*Journal of Survey Statistics and Methodology* (2017) **5**, 350–371

# ESTIMATION OF A TOTAL FROM A POPULATION OF UNKNOWN SIZE AND APPLICATION TO ESTIMATING RECREATIONAL RED SNAPPER CATCH IN TEXAS

BINGCHEN LIU
LYNNE STOKES*
TARA TOPPING
GREG STUNZ



This research was motivated by the desire to more efficiently estimate catch by recreational anglers than current methods do. The method illustrated here combines data from angler self-reports made by a smartphone app and dockside validation samples. The two data sources can be thought of as a capture and recapture, where the parameter of interest is the population total (catch) instead of the population size. We developed several estimators of the total and compared them to one that makes use only of catch observed in the validation sample but not self-reports of catch. All the proposed estimators allow measurement error in the self-reports and do not make any assumptions about their representativeness. The validation sample must be a probability sample for valid inference, and our estimators can accommodate a complex sample design. We provide recommendations about conditions under which one of the estimators discussed may be preferred to another. Finally, we illustrate the method with analysis of data from a pilot project to estimate recreational red snapper catch in the Gulf of Mexico off the coast of Texas.

## 1. INTRODUCTION

The catch of recreational anglers is more difficult for fisheries managers to estimate than that of commercial operations. The reasons include both wider

Bingchen Liu is an Associate Psychometrician at the Educational Testing Service. Lynne Stokes is Professor, Department of Statistical Sciences, Southern Methodist University. Tara Topping is a Research Specialist at the Harte Research Institute for Gulf of Mexico Studies, Texas A&M University - Corpus Christi. Greg Stunz is an Endowed Chair and Professor at the Harte Research Institute for Gulf of Mexico Studies, Texas A&M University - Corpus Christi
*Address correspondence to Lynne Stokes, E-mail: slstokes@smu.edu

doi: 10.1093/jssam/smx006 Advance access publication 25 July 2017

dispersal and less regulatory requirements for participants. Management of recreational fisheries in the United States is the responsibility of the National Oceanic and Atmospheric Administration (NOAA). NOAA's Marine Recreational Information Program (MRIP) produces estimates of recreational catch by species every two months. MRIP's usual estimator is the product of estimators from two complementary surveys of anglers, one by phone or mail to measure "effort" (number of trips) and one face to face at the dockside to measure mean catch per trip. Improvements to the designs of these surveys have been underway for about a decade, guided by input from a National Research Council report (National Research Council 2006). Recently, however, some stakeholders have proposed using angler self-reports to supplement, or perhaps even to replace, other data sources for estimating catch. Self-reporting has become increasingly simple and inexpensive, with the increasing penetration and ease of use of smartphones. The purpose of this paper is to present methodology that will allow direct use of this self-reported data for estimating recreational catch. The proposed estimators combine data from the nonprobability sample of self-reports with that from the dockside probability sample to produce scientifically defensible estimates of recreational catch. Our analysis shows that if angler self-reporting is sufficiently complete and accurate, these new data can make the effort survey unnecessary. This could save money and allow for more timely production of estimates. Thus our methodology is an example of Groves' prediction: "The combination of designed data with organic data is the ticket to the future" (Groves 2011).

Pilot projects in Alabama, Mississippi, and Texas are testing this approach for estimating red snapper catch. Red snapper is a favorite target of recreational anglers in the Gulf of Mexico, and one of its most economically valuable species. Scientists are concerned that the Deepwater Horizon oil spill has affected the health of this long-lived species (Tarnecki and Patterson III 2015) and slowed its recovery from overfishing. To manage the species properly, accurate and timely estimates of catch are needed. Thus, projects seeking to improve the precision and timeliness of data collection for red snapper have been funded by both NOAA and the Gulf Environmental Benefit Fund (GEBF), which distributes funds for mitigation of damages from the spill. The three state projects all combine their traditional dockside survey data with data from reports that anglers make via computer or via smartphone apps known as Snapper Check, Tails 'n Scales, and iSnapper in Alabama, Mississippi, and Texas, respectively. In all three states, anglers are requested to report their catch using their device before they remove the fish from the boat. In Alabama and Mississippi, reporting is mandatory and enforced by fines if the angler fails to comply, while in Texas it is voluntary. Anglers are also incentivized to provide their data by allowing them to access records of their own angling activities maintained by the system.

Survey researchers recently have shown interest in methods for making statistical inferences using data from nonprobability samples (Baker, Brick, Bates, Battaglia, Couper, et al. 2013). The most common approaches involve weighting, using various methods to produce weights reflecting the number of population members each sample member "represents." One method for doing this is to estimate by modeling a pseudo-probability of selection, which typically relies on knowledge of common auxiliary information in the sample and the population to which inference will be made. The population information can come from summaries of the entire population (such as from Census data) or from a valid probability (reference) sample. Pseudo-probability estimation methods include ones first developed for reducing coverage and nonresponse biases in probability samples, such as calibration (e.g., poststratification) and propensity score adjustments (Lee and Valliant 2009). The bias reduction from these methods depends on availability of auxiliary information that is highly correlated with response variables and on an assumption that the units in the sample are representative of those not present within analyst-defined classes. These methods have shown mixed results (Tourangeau, Conrad, and Couper 2013).

Though our data do include a reference sample containing auxiliary data in common with the nonprobability sample, we take a different approach. Both our probability and nonprobability samples contain measures of catch, but the ones from the nonprobability sample are considered less accurate than those reported by dockside samplers. Thus our proposed estimators are based on the probability sample data and use information about reported catch from the nonprobability sample only as auxiliary data. In fact, all the estimators we propose are either ratio estimators themselves or modifications of ratio estimators. Thus neither correction of the nonprobability sample by weighting nor assumptions about its representativeness are required for validity.

An alternative way to view our data collection and methodology is as a generalization of a Capture-Recapture experiment. This point of view provides an intuitive explanation for fisheries scientists who are typically more familiar with this methodology and the assumptions it requires than with ratio estimation. In a Capture-Recapture experiment, $n_1$ fish are caught, marked, and released. These fish are not required to be a probability sample of fish. Then a second sample of $n_2$ fish is captured, and $m$ of them are noted as previously marked. This sample is required to be selected in such a way that all fish (including marked and unmarked) are equally likely to be included. These data can be used to produce an estimator of the unknown population size $N$ by equating the fraction of marked fish in the population and sample:

$$\frac{n_1}{N} = \frac{m}{n_2}.$$

This yields the classical estimator, called the Lincoln-Petersen index, due to the pioneering work of two ecologists (Le Cren 1965):

$$\hat{N} = \frac{n_1 n_2}{m}. \tag{1}$$

Unlike the goal of the usual Capture-Recapture experiment, our goal is to estimate the total $t_y$ of an attribute $y$ (catch) over a population (angler trips), whose size $N$ is unknown. We examine several estimators that generalize the approach of the Lincoln-Petersen index for this purpose. In our application, the sample of self-reports in each of the three red snapper programs plays the role of the capture sample. The dockside sample is viewed as the recapture sample. This validation sample has a complex probability sample design in each of the three programs. The frame of each consists of gulf access points crossed with time blocks, such as four- or six-hour time shifts, so that the sample includes clusters of trips ending at the sampled access point and shift. The clusters are selected into the samples with probabilities that vary, with active fishing times/places more likely to be selected than those with less activity. The catch for all trips sampled is observed and counted by agency personnel. The Lincoln-Petersen Index is based on an assumption that recaptures are a simple random sample of the population, so an adaptation is necessary for valid estimation when the recapture sample design is complex, as ours is. There have been previous applications of the usual Capture-Recapture methodology to estimate $N$ where one or both of the samples are considered to have features of a complex design and weighting is used (Wolter 1986; Alho, Mulry, Wurdeman, and Kim 1993). We adopt a similar approach for our new estimators of total for a population of unknown size.

Pollock, Turner, and Brown (1994) previously considered the problem of estimating a total over a population of unknown size. This application differed from ours in that $y$ was observable only for the units selected in the second sample; that is, there were no angler self-reports of catch for those $n_1$ trips that were included in the capture sample. They proposed as an estimator of total catch

$$\hat{t}_{yp} = \hat{N}\hat{\bar{y}}, \tag{2}$$

with $\hat{N}$ defined in (1) and $\hat{\bar{y}}$ as the sample average of the $y$s from the $n_2$ recaptured units. In our application, the self-reported catch is also available from the capture sample, so it seems reasonable to expect that this information should improve estimation if angler compliance and accuracy are high. Our application also differs from Pollock's scenario in that the validation sample is selected according to a complex design, so that generalizations of expressions (1) and (2) are needed.

In section 2, we introduce three new estimators that use the self-reported data together with that from the validation samples to estimate catch. We develop expressions of the large sample variances of the estimators under a simple random sample (SRS) design for the validation sample. In section 3, we compare these variances with each other and with that of $\hat{t}_{yp}$ under various assumptions about the self-reports. Specifically we consider cases in which reported catch is subject to measurement error, when reporters are not representative but rather are those with extremely high or low catch, and for a range of reporting rates. Section 4 examines estimator performance from simulation studies that are designed to examine the small sample bias of the estimators, as well as the effects of a complex design for the validation sample. One of the simulation settings we use is designed to mimic some features of the data from the Texas iSnapper program. Then we illustrate the method with estimates of red snapper from these data. Discussion follows in section 6.

## 2. ESTIMATORS OF POPULATION TOTALS USING CAPTURE-RECAPTURE METHODS

Let $d_1$ denote the subset of the $N$ population units that self-report their trip and catch. $d_1$ is not assumed to be representative of the population, nor is it a probability sample, but rather is regarded as a domain. Each of the $n_1$ domain units reports a value for $y$, but the $i$th unit's report is denoted by $y_i^*$ to distinguish it from the truth, $y_i$. No assumptions are made about the relationship between $y$ and $y^*$. A validation sample $s_2$ is selected according to a probability design, and $y_i$ is observed for each unit in $s_2$. A subset of $s_2$ will match self-reported trips; these units have both $y$ and $y^*$ available. The goal is to estimate $t_y = \sum_{i=1}^{N} y_i$ using the data from $d_1$ and $s_2$.

The population and sample data can be visualized as shown in Figure 1. The first row represents the reporting domain $d_1$ and includes the trips with $y^*$ available, while the second row contains trips without $y^*$. The first column contains trips in the validation sample, for which $y$ is available; the second column contains the trips without observable $y$. The upper left cell represents the $m$ matched units with observable $y$ and $y^*$; the upper right cell represents the $n_1 - m$ reported ($y^*$ known) but unvalidated trips; the lower left cell represents the $n_2 - m$ validated ($y$ known) but unreported trips. The lower right cell contains units with no data available to the analyst.

We denote the reporting rate, defined as the fraction of trips reported by anglers, as

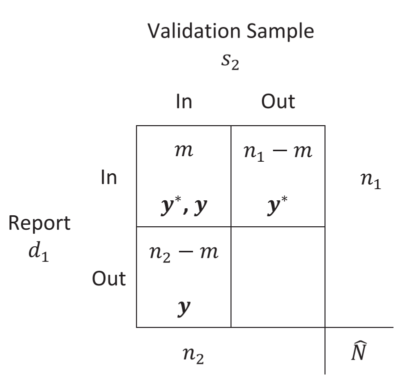


**Figure 1. An Illustration of the Population and Sample Data.**

$$p_1 = n_1/N, \tag{3}$$

and the population mean as $\bar{y} = t_y/N$. Then Pollock's estimator (2) can be generalized for a complex design to

$$\hat{t}_{yp} = \frac{n_1}{\hat{p}_1}\hat{\bar{y}} = n_1\frac{\hat{t}_y}{\hat{n}_1}, \tag{4}$$

where the $r_i's$ are reporting indicators ($r_i = 1$ if the $i^{th}$ unit is included in $d_1$; 0 otherwise); the $w_i's$ are sampling weights for units in $s_2$; $\hat{n}_1 = \sum_{i\in s_2} w_i r_i$; $\hat{p}_1 = (\sum_{i\in s_2} w_i r_i / \sum_{i\in s_2} w_i)$; and $\hat{\bar{y}} = (\sum_{i\in s_2} w_i y_i / \sum_{i\in s_2} w_i)$. Thus $\hat{t}_{yp}$ is recognizable as a ratio estimator with auxiliary variable $r_i$ and ratio

$$B_p = t_y/n_1.$$

Now we propose a new estimator that is also a ratio estimator, but with auxiliary variable $r_i y_i^*$ and ratio denoted by

$$B_c = t_y/\sum_{i=1}^{N} r_i y_i^* = t_y/t_{y^*},$$

where $t_{y*} = \sum_{i\in d_1} y_i^* = \sum_{i=1}^{N} r_i y_i^*$ is the total reported catch. This yields the estimator

$$\hat{t}_{yc} = t_{y*}\frac{\sum_{i\in s_2} w_i y_i}{\sum_{i\in s_2} w_i r_i y_i^*} = t_{y^*}\frac{\hat{t}_y}{\hat{t}_{y^*}}. \tag{5}$$

This estimator can be regarded as the total reported catch inflated by the estimated reporting rate $(\sum_{i\in s_2} w_i r_i y_i / \sum_{i\in s_2} w_i y_i)$, and adjusted for reporting errors by a multiplicative correction factor $(\sum_{i\in s_2} w_i r_i y_i / \sum_{i\in s_2} w_i r_i y_i^*)$. It is a generalization of the Capture-Recapture estimator, where totals of $y$ and $y^*$ replace counts of units in the two data collection periods.

If $y^*$ is accurate and the reporting rate is high, $\hat{t}_{yc}$ would be expected to be more precise than $\hat{t}_{yp}$ due to its highly correlated auxiliary information. If reporting is inaccurate and rare, the reverse would be true. To avoid making the choice of estimator in advance, one could form a linear combination of the two estimators, with weights chosen to minimize variance. This estimator is a special case of what Olkin (1958) called the multivariate ratio estimator, which we denote as

$$\hat{t}_{MR} = (1 - W)\hat{t}_{yp} + W\hat{t}_{yc}. \tag{6}$$

Olkin (1958) showed (eq. 3.1, p. 157) that the optimal weight $W$ for a simple random sample (SRS) can be approximated to order $O(1/n)$ by

$$w_{SRS} = \frac{S_{dp}^2 - S_{dp,dc}}{S_{dp}^2 + S_{dc}^2 - 2S_{dp,dc}}, \tag{7}$$

where $S_{dp}^2, S_{dc}^2$, and $S_{dp,dc}$ denote the variances and covariance of the residuals from the ratio models.

In our application, these residuals are $d_{pi} = y_i - B_p r_i$ and $d_{ci} = y_i - B_c r_i y_i^*$, and their variances and covariances can be expressed as shown in the Appendix in (A.2), (A.3), and (A.4). Using these expressions, (7) simplifies to

$$w_{SRS} = \frac{t_{y^*}}{t_y}\frac{S_{1,yy^*}}{S_{1y^*}^2} = \frac{t_{y^*}}{t_y}\frac{S_{1y}}{S_{1y^*}}R_{1,yy^*}, \tag{8}$$

where $R_{1,yy^*}$, $S_{1,yy^*}$, $S_{1y}$, and $S_{1y^*}$ are the correlation, covariance, and standard deviations of $y$ and $y^*$ in the reporting domain $d_1$. Thus the optimal estimator gives $\hat{t}_{yc}$ the majority of weight ($w_{SRS} > 1/2$) when

$$R_{1,yy^*} > \frac{CV_{1y}}{2p_1 CV_{1y^*}},$$

where $p_1$ is the reporting rate (from (3)), and $CV_{1y}$ and $CV_{1y^*}$ are the coefficients of variation of $y$ and $y^*$ in the reporting domain.

In practice, $w_{SRS}$ must be estimated in order to use $\hat{t}_{MR}$. We consider two estimators. For the first, we replace the components of (8) with estimators calculated from the observed data, as suggested in Olkin (1958). One such estimator is

$$\hat{w}_{SRS,1} = \frac{t_{y^*}}{\hat{t}_{yc}}\frac{s_{1,yy^*}}{s_{1y^*}^2},$$

where $s_{1y^*}^2$ and $s_{1,yy^*}$ are the estimated variance and covariance between $y$ and $y^*$ in the reporting domain, made from the matched sample. (Alternatively, one could use $\hat{t}_{yp}$ or implicitly define an estimator by substituting $\hat{t}_{MR}$ for $t_y$ in (6), or use the observable value of $S_{1y^*}^2$ in the denominator of $\hat{w}_{SRS,1}$. Simulation showed little difference in performance among these alternatives.) We denote the resulting estimator by $\hat{t}_{y1}$. The second estimator we consider is simpler and near optimal when reporting errors are small. Note from (8) that when $y = y^*$, $w_{SRS} = t_{y^*}/t_y$. Thus we estimate $w_{SRS}$ by

$$\hat{w}_{SRS,2} = \frac{t_{y^*}}{\hat{t}_{yc}}.$$

The resulting estimator can be simplified to

$$\hat{t}_{y2} = t_{y^*} + \frac{n_1}{\hat{n}_1}\left(\hat{t}_y - \hat{t}_{y^*}\right) = t_{y^*} + n_1\hat{\bar{\delta}}, \tag{9}$$

where $\delta_i = y_i - r_i y_i^*$ and $\bar{\delta} = (t_y - t_{y^*})/n_1$ is the total population underreport averaged over reporters. In contrast to $\hat{t}_{yc}$, this estimator augments the reported catch by an additive rather than a multiplicative component.

When the validation sample has a complex design, it can be accounted for in $\hat{t}_{yp}$ and $\hat{t}_{yc}$ as shown in (4) and (5), and these estimators combined as in (6). Olkin (1958) generalized (7) to produce an appropriate expression for $W$ when the sample has a stratified design. For a general complex design, however, it is useful to note that the optimal value for $W$ is

$$W = \frac{V(\hat{t}_{yp}) - Cov(\hat{t}_{yp}, \hat{t}_{yc})}{V(\hat{t}_{yp}) + V(\hat{t}_{yc}) - 2Cov(\hat{t}_{yp}, \hat{t}_{yc})}, \tag{10}$$

which reduces to Olkin's expressions for SRS and stratified designs when Taylor series variance approximations are used. Modern survey software can provide estimates of the variances and covariance for most designs, so that explicit expressions are not needed for each design type. The optimal $W$ will not be well approximated by (7) if design effects for the two estimators differ greatly. Then the simplified forms shown in $\hat{w}_{SRS,1}$ and $\hat{w}_{SRS,2}$ and the resulting estimators will no longer be nearly optimal.

In section 3, we focus only on SRS designs. We compare the variances of the estimators to help us understand their relative performance as the completeness, representativeness, and accuracy of reporting change. The goal is to provide guidance on which estimator may be best for different applications. In section 4, we extend the comparison to complex designs via simulation.

## 3. COMPARISON OF ESTIMATORS FOR SIMPLE RANDOM SAMPLES

We compare the approximate variances of $\hat{t}_{yp}$, $\hat{t}_{yc}$, and $\hat{t}_{y2}$, to that of $\hat{t}_{MR}$ under an SRS design for the validation sample and for a range of scenarios for the quality and completeness of the self-reported catch information. We do not consider $\hat{t}_{y1}$ separately since its large sample behavior is that of $\hat{t}_{MR}$. The ratios of the variance expressions in (A.9) through (A.12) are unaffected by the size of the validation sample $n_2$, the population size $N$, or the total itself, $t_y$. They do depend on the reporting rate $p_1$, the correlation and $CV$'s of $y$ and $y^*$ in the reporting domain ($R_{1,yy^*}$, $CV_{1y}$, and $CV_{1y^*}$), and the ratios of the means of $y$ and $y^*$ in the reporting domain to $y$'s mean in the population ($\bar{y}_1/\bar{y}$ and $\bar{y}_1^*/\bar{y}$). Therefore, we present comparisons of the variances of the three estimators to that of $\hat{t}_{MR}$ for the following three scenarios: (1) no errors in reporting and reporters are representative of the population; (2) errors in reporting, but reporters are representative of the population; and (3) no errors in reporting, but reporters are not representative of the population. We examine the loss of precision for $\hat{t}_{yp}$, $\hat{t}_{yc}$, and $\hat{t}_{y2}$, as compared to $\hat{t}_{MR}$.

In scenario 1, we assume that $y = y^*$ ($R_{1,yy^*} = 1$, $\bar{y}_1 = \bar{y}_1^*$, and $CV_{1y} = CV_{1y^*}$) and reporters are representative of the population (defined to mean that $\bar{y}_1 = \bar{y}$ and $CV_{1y} = CV_y$). This will occur on average if reporting is "at random," though randomness is not required for representativeness. Figure 2 displays the ratio of the large sample variance of $\hat{t}_{MR}$ to that of each estimator as a function of reporting rate $p_1$, defined in (3). Panels (A) and (B) show how $CV_y$ (set to 0.32 and 0.55) affects performance. When $CV_y = 0$, $\hat{t}_{yp}$ and $\hat{t}_{yc}$ are equivalent. When $CV_y > 0$, $\hat{t}_{yp}$ is more efficient than $\hat{t}_{yc}$ for small reporting rate but grows less efficient as $p_1$ increases. The crossover point occurs at $p_1 = 1/2$ regardless of $CV_y$, but the advantage for $\hat{t}_{yc}$ grows with $CV_y$. $\hat{t}_{y2}$ is uniformly optimal in this case since $y = y^*$.

Next we examine the performance of the estimators when self-reports are not accurate ($y \neq y^*$), but reporters are representative ($\bar{y}_1 = \bar{y}$ and $CV_{1y} = CV_y$). $\hat{t}_{yp}$ is unaffected by measurement error since it does not use $y^*$. (A.11) shows that errors increase the variance of $\hat{t}_{MR}$ by decreasing $R_{1,yy^*}$, while (A.10) shows that they affect the performance of $\hat{t}_{yc}$ through both $R_{1,yy^*}$ and $CV_{1y^*}$. Since $CV_{1y^*}$ can either increase or decrease when $y \neq y^*$, the effect of measurement error on the variance of $\hat{t}_{yc}$ is not clear. Finally, (A.12) shows that the variance of $\hat{t}_{y2}$ is affected by errors through $R_{1,yy^*}$, $CV_{1y^*}$, and $\bar{y}_1^*/\bar{y}$. Thus we compared the estimators under two measurement error models that impact these parameters differently: the classical measurement error (CME) (Carroll et al. 2006, section 1.2) and the Berkson model (Berkson 1950).

The CME model specifies

$$y^* = y + e, \tag{11}$$

where $e \sim (0, \alpha S_y^2)$, $S_y^2 = \sum_{i=1}^N (y_i - \bar{y})^2/(N-1)$ is the variance of $y$ in the finite population, $y$ and $e$ independent. When (11) holds, $R_{1,yy^*} = 1/\sqrt{1+\alpha}$, $CV_{1y^*} = CV_{1y}\sqrt{1+\alpha}$, and $\bar{y}_1^*/\bar{y} = \bar{y}_1/\bar{y}$. The Berkson model reverses the role of $y$ and $y^*$ and specifies

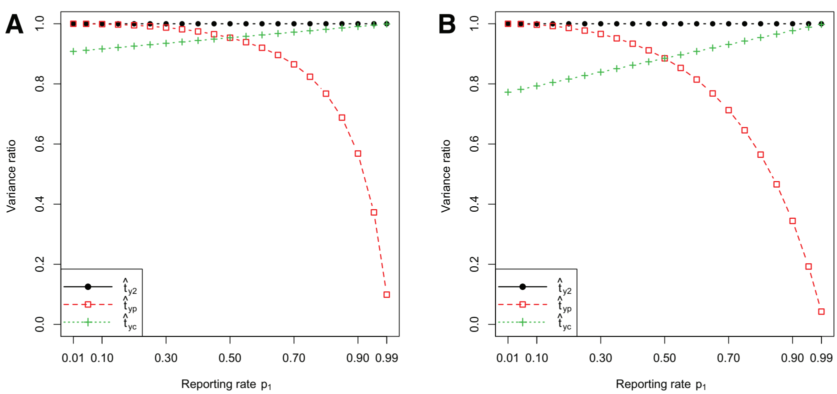


**Figure 2. Relative Precision of Three Estimators to $\hat{t}_{MR}$ as a Function of $p_1$, When There are No Errors, and Representative Reporting.** (A) Variance ratios when $CV_y$=0.32. (B) Variance ratios when $CV_y$=0.55.

$$y = y^* + e, \tag{12}$$

with $e \sim (0, \beta S^2_{y^*})$, $y^*$ and $e$ independent. A Berkson error model is plausible in our application if the trip reporter provides the catch value as the bag limit (maximum legal catch) * # of anglers aboard as an estimation strategy. Under (12), the domain parameters would be $R_{1,yy^*} = 1/\sqrt{1+\beta}$ and $\bar{y}^*_1/\bar{y} = \bar{y}_1/\bar{y}$. Berkson error causes $CV_{1y^*}$ to decrease; $CV_{1y^*} = CV_{1y}/\sqrt{1+\beta}$.

Figure 3 shows the variance ratios of $\hat{t}_{MR}$ to $\hat{t}_{yp}$, $\hat{t}_{yc}$, and $\hat{t}_{y2}$ as functions of $R_{1,yy^*}$, where $CV_y = 0.32$ and $p_1 = 0.7$. The two panels show how their performance changes when the measurement error structure differs; CME is assumed for panel (A) and Berkson error for panel (B). Recall from Figure 2A that $\hat{t}_{yc}$ and $\hat{t}_{y2}$ outperform $\hat{t}_{yp}$ for these settings when $y = y^*$. Figure 3A shows that this advantage is lost when CME afflicts $y^*$, unless the correlation between $y$ and $y^*$ is substantial (about 0.7 for $\hat{t}_{y2}$ and 0.84 for $\hat{t}_{yc}$). When $y^*$ has Berkson error, the relative performance returns to its no-error order. Berkson error does not increase the variance of $\hat{t}_{yc}$ and $\hat{t}_{y2}$ as much as CME does for the same correlation, while both models affect the variance of $\hat{t}_{MR}$ equally. In fact, the large sample variance of $\hat{t}_{y2}$ is identical to that of the optimal estimator in this case, even though errors occur. The point here is that the structure of the measurement error matters for determining which estimator is best, and the preference depends on more than $R_{1,yy^*}$.

Finally, we again assume that $y = y^*$, but reporters are not representative. Then the mean and/or variance of $y$ are different for reporters than for the population, affecting estimator variances through $\bar{y}_1/\bar{y}$ and $CV_{1y}$. To see how much, we must specify a mechanism for determining who reports. We examined two extremes: that reporters have the largest or smallest catch; that is, reporters are assumed to be those in the top (bottom) $100p_1\%$ of $y$'s

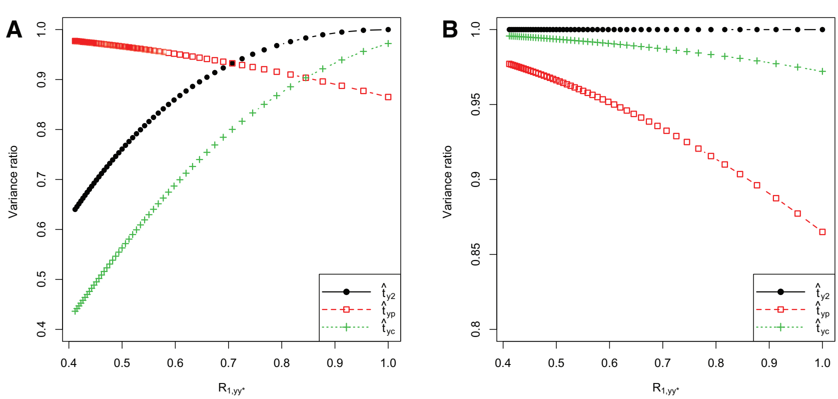


**Figure 3. Ratio of Variance of $\hat{t}_{MR}$ to the Three Estimators as a Function of $R_{1yy^*}$, When $p_1$=0.7, and Reporters are Representative.** (A) Classical measurement error model for $y^*$. (B) Berkson measurement model for $y^*$.

distribution. The effect of these mechanisms on $\bar{y}_1/\bar{y}$ and $CV_{1y}$ depends on the distribution of $y$.

We considered two distributions for $y$, one continuous (zero-truncated normal) and one discrete (zero-truncated Poisson). When $y$ is normal, the distribution of $y$ in the high catch reporting domain is that of an upper tail truncated normal, with truncation point $A = \bar{y} + S_y\Phi^{-1}((1-p_1)(1-\Phi(-\bar{y}/S_y)))$, where $\Phi$ is the standard normal CDF. The low catch domain was defined similarly. Thus the moments of $y$ in the reporting domain are easily calculated (e.g., Johnson and Kotz 1970, pp. 81–3). When $y$ is zero-truncated Poisson, its distribution in the reporting domain is also truncated Poisson, but at a value larger than 0. The moments of the $k$-truncated Poisson are also easily calculated (Johnson and Kotz 1970). Because of the discreteness of this distribution, only some values of $p_1$ are possible for this model.

Figure 4 shows the variance ratios as functions of $p_1$ when the domain contains high removal reporters, where panel (A) shows results for truncated normal $y$ (with $CV_y = 0.32$) and panel (B) for truncated Poisson $y$ (with $\lambda = 1.79$, which yields $CV_y \approx 0.55$). Thus the differences in Figure 2 and Figure 4 illustrate the impact of nonrepresentative reporting only. A comparison shows that high-removal nonrepresentative reporting improves the relative performance of $\hat{t}_{yc}$, especially for small $p_1$. $\hat{t}_{yp}$ alone declines in performance compared to the best estimator as the reporting rate increases.

Table 1 is designed to show how much nonrepresentative reporting affects the absolute and not just the relative variance of the estimators under largest and smallest catch reporting. It displays the ratio of each estimator's variance when reporters are representative and when they are not (larger for the upper and smaller for the lower half of the table) for normal $y$ and two reporting rates:

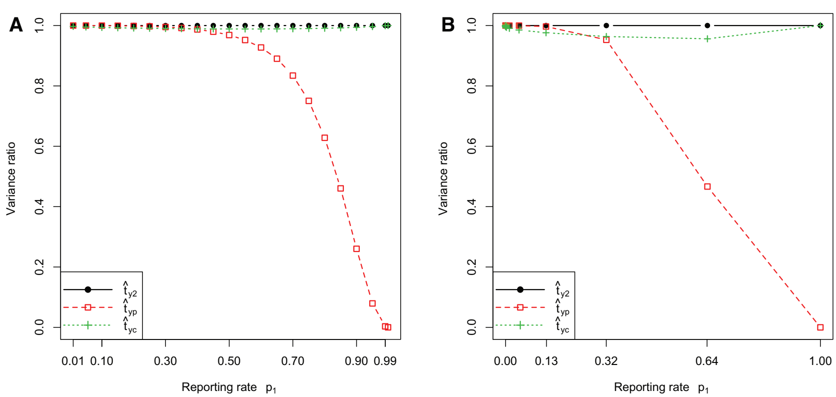


**Figure 4. Ratio of Variance of $\hat{t}_{MR}$ to the 3 Estimators as a Function of $p_1$, When There are No Errors, and Max Catch Reporting.** (A) For zero-truncated Normal distribution with $CV_y$=0.32. (B) For zero-truncated Poisson distribution with $CV_y$=0.55.

**Table 1. Ratios of Variances When Reporters are Representative and Not Representative when $CV_y = .32$**

| | $p_1$ | $\hat{t}_{yp}$ | $\hat{t}_{yc}$ | $\hat{t}_{y2}$ |
|---|---|---|---|---|
| $\frac{Var(\hat{t} \mid \text{representative reporting})}{Var(\hat{t} \mid \text{large removal reporting})}$ | 0.15 | 1.21 | 1.30 | 1.21 |
| | 0.70 | 2.47 | 2.60 | 2.56 |
| $\frac{Var(\hat{t} \mid \text{representative reporting})}{Var(\hat{t} \mid \text{small removal reporting})}$ | 0.15 | 0.85 | 0.87 | 0.85 |
| | 0.70 | 0.63 | 0.57 | 0.57 |

$p_1 = 0.15$ and $p_1 = 0.70$. The table shows that the estimators improve substantially if reporters are those with high values of $y$, and are damaged if only low $y$ units report. This is easy to explain intuitively for $\hat{t}_{yc}$ and $\hat{t}_{y2}$, since they use the reported $y$ values directly, so when large values of $y$ are reported, less uncertainty remains for the unseen domain. It is less obvious for $\hat{t}_{yp}$, since no reported values of $y$ are used in the estimator. The explanation is that when reporters $(r_i = 1)$ have large catch the correlation between $r_i$ and $y_i$ in the population increases, reducing the variance of the ratio estimator $\hat{t}_{yp}$. This shows that reporting should be encouraged, especially for the most avid anglers.

## 4. SIMULATION STUDIES TO EXAMINE INFERENCE FOR POPULATION TOTALS

The bias in $\hat{N}$ and its standard error estimate are known to be substantial when the number of matches between the two capture periods is small. Since all the proposed estimators are related to $\hat{N}$, we wanted to examine their bias and confidence interval coverage, especially for low reporting rates. We were also interested in the performance of the estimators for complex designs as these are common in dockside samples used for validation. Therefore, we conducted simulation studies to examine these issues.

In the first study, we investigated the performance of the estimators for SRS designs under several settings for $(n_2, p_1)$, some of which resulted in a small number of matches. For each estimator, a 95 percent normal theory-based confidence interval was calculated from each simulated sample and its coverage noted. The standard errors were calculated with off-the-shelf survey software (R's Survey package) (Lumley 2004), using the Taylor series-based variance estimates for ratio estimators. The variance estimator for $\hat{t}_{y1}$ was that proposed in Olkin (1958), which ignores the variability in the estimated $W$, that is, we calculated an estimate of $V(\hat{t}_{MR})$ (in A.11). The details of the simulation study and results mentioned in Section 4 are contained in the supplementary materials. The simulation study was conducted to investigate the performance for the proposed estimators under several settings for (n2, p1) when the validation sample is SRS. Both the report has no errors and errors were considered in the simulation study. In general the results showed that confidence interval coverage was near nominal (coverage rate estimates between 93.3 percent and 95.3 percent) for all but one of the 12 settings considered. The one exceptional setting was $(n_2, p_1) = (800, 0.99)$ in a population that showed no errors in reporting. For that case, the coverage of the confidence intervals based on $\hat{t}_{yc}$ and $\hat{t}_{y2}$ was about 90 percent. So the problem with coverage was not caused by the low number of matches in this case.

A second simulation was designed to study performance of the estimators for complex designs. The population and sample designs tested were chosen to mimic some of the features of the data from the Texas iSnapper project. We created the finite population structure by replicating each primary sampling unit (PSU) in the Texas validation sample a number of times that was proportional to its weight to obtain a population of 20,590 trips. The average number of trips per PSU was 12. Then we simulated the "catch" data $y$ for each trip from a zero-truncated Poisson distribution with mean parameter 10. The simulated population total is 205,583. We examined estimators for two forms of "reported" data. For the first, we assumed perfect reporting $(y = y^*)$. For the second, erroneous reports were constructed by first computing $y^* = y + \epsilon$, where $\epsilon$ was simulated as a mean 0 normal random variable, and then $y^*$ rounded to an integer (or to 0 if negative). The variance of the normal random variable was set (by trial and error) so that the correlation between $y$ and $y^*$ in the reporting domain was 0.66. In both cases, the reporting units were simulated to be nonrepresentative, by selecting them randomly from among the units in the largest 70 percent of the $y$ values.

The validation sample was chosen according to a stratified cluster design with PSUs selected with probability proportional to size, where the size measures were those associated with the PSUs in our application data. The strata (weekday and weekend time periods) were defined as in the original data. The fraction of PSUs in the sample that were chosen from the two strata (0.56 from weekday, 0.44 from weekend) match the Texas dockside sample design. Two levels for the number of PSUs sampled (27 and 60) and the reporting rate (0.04, 0.80) were selected for the simulations, and estimates were calculated based on both the perfect and erroneous reports. Sampling was replicated 30,000 times. Then $\hat{t}_{yp}$ and $\hat{t}_{yc}$ were calculated for each sample, along with two hybrid estimators. The first was the complex sample analog of $\hat{t}_{y1}$ which takes the form of $\hat{t}_{MR}$, but with $W$ estimated from (10). The second estimator, which we denote by $\hat{t}_{y2}$, was computed by simply substituting weighted estimates $\hat{n}_1$, $\hat{t}_y, \hat{t}_{y^*}$ in (9). This estimator is not necessarily optimal even if there are no reporting errors since the design effects for the two estimators may differ, but it is still approximately unbiased and is simple to compute. For each simulation setting and replicated sample, the estimator variances were estimated using both the

Taylor series and the jackknife standard error options in *R*'s *Survey* package. Ninety-five percent normal theory-based confidence intervals were computed.

A summary of the results is shown in Table 2. For each variance estimation method/estimator/setting, three statistics describing the results of the 30,000 replicates were computed. First is the proportion of confidence intervals that covered $t_y$, which is reported in the column labeled "Coverage." Next is the average standard error of the estimate, reported in the column labeled "SE." The variance of each estimator was also computed over the replicates of the simulation, as was the average of its replicate variance estimates. The relative bias in the variance estimate was computed as the difference between the estimated and simulated variance divided by the simulated variance. It is reported as a percentage in the table as "RelBias." A negative relative bias means that the variance estimator is biased downward.

The results show that the Taylor series variance estimates do underestimate the true variance for all estimators when the number of PSUs is small ($n_2 = 27$), resulting in confidence interval coverage that is less than nominal. The jackknife estimate of variance performs better and provides closer-to-nominal coverage of the confidence intervals. $\hat{t}_{y1}$ has especially low coverage when $n_2$ is small because it is slightly biased. As predicted from the SRS analysis of section 3, $\hat{t}_{yp}$ outperforms $\hat{t}_{yc}$ for the small reporting rate, and the reverse is true for the large reporting rate. The presence of reporting errors does degrade the precision of all the estimators except $\hat{t}_{yp}$ (which does not use $y^*$), but $\hat{t}_{yc}$ still maintains its advantage. The hybrid estimators show mixed results. When the number of matches is very small (small $p_1$ and $n_2$), $\hat{t}_{y1}$ does not perform well and $\hat{t}_{y2}$ is virtually identical to $\hat{t}_{yp}$ as expected. When the number of matches is large (large $p_1$ and $n_2$), they outperform both $\hat{t}_{yp}$ and $\hat{t}_{yc}$.

## 5. EXAMPLE

Red snapper is one of the most highly targeted species in the northern Gulf of Mexico. Since the 1980s, this fishery has been listed as overfished, and drastic reductions in both season and bag limits have been implemented to help this population fully recover. Despite the stock being classified as overfished, the population is recovering rapidly, and anglers are seeing more red snapper than in previous years. Anecdotal and stock assessment reports both indicate higher abundances of snapper, but these reports continue to be met with reductions in federal season length, which was only 10 days long in 2015. This unique enigma has led to heated conflicts regarding allocation among user groups in the fishery. Much of the concern could be allayed with better data, specifically addressing the uncertainty of recreational catch estimates. With several states trying new management strategies, it is an ideal species to test the feasibility of smartphone "app" technology for private recreational anglers to report their catch. Texas was one of three states with a pilot program using the new technology for estimating

Table 2. Coverage Rate, Standard Error of the Estimate, and Bias of Variance Estimate for Each Estimator Based on 30,000 Replicates

| | | No errors in report | | | | | | | | | Errors are present in report with $R_{1,yy^*} = 0.66$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $p_1 = 0.04, n_2 = 27$ | | | $p_1 = 0.04, n_2 = 60$ | | | $p_1 = 0.80, n_2 = 60$ | | | $p_1 = 0.04, n_2 = 27$ | | | $p_1 = 0.80, n_2 = 60$ | | |
| | | Coverage | SE | RelBias | Coverage | SE | RelBias | Coverage | SE | RelBias | Coverage | SE | RelBias | Coverage | SE | RelBias |
| $\hat{t}_{yp}$ | Taylor | 0.936 | 59,554 | −0.17 | 0.944 | 36,853 | −0.05 | 0.948 | 3,307 | 0.00 | 0.936 | 59,554 | −0.17 | 0.948 | 3,307 | 0.00 |
| | Jackknife | 0.948 | 70,336 | −0.02 | 0.952 | 39,023 | 0.00 | 0.951 | 3,355 | 0.01 | 0.948 | 70,336 | −0.02 | 0.951 | 3,355 | 0.01 |
| $\hat{t}_{yc}$ | Taylor | 0.935 | 61,922 | −0.19 | 0.946 | 37,562 | −0.06 | 0.945 | 2,806 | 0.00 | 0.934 | 64,071 | −0.19 | 0.945 | 3,526 | 0.00 |
| | Jackknife | 0.948 | 74,666 | −0.02 | 0.955 | 39,913 | 0.00 | 0.948 | 2,841 | 0.01 | 0.948 | 78,313 | −0.01 | 0.947 | 3,571 | 0.02 |
| $\hat{t}_{y1}$ | Taylor | 0.885 | 54,606 | −0.15 | 0.924 | 35,696 | −0.07 | 0.940 | 2,748 | −0.02 | 0.885 | 54,660 | −0.17 | 0.940 | 3,029 | −0.02 |
| | Jackknife | 0.914 | 72,361 | 0.12 | 0.935 | 38,026 | 0.00 | 0.943 | 2,783 | −0.01 | 0.910 | 69,456 | 0.06 | 0.938 | 3,072 | −0.01 |
| $\hat{t}_{y2}$ | Taylor | 0.936 | 59,532 | −0.17 | 0.947 | 36,844 | −0.05 | 0.945 | 2,795 | 0.00 | 0.936 | 59,532 | −0.17 | 0.945 | 3,346 | 0.00 |
| | Jackknife | 0.948 | 70,336 | −0.02 | 0.952 | 38,981 | 0.00 | 0.948 | 2,829 | 0.01 | 0.948 | 70,378 | −0.02 | 0.948 | 3,394 | 0.02 |

red snapper catch in 2015. The Harte Research Institute (HRI) created iSnapper to collect catch and effort data from recreational anglers targeting red snapper in state and federal waters off the Texas coast. Here we apply the methods of this paper to estimate the total catch of red snapper in 2015 in Texas by recreational anglers in private boats using data from iSnapper.

Unlike Alabama and Mississippi, self-reporting of catch in Texas was voluntary in 2015. To validate the self-reports, HRI partnered with the Texas Parks and Wildlife Department (TPWD), which routinely samples anglers in dockside surveys using a probability sample of locations and time blocks to produce the mandated estimates of fishing effort by recreational anglers for NOAA. (See NOAA's description of its survey methods at www.st.nmfs.noaa.gov/recreational-fisheries/index.) The time blocks are stratified by weekday and weekend, while the locations have unequal selection probabilities that are proportional to a "pressure" measure, which is meant to capture the average number of anglers using a particular site in past years. To augment the TPWD sample, HRI also sampled in targeted high use marinas and boat ramps during the first six days of the federal red snapper recreational season (June 1–10, 2015). These PSUs were treated as take-all strata for estimation (i.e., given weights of 1). Catch counts were collected from every vessel intercepted during sampled shifts. Vessel registration numbers were recorded and used, along with day and time, to identify matches to trips submitted using the iSnapper app.

The number of intercepted trips in the validation sample was 421, which were clustered in 27 PSUs, with 15 and 12 in the weekday and weekend strata, respectively. The proportion of the trips previously self-reported was estimated to be only $\hat{p}_1 = 0.04$. The estimates of mean catch made from the validation sample for the population and for the self-reporters were 9 and 10, respectively. Thus the self-reporters are not representative, but rather have larger-than-average catch. The CV of catch was estimated to be 0.68. The design effect for the estimate of mean catch from the validation sample alone was about 1.4. The accuracy of reporting was high when measured as a total, with only about a 3.8 percent higher catch reported than observed in the matched sample. However, the correlation between $y$ and $y^*$ was only about 0.66 due to the fact that the erroneous self-reports were small in number but tended to be high outliers.

The estimates of catch from the four estimators, computed as described for the complex design in the previous section, are shown in Table 3. (For context,

**Table 3. Estimated Total Landings of Red Snappers in 2015 Using Four Different Estimators and Two Standard Error Estimates**

| | $\hat{t}_{yc}$ | $\hat{t}_{yp}$ | $\hat{t}_{y1}$ | $\hat{t}_{y2}$ |
|---|---|---|---|---|
| Estimate | 61,659 | 58,686 | 59,422 | 58,789 |
| SE (Taylor) | 17,793 | 17,005 | 16,907 | 16,952 |
| SE (Jackknife) | 21,723 | 21,646 | 21,462 | 21,573 |

the official 2015 estimate of red snapper catch from Texas is 32,062.) Because of the very small self-reporting rate and imperfect correlation, we would expect to find that $\hat{t}_{y1}$ and $\hat{t}_{y2}$ weight $\hat{t}_{yp}$ more heavily than $\hat{t}_{yc}$, and that is what did occur. The jackknife standard errors are larger than the Taylor standard errors and, based on simulation results, are likely to represent the true uncertainty more accurately. However, unlike the simulation results for the small number of matches case, the (jackknife) standard errors are all similar. This could be because of the larger-than-average catch of the self-reporters or some other feature of its distribution that was not captured in the simulated population.

## 6. DISCUSSION

In this paper we have developed a methodology that uses participant-provided data as a supplement to data collected by a probability sample. Such volunteer data are increasingly easy to obtain, but as we have demonstrated here, its utility depends on a variety of factors. For MRIP to consider this method of data collection as a replacement for their current catch and effort surveys for all species, they would need to find a way to ensure an adequate reporting rate. If it were possible to require reporting, as Alabama and Mississippi did for red snapper, this would be easier to accomplish. For example, Alabama's reporting rate was 25 percent for private vessels and 67 percent for charter vessels in 2015 (Alabama DCNR/Marine Resources Division 2015). Alternatively, if the most avid anglers could be incentivized to cooperate at a high rate, even if the less active anglers did not, this too could provide sufficient precision for catch estimates.

The best estimator will depend on the situation. Besides the reporting rate, the accuracy of reporting is likely to vary from one application to another. When a fish is easy to identify, like red snapper, the reporting errors are likely to be fewer than for less familiar species. For catch of those species, anglers are likely to misreport one species for another, causing low correlation between reported and observed catch for both species. Thus if self-reports were to be used for estimating catch of all species, the precision of estimates could vary widely and the best estimator to use could vary also. When the reporting rate is low, the number of matches is small, and the species is easy for anglers to identify, $\hat{t}_{y2}$ is our recommended estimator. Besides being easy to compute with standard survey software and having an intuitively attractive form, it would be near optimal among the estimators considered. When the number of matches is large, $\hat{t}_{y1}$ would be recommended. For species that anglers cannot identify easily, $\hat{t}_{yp}$ may be best. In all our simulations, the jackknife estimators of variance were at least as good, and sometimes better, than the Taylor series ones. Thus we recommend jackknife estimators of variance for any estimator.

As with any estimator, nonsampling error can be even more problematic than sampling error due to the bias it can introduce. The usual sources of nonsampling error, nonresponse, and measurement error are typically less problematic in dockside surveys of anglers than in other types, such as the effort surveys made retrospectively by phone or mail. For example, nonresponse for private anglers in the TPWD survey is about 4 percent. In addition, the interviews are conducted by biologists or technicians, who are well qualified and trained to identify species, so that measurement error is minimized.

However, there are other assumptions required in the implementation of the data collection operation that are in common with those of Capture-Recapture models. The first of these is that there is no matching error, meaning that matching the self-reports to access point encounters is error-free. This holds reasonably well in this application since red snapper angling requires a boat. Catch does not need to be associated with individual anglers, but rather just the boat trip, which can be identified with good accuracy by a unique registration number and date/time of arrival at shore. The time report is critical in order to avoid mismatches due to multiple trips per day, so anglers must be educated about the definition of return time. For species that may be caught without a boat, accurate matching will be more difficult, so no attempts to implement these methods have been made.

More problematic is the Capture-Recapture assumption that the population is closed, meaning that no members enter or leave during the sampling period. This holds only if no angler trips eligible to self-report become inaccessible to selection in the verification sample. But anglers who return from their trip to a private dock, such as one behind a home, are removed from the validation sampling frame since dockside surveys can only occur at publicly accessible locations. The access points in the frame are often referred to as public sites, though some private marinas do allow samplers to conduct dockside surveys on their properties. This is a vexing problem for all recreational angler data collection systems. In the estimation system in current use, no measure of catch is available for trips ending at private sites, though counts of these trips are obtained from the effort survey. The unverifiable assumption that catch per trip is identical for trips ending at public and private sites has to be made in order to obtain an estimate of catch. Though the Capture-Recapture approach does provide some information about catch for the private access point anglers via their self-reports $(y^*)$, we still have no source of data for $y$ for these trips. Thus the estimators will not incur bias only under the unverifiable assumption that reporting rate and accuracy is the same for public and private trips. One possible alternative that has been considered is that the verification sample could add some intercepts that occur before landing, such as at fueling sites or on-the-water encounters. This would add its own problems of assessing probabilities of selection into the verification sample, so that further work is needed on this issue.

The Capture-Recapture model also requires assumptions we describe collectively as independence. This encompasses both independence of selection in the two capture periods (selection into the 2nd sample does not depend on capture in the first), as well as homogeneity of selection probabilities in the verification sample. We have generalized the estimation method so that it accounts for known differences in selection probabilities due to the probability sample. The problem occurs when those differences are not known. For example, if the decision to self-report is influenced by selection into the validation sample, this altered probability cannot be accounted for in estimation, and thus can bias the estimator. This could occur if a returning angler was more likely to report his catch if he could anticipate that he would be in the validation sample (what Capture-Recapture methodologists would refer to as "trap happy," but with the two sampling periods having reversed labels). Care must be taken to prevent this problem by the way the sampling operation and collection of self-report data is implemented. One way to ensure angler reporting status is not influenced by being included in the validation sample would be to conduct the survey out of the view of returning anglers so that the decision to report, which must occur prior to removing fish from the boat in the mandated reporting states, cannot be influenced by the knowledge that they will be interviewed. However, this is difficult to accomplish because the most reliable access to returning anglers is at their landing point. An alternative approach is that used by Mississippi's program, which makes prior authorization for a red snapper trip mandatory. Specifically, anglers must obtain a trip ID ("open" a trip) prior to departure via the website or app. Since this decision is made by the angler prior to the validation sample encounter, it cannot be influenced by the recapture event. The angler is incentivized to provide catch information after return ("close" a trip) by disallowing issuance of a new trip ID to the angler before he/she closes any open trip. Remaining open trips are followed up by the state agency to obtain catch. If neither of these approaches is used, there is a risk of bias due to differing reporting rates for verified and unverified angler trips.

Still, despite these problems, an estimation procedure that uses angler self-reports holds promise for improving the quality and timeliness of the estimates of catch over those currently available. The decreasing response rates for household surveys nationally have been shared by the effort surveys that are part of the current MRIP estimation methodology. Their dockside access point surveys, which are the ones used here as validation samples, enjoy a much higher response rate. Since all data collection is completed at the time that the trip is made, there is potential for a much faster production of estimates than the current MRIP system since the effort survey is conducted retrospectively. Finally, fisheries management agencies report that some angler advocacy groups are anxious to provide data to improve what they perceive as inadequately precise estimates. This methodology provides a valid way to make use of their shared data.

## Supplementary Materials

Supplementary materials are available online at https://academic.oup.com/jssam. They include the details of the simulation study and results mentioned in

Section 4. The simulation study was conducted to investigate the performance for the proposed estimators under several settings for (n2, p1) when the validation sample is SRS.

# APPENDIX

## A.1 VARIANCES OF THE ESTIMATORS

As noted in (2) and (5), $\hat{t}_{yp}$ and $\hat{t}_{yc}$ are ratio estimators, so their variances can be approximated using Taylor linearization. Thus we see (e.g., Lohr 2009, eq. 4.11) that

$$V(\hat{t}_{yp}) = n_1^2 Var(\hat{B}_p) \approx \frac{N^2\left(1-\frac{n_2}{N}\right)}{n_2} S_{dp}^2, \tag{A.1}$$

where $S_{dp}^2 = \sum_{i=1}^{N}(y_i - B_p r_i)^2/(N-1)$. This residual variance can be rewritten as

$$S_{dp}^2 = S_y^2 + \bar{y}^2\left(1+\frac{1}{p_1}\right) - 2\bar{y}\bar{y}_1 \tag{A.2}$$

where $\bar{y} = t_y/N$ and $S_y^2 = \sum_{i=1}^{N}(y_i-\bar{y})^2/(N-1)$ are the mean and variance of $y$ in the entire finite population, $p_1 = n_1/N$ is the fraction of the population in the reporting domain, and $\bar{y}_1 = \sum_{i=1}^{n_1} y_i/n_1$ is the mean of $y$ in this domain. A similar computation yields the variance for $\hat{t}_{yc}$ to have a similar form to (A.1), but with residual variance

$$S_{dc}^2 = S_{dp}^2 + \frac{1}{p_1}\left(\bar{y}/\bar{y}_1^*\right)^2 S_{1y*}^2 - 2\left(\bar{y}/\bar{y}_1^*\right)S_{1,yy^*}, \tag{A.3}$$

where $\bar{y}_1^* = t_{y^*}/n_1$ is the mean of $y^*$ in the reporting domain. The covariance of the two estimators also has the form shown in (A.1), but with residual covariance

$$S_{dp,dc} = S_{dp}^2 - (\bar{y}/\bar{y}_1^*)S_{1,yy^*}. \tag{A.4}$$

Next we consider the variance of the optimally weighted average of these two estimators as defined in (6). Its variance is (Cochran 2007 (6.100))

$$V(\hat{t}_{MR}) = \frac{V(\hat{t}_{yp})V(\hat{t}_{yc}) - Cov^2(\hat{t}_{yp},\hat{t}_{yc})}{V(\hat{t}_{yp}) + V(\hat{t}_{yc}) - 2Cov(\hat{t}_{yp},\hat{t}_{yc})}, \tag{A.5}$$

The covariance of the two ratio estimators is

$$Cov(\hat{t}_{yp}, \hat{t}_{yc}) \approx \frac{N^2\left(1-\frac{n_2}{N}\right)}{n_2} S_{dp,dc} = \frac{N^2\left(1-\frac{n_2}{N}\right)}{n_2}\left\{S_{dp}^2 - \left(\bar{y}/\bar{y}_1^*\right)S_{1,yy^*}\right\}, \quad \text{(A.6)}$$

where $S_{1,yy^*} = \sum_{i=1}^{n_1}(y_i - \bar{y}_1)(y_i^* - \bar{y}_1^*)/(n_1 - 1)$ is the covariance between $y$ and $y^*$ in the reporting domain. Then from (A.1) through (A.6), we have

$$V(\hat{t}_{MR}) \approx \frac{N^2\left(1-\frac{n_2}{N}\right)}{n_2}\left(S_{dp}^2 - p_1 S_{1,yy^*}^2 / S_{1y^*}^2\right), \quad \text{(A.7)}$$

where $S_{1y^*}^2 = \sum_{i=1}^{n_1}(y_i^* - \bar{y}_1^*)^2/(n_1 - 1)$ is the variance of $y^*$ in the reporting domain.

Finally, it can be observed from (9) that $\hat{t}_{y2}$ can be written as a constant ($t_{y^*}$) plus a ratio estimator

$$\hat{t}_{y-ry^*} = n_1 \frac{\sum_{i\in S_2}\left(y_i - r_i y_i^*\right)}{\hat{n}_1}.$$

Therefore, the variance of $\hat{t}_{y2}$ can also be approximated using Taylor linearization, yielding

$$V(\hat{t}_{y2}) \approx \frac{N^2\left(1-\frac{n_2}{N}\right)}{n_2}\left\{S_{dp}^2 + p_1\left(S_{1y^*}^2 - 2S_{1,yy^*}\right)\right\}. \quad \text{(A.8)}$$

In order to facilitate comparison of these variances, it is helpful to rewrite them in canonical form as follows:

$$V(\hat{t}_{yp}) = \frac{t_y^2\left(1-\frac{n_2}{N}\right)}{n_2}\left\{CV_y^2 + \left(1+\frac{1}{p_1}\right) - 2\left(\frac{\bar{y}_1}{\bar{y}}\right)\right\}; \quad \text{(A.9)}$$

$$V(\hat{t}_{yc}) \approx V(\hat{t}_{yp}) + \frac{t_y^2\left(1-\frac{n_2}{N}\right)}{n_2}\left\{\frac{CV_{1y^*}^2}{p_1} - 2\left(\frac{\bar{y}_1}{\bar{y}}\right)R_{1,yy^*}CV_{1y}CV_{1y^*}\right\}; \quad \text{(A.10)}$$

$$V(\hat{t}_{MR}) \approx V(\hat{t}_{yp}) - \frac{t_y^2\left(1-\frac{n_2}{N}\right)}{n_2}\left\{p_1\left(\frac{\bar{y}_1}{\bar{y}}\right)^2 R_{1,yy^*}^2 CV_{1y}^2\right\}; \quad \text{(A.11)}$$

$$V(\hat{t}_{y2}) \approx V(\hat{t}_{yp}) + \frac{t_y^2\left(1-\frac{n_2}{N}\right)}{n_2}\left\{p_1\frac{\bar{y}_1^*}{\bar{y}}CV_{1y^*}\left(\frac{\bar{y}_1^*}{\bar{y}}CV_{1y^*} - 2\frac{\bar{y}_1}{\bar{y}}R_{1,yy^*}CV_{1y}\right)\right\}. \quad \text{(A.12)}$$

## REFERENCES

Alabama DCNR/Marine Resources Division (2015), "Summary of Alabama's 2015 Mandatory Red Snapper Reporting Program," Available at http://www.outdooralabama.com/sites/default/files/preliminary results of alabama's 2015 red snapper reporting.pdf.

Alho, J. M., M. H. Mulry, K. Wurdeman, and J. Kim (1993), "Estimating Heterogeneity in the Probabilities of Enumeration for Dual-system Estimation," *Journal of the American Statistical Association*, 88(423), 1130–1136.

Baker, R., J. M. Brick, N. A. Bates, M. Battaglia, M. P. Couper, J. A. Dever, K. J. Gile, and R. Tourangeau (2013), "Summary Report of the AAPOR Task Force on Non-probability Sampling," *Journal of Survey Statistics and Methodology*, 1(2), 90.

Berkson, J. (1950), "Are There Two Regressions?" *Journal of the American Statistical Association*, 45(250), 164–180.

Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu (2006), *Measurement Error in Nonlinear Models: A Modern Perspective*, Boca Raton, FL: CRC Press.

Cochran, W. G. (2007), *Sampling Techniques*, New York: John Wiley & Sons.

Groves, R. M. (2011), "Three Eras of Survey Research," *Public Opinion Quarterly*, 75(5), 861–871.

Johnson, N. L., and S. Kotz (1970), *Distributions in Statistics: Continuous Univariate Distributions*, *Vol.* 1, New York: Houghton Mifflin.

Le Cren, E. D. (1965), "A Note on the History of Mark-Recapture Population Estimates," *The Journal of Animal Ecology*, 34, 453–454.

Lee, S., and R. Valliant (2009), "Estimation for Volunteer Panel Web Surveys Using Propensity Score Adjustment and Calibration Adjustment," *Sociological Methods & Research*, 37(3), 319–343.

Lohr, S. (2009), *Sampling: Design and Analysis, Pacific Grove*, CA: Duxbury Press.

Lumley, T. (2004), "Analysis of Complex Survey Samples," *Journal of Statistical Software*, 9(1), 1–19.

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

Olkin, I. (1958), "Multivariate Ratio Estimation for Finite Populations," *Biometrika*, 154–165.

Pollock, K., S. Turner, and C. Brown (1994), "Use of Capture-Recapture Techniques to Estimate Population Size and Population Totals When a Complete Frame is Unavailable," *Survey Methodology*, 20, 117–124.

Tarnecki, J. H., and W. F. Patterson III (2015), "Changes in Red Snapper Diet and Trophic Ecology Following the Deepwater Horizon Oil Spill," *Marine and Coastal Fisheries*, 7(1), 135–147.

Tourangeau, R., F. G. Conrad, and M. P. Couper (2013), *The Science of Web Surveys*, New York: Oxford University Press.

Wolter, K. M. (1986), "Some Coverage Error Models for Census Data," *Journal of the American Statistical Association*, 81(394), 337–346.