Oregon Ocean Recreational Boat Survey Point & Variance Estimation

FY 2012 Proposal

Daniel Erickson, Maggie Sommer
Created: 05/13/2015

1. Overview

1.1. Sponsor
Russell Porter

1.2. Focus Group
Survey Design and Evaluation

1.3. Background
Consultant and State Agency final review and determination of the statistical method for computation of variance estimates for the Oregon ORBS (Ocean Recreational Boat Survey)

1.4. Project Description
The Oregon Ocean Recreational Boat Survey (ORBS) provides estimates of angler effort and catch in the boat-based ocean sport fishery out of Oregon ports. Data collected via angler interviews and access-point effort counts are expanded to develop total effort and catch estimates. In 2010, MRIP supported a review of the ORBS survey and estimation methodologies, led by consulting experts in statistical survey design. One of the outcomes of this review was a recommendation to develop an appropriate weighting scheme so that estimation more closely matches how the data is collected and which would allow for the construction of design-based variance estimation procedures.

1.5. Public Description

1.6. Objectives
Jean Opsomer and Bryan Wright undertook initial work on developing weighting and variance estimation schemes. They used the statistical software R, and a small subset of existing ORBS data. Results of this “test” effort were promising, suggesting that the approach was suitable. Substantial additional work is now required to continue development of methodology and R code for variance estimation methodology that will be suitable for use with the entire ORBS dataset, as well as quality control code development and testing. This will enable the construction of confidence intervals around the ORBS estimates, resulting in improved estimates and understanding of their precision.

1.7. References
The proposed project is based on recommendations found in: F.J. Breidt and J.D. Opsomer, Consultant’s Report: Preliminary Review of Oregon Ocean Recreational Boat Survey, Colorado State University, July 27, 2010.

2. Methodology

2.1. Methodology
A statistician with expertise in survey design and R software will lead the effort to develop weighting and variance estimation procedures and programming, after a preliminary stage of familiarization with the ORBS design, dataset, and prior related work. The contractors who conducted the 2010 ORBS review, and appropriate ODFW personnel, will consult and assist.

2.2. Region
Pacific

2.3. Geographic Coverage
Data used are from ports on the entire OR coast. This project to occur primarily in Corvallis, OR.

2.4. Temporal Coverage
Data: year-round. This project to occur in 2012.

2.5. Frequency
Data collection occurs on a daily basis. Estimates made at various temporal scales.

2.6. Unit of Analysis
Angler effort

2.7. Collection Mode
Will use existing data for weighting and variance estimation methodology development & testing

3. Communication

3.1. Internal Communication

Project update meetings will be scheduled every other week during the project duration. These meetings will occur primarily via conference call as project personnel are in different locations. Other internal communication will occur via email, phone, or face-to-face meetings as necessary.

3.2. External Communication

Monthly reports will be submitted to the MRIP OT in the required template. A detailed final report will describe the methods used, results obtained, challenges encountered, and recommendations (if any) for follow-up work or next steps.

4. Assumptions/Constraints

4.1. New Data Collection

N

4.2. Is funding needed for this project?

4.3. Funding Vehicle

Pacific RecFIN Grant

4.4. Data Resources

All individuals included in this proposal have indicated that they are willing and available to work on this project; each will be doing so in addition to other professional commitments. This proposal assumes that all will be able to devote adequate time to this project. Constraints include working with output from an existing program written by an ODFW programmer (retired) which manipulates the raw ORBS interview and effort count data and stores output in an Access database.

4.5. Other Resources

None.

4.6. Regulations

None.

4.7. Other

None.

5. Final Deliverables

5.1. Additional Reports

None.

5.2. New Data Set(s)

The final product of this project will be appropriate weighting and variance estimation procedures.

5.3. New System(s)

Weighting and variance estimation procedures--R code and written documentation.

6. Project Leadership

6.1. Project Leader and Members

First Name	Last Name	Title	Role	Organization	Email	Phone 1	Phone 2
Eric	Schindler	ORBS Project Leader	Team Member	ODFW	eric.d.schindler@state.or.us	541-867-4741	

First Name	Last Name	Title	Role	Organization	Email	Phone 1	Phone 2
Maggie	Sommer	Marine Technical & Data Services Section Leader	Team Leader	ODFW	maggie.sommer@state.or.us	541-867-0300 x227	
Bryan	Wright	Biometrician	Team Member	ODFW	bryan.e.wright@state.or.us	541-757-4186 x225	

7. Project Estimates

7.1. Project Schedule

Task #	Schedule Description	Prerequisite	Schedule Start Date	Schedule Finish Date	Milestone
1	Project start. Project team planning meeting(s)/call(s) to bring contractor up to speed.		07/01/2012	08/31/2012	
2	Weighting & variance estimation methodology development & testing.	1	07/01/2012	06/30/2013	
3	Report writing. Document methodology, results, issues, follow-up recommendations.	2	05/01/2013	06/30/2013	Y

7.2. Cost Estimates

Cost Name	Cost Description	Cost Amount	Date Needed
Corvallis, Oregon Consultant through Oak Management	Cost for recent PhD graduate that is familiar with this project and work on this type task previousl	$45000.00	04/15/2012
Oak Management Consutant Assistance	Costs for consultant time by Ginny Lesser, Oregon State University; Jay Breidt & Jean Opsommer, CSU	$20740.00	04/15/2012
ODFW Portion of Variance and point estimate	Cost for ODFW time working on this task	$13354.00	07/01/2012
TOTAL COST		$79094.00	

8. Risk

8.1. Project Risk

Risk Description	Risk Impact	Risk Probability	Risk Mitigation Approach
Loss of interest or loss of participation by contractor(s).	Inability to complete project.	Low	Establish and maintain communication between all project team members. Stay on schedule.

9. Supporting Documents

"Consultant's Report: Review of Ocean Recreational Boat Survey", page 1

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 ;)

MARINE
RECREATIONAL INFORMATION
PROGRAM

Oregon Ocean Recreational Boat Survey
Point & Variance Estimation

FY 2011 Final Project Report

Maggie Sommer
Oregon Department of Fish & Wildlife
Marine Resources Program

April 25, 2014

OVERVIEW

Background

The Oregon Ocean Recreational Boat Survey (ORBS) conducted by the Oregon Department of Fish and Wildlife (ODFW) provides estimates of angler effort and catch in the boat-based ocean sport fishery out of Oregon ports. Data collected via angler interviews and access-point effort counts are expanded to develop total effort and catch estimates. In 2010, NOAA Fisheries' Marine Recreational Information Program (MRIP) supported a review of the ORBS survey and estimation methodologies, led by consulting experts in statistical survey design. One of the outcomes of this review was a recommendation to develop an appropriate weighting scheme so that estimation more closely matches how the data is collected and which would allow for the construction of design-based variance estimation procedures. Jean Opsomer (MRIP consultant) and Bryan Wright (ODFW biometrician) undertook initial work on developing weighting and variance estimation schemes. They used the statistical software R, and a small subset of existing ORBS data. Results of this "test" effort were promising, suggesting that the approach was suitable.

Project Description

This project was a review of the statistical method for computation of variance estimates for the Oregon ORBS, and development of computer code to accomplish the estimation. The work was done primarily by a contractor, Leigh Ann Starcevich, under the guidance of Virginia Lesser (MRIP consultant), with substantial additional work, consultation and review by ODFW staff members Bryan Wright and Eric Schindler (ORBS Project Leader).

Objectives

The objectives of this project were to continue development of methodology and R code for variance estimation that will be suitable for use with the entire ORBS dataset. Work was to include quality control code development and testing. The ultimate goal was to enable the construction of confidence intervals around the ORBS estimates, resulting in improved estimates and understanding of their precision, for consideration by fishery managers, scientists, and other consumers of ORBS estimate data.

METHODOLOGY

Leigh Ann Starcevich led the effort to develop weighting and variance estimation procedures and programming, after a preliminary stage of familiarization with the ORBS design, dataset, and prior related work.

Existing data from year-round sampling in ports on the entire Oregon coast were used for weighting and variance estimation methodology development and testing. The project work occurred in Corvallis and Newport, Oregon, in 2012-2013.

RESULTS

The contractor provided a report (attached) documenting her analysis of some of the ORBS survey design elements and the estimation procedures, and R code for design-based variance estimation. Recommendations were made for consideration of modifications to the ORBS survey design with the goal of increasing the precision and accuracy of Oregon ocean recreational fishing parameter estimation.

DISCUSSION

Although the intended scope of this project was to develop methodology and R code for variance estimation, the consultant dedicated considerable effort toward analysis of the details of ORBS survey design, and the recommendation of changes in the conduct of ORBS. While a valid consideration in the overall evaluation of the effort and catch estimates produced by ORBS, the survey design itself has been the subject of other thorough reviews by both ODFW staff and external statistical experts.

ODFW received a first draft report and R code near the end of the project period, with little time remaining for work by the consultant. As a result, there was no opportunity for significant changes to the report and the R code in response to comments and requests made by ODFW upon review of the draft deliverables. Substantial revision and testing of the R code by ODFW staff is currently ongoing. Unaddressed issues were incorporated into the report as recommendations for future work. Several additional recommendations are tangential to the key focus area of this project. The recommendations found in the report are copied here, with responses from ODFW:

1. Use randomization of days within Strata. If weekend days are of more interest, stratify by Weekday Group, census the weekend stratum, and use true randomization to select days within the Workweek stratum for unbiased inference.

   ODFW comments: (1) Weekend days and weekdays are of equal interest. In some ports/seasons, more fishing effort occurs on weekend days; therefore, more sampling effort is allocated in order to maintain a relatively constant sample rate. (2) True randomization is not possible due to logistical constraints on the field sampling operations; for example, scheduling in smaller ports where only one sampler is stationed.

2. Allocate sampling effort to all strata, including all Weeks for all Ports within the weeks not defined as Winter ORBS.

   ODFW comments: The expense of expanding the sampling program to allocate sampling effort as recommended here is prohibitive. Fishing effort in unsampled ports/times is estimated and expanded for using established procedures. The MRIP program recently supported a year of supplemental ORBS sampling in order to obtain current effort rates from traditionally unsampled ports/times. Results will be used to update the estimate calculations to ensure that expansion factors are based on current effort and catch data.

3. Track the interview shift times more closely for potential stratification of time blocks within a day.

   ODFW comments: This would create new strata for which no sample data are available and which would require additional steps in the estimation procedures. ODFW believes that this would increase the variance issues and add unnecessary complexity and uncertainty to the estimates.

4. Record days that interviewers encounter no boats to obtain accurate information on survey effort and cost. Consider randomizing interview time blocks within a day.

   ODFW comments: ODFW currently has information on days when no boats were encountered, although it is not simple to access due to the way it is recorded and stored. Randomizing interview time blocks within a day would be subject to similar logistical challenges as those described in Recommendation 2, and might result in a lower sampling rate by directing effort away from the core midday period during which most ocean recreational fishing trips return to port.

5. Assess impact of night returns for halibut and tuna by assessing video counts. Eric Schindler thought this could be initially examined at the ports in Newport and Charleston by comparing actual video counts of late returns and late and early departures to expanded counts.

   ODFW comments: This is feasible in most ports/times, and would require dedication of some staff time. There may be difficulties in some ports such as Charleston due to a longer view distance between the camera and the vessels.

6. Combine inference for strata with lonely PSUs. The proper treatment of this issue requires explicit coding in R. The R *survey* package apparently cannot handle more than one lonely-PSU approach, and domain-level estimates cannot be combined by stratum. If this is too difficult, evaluate the effect of using the "adjust" option for strata with $D_h = 1$.

   ODFW comments: Agreed; further study of this issue needed.

7. Ratio estimation of population totals may improve estimation for private boat trips if exit count data are more precise and accurate than design-based estimates of boat trips from the Horvitz-Thompson estimator.

   ODFW comments: OFDW is currently coding this recommendation. ODFW staff had noted the omission of ratio estimation of totals as one of the key areas needing further work upon review of the draft code/report. It was not implemented in the final code delivered by the consultant and has been one of the primary areas of follow-up development by ODFW.

8. Eric Schindler expressed a need to change the way trip types are recorded so that species-level estimation may be obtained. Currently, angler effort cannot be distinguished below the gross trip type strata and there is no means to assess target species groups when anglers target more than one species group on a trip. Eric Schindler proposed that interviewers collect information on the species anglers are fishing for and then the software would automatically assign the trip type. This would provide further detail by which to assess actual fishing effort for species groups while preserving the existing trip type effort estimation.

ODFW comments: This would be valuable. It would require significant effort in the form of development/modification of code in both the mobile field data collection units (currently Trimble Nomads) used by ORBS samplers, and the ORBS database.

9. Private boat exit counts are currently treated as known but are actually imputed and adjusted. Incorporate these adjustments into the estimation process and account for the additional uncertainty.

   ODFW comments: Very few actual adjustments are made, with the exception of effort data from Astoria-area ports (in order to account for boats that left port but did not enter the ocean and fished only in the Columbia River estuary, vs. ocean trips), and for days when fog significantly impaired the normal effort counts. These adjustments require case-specific decisions by a subject matter expert (ORBS Project Leader) and cannot be automated with acceptable outcomes. Some of the processes referred to as "adjustments" by the consultant are error-checks and result only in the correction of errors when found, rather than any across-the-board modifications to effort data.

CONCLUSIONS

This project resulted in progress toward the creation of variance estimates. ODFW will build upon this progress with further development and testing of the R code, and comparison of the existing estimation design with the design proposed by the consultant. ODFW will evaluate the merits of the proposed design and implement any changes determined to result in improved results. Information on any changes, along with point and variance estimates when available, will be provided to the Pacific States Marine Fisheries Commission's Recreational Fishery Information Network (RecFIN).

ATTACHMENTS

Appendix A: L.A.H. Starcevich, Consultant's Report: Design-based Analysis for the Oregon Recreational Boat Survey: Weighting Approach and Analysis Methods, Corvallis, OR, December, 2013.

REFERENCES

This project was based on recommendations found in:

F.J. Breidt and J.D. Opsomer, Consultant's Report: Preliminary Review of Oregon Ocean Recreational Boat Survey, Colorado State University, July 27, 2010.

Design-based Analysis for the Oregon Recreational Boat Survey

Weighting Approach and Analysis Methods

Leigh Ann Harrod Starcevich

Statistical consultant
PO Box 1032
Corvallis, OR 97339
lah@peak.org

December 2013

Oregon Sampling Project
Marine Resources Program
Oregon Department of Fish and Wildlife

TABLE OF CONTENTS

INTRODUCTION

The Oregon Recreational Boat Survey (ORBS) is an ocean sampling project employed by the Oregon Department of Fish and Wildlife (ODFW) to estimate annual recreational ocean fishing effort and catch. ORBS incorporates bar crossing counts with angler surveys to obtain estimates of angler effort and success. In 2010, the Marine Recreational Information Program (MRIP) funded a review of ORBS by statistical experts. This review resulted in several recommendations on how to improve data collection and estimation, specifically variance estimation so that the precision of estimates may be assessed. This report addresses many of those recommendations and provides weighting and estimation approaches for improved inference from ORBS data.

SAMPLING DESIGN

The ORBS sampling design does not employ classic randomization and is described as "representative and opportunistic" (ORBS 2012). Therefore, obtaining the inclusion probabilities for design-based estimation is not straightforward. Previous ORBS analysis treats this design as a stratified one-stage sampling design with boats pooled across days. Because effort is allocated independently within a day, the proposed analysis is treated as a stratified two-stage random sample. Strata include the Port, Boat Type (charter or private), and Week. A partial-week stratification, known as Expansion Type (ET), is employed if there is a significant change within a week regarding which species are open. For example if the ocean is only open for coho salmon (a key target species) on Thursday through Saturday within a week, then that week would be partitioned into a Monday through Wednesday plus Sunday stratum and a stratum including Thursday through Saturday. Only a single ET split is allowed within any given week. Domains of interest include the Area of catch, Species of catch, and Trip Type (bottomfish, combination – salmon *plus* another target species, dive – spear fishing only, halibut, non-fishing, salmon, or tuna) for each boat trip.

Within each stratum, a two-stage sampling design is assumed in this analysis. Individual *days* serve as primary sampling units (PSUs), and individual *boat trips* within a day define the secondary sampling units (SSUs). In the first-stage sample, the selection of days within each week is not strictly random. Days are prioritized relative to angler effort, so interview effort occurs more frequently and with higher intensity on weekend days than during the week.

In the second-stage sample, ORBS interviewers encounter returning boats with systematic random sampling. Sampling differs between the two boat types. Information on charter boat trips is obtained each morning prior to the interviewer shift. Interview effort is allocated among all charter trip types throughout the week or ET stratum so that each trip type is represented for estimation. A defined frame does not exist for private boats, so the interviewers do not know in advance how many boats they will encounter.

Interviewers use what is referred to as “the next boat protocol” to select private boats to be interviewed. Simply stated, the next boat observed to enter the area where an interviewer is working at the time they are ready to interview another boat is selected to be sampled. Because effort at the second stage is allocated between the two boat types prior to the survey, boat type is treated as a second-stage stratification variable. While the length of interview shifts is known, the sample size of SSUs (boat trips) is unknown. Because the second-stage sample size is unknown prior to the survey and PSU population sizes within each stratum vary, the ratio estimator is used to estimate mean anglers and catch per boat trip. Exit count data are used to expand to estimates of the total anglers and total catch for the population of private boat trips.

Previous data analysis treats the sampling design as a one-stage sample of boats within a stratum and ignores the allocation of effort among days within a stratum. The one-stage approach produces the same inclusion probability for all boats within a stratum which is similar to an average of the two-stage inclusion probabilities. However, accounting for the proportion of boat trips interviewed within a given day provides more accurate estimation within each day. Then the accuracy of higher levels of estimation is improved when day-level estimates are extrapolated to days for which interviews were not conducted. For example, boat trips were censused during 485 of the stratum-days. The number of boat trips ranged from 1 to 23 for these stratum-days, with 5 or fewer boat trips occurring within 95% of these stratum days. The majority of the censused stratum-days occurred between May and September. The one-stage approach would apply an inclusion probability to these boat trips that would reflect an average across all days in the stratum, resulting in overestimation of metrics for these particular days. For estimation across strata, the impact of the one-stage design would likely be small. For lower-level estimation within finely-defined domains, the distinction may be considerable. Estimates from both approaches are provided in an associated spreadsheet, ***ORBS_2011_Estimates_ALL_20131210.xlsx***.

MRIP REVIEW

Previous recommendations were made in the MRIP review by statisticians Jay Breidt and Jean Opsomer (2010). The approach used to address each recommendation, when feasible, is explained in more detail below. The work described in this report is focused on recommendations 3 and 4.

1. *Recommendation*: Avoid reporting at lower levels if possible.

Approach: In a 8/9/12 in Newport meeting with ORBS personnel, reporting at coarser levels was discussed. ORBS personnel remarked that species-level quotas require close monitoring of catch over the season. When salmon or halibut quotas are met, the season is closed immediately. Therefore, catch estimates must be both timely and accurate. For other species, such as bottomfish and rockfish, monthly reporting may be sufficient. However, both Washington and California state reporting is conducted at the weekly level so there may be issues of across-state consistency if reporting is only done at coarser levels of time. Estimates will be available by multiple levels and the appropriate levels may be provided to specific audiences.

2. *Recommendation*: Reallocate effort over time so that all trips have non-zero probability of inclusion in the survey.

Approach: In 2011, sampling was also conducted in strata omitted in previous surveys. These strata included winter weeks in ports previously excluded from the frame of possible trips. The current report focuses on inference only in weeks not defined as "Winter ORBS." However, the information from the 2011 sample could be used to validate assumptions of inference to unsampled weeks in other years.

3. *Recommendation*: Additional weighting to account for randomization of survey time blocks within each day.

Approach: In the 8/9/12 Newport meeting, the practical implications of following this recommendation were discussed. All interviewer shifts occur between 8 AM and 8 PM and approximately overlap the interval occurring between 11 AM and 3 PM. Several interviewers may be used at a single port on a given day and might overlap. Interviewer shifts vary by time of year (available daylight), weather conditions (shifts may be shorter in bad weather), and interviewer duties (video boat counts may also be conducted during a shift). The time of interview is recorded upon data entry, but some interviewers record data by hand and enter later so the time of interview field is not a reliable metric for measuring sampling effort within a day. However, this issue has been resolved for more recent surveys.

These sources of variation make tracking sampling effort within a day very difficult. Therefore, adjustments due to sampling within a day cannot be accurately made. Estimation will assume that the boats observed within a day are a random sample of boats. Eric Schindler expressed some concern that boats returning at night from halibut and tuna fishing trips may not be sampled at an appropriate rate, which would result in underestimation of angler effort and success for those Species and Trip Types and overestimation of other trip types. If accurate information on the distribution of effort throughout the day can be obtained, then this information may contribute to more accurate estimates of effort within time blocks.

4. *Recommendation*: Use the appropriate design-based variance estimators.

Approach: Appropriate variance estimators are based on the assumption of random samples of boats within each port/week/boat type/expansion type stratum combination. Subsequent sections of this report will detail how to obtain design-based variance estimates for estimates of means and totals.

5. *Recommendation*: Use auxiliary data for more efficient estimation.

Approach: Possible sources of auxiliary information include fog and wind metrics for effort and time of day for catch by species. Fog usually does not last all day and its impact varies by Port. Fog may cause video count inaccuracy at Charleston where the bay is wider. Boats are less likely to cross the bars at Garibaldi, Gold Beach, and Winchester Bay when fog is present due to the less-predictable nature of those bars. Wind can also impact fishing effort. Northwesterly wind occurring in early afternoon causes boat returns. These data are not currently implemented in the estimation approach and will require further data collection and modeling to determine if these covariates will improve precision. Incorporating covariates in the estimation process is beyond the scope of this report.

6. *Recommendation*: Automate the subject-matter information when possible.

Approach: Automation of the subject-matter information in the analysis process will require coordination with Eric Schindler. Review of Schindler et al. (2012) indicated that most of the calculations presented in this document are handled by the ORBS project leader or assistant project leader. A list of adjustments detailed in the ORBS and CTSP 2012 Procedures Manual is presented in Table 1. All but two of the adjustments are calculated by the project leader prior to the analysis stage, and boat counters are instructed to not make these corrections so that the actual counts can be reviewed by project staff. Many of these adjustments represent estimation for which variance components are lost because the estimate is reported with raw counts. This will result in underestimation of variance. Currently, the raw data needed to account for the adjustments and their variance components is unavailable for past data sets. However, it is recommended that the raw data be preserved for future data sets so that the adjustment process can be automated and archived and the associated variance can contribute to accurate estimates of variance for assessing precision at higher scales.

Table 1: Adjustments to effort and interview counts and adjustment responsibility

Parameter	Adjustment	Adjusted by
Effort	Private effort adjustments for 1. Late departure (0 – 2% for video counts, ~ 10% for visual counts) 2. Early-returning boats are classified as non-fishing trips and deducted from the exit count 3. Imputed exit counts for unsampled days within a week based on similar season (ET) and day type (weekday/weekend) (Imputed count of "0" if weather/bar conditions are known to prevent effort)	Eric Schindler
Effort	Three-stage Columbia River adjustments for boats that did not enter the ocean or that departed outside the count period	Eric Schindler
Effort	Adjustment for partial day due to fog	Eric Schindler
Interview parameters (Catch, Anglers)	Expansions to: 1. Port-level expansion to portion of the year when sampling occurred and unsampled days within unsampled weeks 2. Expansion for late fishing trips and unsampled minor ports 3. Expansion to portion of the year when no sampling occurred	RecFIN
Interview parameters (Catch, Anglers)	RecFIN expansion to unsampled ports and time periods (not of immediate concern to ODFW and are outside the purview this task)	RecFIN

WEIGHTING AND DESIGN-BASED ESTIMATION

The procedure used to calculate inclusion weights for design-based inference on ORBS parameters is outlined. Issues relevant to estimation are considered, and a set of estimates are examined to compare approaches.

Notation

Let h index one of H strata; let i index the day in the h^{th} stratum, where $i = 1,\ldots,D_h$; and let j index one of N_{hi} boat trips within stratum h and day i. For notation simplicity, the first-stage and second-stage strata are referenced with the same index. Define the following terms:

Z_{hi} = indicator of first-stage sample inclusion;

D_h = the total number of days within the h^{th} stratum;

d_h = the number of days sampled within the h^{th} stratum;

N_{hi} = total effort (exiting boats) for the h^{th} stratum and i^{th} day ($i = 1,..,d_h$);

n_{hi} = total boats interviewed for the h^{th} stratum and i^{th} day;

$Z_{hj(i)}$ = indicator of second-stage sample inclusion given inclusion in the first-stage sample;

Z_{hij} = unconditional indicator of second-stage sample inclusion;

w_{hij} = the inclusion probability of the j^{th} boat trip ($j = 1,..,n_{hi}$) in the h^{th} stratum and i^{th} day;

m_{hij} = number of anglers for the j^{th} boat trip in the h^{th} stratum and i^{th} day; and

y_{hijk} = catch of species k for the j^{th} boat trip in the h^{th} stratum, and i^{th} day.

Inclusion probabilities

Weights are calculated within levels of the combination of stratification variables and reflect a two-stage sample within each stratum level. Let:

$$\pi_{hij} = P\left(Z_{hij} = 1\right) = P\left(Z_{hi} = 1\right) P\left(Z_{hj(i)} = 1 \middle| Z_{hi} = 1\right) = \pi_{hi}\pi_{hj(i)} ,$$

where

$$\pi_{hi} = P\left(\text{day } i \text{ in first-stage sample}\right) = \frac{d_h}{D_h}$$

and

$$\pi_{hj(i)} = P\left(\text{boat trip } j \text{ in second-stage sample} \middle| \text{day } i \text{ in first-stage sample for stratum } h\right) = \frac{n_{hi}}{N_{hi}}.$$

Let $w_{hij} = \dfrac{1}{\pi_{hij}}$ be the inclusion weight of the j^{th} boat trip of the i^{th} day in the h^{th} stratum. Therefore the inclusion weight for the j^{th} boat trip of the i^{th} day in the h^{th} stratum is computed as:

$$w_{hij} = \frac{D_h N_{hi}}{d_h n_{hi}}.$$

Note that exit counts adjustments made by Eric Schindler require that we treat N_{hi} as known. Future work might include an exit count adjustment incorporating an accompanying component of variance.

Stratum-level inference

The design-based estimator of the stratum-level total is given by:

$$\hat{t}_h = \sum_{i=1}^{d_h} \sum_{j=1}^{n_{hi}} w_{hij} y_{hij} \ .$$

The variance of $\hat{t}_h$ is obtained from a Taylor-series approximation (Lohr 1999):

$$Var\left(\hat{t}_h\right) \doteq \left(1 - \frac{d_h}{D_h}\right)\frac{s_h^2}{d_h} + \frac{1}{d_h D_h}\sum_{i=1}^{d_h}\left(1 - \frac{n_{hi}}{N_{hi}}\right) N_{hi}^2 \frac{s_{hi}^2}{n_{hi}} \ ,$$

where $s_h^2 = \dfrac{1}{d_h - 1}\sum_{i=1}^{d_h} N_{hi}\left(\bar{y}_{hi} - \hat{\bar{y}}_h\right)^2$ and and $s_{hi}^2 = \dfrac{1}{n_{hi} - 1}\sum_{i=1}^{d_h} N_{hi}\left(y_{hij} - \bar{y}_{hi}\right)^2$. The estimate of the population total and its variance are obtained by summing the stratum-level estimates and variance estimates, respectively.

Because the PSUs vary in size within a stratum, a ratio estimator is used to estimate means of outcomes of interest. The ratio estimator of the stratum-level mean is given by:

$$\hat{\bar{y}}_h = \frac{\hat{t}_h}{\hat{N}_h} \ ,$$

where $\hat{N}_h = \sum_{i=1}^{d_h}\sum_{j=1}^{n_{hi}} w_{hij}$ is the design-estimated number of boat trips within the h^{th} stratum.
The variance of $\hat{\bar{y}}_h$ is given by:

$$Var\left(\hat{\bar{y}}_h\right) \doteq \frac{1}{\bar{N}_h^2} Var\left(\hat{\tau}_h\right),$$

where $\bar{N}_h = \frac{\sum_{i=1}^{d_h} N_{hi}}{d_h}$.

When the days within a week are censused $\left(d_h = D_h\right)$, the variance simplifies to the following:

$$Var\left(\hat{\bar{y}}_h\right) = \frac{1}{\left(\bar{N}_h D_h\right)^2} \sum_{i=1}^{D_h} \left(1 - \frac{n_{hi}}{N_{hi}}\right) N_{hi}^2 \frac{s_{hi}^2}{n_{hi}} = \frac{1}{N_h^2} \sum_{i=1}^{D_h} \left(1 - \frac{n_{hi}}{N_{hi}}\right) N_{hi}^2 \frac{s_{hi}^2}{n_{hi}},$$

where $N_h = \sum_{i=1}^{d_h} N_{hi}$. Note that this variance is analogous to the variance of a stratified random sample with the censused PSUs serving as strata.

Domain Estimation with the Ratio Estimator

Inference on effort and catch for domains are also of interest. Domains of interest include trip type, catch area, and species of catch. Let l index one of L levels of a domain of interest. Define the following terms:

N_{hil} = total effort (boat trips) for the h^{th} stratum, i^{th} day, and l^{th} domain;
n_{hil} = total boats interviewed for the h^{th} stratum, i^{th} day, and l^{th} domain;
w_{hij} = the inclusion weight of the j^{th} boat trip in the h^{th} stratum and i^{th} day;
m_{hijl} = number of anglers for the j^{th} boat trip in the h^{th} stratum, i^{th} day, and l^{th} domain; and
y_{hijl} = catch or number of anglers for the j^{th} boat trip in the h^{th} stratum, i^{th} day, and l^{th} domain.

Note that domains are defined at the boat-trip level, so domain inclusion is assessed only for the second-stage sample. The sample size within each domain in stratum h is a random variable, so the ratio estimator is used. The ratio estimator of the mean outcome for a domain *l* within stratum $h\left(\hat{\bar{y}}_{hl}\right)$ is defined for complex survey design as (Lohr, 1999):

$$\hat{\bar{y}}_{hl} = \frac{\hat{t}_{hl}}{\hat{N}_{hl}} = \frac{\sum_{i=1}^{d_h}\sum_{j=1}^{n_{hi}} w_{hij} X_{hijl} y_{hij}}{\sum_{i=1}^{d_h}\sum_{j=1}^{n_{hi}} w_{hij} X_{hijl}},$$

where $\hat{t}_{hl}$ is the estimated total within domain *l* for the outcome of interest, y_{hij} ; $\hat{N}_{hl}$ is the estimated total number of boat trips in the h^{th} stratum and l^{th} domain; and X_{hijl} indicates inclusion in the l^{th} domain for the j^{th} boat trip in the i^{th} day.

The ratio estimator is biased, but the bias is small when the sample size within domains is adequate, the variation in the number of boats surveyed within a week is small, and the number of anglers is positively correlated with the number of boats (Lohr 1999). The estimator of the asymptotic variance of the ratio estimator of the mean $\left(\hat{\bar{y}}_{hl}\right)$ for a domain *l* within stratum *h* is obtained with a Taylor series approximation (Lohr, 1999). Define the linearized outcome of interest as $q_{hij} = \frac{1}{t_x}\left(y_{hij} - \hat{\bar{y}}_h x_{hij}\right)$. Then the variance of the estimate of the total of the linearized variable approximates the MSE of the ratio estimator as follows:

$$\hat{Var}\left(\hat{\bar{y}}_{hl}\right) \doteq \frac{1}{\hat{N}_{hl}^2}\sum_{i}^{d_h}\sum_{i'}^{d_h}\sum_{j}^{n_{hil}}\sum_{j'}^{n_{hil}} \breve{\Delta}_{ii'jj'} \frac{y_{hij} - \hat{\bar{y}}_{hl} x_{hij}}{\pi_{hij}} \frac{y_{hi'j'} - \hat{\bar{y}}_{hl} x_{hi'j'}}{\pi_{hi'j'}},$$

where $\hat{N}_{hl}$ is the design-based estimator of total boat trips occurring in stratum *h* and domain *l* and $\breve{\Delta}_{ii'jj'}$ is the design-expanded covariance of q_{hij} and $q_{hi'j'}$ (Särndal, Swensson, and Wretman 1992) defined as:

$$\breve{\Delta}_{ii'jj'} = \begin{cases} 1 - \dfrac{d_h n_{hi}}{D_h N_{hi}}, & i = i' \text{ and } j = j' \\ \dfrac{\left(\dfrac{d_h n_{hi} \left(n_{hi} - 1 \right)}{D_h N_{hi} \left(N_{hi} - 1 \right)} - \left(\dfrac{d_h n_{hi}}{D_h N_{hi}} \right)^2 \right)}{\dfrac{d_h n_{hi} \left(n_{hi} - 1 \right)}{D_h N_{hi} \left(N_{hi} - 1 \right)}}, & i = i' \text{ and } j \neq j' \\ \dfrac{\left(\dfrac{d_h \left(d_h - 1 \right) n_{hi} n_{hi'}}{D_h \left(D_h - 1 \right) N_{hi} N_{hi'}} - \dfrac{d_h^2 n_{hi} n_{hi'}}{D_h^2 N_{hi} N_{hi'}} \right)}{\dfrac{d_h \left(d_h - 1 \right) n_{hi} n_{hi'}}{D_h \left(D_h - 1 \right) N_{hi} N_{hi'}}}, & i \neq i' \end{cases}.$$

These approaches are applied in the R *survey* package (Lumley 2012) which will be used for data analysis with the ORBS survey data. Note that the general indexing of both first- and second-stage strata is still used here.

Estimation of the population total for Boat Types

ORBS obtains exit count data from charter boat companies specific to each Port, Week, Expansion Type, and Trip Type. This additional information on Charter boat types may be used to obtain more accurate estimates of angler and catch totals for this domain and aggregates of these domain levels. This information is not available for Private boat trips, so estimation at these levels is handled differently. For levels of estimation in which the domain population size within each stratum, N_{hl}, is known, a ratio estimator of the population total (Särndal et al. 1992, p. 391; Lumley 2010) is used.

For Charter boat types, the domain mean within each stratum h is obtained with ratio estimation as follows:

$$\hat{\bar{y}}_{hl} = \frac{\hat{t}_{hl}}{\hat{N}_{hl}} = \frac{\sum_{i=1}^{d_h} \sum_{j=1}^{n_{hi}} w_{hij} X_{hijl} y_{hij}}{\sum_{i=1}^{d_h} \sum_{j=1}^{n_{hi}} w_{hij} X_{hijl}},$$

where X_{hijl} is an indicator of inclusion in domain *l*. The true domain population size is then used to obtain a more accurate estimate of the domain total, $\tilde{t}_l$, and its variance where:

$$\tilde{t}_{hl} = N_{hl}\hat{\bar{y}}_{hl} = \frac{N_{hl}\hat{t}_{hl}}{\hat{N}_{hl}},$$

$$\hat{Var}\left(\tilde{t}_{hl}\right) = N_{hl}^2\,\hat{Var}\left(\hat{\bar{y}}_{hl}\right),$$

and $\hat{Var}\left(\hat{\bar{y}}_{hl}\right)$ is obtained as the variance of a ratio estimator. Estimates of population totals across strata are obtained by obtaining ratio estimates of the mean at the appropriate levels then multiplying by the population size at the appropriate levels.

Because the population size is not known for Private boat trips, estimates of population totals for private boats are obtained as a standard design-based estimate of the total, as follows:

$$\hat{t}_{hl} = \hat{N}_{hl}\hat{\bar{y}}_{hl} = \sum_{i=1}^{d_h}\sum_{j=1}^{n_{hi}} w_{hij} X_{hijl}\, y_{hij}.$$

Estimation when a single PSU or SSU is sampled

When only a single PSU is surveyed within a stratum, the variance cannot be estimated directly. There are several cases in which this might occur for ORBS. First, a stratum may only consist of a single day, referred to as a "certainty PSU," that is censused at the first stage, with the only variance contribution arising from the second-stage sampling. Second, a sampled stratum may only include a single PSU when several days were available. Finally, a stratum might be surveyed more than one day but one of the boat types may have only been interviewed on one of those days, effectively reducing the sample size of PSUs to 1 for that stratum and boat type.

The *survey* package in R provides several approached to resolving the "lonely PSU" problem. The "certainty" option treats the stratum as censused and computes the variance of the estimate as 0. The remaining data contain only strata with at least two days, including boats trips for strata sampled only on a single day. The "adjust" option is used for strata with only one interview day in the sample, and the variance component is conservatively obtained by calculating variance relative to the sample grand mean (Lumley 2012).

A summary of PSU population sizes and sample sizes (Table 2) indicates 19 strata containing certainty PSUs and 62 strata (in bold) with more than one population PSU but only a single sampled PSU. The optimal approach is separate inference for lonely PSUs (D_h = 1 and d_h = 1) and strata with more than one PSU but only a sample of one day (D_h > 1 and d_h = 1). Combining inference for these two cases has proved problematic when domain estimation is of interest. The "adjust" option for dealing with a single sampled PSU calculates the variance relative to the estimate of the grand mean rather than the stratum-level mean, and the grand mean cannot be calculated across strata when estimation is split across strata. The R *survey* package cannot currently accommodate this scenario by allowing specification of different options by strata. Therefore, the optimal approach must be explicitly programmed in R, which is beyond the scope of the current task list. The recommended approach is to use the "adjust" option for the entire data set. The implication is that the 19 strata containing only a single PSU will contribute a non-zero component to the variance of the domain estimate, resulting in a conservative approach. An examination of the impact of this approach revealed little effect on inference and confidence intervals for several levels of estimation. However, future inference may benefit from a more careful treatment of lonely PSUs.

Table 2: Summary of PSU population sizes and sample sizes

	D_h						
d_h	1	2	3	4	5	6	7
1	19	**3**	**13**	**14**	**4**	**7**	**21**
2	0	21	16	21	5	3	21
3	0	0	40	12	6	1	25
4	0	0	0	24	5	2	31
5	0	0	0	0	4	0	22
6	0	0	0	0	0	7	18
7	0	0	0	0	0	0	36

ANALYSIS WITH R STATISTICAL SOFTWARE

Design-based analysis based on the stratified two-stage sampling design is conducted with R software (2013) and the *survey* package (Lumley 2012). The *survey* package allows specification of the sampling design and inclusion probabilities, including multiple stages of sampling, stratification at each stage, and domain estimation for specified variables. Estimation employs the ratio estimator, so random sample sizes for domain estimates are treated appropriately.

R code for the analysis of the 2011 ORBS data is provided in Appendix A, and most of the code prepares the data for analysis. Formatting code is followed by analysis code for several levels of analysis, including subsets of data, the case of adjusting Charter Boat Type estimates for known population size, and an analysis that matches a table previously provided by Bryan Wright. Several tables provided by Eric Schindler are required for accurately defining the sampling frame (excluding Winter ORBS and identifying the sampling frame of dates by stratum) and obtaining inclusion probabilities for design-based analysis. All files necessary to conduct the 2011 analysis are provided in a ZIP file, *ORBS_2011_AnalysisFiles.zip*. Similar background files will be needed to format data from other years.

DISCUSSION

Weekends provide more angler opportunity than weekdays, and certain days of the week might be dedicated to other tasks at some ports. For these reasons, consideration is given to calculating design weights within weekday groups defined as weekend days and workweek days. Stratification by Weekday Group was outlined by this author in a 9/26/13 memo titled "ORBS Weighting Approach – Examining Weekday Groups." Concurrent work by ODFW personnel did not find indications of bias from inference that did not take into account Weekday Groups, but this topic will be given further consideration.

Ratio estimator of the population total may be used to obtain estimates of catch for charter boat trips because the total number of boat trips is known. The ratio estimator of the total may result in a more accurate and precise estimate of the total if the estimated number of boat trips is biased or estimated imprecisely (Särndal et al. 1992, p. 391; Lumley 2010). Estimation for private boat trips may also be improved by ratio estimation if exit counts are more accurate than estimates of boat trips obtained from Horvitz-Thompson estimation. An example of combining inference for Horvitz-Thompson estimates of total catch for private boat trips and ratio estimator total catch estimates is provided in Appendix A.

Estimates obtained from the R code provided in Appendix A are available in an accompanying spreadsheet, ***ORBS_2011_Estimates_ALL_20131210.xlsx***. This spreadsheet also contains the original ORBS estimates of total fish caught and released (by Port, Week, Expansion Type, Boat Type, Trip Type, Area, and Species) obtained assuming a stratified one-stage cluster sample. The estimates of total catch from the analysis reflecting the two-stage stratified design are generally larger than for analysis assuming the one-stage stratified design. Cases when the one-stage estimates greatly exceeded the two-stage estimates occurred for estimates of total Black Rockfish and Dungeness Crab catch in Ports 10, 24, 34, and 42 during Weeks 25 through 38.

In a few cases, either one-stage or two-stage estimates were not available for comparison. Many of these cases occurred in Port 38 where fishing activity was recorded as occurring in either Area 5 or Area 6, and apparently these Area designations did not match between the two sets of estimates. In a few other cases, two-stage estimates were not available for domains that did have one-stage ORBS estimates, likely due to the effect of pooling across days in the one-stage analysis.

RECOMMENDATIONS

This author offers the following recommendations, while acknowledging that some will require considerable effort and may be impractical given interviewer responsibilities and sampling effort limitations:

1. Use randomization of days within Strata. If weekend days are of more interest, stratify by Weekday Group, census the weekend stratum, and use true randomization to select days within the Workweek stratum for unbiased inference.
2. Allocate sampling effort to all strata, including all Weeks for all Ports within the weeks not defined as Winter ORBS.
3. Track the interview shift times more closely for potential stratification of time blocks within a day.
4. Record days that interviewers encounter no boats to obtain accurate information on survey effort and cost. Consider randomizing interview time blocks within a day.
5. Assess impact of night returns for halibut and tuna by assessing video counts. Eric Schindler thought this could be initially examined at the ports in Newport and Charleston by comparing actual video counts of late returns and late and early departures to expanded counts.
6. Combine inference for strata with lonely PSUs. The proper treatment of this issue requires explicit coding in R. The R *survey* package apparently cannot handle more than one lonely-PSU approach, and domain-level estimates cannot be combined by stratum. If this is too difficult, evaluate the effect of using the "adjust" option for strata with $D_h = 1$.
7. Ratio estimation of population totals may improve estimation for private boat trips if exit count data are more precise and accurate than design-based estimates of boat trips from the Horvitz-Thompson estimator.

8. Eric Schindler expressed a need to change the way trip types are recorded so that species-level estimation may be obtained. Currently, angler effort cannot be distinguished below the gross trip type strata and there is no means to assess target species groups when anglers target more more than one species group on a trip. Eric Schindler proposed that interviewers collect information on the species anglers are fishing for and then the software would automatically assign the trip type. This would provide further detail by which to assess actual fishing effort for species groups while preserving the existing trip type effort estimation.
9. Private boat exit counts are currently treated as known but are actually imputed and adjusted. Incorporate these adjustments into the estimation process and account for the additional uncertainty.

CONCLUSIONS

ORBS demonstrates an admirable commitment to obtaining accurate and precise estimates of ocean fishing effort and success. Improvements in the analysis of ORBS data include the calculation of design-based variance estimates and consideration of issues related to differences in effort and catch between weekdays and weekends. Considerations of improvements such as additional randomization and stratification in the survey design may lead to increased precision and accuracy in estimates of Oregon ocean fishing parameters.

REFERENCES

Breidt, J. and J. Opsomer (2010). Preliminary Review of Ocean Recreational Boat Survey. Consultants' report, July 27, 2010.

Lohr, S.L. (1999). *Sampling: Design and Analysis.* California: Brooks/Cole Publishing Company.

Lumley, T. (2010). *Complex Surveys: A Guide to Analysis Using R*. New Jersey: Wiley and Sons, Inc.

Lumley, T. (2012) "survey: analysis of complex survey samples". R package version 3.28-2.

ORBS and CTSP 2012 Procedures Manual (2012). Oregon Sampling Project, Marine Resources Program, Oregon Department of Fish and Wildlife.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Särndal, C-E, B. Swensson, and J. Wretman (1992). *Model Assisted Survey Sampling*. New York: Springer.

Schindler, E., M. Freeman, and B. Wright (2012). Sampling Design of the Oregon Department of Fish and Wildlife's Ocean Recreational Boat Survey (ORBS). ODFW Marine Resources Program report, March 14, 2012.

APPENDIX A: R CODE FOR 2011 ORBS ANALYSIS

```
setwd("C:\\Consulting\\ORBS\\")
setWindowTitle("ORBS analysis")
memory.limit(4000)

# CLEAR WORKSPACE
#rm(list = ls(all=TRUE))

# LOAD PACKAGES
library(survey)
library(reshape)

#----------------------------------------------------------------------
# IMPORT FRAME INFORMATION
#----------------------------------------------------------------------
# Get ExpType by Date and Port for unsurveyed dates
# ORBS data only contains dates for days in which interviews
# and exit counts were conducted
# Table provided by Eric Schindler

ET_2011_orig<-read.csv("ET 2011 - ES corrections.csv", header=TRUE)
names(ET_2011_orig)
dim(ET_2011_orig)

# Create Port field
ET_2011_1<-melt(ET_2011_orig, id.vars="Date", measure.vars=
      names(ET_2011_orig)[2:13])
ET_2011_1$Date<- as.POSIXct(as.character(ET_2011_1$Date),
format="%m/%d/%Y")
ET_2011_1$WeekDay<-factor(weekdays(ET_2011_1$Date), levels=c("Monday",
      "Tuesday","Wednesday","Thursday","Friday", "Saturday","Sunday"))
dim(ET_2011_1)                # [1] 4380    5
head(ET_2011_1)

# Format Port field
ET_2011_1$Port<- gsub("X", "", as.character(ET_2011_1$variable))
ET_2011_1$Stratum.PSU.Ind<-1        # 1 for each day within stratum
names(ET_2011_1)[3]<-"ExpansionType"

# Get Week field
# Table provided by Eric Schindler
WeekDefn<-read.csv("WeekDefn.csv", header=TRUE)
names(WeekDefn)  # "Date"  "Year"  "Month" "Day"   "Week"
WeekDefn$DateOrig<-WeekDefn$Date
WeekDefn$Date<- as.POSIXct(as.character(WeekDefn$DateOrig),
      format="%m/%d/%Y")      # format Date field

ET_2011_2<-merge(ET_2011_1, WeekDefn)  # ET_2011_2 = frame of PSU's by date
dim(ET_2011_2)
head(ET_2011_2)

# Calc Dh
ET_2011_frame<-aggregate(ET_2011_2$Stratum.PSU.Ind,
list(ET_2011_2$Port, ET_2011_2$Week, ET_2011_2$ExpansionType), sum)
```

```
names(ET_2011_frame)<-c("Port", "Week", "ExpansionType", "Dh")
ET_2011_frame <- ET_2011_frame [order(ET_2011_frame [,1], ET_2011_frame
[,2], ET_2011_frame [,3], ET_2011_frame [,4]),]
nrow(ET_2011_frame)   # Number of strata

# Merge frame info with info on which weeks correspond to
# ORBS estimation and omits Winter ORBS
# Table provided by Eric Schindler

WinterORBS_orig<-read.csv("Week - Month 2009 and 2011.csv",
      header=TRUE)
names(WinterORBS_orig)
WinterORBS_1<-melt(WinterORBS_orig, id.vars=c("Year", "Week"),
      measure.vars= names(WinterORBS_orig)[3:14])
WinterORBS_1$Port<- gsub("X", "", as.character(WinterORBS_1$variable))
WinterORBS_1<-WinterORBS_1[,-3]    # Remove old Port field
names(WinterORBS_1)[3]<-"Status"
names(WinterORBS_1)    # "Year"   "Week"   "Status" "Port"
table(WinterORBS_1$Status)

# exclude Winter ORBS and unsampled weeks
ORBSWeeks<- WinterORBS_1[WinterORBS_1$Status=="Week",]
ORBSWeeks_2011<- ORBSWeeks [ORBSWeeks $Year==2011,2:4]
dim(ORBSWeeks_2011)     # 318   3

# merge to exclude weeks not in ORBS
ET_2011_frame_ORBS<-merge(ET_2011_frame, ORBSWeeks_2011)
dim(ET_2011_frame_ORBS)       # 449   5
dim(ET_2011_frame)            # 783   4
ET_2011_frame_ORBS $StratumCode<-1:449    # Label strata
ET_2011_frame_ORBS<- ET_2011_frame_ORBS[order(ET_2011_frame_ORBS$Port,
ET_2011_frame_ORBS$Week, ET_2011_frame_ORBS$ExpansionType),]
tapply(ET_2011_frame_ORBS $Dh, ET_2011_frame_ORBS $ExpansionType,
range)

#-----------------------------------------------------------------------
# IMPORT EFFORT/EXIT COUNT FILE (=DESIGN DATA)
#-----------------------------------------------------------------------

# IMPORT AND REVIEW DATA
# Charter and private boat trips by Week, Port, Date, Expansion Type
eff_2011<-read.csv("tbEffOcean_2011.csv")
head(eff_2011)
str(eff_2011)
summary(eff_2011)

# Examine sample coverage
with(eff_2011, table(Week, Port))  # days by week and port – all 0 or 7

# Csum charters
eff_2011$Charter<-rowSums(eff_2011[,5:11])    # total Charter trips

# Create datetime variable
eff_2011$Date<-as.POSIXct(as.character(eff_2011$Date), format="%m/%d/%Y
      %H:%M:%S" )
```

```
# Identify day of the week
eff_2011$WeekDay<-factor(weekdays(eff_2011$Date), levels=c("Monday",
      "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday",
      "Sunday"))

# Data cleaning
# E.S. suggested changing "D" to "W" for Port 40 and Week 40
eff_2011$ ExpansionType [(eff_2011$Port==40)&( eff_2011$Week==40)&(
eff_2011$ExpansionType=="D")]<-"W"

dim(eff_2011)                       # 2492   14
eff_2011_ORBS1<-merge(eff_2011, ET_2011_frame_ORBS, by=c("Port",
"Week", "ExpansionType"))
dim(eff_2011_ORBS1)                 # 2205   17


#----------------------------------------------------------------------
# CREATE BOATTYPE AND TRIPTYPE FIELDS FOR EFFORT
#----------------------------------------------------------------------

eff_2011_ORBS<-melt(eff_2011_ORBS1, id.vars=c("Port", "Week",
      "ExpansionType","Date"), measure.vars=c("ChrSal", "ChrBot",
      "ChrCmb", "ChrTun", "ChrHal", "ChrDiv", "ChrNon", "Private"))
eff_2011_ORBS$BoatType<-
ifelse(eff_2011_ORBS$variable=="Private","P","C")
head(eff_2011_ORBS)
dim(eff_2011_ORBS)            # 17640     7

eff_2011_ORBS$TripType<-NA
eff_2011_ORBS$TripType[eff_2011_ORBS$variable=="ChrSal"]<-"S"
eff_2011_ORBS$TripType[eff_2011_ORBS$variable=="ChrBot"]<-"B"
eff_2011_ORBS$TripType[eff_2011_ORBS$variable=="ChrCmb"]<-"C"
eff_2011_ORBS$TripType[eff_2011_ORBS$variable=="ChrTun"]<-"T"
eff_2011_ORBS$TripType[eff_2011_ORBS$variable=="ChrHal"]<-"H"
eff_2011_ORBS$TripType[eff_2011_ORBS$variable=="ChrDiv"]<-"D"
eff_2011_ORBS$TripType[eff_2011_ORBS$variable=="ChrNon"]<-"N"

names(eff_2011_ORBS)[ names(eff_2011_ORBS)=="value"]<-"ExitCount"
eff_2011_ORBS<- eff_2011_ORBS[,-5]  # Remove "variable"
head(eff_2011_ORBS)

# Aggregate across dates by strata – use for calculating incl prob
eff_2011_Strata<-aggregate (eff_2011_ORBS$ExitCount,
      list(Port=eff_2011_ORBS$Port, Week=eff_2011_ORBS$Week,
      ExpansionType =eff_2011_ORBS$ExpansionType, BoatType =
      eff_2011_ORBS$BoatType), sum)
names(eff_2011_Strata)[ names(eff_2011_Strata)=="x"]<-"Nhi"
dim(eff_2011_Strata)          # 890   5
```

```
#---------------------------------------------------------------------
# IMPORT INTERVIEW FILE (=OBSERVAION UNIT OR ELEMENT)
#---------------------------------------------------------------------

# Import and review data
intrv_2011<-read.csv("tbIntrv_2011.csv")
dim(intrv_2011)
head(intrv_2011)
str(intrv_2011)
summary(intrv_2011)

# Data cleaning – las
intrv_2011$TripType<-toupper(intrv_2011$TripType)           # all caps
intrv_2011$Fishery <-toupper(intrv_2011$Fishery)            # all caps

# Ocean interviews only
intrv_2011<- intrv_2011[intrv_2011$Fishery=="O",]
dim(intrv_2011)               # 16883    18

# Create datetime variables
intrv_2011$Date <-as.character(intrv_2011$Date)
intrv_2011$Date<-as.POSIXct(intrv_2011$Date, format="%m/%d/%Y
      %H:%M:%S")

dim(intrv_2011)
intrv_2011<-merge(intrv_2011, WeekDefn[,c(1,5)])
dim(intrv_2011)

# Identify day of the week
intrv_2011$WeekDay<-factor(weekdays(intrv_2011$Date),
      levels=c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday",
      "Saturday", "Sunday"))

# Data cleaning
intrv_2011<- intrv_2011[,c(1:9,11,12,15:20)]  # Select needed fields
head(intrv_2011)
dim(intrv_2011)         # 16883    17

# remove Winter ORBS
intrv_2011_ORBS<- merge(intrv_2011, ET_2011_2)
dim(intrv_2011_ORBS)    # 16883    24

# Merge to remove winter ORBS
intrv_2011_ORBS<- merge(intrv_2011_ORBS, ET_2011_frame_ORBS)
dim(intrv_2011_ORBS)    # 16089    27    All ORBS

# Create unique ID for BoatTrip
intrv_2011_ORBS $BoatTrip<-paste(paste(intrv_2011_ORBS $BoatNumber,
      intrv_2011_ORBS $IntvNum, sep="_"), intrv_2011_ORBS$SID, sep="_")
```

```
#---------------------------------------------------------------------
# CALCULATE INCLUSION PROBS
#---------------------------------------------------------------------
# Get nhi – number of interviews by Stratum, BoatType, Date
intrv.PSU<-aggregate(intrv_2011_ORBS$Stratum.PSU.Ind, list(
Port =intrv_2011_ORBS$Port,
Week =intrv_2011_ORBS$Week,
ExpansionType =intrv_2011_ORBS$ExpansionType,
BoatType =intrv_2011_ORBS$BoatType,
Date =intrv_2011_ORBS$Date), sum)
names(intrv.PSU)[ names(intrv.PSU)=="x"]<-"nhi"
head(intrv.PSU)
intrv.PSU$Port<-as.character(intrv.PSU$Port)

# Get dh –get unique Dates within a stratum, then count PSU sample
intrv.Stratum.PSU<-aggregate(intrv.PSU $nhi, list(
Port = intrv.PSU $Port,
Week = intrv.PSU $Week,
ExpansionType= intrv.PSU $ExpansionType,
Date = intrv.PSU $Date), length)

# sum the number of unique days per stratum (dh)
intrv.Stratum <-aggregate(intrv.Stratum.PSU $Date, list(
Port = intrv.Stratum.PSU $Port,
Week = intrv.Stratum.PSU $Week,
ExpansionType= intrv.Stratum.PSU $ExpansionType), length)

# name dh variable and sort
names(intrv.Stratum)[ names(intrv.Stratum)=="x"]<-"dh"
intrv.Stratum <- intrv.Stratum[order(intrv.Stratum$Port,
intrv.Stratum$Week, intrv.Stratum$ExpansionType),]
head(intrv.Stratum)
intrv.Stratum$Port<-as.character(intrv.Stratum$Port)

# Merge frame and sample info to get inclusion probs
# Stage 1 inclusion probability calculations
ORBS_Stage1<-merge(ET_2011_frame_ORBS, intrv.Stratum, by=c("Port",
"Week", "ExpansionType"), all=TRUE)
dim(ORBS_Stage1)  #  449   7
ORBS_noStage1<-ORBS_Stage1[is.na(ORBS_Stage1$dh),]

ORBS_Stage1$dh[is.na(ORBS_Stage1$dh)]<-0
dim(ET_2011_frame_ORBS) # 449   6
ORBS_Stage1<-ORBS_Stage1[!is.na(ORBS_Stage1$dh),]
dim(ORBS_Stage1[ORBS_Stage1$dh==0,])      # 47  7

dim(ORBS_Stage1)        # 449   7
dim(ET_2011_frame_ORBS) # 449   6
dim(intrv.Stratum)      # 402   4  Lose 47 strata
range(ORBS_Stage1$dh)   # 0 7
sum(ORBS_Stage1 $Dh[ORBS_Stage1 $dh==0])        # 203
# number of unsampled Stratum-Days (PSUs) = 203

# Calc stage 1 inclusion probability
ORBS_Stage1$Stage1Prob<- ORBS_Stage1$dh/ ORBS_Stage1$Dh
range(ORBS_Stage1$Stage1Prob)
# 0 1
```

```
# Stage 2 inclusion probability calculations
Eff_2011<-aggregate(eff_2011_ORBS$ ExitCount, list(Port= eff_2011_ORBS$
Port, Week= eff_2011_ORBS$ Week, ExpansionType= eff_2011_ORBS$
ExpansionType, BoatType= eff_2011_ORBS$ BoatType, Date= eff_2011_ORBS$
Date), sum)
names(Eff_2011)[ names(Eff_2011)=="x"]<-"Nhi"

ORBS_Stage2 <-merge(Eff_2011, intrv.PSU)
dim(Eff_2011)                # 4410    6
dim(ORBS_Stage2 ) # 2047    8
dim(merge(Eff_2011, ORBSWeeks[ORBSWeeks$Year==2011,]))      # 4410    8
dim(Eff_2011[Eff_2011$Nhi>0,])
#[1] 2649    6    # Have exit count data for days with no interviews

# Correct missing boats such that nhi>Boats
# Eric Schindler agreed that this is appropriate
ORBS_Stage2$Nhi[ORBS_Stage2$nhi>ORBS_Stage2$Nhi] <-
ORBS_Stage2$nhi[ORBS_Stage2$nhi > ORBS_Stage2$Nhi]
ORBS_Stage2$Stage2Prob<- ORBS_Stage2$nhi / ORBS_Stage2$ Nhi
range(ORBS_Stage2$Stage2Prob)  # 0.009615385 1.000000000

# Merge Stage 1 and Stage 2 info to compute InclProb
ORBS_InclProbs<-merge(ORBS_Stage1, ORBS_Stage2, by=c("Port", "Week",
"ExpansionType"))
dim(ORBS_InclProbs)

ORBS_InclProbs$InclProb<- ORBS_InclProbs $ Stage1Prob * ORBS_InclProbs $
Stage2Prob
range(ORBS_InclProbs$InclProb)         # 0.00952381  1.000000000



#-----------------------------------------------------------------------
# MERGE DATA AND INCLUSION PROBS
#-----------------------------------------------------------------------
ORBS_2011<-merge(intrv_2011_ORBS, ORBS_InclProbs, by=c("Port", "Week",
"ExpansionType", "BoatType", "Date", "Status", "StratumCode", "Dh"))
dim(ORBS_2011)                #[1] 16086    34
ORBS_2011$Stratum<-interaction(ORBS_2011$Port, ORBS_2011$Week,
      ORBS_2011$ExpansionType, ORBS_2011$BoatType, drop=TRUE)
```

```
#-----------------------------------------------------------------------
# IMPORT ENCOUNTER (CATCH) FILE (=RESPONSE DATA) AND SPECIES CODE NAMES
#-----------------------------------------------------------------------

# Import and review data
enc<-read.csv("tbEnc_2011.csv")
head(enc)
summary(enc)
str(enc)

# Create datetime variable
enc$Date<-as.POSIXct(as.character(enc$Date), format="%m/%d/%Y
      %H:%M:%S")

# Import and review data
tbSpp<-read.csv("tbSpeciesCodes_2011.csv")
head(tbSpp)
summary(tbSpp)
str(tbSpp)
names(tbSpp)[1]<-"Species"

#-----------------------------------------------------------------------
# MERGE FISH NUMBERS WITH FISH NAMES
#-----------------------------------------------------------------------
enc2<-merge(enc, tbSpp)
head(enc2)
str(enc2)
summary(enc2)

#-----------------------------------------------------------------------
# MERGE FISH DATA WITH ALL OTHER DATA
#-----------------------------------------------------------------------
ORBS_2011<-merge(ORBS_2011, enc2, by=c("Date", "Port", "IntvNum",
      "SID"))
dim(ORBS_2011)                # 33687    42

# data cleaning
ORBS_2011$Released[is.na(ORBS_2011$Released)]<-0

# Remove species = PIT-tagged Black RF
ORBS_2011 <- ORBS_2011[ORBS_2011$Species!=1,]
names(ORBS_2011)
dim(ORBS_2011)                # 32140    42

# sort
ORBS_2011<- ORBS_2011[order(ORBS_2011$Port, ORBS_2011$Week,
ORBS_2011$ExpansionType),]

# create PSU stratum ID
ORBS_2011$PSU.Stratum<-interaction(ORBS_2011$Port, ORBS_2011$Week,
      ORBS_2011$ExpansionType, drop=TRUE)
ORBS_2011$Stratum.Day<-paste(ORBS_2011$PSU.Stratum, ORBS_2011$Date,
sep="_")
ORBS_2011$Count<-1
```

```
#-----------------------------------------------------------------------
# DESIGN-BASED ESTIMATION – requires the survey package
#-----------------------------------------------------------------------

library(survey)

# Specify sampling design for analysis in survey package
ORBS_2011_2stage_design <-svydesign(id=~ Stratum.Day + BoatTrip,
probs=~InclProb, strata= ~ PSU.Stratum  + BoatType, nest=TRUE, data=
ORBS_2011, variables = ~ Port + Week + Area+ Anglers +  BoatType +
TripType + ExpansionType + Species +  Catch +Released +Count, fpc = ~
Dh + Nhi)

# Set options to adjust lonely PSUs
options(survey.adjust.domain.lonely=TRUE)
options(survey.lonely.psu= "adjust")      # use grand mean for variance

# Obtain total estimates to match format in ORBS table
#  tbEstCatchRel.xlsx

Est_TotalCatch_ORBS <- svyby(~Catch, by= ~ Week + Port + Area +
BoatType + TripType + ExpansionType + Species,
design = ORBS_2011_2stage_design,
FUN = svytotal,
verbose=TRUE)
rownames(Est_TotalCatch_ORBS)<-NULL
names(Est_TotalCatch_ORBS )[8:9]<-c("EstCatch", "SE.EstCatch")

Est_TotalReleased_ORBS <- svyby(~Released, by= ~ Week + Port + Area +
BoatType + TripType + ExpansionType + Species,
design = ORBS_2011_2stage_design,
FUN = svytotal,
verbose=TRUE)
rownames(Est_TotalReleased_ORBS)<-NULL
names(Est_TotalReleased_ORBS)[8:9]<-c("EstReleased", "SE.EstReleased")

# Format Estimates to match ORBS table, tbEstCatchRel.xlsx

SampledCatch <-with(ORBS_2011, aggregate(Catch, list(Week=Week,
      Port=Port, Area=Area, BoatType=BoatType, TripType=TripType,
      ExpansionType=ExpansionType, Species=Species), sum))
names(SampledCatch)[8]<-"SampledCatch"
SampledReleased <-with(ORBS_2011, aggregate(Released, list(Week=Week,
      Port=Port, Area=Area, BoatType=BoatType, TripType=TripType,
      ExpansionType=ExpansionType, Species=Species), sum))
names(SampledReleased)[8]<-"SampledReleased"
SampledTagged <-with(ORBS_2011, aggregate(Tagged, list(Week=Week,
      Port=Port, Area=Area, BoatType=BoatType, TripType=TripType,
      ExpansionType=ExpansionType, Species=Species), sum))
names(SampledTagged)[8]<-"SampledTagged"
SampledInterviews <-with(ORBS_2011, aggregate(BoatTrip, list(Week=Week,
      Port=Port, Area=Area, BoatType=BoatType, TripType=TripType,
      ExpansionType=ExpansionType, Species=Species), length))
names(SampledInterviews)[8]<-"SampledInterviews"

dim(SampledCatch)
SampledFields<-merge(SampledCatch, SampledReleased)
```

```
dim(SampledFields)
SampledFields<-merge(SampledFields, SampledTagged)
dim(SampledFields)
SampledFields<-merge(SampledFields, SampledInterviews)
dim(SampledFields)

Est_TotalCatch_ORBS_Sampled<-merge(Est_TotalCatch_ORBS, SampledFields,
by=names(Est_TotalCatch_ORBS)[1:7], all=TRUE)
Est_TotalCatch_ORBS_Sampled$SampledReleased
[is.na(Est_TotalCatch_ORBS_Sampled$SampledReleased)]<-0
Est_TotalCatch_ORBSReleased_Sampled<-merge(Est_TotalReleased_ORBS,
Est_TotalCatch_ORBS_Sampled, by=
names(Est_TotalCatch_ORBS_Sampled)[1:7])

# Arrange columns
ORBS_2011_Estimates_ORBS<-data.frame(Year=rep(2011,
nrow(Est_TotalCatch_ORBSReleased_Sampled)),
Week= Est_TotalCatch_ORBSReleased_Sampled $Week,
Port= Est_TotalCatch_ORBSReleased_Sampled $Port,
Fishery= rep("O", nrow(Est_TotalCatch_ORBSReleased_Sampled)),
Area= Est_TotalCatch_ORBSReleased_Sampled $Area,
BoatType= Est_TotalCatch_ORBSReleased_Sampled $BoatType,
TripType= Est_TotalCatch_ORBSReleased_Sampled $TripType,
ExpansionType= Est_TotalCatch_ORBSReleased_Sampled $ExpansionType,
Species= Est_TotalCatch_ORBSReleased_Sampled $Species,
SampledCatch= Est_TotalCatch_ORBSReleased_Sampled $SampledCatch,
SampledReleased= Est_TotalCatch_ORBSReleased_Sampled $SampledReleased,
SampledTagged= Est_TotalCatch_ORBSReleased_Sampled $SampledTagged,
SampledInterviews= Est_TotalCatch_ORBSReleased_Sampled
$SampledInterviews,
EstCatch= Est_TotalCatch_ORBSReleased_Sampled $EstCatch,
SE.EstCatch= Est_TotalCatch_ORBSReleased_Sampled $SE.EstCatch,
EstReleased= Est_TotalCatch_ORBSReleased_Sampled $EstReleased,
SE.EstReleased= Est_TotalCatch_ORBSReleased_Sampled $SE.EstReleased
)

# Merge with unadjusted ORBS estimates
tbEstCatchRel_2011<-read.csv("tbEstCatchRel_2011.csv", header=TRUE)
names(tbEstCatchRel_2011)[6:7]<-c("BoatType", "TripType")
ORBS_2011_Estimates_ALL<-merge(ORBS_2011_Estimates_ORBS,
tbEstCatchRel_2011, by= names(tbEstCatchRel_2011) [1:9], all=TRUE)
# Remove weeks not in ORBS
ORBS_2011_Estimates_ALL<-merge(ORBS_2011_Estimates_ALL,
ORBSWeeks[,c(1,2,4)])

# sort
ORBS_2011_Estimates_ALL<-
ORBS_2011_Estimates_ALL[order(ORBS_2011_Estimates_ALL$Week,
ORBS_2011_Estimates_ALL$Port, ORBS_2011_Estimates_ALL$Area,
ORBS_2011_Estimates_ALL$BoatType, ORBS_2011_Estimates_ALL$TripType,
ORBS_2011_Estimates_ALL$ExpansionType,
ORBS_2011_Estimates_ALL$Species),]

# Export to CSV file
write.csv(ORBS_2011_Estimates_ALL,
"ORBS_2011_Estimates_ALL_20131210.csv")
save.image("2011_DataAnalysis_20131210.RData")
```

```
#-----------------------------------------------------------------------
# Calculate total catch by species by Port and Boat Type
# across Weeks and Expansion Types
# Calculate separately for Charter and Private boats then combine
# Calculate RATIO ESTIMATOR OF TOTAL CATCH for Charter Boat Type
# then adjust with known Charter boat trip
#-----------------------------------------------------------------------
# Specify sampling design for analysis in survey package
ORBS_2011_2stage_design.C <-subset(ORBS_2011_2stage_design,
      BoatType=="C")
ORBS_2011_2stage_design.P <-subset(ORBS_2011_2stage_design,
      BoatType=="P")

# Set lonely-PSU options
options(survey.adjust.domain.lonely=TRUE)
options(survey.lonely.psu= "adjust")      # use grand mean for variance

# Design-based estimation
Est_MeanCatch.Port.C <- svyby(~Catch, by= ~ Port + Species,
design = ORBS_2011_2stage_design.C,
FUN = svymean,
verbose=TRUE)
rownames(Est_MeanCatch.Port.C)<-NULL
dim(Est_MeanCatch.Port.C)                 # 299   4

# Obtain N.hl for each domain for Charter boats and estimate total
BoatMat.C<- unique(data.frame(Port =ORBS_2011$Port[ORBS_2011$BoatType
=="C"], Date=ORBS_2011$Date[ORBS_2011$BoatType=="C"], BoatTrip=
ORBS_2011$BoatTrip[ORBS_2011$BoatType=="C"], IntvNum=ORBS_2011$
IntvNum[ORBS_2011$BoatType=="C"]))
dim(BoatMat.C)                # 2229    4

# Calculate number of Charter Boats, N.C
N.C<-aggregate(rep(1,nrow(BoatMat.C)), list(Port= BoatMat.C$Port), sum)
      # Total boats trips by Port
names(N.C)[names(N.C)=="x"]<-"N.C"
Est_MeanCatch.Port.C<-merge(Est_MeanCatch.Port.C, N.C)
dim(Est_MeanCatch.Port.C)                 # 299   5

# Calculate ratio estimator of total for Charter boat trips
Est_TotalCatch.Port.C.adj<-data.frame(Est_MeanCatch.Port.C,
TotalCatch=Est_MeanCatch.Port.C$Catch* Est_MeanCatch.Port.C$N.C,
SE.TotalCatch= Est_MeanCatch.Port.C$se*Est_MeanCatch.Port.C$N.C)

# Compare to Horvitz-Thompson estimates of total catch
Est_TotalCatch.Port.C <- svyby(~Catch, by= ~ Port + Species,
design = ORBS_2011_2stage_design.C,
FUN = svytotal,
verbose=TRUE)
rownames(Est_TotalCatch.Port.C)<-NULL
CompareEsts<-
merge(Est_TotalCatch.Port.C.adj[,c(1:2,6:7)],Est_TotalCatch.Port.C,
by=c("Port", "Species"))
CompareEsts[,3:6]<-round(CompareEsts[,3:6],0)
names(CompareEsts)[5:6]<-c("TotalCatch.HT", "SE.TotalCatch.HT")
# Note differences in estimated total catch caused by bias in the
# estimated total boat trips
```

```
# Calculate total catch for Private Boat Trips
Est_TotalCatch.Port.P <- svyby(~Catch, by= ~ Port + Species,
design = ORBS_2011_2stage_design.P,
FUN = svytotal,
verbose=TRUE)
rownames(Est_TotalCatch.Port.P)<-NULL

names(Est_TotalCatch.Port.C.adj)<- names(Est_TotalCatch.Port.P)
Est_TotalCatch.Port.BT<-rbind(data.frame(BoatType="C",
Est_TotalCatch.Port.C), data.frame(BoatType="P",
Est_TotalCatch.Port.P))
Est_TotalCatch.Port.BT<- Est_TotalCatch.Port.BT
[order(Est_TotalCatch.Port.BT $Port, Est_TotalCatch.Port.BT $Species),]
rownames(Est_TotalCatch.Port.BT)<-NULL

# Sum across Boat Type
Est_TotalCatch.Port.Catch<-aggregate(Est_TotalCatch.Port.BT$Catch,
list(Port =Est_TotalCatch.Port.BT$Port, Species=
Est_TotalCatch.Port.BT$Species), sum)
names(Est_TotalCatch.Port.Catch)[
names(Est_TotalCatch.Port.Catch)=="x"]<-"Catch"
Est_TotalCatch.Port.SE<-aggregate(Est_TotalCatch.Port.BT$se, list(Port
=Est_TotalCatch.Port.BT$Port, Species= Est_TotalCatch.Port.BT$Species),
sum)
names(Est_TotalCatch.Port.SE)[ names(Est_TotalCatch.Port.SE)=="x"]<-
"SE"

# obtain Spc Names
Est_TotalCatch.Port<-merge(Est_TotalCatch.Port.Catch,
Est_TotalCatch.Port.SE)
Est_TotalCatch.Port<-merge(Est_TotalCatch.Port, tbSpp[,c(1,3)])

# Sort results
Est_TotalCatch.Port<- Est_TotalCatch.Port[,c(2,1,5,3,4)]
Est_TotalCatch.Port<-
Est_TotalCatch.Port[order(Est_TotalCatch.Port$Port,
Est_TotalCatch.Port$ SpcName),]

# Export output to CSV file
write.csv(Est_TotalCatch.Port, "Est_TotalCatch_Port_20131210.csv")
```

```
#--------------------------------------------------------------------
# Estimate number of boat trips by Port for Charter boat type
#--------------------------------------------------------------------
# Note that Angler and Boat data are duplicated for every Species
# Create boat-level data set that pools across species

ORBS_2011_BoatTrip<-aggregate(ORBS_2011$Catch, by=list(Date
=ORBS_2011$Date, Port =ORBS_2011$Port, Week =ORBS_2011$Week,
ExpansionType =ORBS_2011$ ExpansionType, Stratum =ORBS_2011$Stratum,
PSU.Stratum =ORBS_2011$PSU.Stratum, Area= ORBS_2011$ Area, Anglers=
ORBS_2011$ Anglers, BoatTrip= ORBS_2011$ BoatTrip, BoatType=
ORBS_2011$BoatType, TripType= ORBS_2011$TripType, InclProb=
ORBS_2011$InclProb), sum)
dim(ORBS_2011_BoatTrip)             # 12451    13
names(ORBS_2011_BoatTrip)[ names(ORBS_2011_BoatTrip)=="x"]<-"Catch"

# Specify sampling design for analysis in survey package
ORBS_2011_BoatTrip.C<- ORBS_2011_BoatTrip [ORBS_2011_BoatTrip
$BoatType=="C",]
ORBS_2011_BoatTrip.C $BoatTrip<-1
ORBS_2011_2stage_BT.C<-svydesign(id=~ Date+BoatTrip, probs=~InclProb,
      strata= ~ PSU.Stratum + BoatType, nest=TRUE, data=
      ORBS_2011_BoatTrip.C, variables = ~ Port + Week + Area+ Anglers +
      BoatType + TripType + ExpansionType + BoatTrip)

# Set lonely-PSU options
options(survey.adjust.domain.lonely=TRUE)
options(survey.lonely.psu= "adjust")      # use grand mean for variance

# Design-based estimation
Est_TotalBoatTrips.Port.C <- svyby(~BoatTrip, by= ~ Port,
      design = ORBS_2011_2stage_BT.C,
      FUN = svytotal,
verbose=TRUE)
rownames(Est_TotalBoatTrips.Port.C)<-NULL

# Compare with true counts, N.C
merge(N.C, Est_TotalBoatTrips.Port.C)
save.image("2011_DataAnalysis_20131210.RData")
```

```
#---------------------------------------------------------------------
# Calculate total catch by Port and Species
# only for ExpansionType = "W"
#---------------------------------------------------------------------

# Specify sampling design for analysis in survey package
ORBS_2011_2stage_design.W <-subset(ORBS_2011_2stage_design,
      ExpansionType=="W")

# Set lonely-PSU options
options(survey.adjust.domain.lonely=TRUE)
options(survey.lonely.psu= "adjust")

# Design-based estimation
TotalCatch_Spc_W <- svyby(~Catch, by= ~ Port + Species,
      design = ORBS_2011_2stage_design.W,
      FUN = svytotal,
      verbose=TRUE)

rownames(TotalCatch_Spc_W)<-NULL

# Obtain species names
dim(TotalCatch_Spc_W)
TotalCatch_Spc_W <-merge(tbSpp, TotalCatch_Spc_W)
dim(TotalCatch_Spc_W)
TotalCatch_Spc_W <- TotalCatch_Spc_W [order(TotalCatch_Spc_W
$SpcName),]

# Export output to CSV file
write.csv(TotalCatch_Spc_W, "RatioEst_Catch_Spc_W_20131210.csv")

save.image("2011_DataAnalysis_20131210.RData")
```

```
#---------------------------------------------------------------------
# Estimation of Angler totals by Stratum, Boat Type, and Trip Type
#---------------------------------------------------------------------

# Note that Angler and Boat data are duplicated for every Species
# Create boat-level data set that pools across species

ORBS_2011_BoatTrip<-aggregate(ORBS_2011$Catch, by=list(Date
=ORBS_2011$Date, Port =ORBS_2011$Port, Week =ORBS_2011$Week,
ExpansionType =ORBS_2011$ ExpansionType, Stratum =ORBS_2011$Stratum,
PSU.Stratum =ORBS_2011$PSU.Stratum, Area= ORBS_2011$ Area, Anglers=
ORBS_2011$ Anglers, BoatTrip= ORBS_2011$ BoatTrip, BoatType=
ORBS_2011$BoatType, TripType= ORBS_2011$TripType, InclProb=
ORBS_2011$InclProb), sum)
dim(ORBS_2011_BoatTrip)              # 12451    13
names(ORBS_2011_BoatTrip)[ names(ORBS_2011_BoatTrip)=="x"]<-"Catch"

# Specify sampling design for analysis in survey package
ORBS_2011_2stage_design_BT<-svydesign(id=~ Date+BoatTrip,
      probs=~InclProb, strata= ~ PSU.Stratum + BoatType, nest=TRUE,
      data= ORBS_2011_BoatTrip, variables = ~ Port + Week + Area+
      Anglers +  BoatType + TripType + ExpansionType)

# Set lonely-PSU options
options(survey.adjust.domain.lonely=TRUE)
options(survey.lonely.psu= "adjust")

# Design-based estimation
TotalAnglers_Strata.BT.TT <- svyby(~Anglers, by= ~ Port + Week
+ExpansionType + BoatType + TripType,
      design = ORBS_2011_2stage_design_BT,
      FUN = svytotal,
      verbose=TRUE)

rownames(TotalAnglers_Strata.BT.TT)<-NULL

# Export output to CSV file
write.csv(TotalAnglers_Strata.BT.TT,
"TotalAnglers_Strata.BT.TT_20131210.csv")
save.image("2011_DataAnalysis_20131210.RData")
```