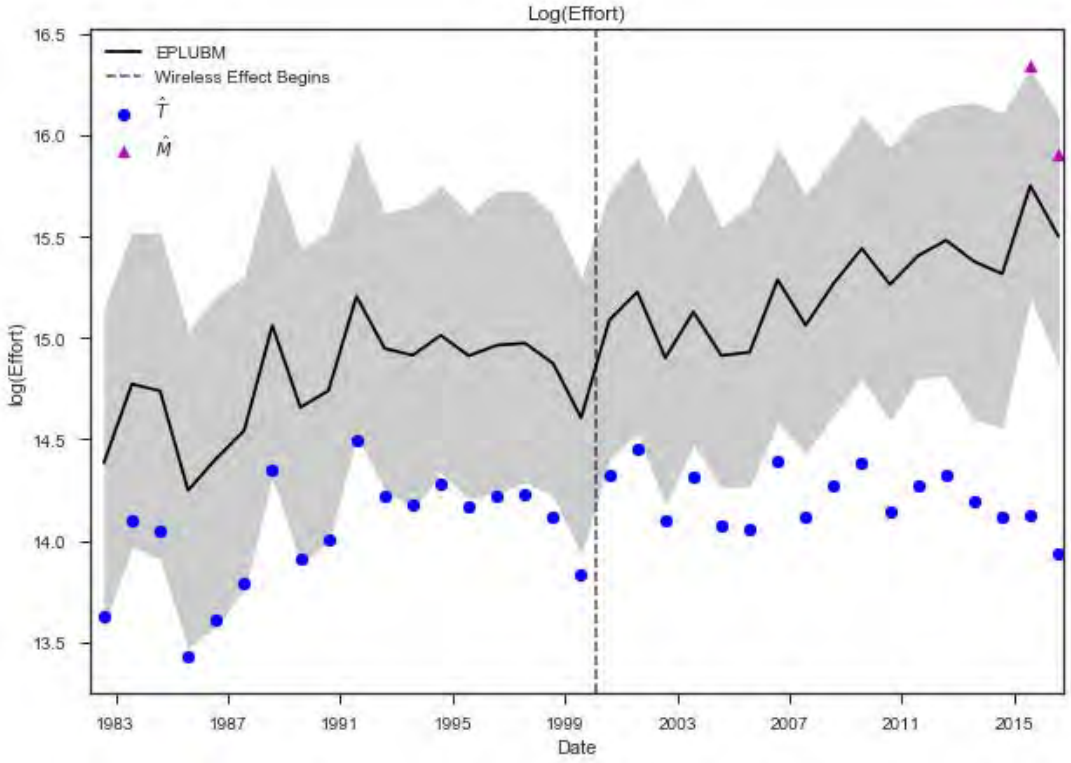


Figure 1: Florida Shore Mode Wave 4

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.

2.2Evaluation 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 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.

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.

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.

Pre-Meeting Preparations

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.

Proceedings

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:

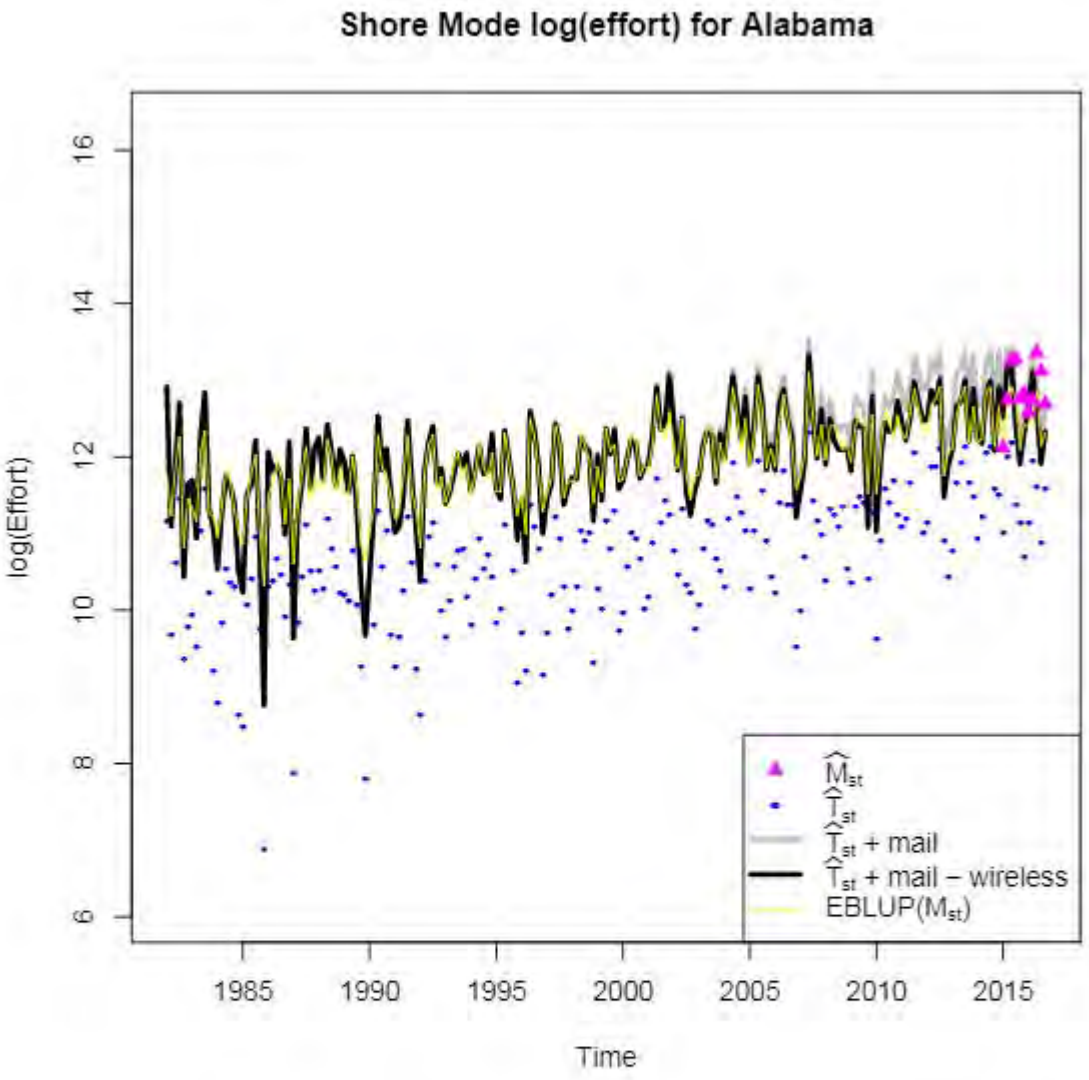


Figure 3: EBLUP's $\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \hat{\psi}\right)\right\}$ (gold curve) of mail targets $\{M_{st}\}$ for shore fishing log-effort in Alabama. Blue dots are telephone log-effort estimates $\{\widehat{T}_{st}\}$ and pink triangles are mail log-effort estimates $\{\widehat{M}_{st}\}$. 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 $\{\widehat{T}_{st} + \boldsymbol{b}'_{st}\widehat{\boldsymbol{\mu}} - w_{st}\boldsymbol{c}'_{st}\widehat{\boldsymbol{\gamma}}\}$ that adjusts for mail and wireless.

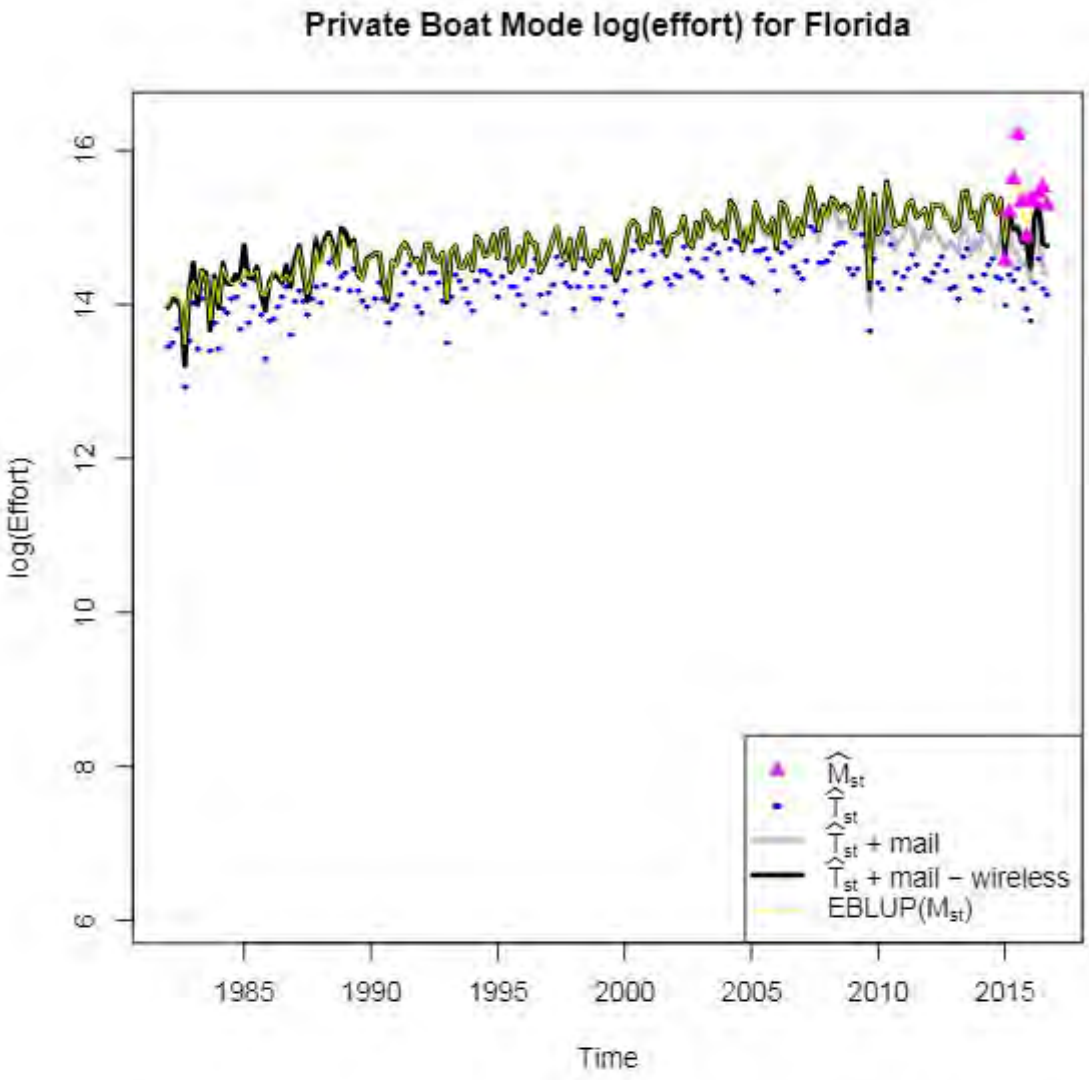


Figure 4: EBLUP's $\left\{\phi_{st}\left(\widehat{\boldsymbol{\beta}}, \hat{\psi}\right)\right\}$ (gold curve) of mail targets $\{M_{st}\}$ for private boat fishing in Florida. Blue dots are telephone log-effort estimates $\{\widehat{T}_{st}\}$ and pink triangles are mail log-effort estimates $\{\widehat{M}_{st}\}$. 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.

`

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

Eblup comparisons.docx

MRIP FES website link

MRFSS Fish Hunt Comps.xlsx

FESCALIBRATIONNOTESDay2.docx

FES Errors.pptx

EBLUPS.csv

Model_Fits.txt

EBLUPS_Variable_Names.csv

Mode_3_logeffort_poly_fixed.pdf

FESCALIBRATIONNOTESDay1.docx

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.

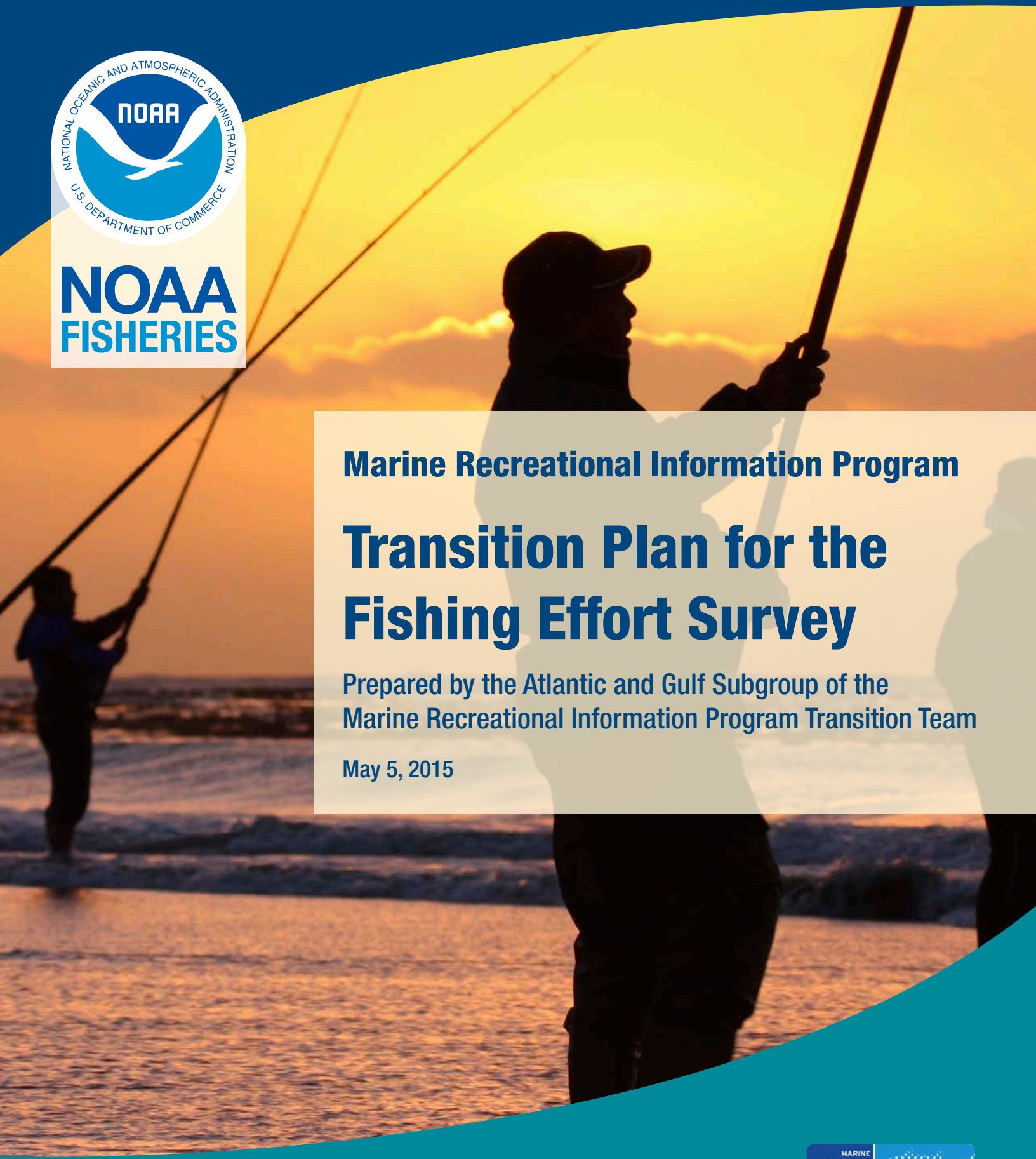
Recommendations for Future Peer Reviews

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.



NOAA FISHERIES

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

MARINE RECREATIONAL INFORMATION PROGRAM

Table of Contents

I. Executive Summary

In this plan, an Atlantic and Gulf coast Subgroup (henceforth, Subgroup) of the Marine Recreational Information Program's (MRIP) Transition Team details a comprehensive three-year timeline for transitioning from the current Coastal Household Telephone Survey (CHTS) conducted on the Atlantic coast and in the Gulf of Mexico to a new mail survey design for estimating marine recreational shore and private boat fishing effort, known as the Fishing Effort Survey (FES). All members of the Subgroup agree the timeline presented in this document is the most efficient and scientifically sound approach to implement the FES.

Since 2008, MRIP has conducted six pilot studies to determine the most accurate and efficient survey to estimate marine recreational fishing effort on the Atlantic and Gulf coasts. The most recent study, conducted in four states (Massachusetts, New York, North Carolina, and Florida) in 2012-2013, compared a new mail survey design for estimating recreational shore and private boat fishing effort with the CHTS design that has been used on the Atlantic coast and in the Gulf of Mexico since 1979. MRIP subjected the final report from the pilot project to external peer review in 2014 and certified the new survey design in February 2015 as a suitable replacement for the CHTS. The FES is much less susceptible to potential sources of bias than the CHTS. The new mail survey design can reach more anglers, achieve higher response rates, and is less prone to possible recall errors. The pilot project results indicated that FES estimates are on average 2.6 times higher than CHTS estimates for private boat fishing and 6.1 times higher for shore fishing. Because there are consistent differences in the results of the two surveys, the National Oceanic and Atmospheric Administration's (NOAA) National Marine Fisheries Service (NOAA Fisheries) recognized the FES cannot be implemented immediately as a replacement for the CHTS, and a well thought out transition plan is 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;
- Creates a replicable process for implementing new or improved scientific methods into fisheries science, stock assessment, and management;
- Supports the Recreational Fisheries Policy goals and guiding principles to foster, support, and enhance a broadly accessible and diverse array of sustainable saltwater recreational fisheries and builds stakeholder support, understanding, and engagement in implementing the new survey; and
- Advances the mission of NOAA Fisheries to ensure the sustainability of our Nation's living marine resources.

Role of Transition Team

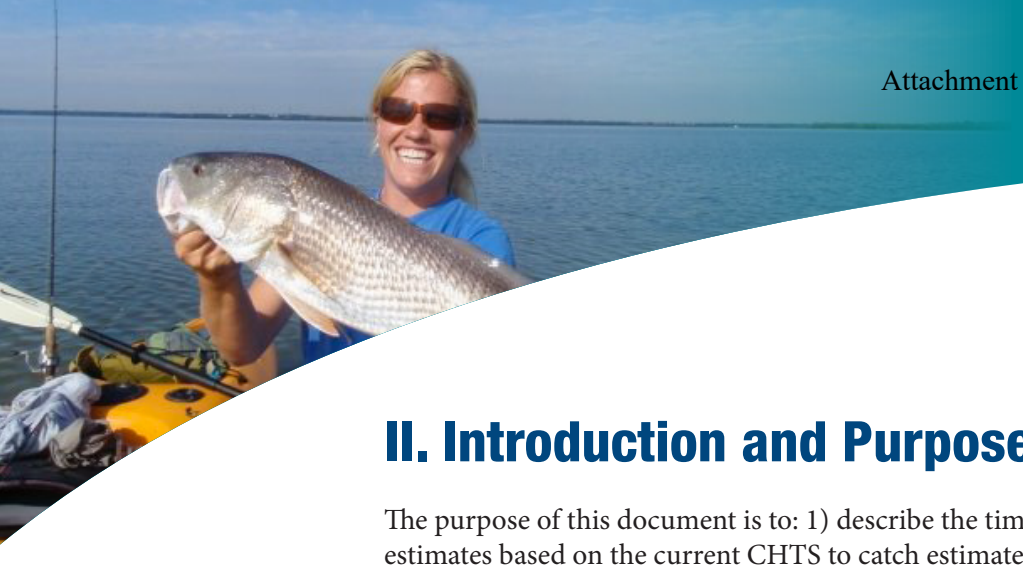
In response to recreational fishing survey design improvements and a recognized need to appropriately transition from current to new surveys, a MRIP Transition Team was formed to develop and recommend standardized processes for transitioning from historical estimates to estimates derived from improved sampling and estimation designs. The Transition Team comprises representatives from NOAA Fisheries, the regional fishery management councils, the interstate marine fisheries commissions, and several state agencies. In order for a new survey method to be implemented, historical catch statistics would first need to be converted into the same 'currency' as the new estimates; MRIP charged the Transition Team with the planning and execution of appropriate transition plans to ensure this happens. It is critical to establish processes that will enable scientists and fishery managers to make "apples to apples" comparisons between new and historical catch statistics, providing a framework that decision-makers can use for integrating new data into science and management activities at the regional and state level. The Team will play an important role in coordinating consistent approaches and methods for Councils, Interstate Commissions, and NOAA Fisheries Regions to apply to recreational catch estimates derived from new or improved survey designs for:

- Determining the status of exploited stocks;
- Setting annual catch limits;
- Monitoring catch against catch limits;
- Assessing the need for and selection of accountability measures; and
- Conducting analyses leading to the adoption of recreational fishing regulations.

In 2014, MRIP formed a Transition Team composed of representatives from NOAA Fisheries, the regional fishery management councils, the interstate marine fisheries commissions, and several state agencies to develop appropriate plans for transitioning from legacy survey designs to new, improved survey designs. A subset of the Transition Team representing the Atlantic coast and the Gulf of Mexico, or Subgroup, was formed to consider different timelines for the number of years of side-by-side benchmarking required before catch estimates based on the FES would be used for management decisions and the CHTS can be terminated. After reviewing the pros and cons of alternatives, the Subgroup recommended the three-year timeline described in this Transition Plan and approved by NOAA Fisheries leadership.

This Transition Plan outlines the necessary steps and activities needed to ensure a smooth transition to the new survey method, while taking the necessary time and effort to properly incorporate new estimates into the science and management processes. During the transition period, fishery management agencies will continue to use analyses based on the CHTS data as the "best available" science to effectively manage the health of fish stocks and marine ecosystems.

The Transition Plan calls 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 revise historical catch statistics after the second year. With this timeline, new estimates will be incorporated into stock assessments during 2017 using preliminary calibrations of historical landings, and setting of new Annual Catch Limits (ACLs) in 2018 for at least some stocks (See Appendix 1). The Plan does not allow for any extension of the benchmarking beyond three years, so the needed changes in stock assessment schedules can be set.



II. Introduction and Purpose

The purpose of this document is to: 1) describe the timeline for transitioning from catch estimates based on the current CHTS to catch estimates based on the new FES; and 2) stress the importance of a planned step-by-step process for transitioning from the use of catch statistics produced by a legacy survey design to the use of catch statistics produced by a new, improved survey design.

Introduction

MRIP has been developing, testing, and evaluating ways to improve the survey designs used to monitor fishing effort and catch in marine recreational fisheries. The goal has been to provide new methods that are less prone to possible sources of bias and can be adapted for use with increased sampling to provide greater statistical precision in estimates of recreational catches, as well as desired levels of temporal and geographic resolution as additional resources become available.

Immediate implementation of any new survey design will most likely cause a disruption to fishery management processes for at least some stocks because cumulative catch estimates based on the new design may not be comparable to the current Annual Catch Limits (ACLs). Fish stock assessments and the ACLs set based on them rely heavily on accurate time series of both commercial and recreational fishery catch statistics. The statistics provided for recreational fisheries must be comparable across the time series to ensure accurate accounting of fishing mortality each year and accurate monitoring of year-to-year trends in the fishery. The stock assessments we have used to set current ACLs have incorporated the time series of recreational fishery catch statistics produced by our legacy survey designs, which include the CHTS.

Because new survey designs are likely to produce consistently different statistical estimates than the legacy designs they replace, we should expect that catch estimates based on new designs will not immediately be the "best available" for use in making fishery management decisions. It will be necessary to continue use of the legacy design for catch estimates until continuity is established with data sets generated from the new survey design.

Role of Transition Team

In response to recreational fishing survey design improvements and a recognized need to appropriately transition from current to new surveys, a MRIP Transition Team was formed to develop and recommend standardized processes for transitioning from historical estimates to estimates derived from improved sampling and estimation designs. The Transition Team comprises representatives from NOAA Fisheries, the

regional fishery management councils, the interstate marine fisheries commissions, and several state agencies. In order for a new survey method to be implemented, historical catch statistics would first need to be converted into the same 'currency' as the new estimates; MRIP charged the Transition Team with the planning and execution of appropriate transition plans to ensure this happens. It is critical to establish processes that will enable scientists and fishery managers to make "apples to apples" comparisons between new and historical catch statistics, providing a framework that decision-makers can use for integrating new data into science and management activities at the regional and state level. The Team will play an important role in coordinating consistent approaches and methods for Councils, Interstate Commissions, and NOAA Fisheries Regions to apply to recreational catch estimates derived from new or improved survey designs for:

- Determining the status of exploited stocks;
- Setting annual catch limits;
- Monitoring catch against catch limits;
- Assessing the need for and selection of accountability measures; and
- Conducting analyses leading to the adoption of recreational fishing regulations.

General Transition Approach

There are several steps that must be taken before estimates based on any new design can be used effectively in the management process.

1. **Benchmarking:** The newly designed survey should be conducted side-by-side with the legacy survey to allow measurement and evaluation of consistent differences in the statistical estimates produced. During this benchmarking period, statistical estimates produced by the legacy design are the "best available" for use in monitoring catches relative to ACLs and making management decisions.
2. **Calibration model development:** Consistent differences between new design and legacy design estimates should be evaluated to determine possible sources of bias in the legacy design to explain those differences. In addition, literature research should be conducted to assess how biases identified in the legacy design would most likely have changed over time. Based on the information gained, one or more calibration models should be developed and evaluated 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 an external peer review.
3. **Re-estimation of historical catch statistics:** Once a calibration model has been proposed, peer reviewed, and approved, the model should be used to generate a corrected time series of recreational catch statistics. The revised time series should immediately be made available to stock assessment scientists and fishery managers.
4. **Incorporation of new estimates into stock assessments:** The revised catch statistics should be incorporated into stock assessments as soon as possible to provide the most accurate assessments of stock status and provide new ACLs for use in fisheries management. Stocks with very substantial mortality levels due to recreational fishing (high proportion of total mortality relative to that caused by commercial fishing)

should be identified as "key stocks" and prioritized for assessment scheduling. Depending on the magnitude of the estimation changes and potential disruption of the management process, assessments scheduled for key stocks may have to be moved to earlier dates while those scheduled for non-key stocks are moved to later dates.

5. **Incorporation of new estimates and ACLs into management actions:** Once revised catch statistics and new assessment results become available, management should begin to use both for decision making as soon as possible. If revised statistics are available but new assessments are not, then managers may need to continue using the statistics based on the legacy design until new assessment results are available. In years when the legacy design is no longer being conducted, the approved calibration model would be used to convert catch estimates based on the new design into estimates that are compatible with the legacy design for use in management.

New Fishing Effort Survey

In 2012, MRIP conducted a pilot study in four states that compared a new mail survey design for estimating recreational shore and private boat fishing effort with the CHTS design that has been used on the Atlantic coast and in the Gulf of Mexico since 1979. MRIP subjected the final report from the pilot project to external peer review in 2014 and certified the new survey design in February 2015 as a suitable replacement for the CHTS. MRIP has named the new survey the Fishing Effort Survey (FES), and it is much less susceptible to potential sources of bias than the CHTS. It can reach more anglers, can achieve higher response rates, and is less prone to possible recall errors.

The FES was tested in Massachusetts, New York, North Carolina, and Florida to estimate fishing effort in September 2012 through December 2013. The design is a single phase, dual-frame, self-administered mail survey. The two frames used for sampling coastal state residents are: 1) the U.S. Postal Service (USPS) Delivery Sequence File, which includes all residential addresses serviced by the USPS; and 2) the list of mailing addresses provided by licensed or registered anglers that is derived from the MRIP National Saltwater Angler Registry (NSAR). To improve survey efficiency, the design matches samples of USPS addresses to the list of NSAR addresses. Matching addresses are sampled at a higher rate and the resultant data are appropriately weighted. The pilot project results indicated the mail survey estimates are on average 2.6 times higher than CHTS estimates for private boat fishing and 6.1 times higher for shore fishing. More detailed information comparing the FES to the CHTS can be found at: http://www.st.nmfs.noaa.gov/Assets/recreational/pdf/2012-FES_w_review_and_comments_FINAL.pdf.

Need for FES Transition Plan

Because there are consistent differences in the results of the two surveys, NOAA Fisheries determined that the FES should not be implemented immediately as a replacement for the CHTS. A well thought out Transition Plan is essential to ensure the FES is appropriately phased in with minimum disruptions to stock assessment and fisheries management processes. Research studies should continue in parallel with this transition process to better understand and explain differences between the simultaneous estimates produced by the FES and the CHTS. Stakeholders will want to know why catch estimates are being

revised and will need a clear explanation of why the new numbers are more accurate than the ones replaced.

To develop a Transition Plan for implementation of the FES on the Atlantic and Gulf of Mexico (Gulf) coasts, the Subgroup met weekly, discussing all potential options of the best approach for a smooth transition and incorporation of new estimates into the stock assessment and management processes. This plan provides a description of the potential methods to be used to:

- Compare legacy estimates to estimates produced by using the new FES in a statistically robust manner;
- Determine when calibration or other means of linking legacy data sets with the new FES estimates is feasible and necessary, and identify the requirements and methods for making such linkages; and
- Minimize disruptions to stock assessments, catch monitoring, and management regulations, and facilitate decisions on when and how implementation of the FES is introduced.

Transition Planning and Best Scientific Information Available

The Magnuson-Stevens Fisheries Conservation and Management Act is the principal law governing marine fisheries in the U.S., and it includes ten National Standards to guide fishery conservation and management. One of these standards, referred to as National Standard 2, guides scientific integrity and states that "conservation and management measures shall be based upon the best scientific information available." The Magnuson-Stevens Fishery Conservation and Management Reauthorization Act of 2007 added provisions to improve the use of science in decision-making. One of the revisions specified that the Secretary of Commerce and councils must establish a peer review process for scientific information used to advise councils on the conservation and management of fisheries.

Catch estimates based on the new FES design will only be the "best available" for management use after historical catch estimates have been appropriately adjusted to the new design and incorporated into stock assessments and the setting of management measures. Until historic catch date is adjusted to be compatible with the FES and results incorporated into ACLs and other management reference points, estimates based on the CHTS will continue to be the "best available."

III. Fishing Effort Survey Transition Timeline

The Subgroup determined a three-year timeline was needed for the transition from the CHTS to the FES. The timeline covers the benchmarking and calibration of historical catch estimates and includes some detail on the stock assessment and management processes that must follow. However, more information will be added as decisions are made on the scheduling of those processes by the appropriate fishery management agencies. This Transition Plan for the FES is a living document and will be updated as needed.

With this approach, the FES will run side-by-side with the CHTS from 2015 to 2017, with full use of FES estimates and termination of the CHTS no earlier than 2018. For the first two years of side-by-side benchmarking, NOAA Fisheries scientists will work to develop a model for calibrating the two sets of estimates. In 2017, calibrated historical time series estimates will be used for incorporation into stock assessments and setting of ACLs for key stocks in 2018 and beyond.

FES Transition Timeline

- **February 2015: Benchmarking (Side-by-side conduct of FES and CHTS)** began on February 20, 2015, and will continue for three full years (2015-2017).

- **June 2015—late-December 2015:** NOAA Fisheries/Science and Technology (ST) will work with the Northeast Fisheries Science Center (NEFSC) and Southeast Fisheries Science Center (SEFSC) to test the possible effects of different calibration scenarios on assessments for a couple of key stocks in each region.
  - ST staff will provide two or three simple alternative models for hind-casting how differences between CHTS and FES estimates may have changed between 1981 and the present. These models will be based on very simple assumptions of how several factors causing differences either stayed the same or changed.
  - NEFSC and SEFSC staff will use these simple calibration models in combination with different assumptions made regarding the magnitude of current differences between CHTS and FES estimates at the subregional level to create a number of revised time series of catches for the selected stocks. The effects of incorporating different revised time series into assessments will then be tested to get some idea of the potential range of possible outcomes.
  - A common basis for the testing will be essential to compare possible effects of the different assumed calibrations across regions and stocks.

- **September 2015—early 2017: Calibration model development.**
  - NOAA Fisheries/ST staff will begin developing an appropriate calibration model for re-estimating recreational catch statistics.

- **October 2015: CHECKPOINT 1—**The Transition Team and NOAA Fisheries will review preliminary side-by-side estimates for January-August.
  - NOAA Fisheries/ST will draft a report comparing effort estimates among states for January-August, as well as comparing the 2015 estimates for Massachusetts, New York, North Carolina, and Florida to estimates from the 2012/2013 FES pilot study to assess inter/intra-state variation and magnitude of difference between the CHTS and FES.
  - The Transition Team and NOAA Fisheries will use this **CHECKPOINT** as an initial chance to look at the potential impact of the new survey methodology and to start planning accordingly.

- **January 2016: Side-by-side testing of the CHTS and FES continues.**

- **mid-April 2016: CHECKPOINT 2—**FINAL estimates from the 2015 side-by-side testing of the CHTS and FES will be available.
  - The Transition Team and NOAA Fisheries will review the final estimates from 2015 and continue to assess the potential impacts of the new estimates and prepare accordingly.
  - At this point, there will be two years of July-December FES data for Massachusetts, New York, North Carolina, and Florida (the four 2012/2013 FES pilot study states) for comparison.

- **May—June 2016:** A **Progress Report** will be developed to provide an update on the status of the transition, review of final 2015 effort estimate comparisons of the FES and CHTS, and review any issues that may have arisen during the first year of benchmarking.

- **May 2016—April 2017: Management and stock assessment preparations** will be made to anticipate calibration in 2017.

- **late-January—March 2017: Peer review of the calibration model.**
  - The external peer review of the calibration model will take an additional 2-3 months to complete.
  - The peer review will occur separate of a data review workshop and address only the model itself, not the application. The application of the calibration model will be reviewed at one or more data review workshops conducted for planned stock assessments (updates or benchmarks).
  - Only after the model has been accepted can the calibration be applied to revise the time series of catch statistics (if approved in July, could have revised catch statistics as early as September).

- **mid-April 2017: READY FOR CALIBRATION—**FINAL estimates from the 2015-2016 side-by-side testing of the CHTS and FES will be available.
  - The Transition Team and NOAA Fisheries will review the final estimates from 2015-2016 and move forward with calibrating and incorporating new estimates into stock assessments.
  - Side-by-side benchmarking will be ongoing through the remainder of 2017.
  - At this point, there will be three years of July-December FES data for Massachusetts, New York, North Carolina, and Florida for comparison and two full years from all other Atlantic and Gulf coast states.

- **May—October 2017: Re-estimation of historical catch.**
  - NOAA Fisheries/ST will begin using the approved calibration model to re-estimate the historical time series of catch estimates.

- **May—June 2017:** A **Progress Report** will be developed to provide an update on the status of the transition, review of final 2016 effort estimate comparisons of the FES and CHTS, and review any issues that may have arisen during the second year of benchmarking.

- **June 2017: Data review workshop.**
  - A review of the new effort and catch estimates will occur prior to incorporating them into stock assessments.
  - The goal is to have an all-inclusive review, including representatives from all affected Scientific and Statistical Committees (SSCs).

- **July—November 2017: Incorporation of new estimates into stock assessments.**
  - The new calibrated catch estimates will be incorporated into key stock assessments.
  - The Subgroup ranked all stocks according to both percentage of recreational catch and regional importance. Only a selected number of the top-ranked stocks will be considered "key" stocks.

- Based on staff availability and resources, as many key stocks as possible will be re-assessed at this time.
- Assessment updates, where possible, will: 1) not take into account any other new factors; 2) not review the FES itself; and 3) produce new results as expeditiously as possible.
- The purpose is to adequately prepare for what could potentially be very large changes in assessment results, even in terms of status determinations and rebuilding rates.
- 'Fine-tuning' of the calibration model and historical catch statistics can be done after the third year of benchmarking and data are available mid-2018.

➢ **November 2017: Begin incorporation of new estimates and ACLs into management actions.**

- Assessment updates based on data collected in years 1 and 2 will be used to set ACLs for 2018 and beyond in FES currency.
- Projections of 2018+ catches will be based on revised historical catch statistics.
- FES-based catch statistics will be used for monitoring of catches in 2018 and beyond.
- FES-based catch statistics would be converted into CHTS-based statistics for use in management of stocks for which the 2018+ ACLs are based on older assessments that do not incorporate FES data. Such statistics would be estimated by applying a reverse calibration to FES estimates.

➢ **January 2018: Terminate side-by-side testing** after final November-December 2017 data collection.

➢ **May 2018—October 2018: Fine-tune the historical time series** based on a revised calibration that uses three years of side-by-side benchmarking comparisons after final estimates from 2017 become available.

➢ **June—October 2018: Continue updating stocks that were not assessed in 2017** based on rankings.

This three-year timeline is needed to: 1) ensure development of a stable calibration that can account for interannual variability in the differences between FES and CHTS estimates; 2) allow time needed to prepare for stock assessment and management processes that will need to be accelerated to complete the incorporation of revised historical catch statistics within the desired time frame; and 3) allow further research needed to more fully explain the differences between the FES and CHTS effort estimates.

IV. Potential Stock Assessment Impacts and Schedule

As mentioned in the previous section, starting mid-2015, staff at the NEFSC and SEFSC will begin preliminary evaluations of the effects of hypothetical FES:CHTS calibration scenarios on the outcomes of the assessments for several key stocks in each region. This exercise will help identify any technical complications that may arise as well as the potential implications for management advice. This in turn will help inform the rate at which the remaining assessments can be updated and the level of review that may be warranted.

The Subgroup suggested using relatively simple FES:CHTS calibration scenarios that are based on inferred temporal changes in coverage by the CHTS and different magnitudes of current estimation differences based on the 2012/2013 FES pilot study. They also pointed out that the time series of adjusted catches should be extended as far back in time as possible, as there could be significantly different effects on the stock assessments if the CHTS effectiveness was constant or trended over time (e.g., owing to increased cell phone usage in recent years).

Once the results of these preliminary analyses are complete, it will be necessary to decide whether to do as many assessments as resources allow all at once, or to spread the assessments out with the normal schedule. It is likely that a sequential implementation of the revised data for the assessments would create some difficulties for managers because different species would be regulated based on the new and old estimates at the same time. The timeline and workload will need to account for both state and federally managed species and may differ among regions.

Overall, the costs and timing of revised stock assessments will depend on 1) the magnitude of changes and complexity of developing a temporally varying calibration model; 2) technical details of incorporating revised estimates into stock assessment models; 3) availability of resources to focus only on this effort; 4) the number of species to be assessed; and 5) the regional review processes. It is likely that not all stocks may be updated easily and there is the need to prioritize assessments (Appendix 2) and that, depending on the complexity, it may take two years or longer to run assessments. Additionally, there may be a regional split in the timeline due to differing assessment processes in each region.

At the earliest, revised catch statistics would be ready for use in stock assessments by October of the second year of side-by-side benchmarking (2017). If priority stocks are to be assessed first, the impacts of a gradual assessment schedule must be clearly

communicated. Some stocks will be managed under the lower estimates (possibly with retrospective adjustments) while others will be at the higher, new FES estimates. The Subgroup has developed preliminary recommendations for stock assessment priorities. The priority rankings combine input from the management and science sub-groups and are intended to be used for future planning by the fishery management bodies and regional stock assessment processes. A table of key stocks with initial priority rankings can be found in Appendix 2.

V. Potential Management Impacts and Schedule

The potential management impacts in the short term and long term are likely to be quite substantial given current management schedules for both federal and commission managed stocks with a recreational fishing component. Given the potential scheduling issues and increased staff workload, the Subgroup concluded the stocks with the largest recreational catch component should be done first, followed by other identified stocks based on their ranking and recommends addressing as many stocks as possible. Additionally, developing management schedules will be difficult and may vary within and among regions, with some stocks being more heavily impacted than others.

The user group and public perceptions of MRIP and any changes that may result from the revised recreational fishing effort estimates will be substantial, with some constituents having increased negative feelings regarding additional changes (i.e., the perception already exists among some that current MRIP estimates are already unrealistically high). It cannot be easily determined which is more likely, the pressure to incorporate estimates sooner than later or accepting that taking longer would ensure a more stable calibration. A clear communications strategy is vital to the success of the planned transition and must be incorporated. A broad communications strategy is provided in Appendix 3.

With the proposed schedule, the time for making decisions on 1) setting acceptable biological catches (ABC) and ACLs for stocks affected by the transition and 2) making potential allocation adjustments will need to be much shorter/compressed than the typical process. The transition timeline for management is abbreviated and optimistic (see Appendix 1), i.e., there is no room for delay or to work through unknown issues that may arise. For Fishery Management Councils it takes approximately one year to implement ACLs based on new assessment results; however, there have been a few occasions when this has occurred mid-year. For Commissions, their management cycle could allow for almost immediate change to their stocks' quotas and they have the ability to modify mid-year.

A critical issue that must be anticipated is even if an assessment with newer higher recreational data produces higher ABCs, in an allocated fishery part of the increase will go to the recreational sector and part of the increase will go to the commercial sector. Thus, without allocation adjustments, substantial recreational restrictions might be triggered even if ABCs get substantially increased. There could be stock assessments that trigger allocation reviews and other assessments that do not, depending on the nature of existing allocations and the results of new assessments. There was also concern on how unassessed stocks would be affected. The Subgroup's best

estimate for the quickest turn-around for completing allocation reviews that Councils consider urgent is 1-2 years, and a minimum of one year for Commissions for an amendment and six months for a Commission addendum.

If revised estimates result in stock status determinations being changed to "overfishing" or "overfished," workload will increase for NOAA Fisheries, Councils, and Commissions to develop and react to rebuilding plans and requirements to end overfishing.

As the primary transition year, 2018, approaches, fishery managers will need to evaluate if ACLs can be set for each stock by using FES-compatible updated assessment information. If so, then MRIP catch estimates based on 2018 FES effort estimates will be used to account for the ACLs and to determine if accountability measures are triggered.

For stocks that do not have updated FES-compatible assessments and ACLs in 2018, accounting for catch and management of ACLs and AMs will be done by using reverse-calibration of the FES-based MRIP catch estimates, so that the catch data used for management is consistent with the legacy catch data used to set the ACLs.

VI. Identification of Unknowns

As the transition proceeds, the Subgroup identified several unknowns that will be important to monitor due to the potential effects they may have on planned schedules.

1. Developing the calibration model could prove to be more challenging than expected, with the various complexities of coverage differences (e.g., decreased landline households), response rates declining in phone surveys, measurement error in the proportion of households reporting fishing, etc.
    a. This could impact the timeline for development of the model, currently planned to start September 2015 and continue into early 2017.
    b. There will be future work with statistical consultants to update the group on these issues once more is known.
2. However, with future budgets unknown, the decision to stop side-by-side testing may be forced. Currently, the CHTS costs approximately $1.8M per year and the FES is estimated to cost $1.3M per year during the benchmarking period.
3. There could potentially be a legal challenge to either incorporate estimates sooner or to delay in order to get better precision; e.g., the Red Snapper court decision (Guindon v. Pritzker; March 26, 2014).
4. Congressional direction may mandate either incorporating estimates sooner or a delay to get better precision.

After the first and second years of benchmarking, any issues that arose will be outlined in a progress report.

VII. Further Experiments

It is necessary to understand what factors are causing the large differences in estimates between the CHTS and the FES found in the pilot study. Starting in 2015, NOAA Fisheries will continue to investigate what potential causes could have affected recreational fishing effort estimates. Any additional studies could occur concurrently with the benchmarking and calibration. To the maximum extent feasible, NOAA Fisheries will need to be able to explain why the FES estimates are higher than the CHTS, as well as why they are likely to be more accurate. Working toward better understanding of possible sources of bias in both the CHTS and FES is necessary to ensure that NOAA Fisheries is providing the most accurate and best available science for estimating recreational fishing catch and effort.

VIII. Lessons Learned from Improved Catch Survey Implementation

The Subgroup recognizes there is an opportunity to learn from prior MRIP calibration efforts that were deemed necessary to account for consistent changes in estimates that have resulted from implementing other survey design improvements. It will be important to monitor progress in the development and application of these other calibration approaches to determine how best to move forward with developing and applying a calibration of the CHTS to the FES.

In 2012, an MRIP Calibration Workshop was held to determine the most appropriate way to account for any consistent changes in 2004-2011 catch statistics that resulted from the implementation of an improved estimation method for the Access Point Angler Intercept Survey (APAIS) conducted on the Atlantic and Gulf coasts. Workshop participants included representatives from NOAA Fisheries, the regional fishery management councils, the interstate marine fisheries commissions, and several state agencies. Participants concluded a simple ratio calibration approach based on the 2004-2011 comparisons was appropriate to use for re-estimating catches in earlier years (1981-2003). In this case, eight years of side-by-side estimates were available for benchmarking, and the calibration model was based on the average annual new:old ratios. Since then, this ratio calibration has been applied to update recreational catch statistics for all stocks prior to incorporation into stock assessments.

It may be useful to conduct a study of the sensitivity of this first APAIS ratio calibration to the number of years of side-by-side comparisons used. This may help to evaluate the importance of accounting for interannual variability when calibrating between alternative estimators. The extent to which the ratio calibration changes as more years of side-by-side estimates are added to the calculation of mean ratios for a number of recreationally important stocks could be examined. For any given stock, it is very likely that changes in the mean ratio will tend to decrease as more years are added to the analysis, leading to greater stability of the calibration. This proposed study could be instructive in determining how to go about evaluating the number of benchmarking years needed to provide a relatively stable calibration between the FES and the CHTS.

In 2014, a Calibration Workshop was held to evaluate the potential consistent effects of implementing a new sampling design for the Access Point Angler Intercept Survey (APAIS) on the Atlantic and Gulf coasts in 2013. Workshop participants included three expert statistical consultants and representatives from NOAA Fisheries, the regional fishery management councils, the interstate marine fisheries commissions, and several state agencies. The participants determined that analyses conducted by the NOAA

Fisheries Office of Science and Technology showed there was sufficient evidence that the more complete temporal coverage of the new design resulted in consistent increases or decreases in APAIS angler catch rate statistics for at least some species. They developed three different calibration models to evaluate for possible use in correcting the pre-2013 legacy catch statistics. In this case, no side-by-side benchmarking comparisons could be made. However, the statistical consultants concluded the simplest of the three proposed models was appropriate for use in the short term until more data collected with the new APAIS design could be used to complete evaluation of the other two proposed calibration models. The simple ratio calibration approach has been used to revise historical catch statistics and incorporate them into stock assessment updates for Gulf Red Snapper, Gulf Red Grouper, and other key stocks in 2015. Once an evaluation of the other two proposed calibration models has been completed, one of the three methods will be selected as the best for use in re-estimation of historical catches and incorporation of new estimates into stock assessments and management.

One important lesson learned in the 2014 APAIS design change calibration effort was that the development of an appropriate calibration would have been much simpler if data from a side-by-side benchmarking of the new and old APAIS designs had been available for all states. Without such data, this particular calibration relies on a number of assumptions about how 2013 catch statistics would have differed if based on conduct of the old APAIS sampling design. If side-by-side data were available, the differences in estimates caused by a difference in temporal coverage and other factors could have been directly measured. This underlines the importance of conducting the FES alongside the CHTS to get a good measure of consistent differences in their resulting estimates of fishing effort.

As work continues to evaluate the three alternative APAIS design change calibration models, it will be possible to examine the sensitivity of the three proposed models to the number of years used for comparisons of the temporal coverage of sampling under the new and old designs. For example, it will be possible to look at how the simple ratio calibration used in 2014-2015 may have changed with the inclusion of a second year (2013-2014) of data collected using the new APAIS design. Such sensitivity studies could be important for understanding the potential effects of interannual variability on the results of any calibration approach. Such studies will potentially help the Transition Team understand the potential consequences of attempting a calibration with less than two years of side-by-side data.

It may be important to integrate any further APAIS calibration efforts with the FES calibration and transition planning. The Terms of Reference for the 2014 APAIS Calibration Workshop stated it would be important to coordinate any new calibration accounting for the APAIS sample design change with the previous APAIS estimation change calibration and any future calibrations for other substantial MRIP survey design changes for the Atlantic and Gulf states. If at some point in late 2015 or 2016 one of the other two APAIS calibration models is chosen over the simple one used for 2015 assessment updates, it may be desirable to wait and apply that selected APAIS calibration in combination with the FES calibration planned for 2017.

Certainly, any attempts to integrate calibrations developed for different survey design changes would have to determine if there may be interacting effects of those changes. As the FES calibration is developed, the Transition Team will make sure analyses are

conducted to evaluate possible interactions with the APAIS estimation change and APAIS design change calibrations. If the calibrations turn out to be independent of one another, then it may be safe to apply them sequentially to obtain the most correct adjustments needed for historical catch statistics. However, if they are not independent, then developing an appropriate overall calibration that accounts for all changes could turn out to be a very complex process requiring considerably more time to find the right solution.

IX. Appendices

APPENDIX 1: Gantt Chart with descriptive timeline of the FES Transition Plan.

APPENDIX 2: Key stocks with initial priority ranking.

APPENDIX 3: Overview of the long-term communications strategy for the transition from the Coastal Household Telephone Survey (CHTS) to the Fishing Effort Survey (FES).

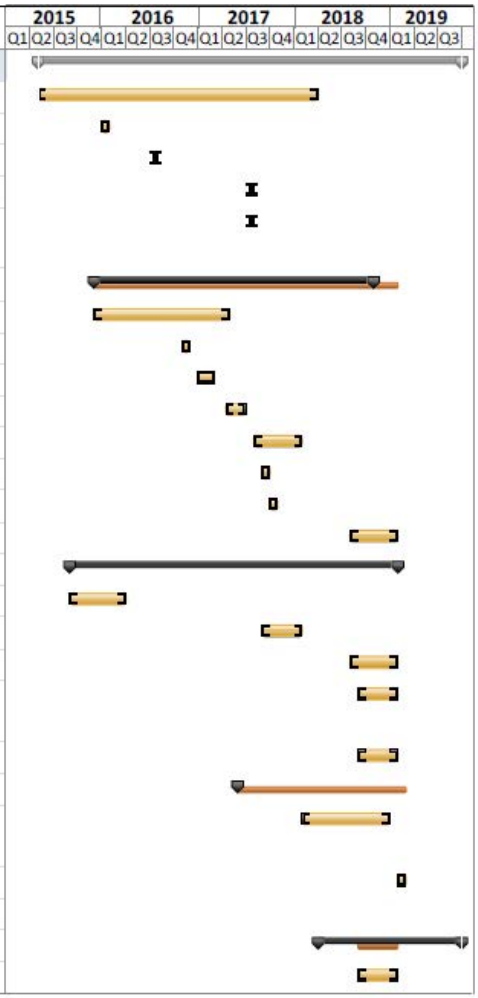
APPENDIX 4: List of Previous Pilot Studies and Links to Final Reports.

APPENDIX 5: List of the Marine Recreational Information Program Transition Team's Atlantic and Gulf Subgroup Representatives.

APPENDIX 1

Gantt Chart with descriptive timeline of the FES Transition Plan.

Task Name	Start	Finish
Fishing Effort Survey Transition Timeline	**Sun 2/1/15**	**Sun 6/30/19**
Benchmarking (side-by-side FES/CHTS)	Sun 2/1/15	Sun 12/31/17
Checkpoint-1	Thu 10/1/15	Fri 10/30/15
Checkpoint-2	Fri 4/15/16	Sat 4/30/16
Ready for Calibration	Sat 4/15/17	Sun 4/30/17
Final estimates and stock assessment scenarios	Sat 4/15/17	Sun 4/30/17
Calibration	**Tue 9/1/15**	**Mon 7/30/18**
Development	Tue 9/1/15	Mon 1/30/17
Check results of pilot studies	Mon 8/1/16	Tue 8/30/16
List of stocks	Sat 10/1/16	Wed 11/30/16
Peer review	Sun 1/15/17	Fri 3/31/17
Re-estimation of historical catch	Mon 5/1/17	Tue 10/31/17
Data Review workshop	Thu 6/1/17	Fri 6/30/17
Communicate results	Sat 7/1/17	Sun 7/30/17
Update	Tue 5/1/18	Wed 10/31/18
Stock assessment	**Mon 6/1/15**	**Wed 10/31/18**
Stock assessment scenarios	Mon 6/1/15	Thu 12/31/15
Updates based on 2nd year	Thu 6/1/17	Tue 10/31/17
Fine tune historical times series	Tue 5/1/18	Wed 10/31/18
Continuing update stock assessments not done in 2017	Fri 6/1/18	Wed 10/31/18
Updates based on 3rd year	Fri 6/1/18	Wed 10/31/18
ACL setting schedule	**Tue 2/28/17**	**Tue 2/28/17**
Begin incorporation of new estimates/ACLs on Year 1-2 data	Wed 11/1/17	Sun 9/30/18
Projections of 2018 catches	Thu 11/1/18	Fri 11/30/18
FES-based catch stats monitoring Year 1		
Allocation review	**Mon 1/1/18**	**Sun 6/30/19**
Update in 2018	Fri 6/1/18	Tue 10/30/18



Note: All dates are approximate and subject to change.

APPENDIX 2

Key stocks with initial priority ranking.

Jurisdiction	Center	FMP	Status Stock	Ranking (1-3) (1 = highest priority)	Percent Recreational
GMFMC	SEFSC	Reef Fish Resources of the Gulf of Mexico	Gag - Gulf of Mexico	1	61
GMFMC	SEFSC	Reef Fish Resources of the Gulf of Mexico	Greater amberjack - Gulf of Mexico	1	73
GMFMC	SEFSC	Reef Fish Resources of the Gulf of Mexico	Red snapper - Gulf of Mexico	1	49
GMFMC	SEFSC	Reef Fish Resources of the Gulf of Mexico	Gray triggerfish - Gulf of Mexico	1	79
GMFMC	SEFSC	Reef Fish Resources of the Gulf of Mexico	Red grouper - Gulf of Mexico	2	24
GMFMC	SEFSC	Reef Fish Resources of the Gulf of Mexico	Gray snapper - Gulf of Mexico	2	68
GMFMC	SEFSC	Reef Fish Resources of the Gulf of Mexico	Cubera snapper - Gulf of Mexico	3	51
GMFMC	SEFSC	Reef Fish Resources of the Gulf of Mexico	Gulf of Mexico Deep Water Grouper Complex	3	35
GMFMC	SEFSC	Reef Fish Resources of the Gulf of Mexico	Gulf of Mexico Mid-Water Snapper Complex	3	51
GMFMC	SEFSC	Reef Fish Resources of the Gulf of Mexico	Gulf of Mexico Shallow Water Grouper Complex	3	35
GMFMC	SEFSC	Reef Fish Resources of the Gulf of Mexico	Lane snapper - Gulf of Mexico	3	75
GMFMC	SEFSC	Reef Fish Resources of the Gulf of Mexico	Snowy grouper - Gulf of Mexico	3	35
GMFMC	SEFSC	Reef Fish Resources of the Gulf of Mexico	Vermilion snapper - Gulf of Mexico	2	NA
GMFMC	SEFSC	Reef Fish Resources of the Gulf of Mexico	Yellowedge grouper - Gulf of Mexico	3	35
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Sandbar shark	1	50
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Silky shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Tiger shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Blacktip shark - Atlantic	1	44
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Blacktip shark - Gulf	1	37

Jurisdiction	Center	FMP	Status Stock	Ranking (1-3) (1 = highest priority)	Percent Recreational
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Bull shark	2	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Spinner shark	2	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Lemon shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Nurse shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Scalloped hammerhead shark	1	86
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Smooth hammerhead shark	2	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Great hammerhead shark	2	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Atlantic sharpnose shark - Atlantic	2	12
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Atlantic sharpnose shark - Gulf	3	2
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Blacknose shark - Atlantic	3	3
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Blacknose shark - Gulf	2	8
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Bonnethead shark - Atlantic	3	9
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Bonnethead shark - Gulf	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Finetooth shark	2	31
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Blue shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Oceanic whitetip	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Porbeagle shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Shortfin mako	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Common thresher shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Atlantic angel shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Basking shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Bigeye sand tiger shark	3	NA

Jurisdiction	Center	FMP	Status Stock	Ranking (1-3) (1 = highest priority)	Percent Recreational
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Bigeye sixgill shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Bigeye thresher shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Bignose shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Caribbean reef shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Caribbean sharpnose shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Dusky shark	1	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Galapagos shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Longfin mako shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Narrowtooth	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Night shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Sand tiger shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Sevengill shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Sixgill shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Smalltail shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Whale shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	White shark	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Smooth dogfish shark - Atlantic	1	34
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Smoothhound complex- Gulf	3	1.3
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Bluefin tuna - Western Atlantic	3	20
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Swordfish - Atlantic	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Yellowfin tuna - Atlantic	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Bigeye tuna - Atlantic	3	NA

Jurisdiction	Center	FMP	Status Stock	Ranking (1-3) (1 = highest priority)	Percent Recreational
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Albacore - Atlantic	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Skipjack - Atlantic	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	White marlin - Western Atlantic	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Blue marlin - North Atlantic	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Roundscale spearfish - North Atlantic	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Sailfish - Atlantic	3	NA
HMS	SEFSC	Consolidated Atlantic Highly Migratory Species	Longbill spearfish - Western Atlantic	3	NA
MAFMC	NEFSC	Atlantic Mackerel, Squid and Butterfish	Atlantic mackerel - Gulf of Maine / Cape Hatteras	3	6.2
MAFMC	NEFSC	Bluefish	Bluefish - Atlantic Coast	1	83
MAFMC	NEFSC	Summer Flounder, Scup and Black Sea Bass	Black sea bass - Mid-Atlantic Coast	1	51
MAFMC	NEFSC	Summer Flounder, Scup and Black Sea Bass	Scup - Atlantic Coast	2	22
MAFMC	NEFSC	Summer Flounder, Scup and Black Sea Bass	Summer flounder - Mid-Atlantic Coast	1	40
MAFMC	NEFSC		Blueline tilefish – Mid-Atlantic Coast	2	NA
MAFMC	NEFSC		Golden tilefish	3	NA
NEFMC	NEFSC	Northeast Multispecies	Atlantic cod - Gulf of Maine	1	33.7
NEFMC	NEFSC	Northeast Multispecies	Haddock - Gulf of Maine	1	27.5
SAFMC	SEFSC	Dolphin and Wahoo Fishery of the Atlantic	Dolphinfish - Southern Atlantic Coast	3	87
SAFMC	SEFSC	Dolphin and Wahoo Fishery of the Atlantic	Wahoo - Southern Atlantic Coast	3	95.7
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	Black sea bass - Southern Atlantic Coast	1	57
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	Gag - Southern Atlantic Coast	1	49

Jurisdiction	Center	FMP	Status Stock	Ranking (1-3) (1 = highest priority)	Percent Recreational
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	Red snapper - Southern Atlantic Coast	1	71.93
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	Snowy grouper - Southern Atlantic Coast	2	5
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	Atlantic spadefish - Southern Atlantic Coast	3	87.1
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	Bar jack - Southern Atlantic Coast	3	67.42
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	Blue runner - Southern Atlantic Coast	3	85.4
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	Blueline tilefish - Southern Atlantic Coast	2	52.61
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	Gray snapper - Southern Atlantic Coast	1	80
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	Gray triggerfish - Southern Atlantic Coast	1	54.61
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	Greater amberjack - Southern Atlantic Coast	2	59.34
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	Hogfish - Southern Atlantic Coast	1	66.97
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	Lane snapper - Southern Atlantic Coast	3	87.79
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	Red grouper - Southern Atlantic Coast	2	56
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	Red porgy - Southern Atlantic Coast	2	50
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	Scamp - Southern Atlantic Coast	1	30.64
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	South Atlantic Deepwater Complex	3	Varies
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	South Atlantic Grunts Complex	3	Varies
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	South Atlantic Jacks Complex	3	Varies
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	South Atlantic Porgy Complex	3	Varies

Jurisdiction	Center	FMP	Status Stock	Ranking (1-3) (1 = highest priority)	Percent Recreational
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	South Atlantic Shallow Water Snapper-Grouper Complex	3	varies
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	South Atlantic Snappers Complex	3	Varies
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	Tilefish - Southern Atlantic Coast	3	3
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	Vermilion snapper - Southern Atlantic Coast	1	32
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	White grunt - Southern Atlantic Coast	2	67.33
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	Wreckfish - Southern Atlantic Coast	3	5
SAFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region	Yellowedge grouper - Southern Atlantic Coast	3	3.81
SAFMC / GMFMC	SEFSC	Coastal Migratory Pelagic Resources of the Gulf of Mexico and South Atlantic	King mackerel - Gulf of Mexico	1	68
SAFMC / GMFMC	SEFSC	Coastal Migratory Pelagic Resources of the Gulf of Mexico and South Atlantic	King mackerel - Southern Atlantic Coast	1	62.9
SAFMC / GMFMC	SEFSC	Coastal Migratory Pelagic Resources of the Gulf of Mexico and South Atlantic	Spanish mackerel - Gulf of Mexico	1	43
SAFMC / GMFMC	SEFSC	Coastal Migratory Pelagic Resources of the Gulf of Mexico and South Atlantic	Spanish mackerel - Southern Atlantic Coast	1	45
SAFMC / GMFMC	SEFSC	Coastal Migratory Pelagic Resources of the Gulf of Mexico and South Atlantic	Cobia - Southern Atlantic Coast	1	92
SAFMC / GMFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region / Reef Fish Resources of the Gulf of Mexico	Black grouper - Southern Atlantic Coast / Gulf of Mexico	2	63.12
SAFMC / GMFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region / Reef Fish Resources of the Gulf of Mexico	Mutton snapper - Southern Atlantic Coast / Gulf of Mexico	1	57
SAFMC / GMFMC	SEFSC	Snapper-Grouper Fishery of the South Atlantic Region / Reef Fish Resources of the Gulf of Mexico	Yellowtail snapper - Southern Atlantic Coast / Gulf of Mexico	1	47.44

Jurisdiction	Center	FMP	Status Stock	Ranking (1-3) (1 = highest priority)	Percent Recreational
SAFMC / GMFMC	SEFSC	Coastal Migratory Pelagic Resources of the Gulf of Mexico and South Atlantic	Cobia - Gulf of Mexico	3	NA
ASMFC		Tautog FMP	Tautog	1	NA
ASMFC	NEFSC	Atlantic Striped Bass FMP	Striped bass	1	NA
ASMFC		Weakfish FMP	Weakfish	2	NA
ASMFC		Omnibus Amendment for Spot, Spotted Seatrout, and Spanish Mackerel	Spot	3	NA
ASMFC		Red Drum FMP	Red drum	1	NA

APPENDIX 3

Overview of the long-term communications strategy for the transition from the Coastal Household Telephone Survey (CHTS) to the Fishing Effort Survey (FES).

A more detailed communications plan specific to the Fishing Effort Survey (FES) and this transition plan is being developed. The following communications plan is a high-level overview of what MRIP will be doing to ensure successful communication of the FES transition plan.

Communications Goals:

- Promote an open and productive dialogue among data partners and other stakeholders to facilitate the progress of the Transition Team toward meeting its goals.
- Communicate the broad, regionally-based perspectives that informed the development of this plan.
- Ensure consistency and quality of information about the rationale behind this Transition Plan among all audiences through pickup and reuse of outreach materials by data partners and other stakeholders.

Strategy:

Continually leverage existing communications opportunities and create new opportunities through outreach to and dialogue with our data partners and other stakeholders.

This is a two-pronged approach aimed at building trust in MRIP and NOAA Fisheries with our partners and stakeholders. Primarily, we focus on continuing our discussions and presentation of MRIP improvements and the FES transition so that our partners are fully informed about and engaged in every step of the transition process. This strategy relies on two-way communications to ensure that we are providing the information our audiences most need, and doing so in a way that most effectively conveys that information. All of this will be done while connecting FES to broader recreational fisheries messages.

Prong 1: Take advantage of national and regional opportunities to engage with partners

The initial roll out of the FES design resulted in a "wait and see" attitude among our partners and stakeholders. Over the next months a number of decisions regarding transition planning will be made and over the next years we will be hitting many key milestones. We want to take advantage of meetings and other events to build upon the success of the initial roll out and lay the groundwork for the next milestones.

Prong 2: Create opportunities for outreach and engagement.

The work of the Transition Team will provide numerous potential points at which new information and continued research will become opportunities to educate and engage our audiences. The MRIP Communication and Education Team (CET) will work with the Transition Team to identify these opportunities.

Overarching: Connect FES-specific messaging to broader MRIP and recreational fishing messaging.

We will use FES-specific messaging as a platform for building a broader understanding of MRIP as a complete process and "bottom-up" support by engaging and empowering staff to communicate about MRIP. This involves connecting FES communications to broader MRIP and NOAA Fisheries communications efforts and to regionally-specific recreational fishing communications goals. The national and regional MRIP communications teams will be the primary avenue for this.

Audiences:

NOAA audiences

- NOAA and NOAA Fisheries Leadership
- F/ST, F/SF
- Science Center (SC) and Regional Office (RO) recreational fisheries staff, regional recreational fisheries coordinators, recreational communication leads
- MRIP teams

Management Groups and Subgroups (e.g,. SSCs and Advisory Panel members)

- Fishery Management Council members and staff for all regions
- Marine Fisheries Commission members and staff for all regions
- State Resource Agencies (primarily Atlantic and Gulf coasts)
- Members of NOAA's Marine Fisheries Advisory Committee (MAFAC) with an emphasis on the MAFAC Recreational Fisheries Subcommittee and Working Group

Congressional Stakeholders

Recreational Fishing Community

Commercial Fishing Community Leaders

Environmental Community Leaders

National and Regional Media

Audience[1]	Goals	Strategy
Regional Fishery Management Council and Marine Fisheries Commission members and staff for all regions, with emphasis on the Atlantic and Gulf regions	• Understand and support decision to move to new survey methods. • Understand and support transition strategy. • Be equipped to communicate new survey and transition strategy to constituents.	• Be in constant touch to keep members and staff informed about the transition process, the reasons behind the decisions, and where we are in the process (i.e. attend meetings, hold webinars). • Provide members and staff with customized communications materials to respond to constituent inquiries. • Enhance the functionality of the MRIP website to serve as a go-to resource for a range of audiences. • Conduct stakeholder research to test the effectiveness of messaging and tailor the messaging to specific audiences. • Use regional communications teams to find additional outreach opportunities.
State Resource Agencies, with emphasis on the Atlantic and Gulf Coasts	• Understand and support decision to move to new survey methods. • Understand and support transition strategy. • Be active spokespersons among fishing and stakeholder communities about the new effort survey, the transition strategy, and the impacts on fisheries, fishermen and coastal communities.	• Be in constant touch to keep agencies informed about the transition process, the reasons behind the decisions, and where we are in the process (i.e. attend meetings, hold webinars). • Provide agencies with customized communications materials to respond to constituent inquiries. • Enhance the functionality of the MRIP website to serve as a go-to resource. • Conduct stakeholder research to test the effectiveness of messaging and tailor the messaging to specific audiences. • Increase MRIP's awareness of relevant regional issues. • Use regional communications teams to find additional outreach opportunities.

Audience[1]	Goals	Strategy
Members of NOAA's Marine Fisheries Advisory Committee (MAFAC) with an emphasis on the MAFAC Recreational Fisheries Subcommittee and Working Group	• Understand and support decision to move to new survey methods. • Understand and support transition strategy. • Be active spokespersons among fishing and stakeholder communities about the new effort survey, the transition strategy, and the impacts on fisheries, fishermen and coastal communities.	• Keep members up to date about the transition process, the reasons behind the decisions, and where we are in the process (i.e. provide materials, hold webinars). • Provide members with the key messages and targeted materials to respond to inquiries from stakeholders.
Recreational fishing community stakeholders	• Understand improvements to new survey and the transition strategy. • Understand immediate-term and potential long term implications of new effort survey for fishermen. • Understand role of state licensing in the mail survey. • Participate in mail survey if they receive it.	• *Working with the NOAA Fisheries Recreational Engagement Initiative Team to expand* • Enhance the functionality of the MRIP website to serve as a go-to resource for a range of audiences. • Use regional communications teams to find additional outreach opportunities.
Environmental community stakeholders	• Understand and support the new effort survey and transition approach. • Specifically, understand that higher effort estimates do not on their face indicate overfishing.	• *Will work with NOAA Fisheries Office of Communications to identify*
National and regional media	• Accurately characterize the findings of the pilot study, the approach to implementing it, the transition strategy, and the implications for fishery health, fishermen, coastal communities and other stakeholders.	• *Will work with NOAA Office of Public and Constituent Affairs to identify*

Audience[1]	Goals	Strategy
Commercial fishing community operating under MSA/regional FMP's	• Understand and support strategy for phasing in new effort survey estimates alongside old effort survey estimates. • Be engaged participants in the management decision-making process as calibrations and adjustments are made to estimates and stock assessments.	• *Will work with NOAA Fisheries Office of Sustainable Fisheries to identify*
Congressional members and staff who have fishing constituencies (rec and commercial)	• Understand and support decision to move to new survey methods. • Understand and support transition strategy. • Be equipped to answer questions about new survey and transition strategy from constituents.	• *Will work with NOAA Office of Legislative Affairs to identify*
NOAA and NOAA Fisheries Leadership, F/ST, F/SF, SC and RO recreational fisheries staff, regional recreational fisheries coordinators, recreational communication leads, MRIP teams	• Understand and support decision to move to new survey methods. • Understand and support transition strategy. • Be equipped to communicate new survey and transition strategy to partners and constituents.	• Initiate steps to improve internal communication. • Provide leadership and staff with big picture information they can carry to constituents and more detailed information they can use to enhance coordination with MRIP. • Increase MRIP's awareness of relevant national and regional programs and events.

Additional outreach and materials targeted to specific groups and key messages will be developed through work with members of the national and regional MRIP Communications and Education Team. This will include identification of key regional stakeholder groups that should be targeted, a regional needs assessment of FES communication materials and resources, and identification of areas where regional communications staff will take the lead instead of MRIP.

A more extensive database is being developed which lays out goals, challenges, relative priority, ideal frequency, POC, key events, and specific messages, materials, and tactics for those events.

APPENDIX 4

List of Previous Pilot Studies and Links to Final Reports.

The following is a list of the pilot projects that led to the final survey design of the Fishing Effort Survey (FES). Included for each pilot is a link to access the final report.

Development of a Dual-Frame Methodology for Estimating Marine Recreational Fishing Effort

http://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/acom/2009/WKSMRF/WKSMRF%202009.pdf

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

https://www.st.nmfs.noaa.gov/mdms/public/finalReport.jsp?ReportID=355

Dual-Frame Mail Survey: Enhancing Survey Mail Response Rates

https://www.st.nmfs.noaa.gov/mdms/public/finalReport.jsp?ReportID=362

Continued Development and Testing of Dual-Frame Surveys of Fishing Effort

https://www.st.nmfs.noaa.gov/mdms/public/finalReport.jsp?ReportID=831

APPENDIX 5

List of the Marine Recreational Information Program Transition Team's Atlantic and Gulf Subgroup Representatives.

Members

Galen Tromble (co-chair)	NOAA Fisheries, Office of Sustainable Fisheries
Dave Van Voorhees (co-chair)	NOAA Fisheries, Office of Science & Technology
Kevin Anson	Alabama Department of Conservation and Natural Resources
Mel Bell	South Carolina Department of Natural Resources
Gregg Bray	Gulf States Marine Fisheries Commission
Kevin Chu	NOAA Fisheries, Greater Atlantic Regional Fisheries Office
Richard Cody	Florida Fish & Wildlife Conservation Commission
Rita Curtis	NOAA Fisheries, Office of Science & Technology
Matt Hill	Mississippi Department of Marine Resources
Moira Kelly	NOAA Fisheries, Greater Atlantic Regional Fisheries Office
Toni Kerns	Atlantic States Marine Fisheries Commission
Kathy Knowlton	Georgia Department of Natural Resources
Laura Lee	North Carolina Department of Environment & Natural Resources
Jason McNamee	Rhode Island Department of Environmental Management
Clay Porch	NOAA Fisheries, Southeast Fisheries Science Center
Paul Rago	NOAA Fisheries, Northeast Fisheries Science Center
Andy Strelcheck	NOAA Fisheries, Southeast Regional Office
Steve Turner	NOAA Fisheries, Southeast Fisheries Science Center

Participants

John Carmichael	South Atlantic Fishery Management Council
Jamie Cournane	New England Fishery Management Council
Jason Didden	Mid-Atlantic Fishery Management Council
Mike Errigo	South Atlantic Fishery Management Council
John Froeschke	Gulf of Mexico Fishery Management Council
April Bagwill	NOAA Fisheries Affiliate, Office of Science & Technology
Richard Methot	NOAA Fisheries, Senior Scientist for Stock Assessments
Chris Wright	NOAA Fisheries, Office of Sustainable Fisheries

Marine Recreational Information Program
Fishing Effort Survey Transition
Progress Report

October 28, 2016

The MRIP Fishing Effort Survey (FES) was implemented in January, 2015 to estimate shore and private boat fishing effort for states in the Atlantic and Gulf of Mexico regions. The FES design, which was tested in MA, NY, NC and FL in 2013, has been identified as a more efficient and accurate approach for monitoring recreational fishing effort than the Coastal Household Telephone Survey (Andrews et al., 2014). Testing of the FES suggested that the design is less susceptible to survey errors than the CHTS and demonstrated that FES estimates were considerably larger than CHTS estimates.

Given the magnitude of differences between FES and CHTS effort estimates, NOAA Fisheries developed and executed a Transition Plan to facilitate the transition from the CHTS to the FES. The Transition Plan includes a three-year benchmarking period during which the FES and CHTS will be conducted concurrently in all Atlantic and Gulf coast states. This document describes results from the first full year (wave 1, 2015 – wave 6, 2015) of the benchmarking period.

Response Rates

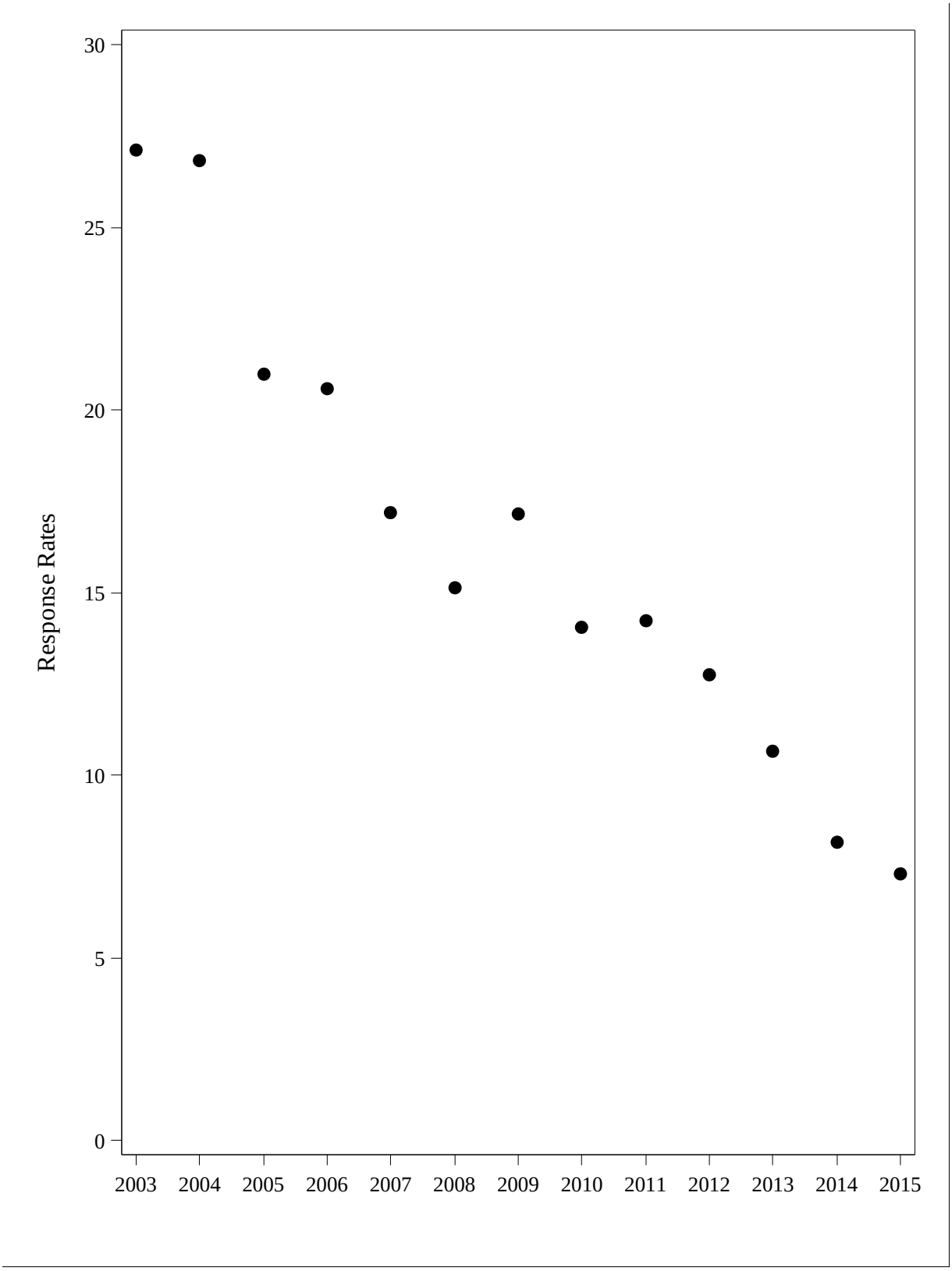
Table 1 provides final response rates for the 2015 CHTS and FES. During 2015, the 2015 FES achieved an overall response rate of 35.1%. Among states, response rates ranged from 32.3% in NJ to 44.7% in ME. In contrast, the CHTS achieved an overall response rate of 7.3%, and response rates ranged from 4.6% in RI to 11.2% in ME. The overall CHTS response rate for 2015 is consistent with the steady decline in response rates that has been observed over the past 10+ years (Figure 1).

Table 1. Weighted response rates overall and by state for the 2015 FES and FES Pilot Study.

State	2015 CHTS	2015 FES
AL	11.0	35.2
CT	8.2	35.0
DE	8.0	37.1
FL	7.5	34.3
GA	9.1	32.6
LA	8.0	32.5
ME	11.2	44.7
MD	4.8	36.6
MA	5.7	37.6
MS	9.1	34.9
NH	8.8	39.1
NJ	7.1	32.3
NY	6.6	33.6
NC	9.2	37.2
RI	4.6	38.1
SC	9.9	38.3
VA	7.6	38.3
Overall	7.3	35.1

Note: American Association for Public Opinion Research Response Rate 2 (AAPOR RR2). Response rate formula excludes ineligible addresses.

Figure 1. Annual CHTS response rates (AAPOR RR2) from the period 2003-2015.



FES/CHTS Estimate Comparisons

Overall, the FES estimate of total shore and private boat fishing effort across all states and waves (waves 1-6, 2015) is 4.7 times larger than the corresponding CHTS estimate (245,000,000 angler trips vs. 52,000,000 angler trips). This result is similar to pilot study results, where the overall FES estimate was 4.1 times larger than the CHTS estimate. We believe that the larger difference between FES and CHTS estimates in 2015 is the result of the expanded coverage of the FES to all Atlantic and Gulf of Mexico states in 2015, as well as the continued deterioration of the CHTS between 2013 and 2015 (i.e. declining response rates and coverage of landline telephone service).

As in the pilot study, overall differences between FES and CHTS estimates are larger for shore fishing (6.2X) than for private boat fishing (3.3X) (Figure 2). Differences between FES and CHTS estimates are larger for shore fishing (Figures 3a-6a) than private boat fishing (Figures 3b-6b) in all states. Differences between FES and CHTS estimates range from a factor of 2.2 for private boat fishing in Louisiana and Alabama to a factor of 11.1 for shore fishing in Georgia.

Figure 2. 2015 FES and CHTS effort estimates and the ratio of FES to CHTS estimates by fishing mode across all states and waves (wave 1-6, 2015).

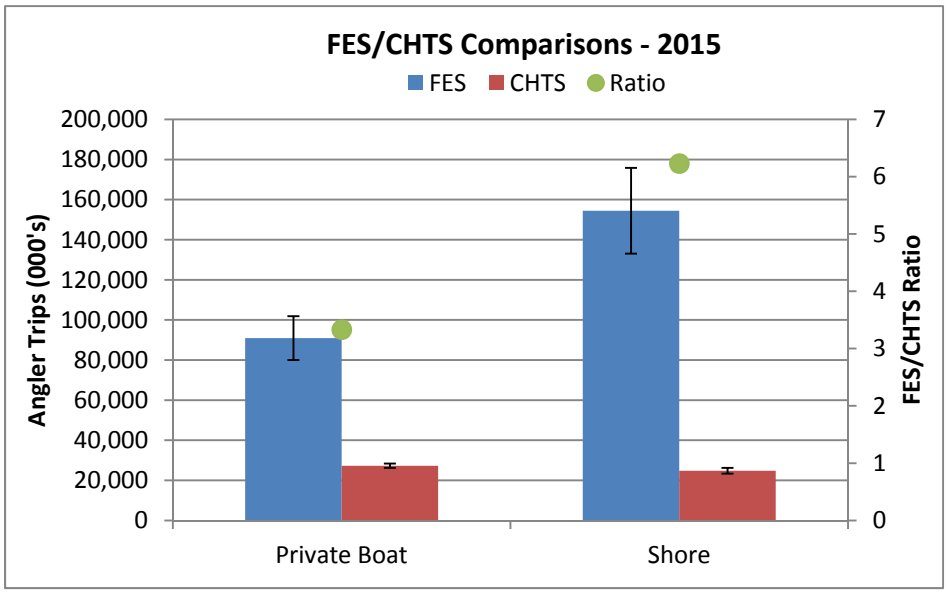


Figure 3a. 2015 FES and CHTS shore fishing effort estimates by state, North Atlantic subregion

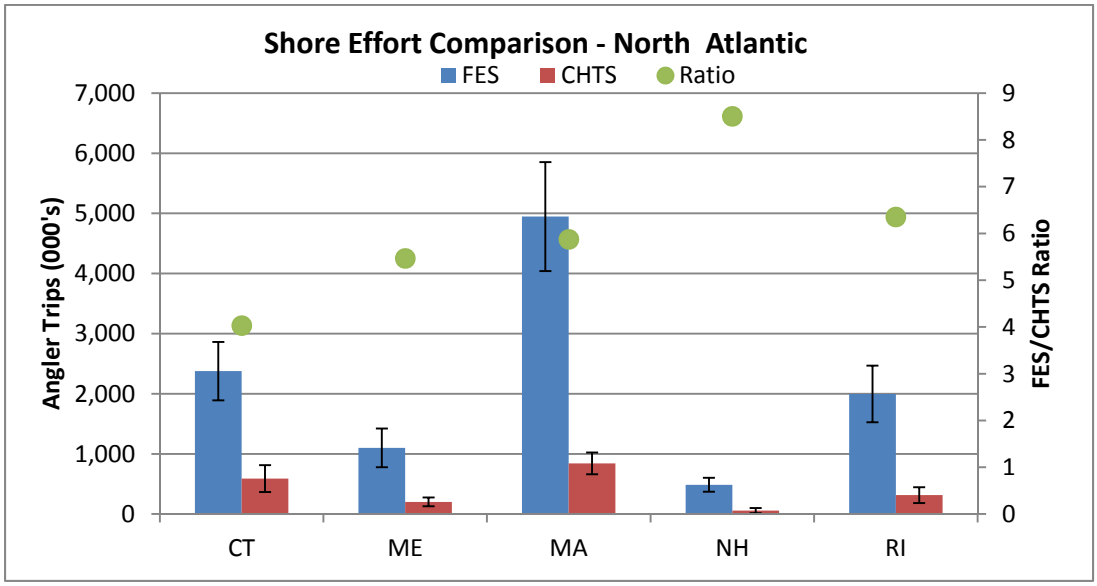


Figure 3a. 2015 FES and private boat fishing effort estimates by state, North Atlantic subregion

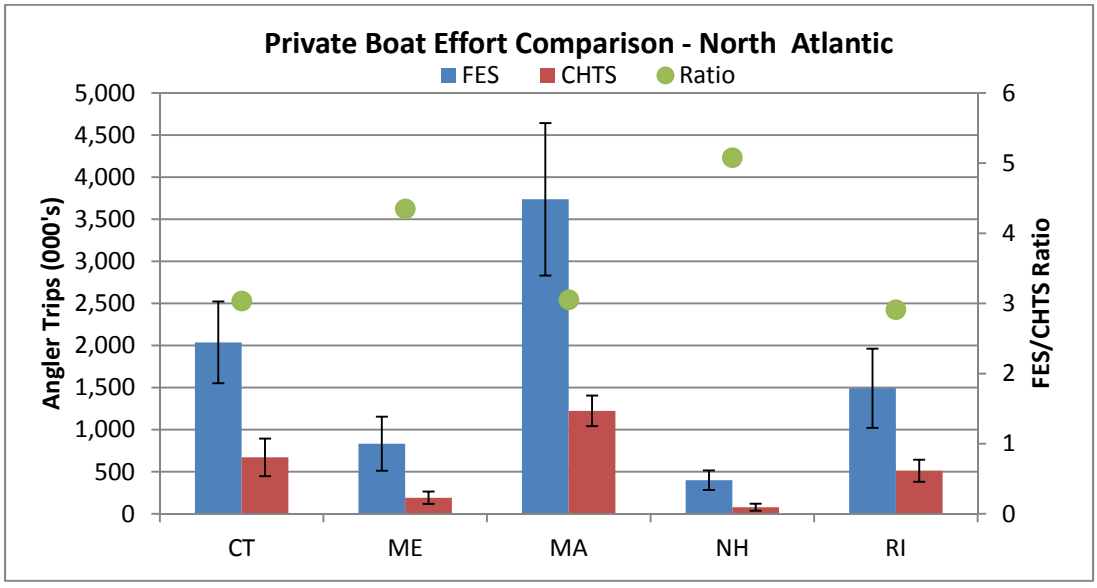


Figure 4a. 2015 FES and CHTS shore fishing effort estimates by state, Mid Atlantic subregion

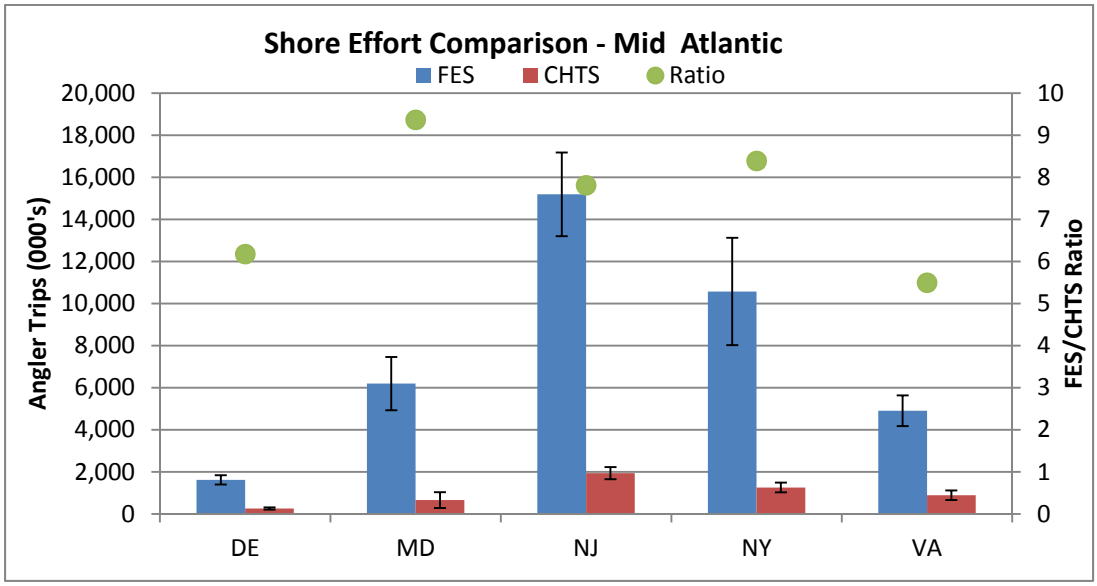


Figure 4b. 2015 FES and private boat fishing effort estimates by state, Mid Atlantic subregion

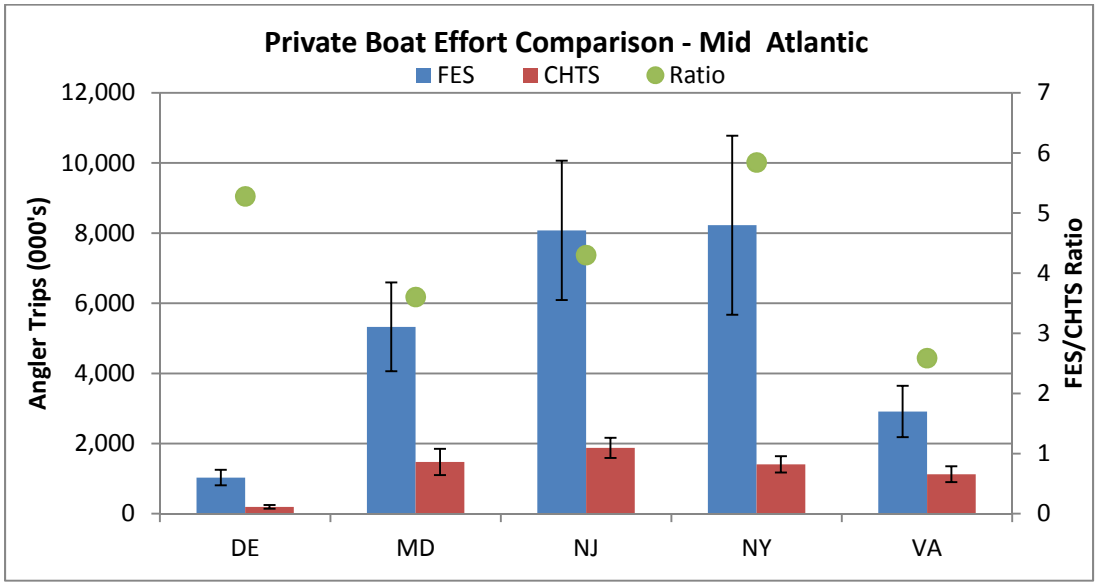


Figure 5a. 2015 FES and CHTS shore fishing effort estimates by state, South Atlantic subregion

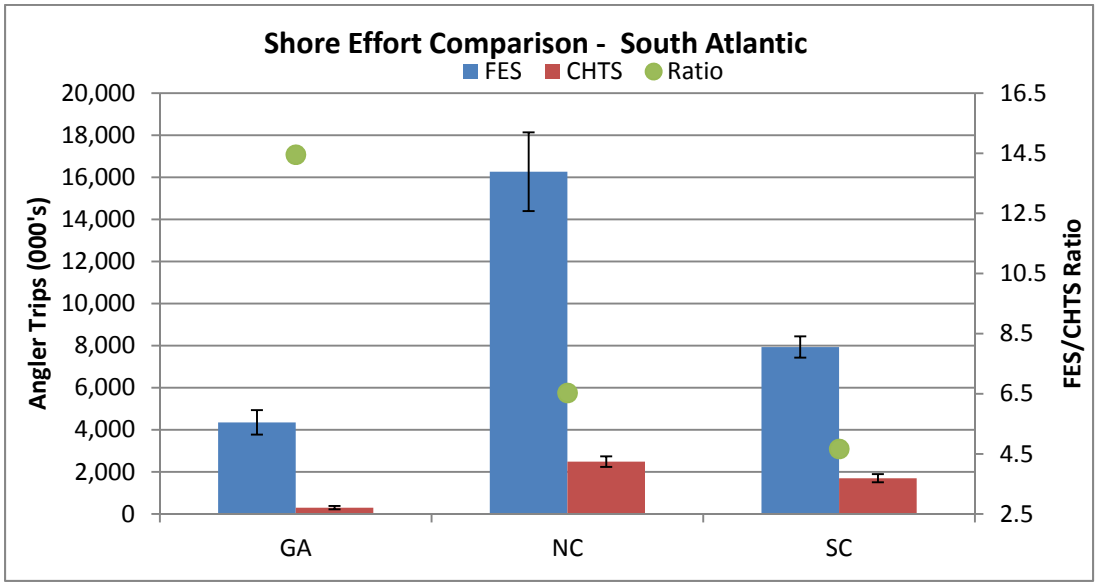


Figure 5b. 2015 FES and private boat fishing effort estimates by state, South Atlantic subregion

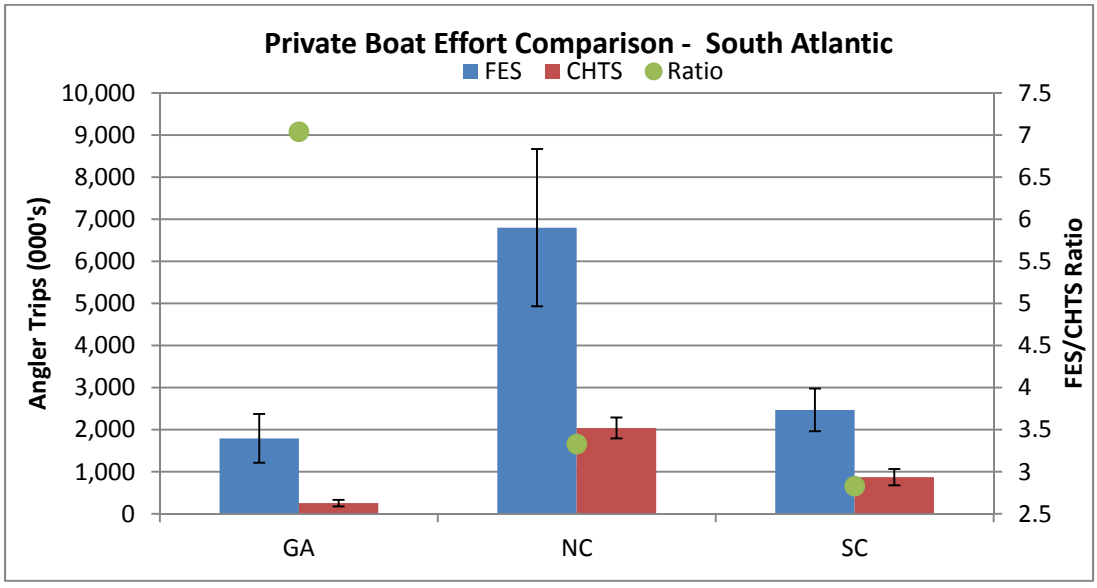


Figure 6a. 2015 FES and CHTS shore fishing effort estimates by state, Gulf of Mexico subregion

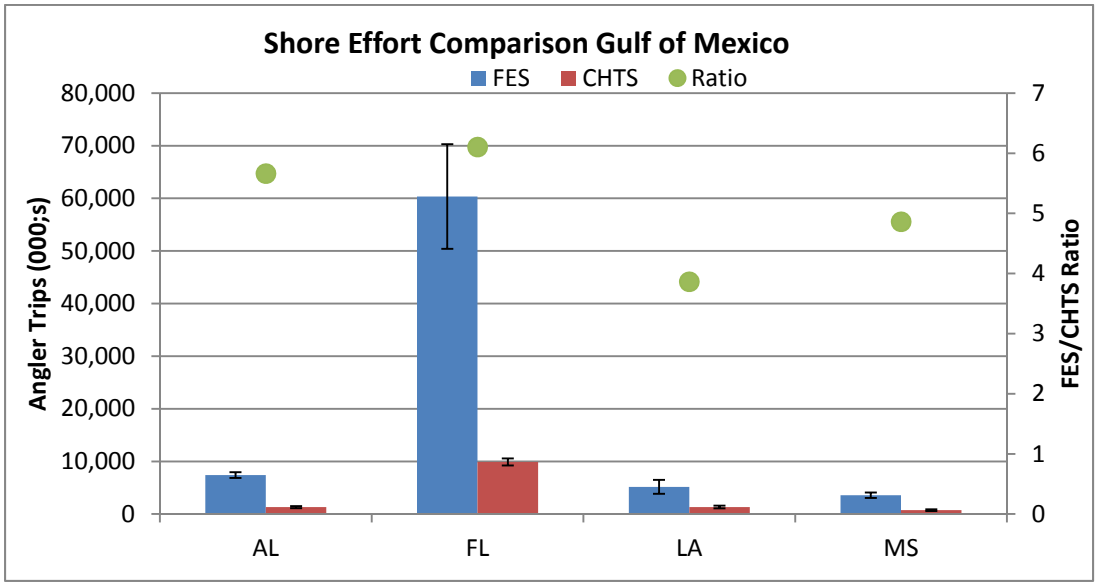
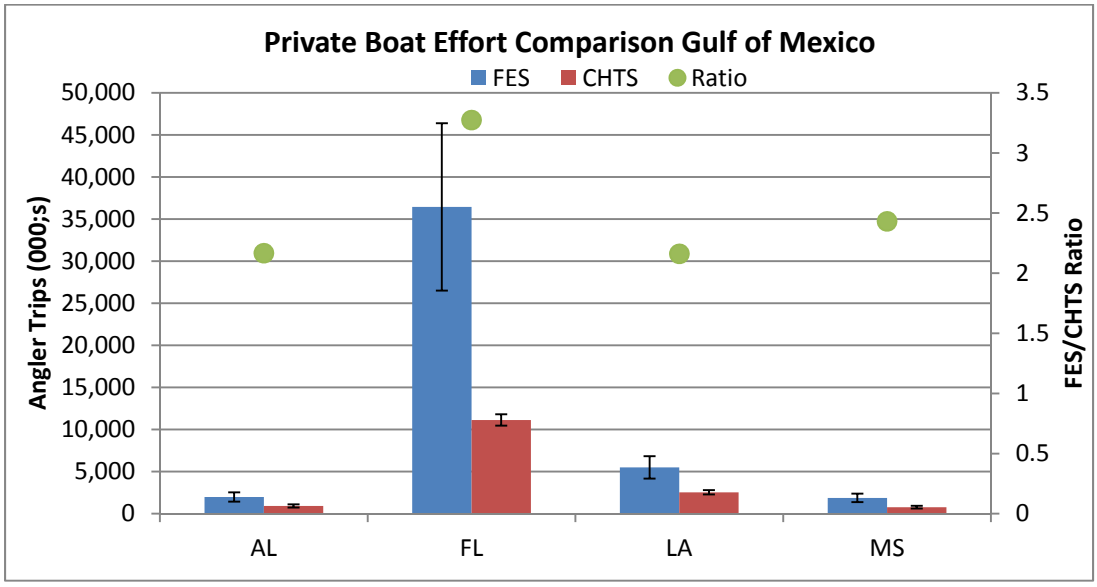


Figure 6b. 2015 FES and private boat fishing effort estimates by state, Gulf of Mexico subregion



References

Andrews, R., J.M. Brick, and N.M. Mathiowetz. 2014. Development and testing of recreational fishing effort surveys, testing a mail survey design. Available: http://www.st.nmfs.noaa.gov/Assets/recreational/pdf/2012-FES_w_review_and_comments_FINAL.pdf.