

Consultant's Report: Review of Ocean Recreational Boat Survey

F. Jay Breidt* and Jean D. Opsomer†
Colorado State University

July 27, 2010

1 Introduction

During the two-day meeting in Newport on July 19-20, 2010, we met with Oregon Fish and Wildlife staff to discuss the Oregon Department of Fish and Wildlife's Ocean Recreational Boat Survey (ORBS). In this document, we will provide our initial reaction to the survey procedures we learned about during the meeting. Our goal is to initiate a discussion on a range of possible improvements to the ORBS.

We begin by briefly summarizing our understanding of the survey itself. ORBS provides timely estimates of effort and catch, which are used for in-season management of key fisheries. The sampling effort for ORBS is focused on March through October and on major ports, with lower sampling rates outside of the main season and outside of the major ports. Currently, some times at some ports have no chance of selection into the sample.

Sampling is finely stratified, in space (by port) and in time (by week, and season within week when seasons open or close mid-week). Fishing effort is collected from a combination of sources, depending on the port. Effort is collected separately for charters (commercial guides with an identified office space) and other guides and private boats. Trip counts by type are collected for the entire week from charters. Efforts for private boats and guides are

*jbreidt@stat.colostate.edu.

†jopsomer@stat.colostate.edu.

obtained from bar crossing counts, where applicable, or from counts of empty slips and boat trailers. In some locations, bar crossing counts are obtained from an observer, and others are obtained by review of video. The ability to obtain daily effort estimates through monitoring of departures of fishing vessels is very helpful in obtaining accurate estimators of catch.

Dockside interviews are conducted to obtain catch information. Field crews by port range from 1–3 samplers. They are assigned blocks of time during which to conduct interviews, and record catch for all anglers on selected boats. Boats are selected within a time block in a systematic fashion.

According to ORBS staff, samplers have access to private landings, and night fishing is extremely rare. Thus, two potentially problematic issues that may lead to bias in other fisheries surveys seem to be largely non-issues in ORBS.

Our first reaction to ORBS is that it has many attractive features that simplify its analysis, relative to other fisheries surveys in our experience:

- large and thorough sampling effort
- fine spatial and temporal stratification
- required compliance by anglers
- census of charter efforts
- (almost) direct measures of effort, due to geography; relatively few sites are suitable for launching ocean boats
- possible bias due to lack of access to private sites seems to be a non-issue
- possible bias due to unsampled night fishing seems to be mostly a non-issue

In the remainder of this report, we outline our recommendations for possible improvements to the ORBS, as well as a number of areas where further study might be warranted.

2 Preliminary Findings and Recommendations

2.1 Small Area Estimation

The classical problem of small area estimation is to use a model to “borrow strength” across space and/or time to get estimates at a fine spatio-temporal resolution, meaning finer than the resolution supported by only the sample data occurring within the spatio-temporal cell. Our impression is that ORBS is reporting estimates at the level of small areas, but without small area models or estimates of precision. Given the objectives of ORBS, reporting at such a fine resolution seems unnecessary:

- quotas are coastal or regional, not port-specific
- quotas are seasonal, not weekly.

Our preliminary recommendation is to avoid “volunteering” to report port-week-species level estimates, and backing away from such reports wherever practical. Estimates for many objectives of interest are already being achieved with high levels of precision, and these successes should be emphasized.

2.2 Sample Size and Issues of Probability Sampling

For all of the major ports in high season, ORBS has a major sampling effort, dedicated to achieving 20% sampling fraction by port/week. From general sampling principles, it makes sense to target high-volume sites and times. Our reaction was, however, that even allowing for the mandatory 20% capture, it might be possible to reallocate some of the sampling effort to achieve other purposes. These could include gathering more information for smaller ports, for months outside the main season, or for rarer species.

Currently, some ports and some months have zero probability of selection. In sampling terminology, this is an undercoverage problem, which leads to the possibility of bias in estimation of some target parameters if the “uncovered” part of the population differs from the “covered”, sampled part of the population. Even if the uncovered part of the population is similar to the covered part *now*, bias due to undercoverage can arise over time in a dynamic population. For example, while boats may almost never go out from some ports in winter *now*, this may change as anglers obtain better gear (e.g.,

GPS) or target different species in the future. One example that ODFW has already encountered is the targeting of tuna by recreational anglers. A second example is changing site characteristics, such as the changing erosion state of a beach meaning that sometimes it is possible and sometimes not possible to launch from shore.

For an “uncovered” part of the population, there is by definition no possibility of information obtained in a sample, so only extrapolation from the covered part of the population is possible. Lynn Mattes presented an excellent example of the possible problems of such extrapolation. In her example, smaller ports not sampled in winter had an *ad hoc* adjustment using

$$\frac{(\text{summer for small port})}{(\text{summer for big port})}(\text{winter for big port}) \equiv (\text{winter for small port}).$$

Intuitively, there are clear problems with this assumption, but there is no way to fix the problems on the basis of sample data. With a probability sample, even one with a small sample size, it would be possible to develop a solution.

Our recommendation is, whenever possible, to move in the direction of a full probability sample of the population of interest by reallocating resources beyond those needed to achieve sufficient precision for the large ports in the main season. This could be done with a relatively small reallocation of the full sampling effort.

2.3 Weighting

Ideally, weights for estimation of means and totals in a probability sample are obtained as inverses of inclusion probabilities. Such weights guarantee unbiased estimation: the average of the estimator over all possible samples from the population is exactly the population parameter. For any given sample, the estimate may be higher or lower than the population parameter, but the variation of the estimates around the true population parameter can also be estimated from the sample data. This makes it possible to construct valid confidence intervals, which contain the true population parameter in a large and predetermined fraction of all possible samples.

In ORBS, there are typically two stages of selection. In the first stage, blocks of time within days within a week are selected for assignment to interviewers. Blocks appear because the interviewer assignments are typically shorter than the full fishing day. It is not clear to what extent blocks of time

are randomly selected. In the second stage, interviewers select boat trips within assigned time blocks. It is not possible for interviewers to enumerate boat trips within an assignment and randomly sample from the list, so systematic procedures are employed. To the extent that interviewers follow the “next boat” protocol, it may be reasonable to approximate the second stage of selection as simple random sampling without replacement from all boat trips returning during the time block.

Consider a particular port-week and let π_{db} = probability of selecting block b on day d . Let N_{db} denote the total number of returning boats in block b on day d and let n_{db} denote the intercepted number of returning boats. If N_{db} were observed, then a set of weights could be constructed as

$$w_{db} = \frac{1}{\pi_{db}} \frac{N_{db}}{n_{db}}.$$

In these weights, the second factor expands the n_{db} intercepted trips in the day-block to the N_{db} total trips in the day-block. The first factor expands the sampled day-blocks to the set of all day-blocks.

Weights $\{w_{db}\}$ are currently infeasible, since N_{db} is not observed. Instead, $N_d = \sum_b N_{db}$ = total effort for the day is observed, or at least well-estimated through a separate measurement (census of trips for charters, bar crossing count for private trips). Hence, a number of alternative weighting schemes could be investigated, for instance by using day weights (which would ignore the blocks within the days) or even multi-day weights (pooling blocks and days within a week). The latter is most similar to the method currently in use.

A critical consideration will be whether the bias due to using approximate weights is sufficiently small to be ignorable. The biases (and variances) of these various approaches depend on within-day (across blocks in the day) versus within-week (across days of the week) variation. It should be possible to characterize the bias analytically, approximate the variance, and derive variance estimation strategies for each. The various methods could also be compared via simulation, and using historical data.

2.4 Variance Estimation

Once an appropriate weighting scheme for the data is developed, it will be possible to construct a design-based variance estimation procedure. As noted

above, the design specified for purposes of weighting and variance estimation will be an approximation to the actual design implemented in the field.

An early goal in the review process will then be to put current data into the framework of a data set with the following elements:

- stratum identifiers (these can be collapsed strata for the purposes of variance estimation)
- primary sampling unit identifier: day or block within day (for proper two-stage variance estimation)
- sampling weight
- sampling fractions within strata (taking advantage of finite population corrections)
- response variables

Once the data set is in this form, point and variance estimation can be conducted using existing statistical software, including the `survey` package in R or `proc surveymeans` in SAS, among others. Use of existing software eliminates the need for a new programming effort and ensures that well-documented best practices are being employed.

2.5 Auxiliary Data

Effort estimation in particular may benefit from the use of auxiliary data. In some ports in high season, there is essentially a census of effort for charters, and in other ports and times, charter data may be available. Charter data may have some explanatory power for non-charter effort, and it would be worth exploring this possibility whenever a charter census is conducted along side a non-charter sample. In addition, weather, bar conditions, ocean conditions, and (where relevant) river conditions may have some explanatory power for effort, particularly in the off-season when other information may be difficult and costly to obtain. Note that even if regression relationships are imperfect, auxiliary data may be very useful in producing more efficient estimators using “model-assisted estimation.” Like direct survey estimates, model-assisted estimators are design-unbiased or nearly so, and allow for consistent variance estimation and proper confidence interval construction (even if the regression model is imperfect). If the regression model has reasonable

explanatory power, the model-assisted estimator has smaller variance and narrower confidence intervals than the direct estimator that ignores auxiliary data.

2.6 Codifying Subject-Matter Expertise

Our impression from this preliminary review is that some key parts of the estimation process require manual input from a subject-matter expert, Eric Schindler. These include, for example, decisions on whether to eliminate early-returning trips. Wherever possible, it would be better to replace these manual adjustments by developing rule-based procedures and implementing them in the estimation code. This yields a reproducible and transferable methodology that is documented in the estimation software. Further documentation is also desirable. Development and documentation of such rule-based procedures shields the organization from potential criticism, and has the potential side benefit of allowing rigorous simulation testing of estimation methods. This would not be possible if every replication in a large simulation experiment required manual input!

3 Conclusion

The Oregon Department of Fish and Wildlife has done an excellent job conducting and improving ORBS, as noted at the beginning of this report. The recommendations for further improvements in the six subsections cover a range of issues, some of which would require further investigation. In particular, the possible sample redesign to capture less-frequented ports and times (§2.2), the most appropriate approach for weighting (§2.3) and the use of auxiliary data to increase the precision of estimators (§2.5) will require further study in order to determine how to best implement them.

Survey Review Final Status
Marine Recreational Information Program

Provider Name: **Maggie Sommer**

Survey: **Oregon Recreational Boat Survey (ORBS)**

Date of Review: **7/27/10**

Date of Final Response: **1/27/12**

Provider Instructions: Read the review and provide feedback if desired. Feedback includes accuracy, usefulness, and potential to implement recommendations. Comments on the review process are also welcome.

1. Accept final report: Yes No

2. Submitted MRIP proposal(s) in response to review: Yes No

3. Formal Feedback Provided: Yes No
 - 3a. Type of formal feedback provided: Corrections Comments

 - 3b. Corrections incorporated in final report: Yes No

 - 3c. Comments attached: Yes No

Notes:

Comments written by Eric Shindler

ODFW Comments on this report

2.1 Small Area Estimation

How would moving away from a port based expansion encompass the coded wire tag expansion needs? Resolution of CWT recoveries at the port level (especially for the recreational fishery) is very valuable.

2.2 Sample Size and Issues of Probability Sampling

I don't know if I was able to adequately explain the staffing needs relative to the effort during our meeting. Staffing is lined up for the season based on anticipated effort levels, port characteristics, and meeting the 20% minimum rate during the salmon seasons. Pulling a whole sampler from one port will likely result in not meeting the 20%.

Apparently, my description of the reason why boats would launch at Tierra del Mar vs. Pacific City created a overtly dramatic image of an effort shift. In a hope to clarify this, boats only launch at Tierra del Mar if the beach at Pacific City is in poor shape. However, only a very small part of the Pacific City fleet will ever launch at Tierra del Mar, and the Pacific City sampler is instructed to address such launches should they occur.

Within our current frame and funding, as noted above in an earlier sticky, this is not practical as assigned resources in the prime season are planned to meet the minimum. There may be a means by which to reassign some existing winter samplers where the strata is at the month level to collect some data from smaller ports. In the end, I still see a need for more resources to address the off-season small port issues.

2.3 Weighting

Based on the discussion at the meeting and the expressed importance of weighting, we will need to investigate the feasibility of weighting, and whether this is the best approach to resolving perceived bias issues with sampling assignments.

2.5 Auxiliary Data

The model approach using alternate data sources is, at least, very intriguing. There certainly could be any number of possible directions to explore along

these lines... but they will require both reliable data sources, likely from outside of ODFW, and the time to evaluate.

2.6 Codifying Subject-Matter Expertise

I probably made myself out to be more of an expert than I am... and the manual inputs to the process are not nearly as time of effort consuming as they probably appear. However, the point that rules should be documented to make the methods more transparent, repeatable, and possibly automated is well taken, and see this as a task to put at the top of the "to do list" for this winter... after all, you never know when a key player might get hit by a bus ;)