

DRAFT

**MONITORING AND EVALUATION OF SMOLT MIGRATION  
IN THE COLUMBIA BASIN**

**VOLUME VII**

Evaluation of the Compliance Testing Framework for RPA Improvement as Stated in  
the 2000 Federal Columbia River Power System (FCRPS) Biological Opinion

Prepared by:

John R. Skalski

and

Roger F. Ngouenet

School of Aquatic and Fishery Sciences

University of Washington

1325 Fourth Avenue, Suite 1820

Seattle, WA 98101-2509

Prepared for:

U.S. Department of Energy

Bonneville Power Administration

Division of Fish and Wildlife

P.O. Box 3621

Portland, OR 97208-3621

Project No. 91-051-00

Contract No. 96BI-91572

May 2001

Monitoring and Evaluation of Smolt Migration in the Columbia Basin

**Other Publications in this Series**

**Volume I:** Townsend, R. L., J. R. Skalski, and D. Yasuda. 1997. Evaluation of the 1995 predictions of run-timing of wild migrant subyearling chinook in the Snake River Basin using program RealTime. Technical Report (DOE/BP-35885-11) to BPA, Project 91-051-00, Contract 91-BI-91572.

**Volume II:** Townsend, R. L., J. R. Skalski, and D. Yasuda. 1998. Evaluation of the 1996 predictions of run-timing of wild migrant subyearling chinook in the Snake River Basin using program RealTime. Technical Report (DOE/BP-91572-2) to BPA, Project 91-051-00, Contract 91-BI-91572.

**Volume III:** Townsend, R. L., J. R. Skalski, and D. Yasuda. 2000. Evaluation of the 1997 predictions of run-timing of wild migrant yearling and subyearling chinook and sockeye in the Snake River Basin using program RealTime. Technical Report (accepted) to BPA, Project 91-051-00, Contract 91-BI-91572.

**Volume IV:** Burgess,C., R. L. Townsend, J..R. Skalski, and D. Yasuda. 2000. Evaluation of the 1998 predictions of the run-timing of wild migrant yearling and subyearling chinook and steelhead, and hatchery sockeye in the Snake River Basin using program RealTime. Technical Report (submitted) to BPA, Project 91-051-00, Contract 96BI-91572.

**Volume V:** Burgess,C., J..R. Skalski. 2000. Evaluation of the 1999 predictions of the run-timing of wild migrant yearling and subyearling chinook salmon and steelhead trout, and hatchery sockeye salmon in the Snake River Basin using program RealTime. Technical Report (submitted) to BPA, Project 91-051-00, Contract 96BI-91572.

**Volume VI:** Burgess,C., J..R. Skalski. 2000. Evaluation of the 2000 predictions of the run-timing of wild migrant chinook salmon and steelhead trout, and hatchery sockeye salmon in the Snake River Basin, and combined wild and Hatchery Salmonids migrating to Rock Island and McNary Dams using program RealTime. Technical Report (submitted) to BPA, Project 91-051-00, Contract 96BI-91572.

**Other Publications Related to this Series**

Other related publications, reports and papers available through the professional literature or from the Bonneville Power Administration (BPA) Public Information Center - CKPS-1, P.O. Box 3621, Portland, OR 97208.

Lady, J., and J. R. Skalski. 1998. Estimators of stream residence time of Pacific salmon (spp. *Oncorhynchus*) based on release-recapture data. *Canadian Journal of Fisheries and Aquatic Sciences* 55:2580-2587.

## DRAFT

- Lowther, A. B., and J. R. Skalski. 1998. A multinomial likelihood model for estimating survival probabilities and residualization for fall chinook salmon (*Oncorhynchus tshawytscha*) smolt using release-recapture methods. *Journal of Agricultural Biology and Environmental Statistics* 3:223-236.
- Mathur, D., P. G. Heisey, E. T. Euston, and J. R. Skalski. 1996. Turbine passage survival estimates for chinook salmon smolt (*Oncorhynchus tshawytscha*) at a large dam on the Columbia River. *Canadian Journal of Fisheries and Aquatic Sciences* 53:542-549.
- Ryding, K. E., and J. R. Skalski. 1999. Multivariate regression relationships between ocean conditions and early marine survival of coho salmon (*Oncorhynchus kisutch*). *Canadian Journal of Fisheries and Aquatic Sciences* 56:2374-2384.
- Skalski, J. R. 1998. Estimating season-wide survival rates of outmigrating smolt in the Snake River, Washington. *Canadian Journal of Fisheries and Aquatic Sciences* 55:761-769.
- Skalski, J. R., S. G. Smith, R. N. Iwamoto, J. G. Williams, and A. Hoffmann. 1998. Use of PIT-tags to estimate survival of migrating juvenile salmonids in the Snake and Columbia Rivers. *Canadian Journal of Fisheries and Aquatic Sciences* 55:1484-1493.
- Skalski, J. R. 1996. Regression of abundance estimates from mark-recapture surveys against environmental covariates. *Canadian Journal of Fisheries and Aquatic Sciences* 53:196-204.
- Townsend, R. L., D. Yasuda, and J. R. Skalski. 1997. Evaluation of the 1996 predictions of run timing of wild migrant spring/summer yearling chinook in the Snake River Basin using Program RealTime. Technical Report (DOE/BP-91572-1) to BPA, Project 91-051-00, Contract 91-BI-91572.
- Townsend, R. L., P. Westhagen, D. Yasuda, J. R. Skalski, and K. Ryding. 1996. Evaluation of the 1995 predictions of run timing of wild migrant spring/summer yearling chinook in the Snake River Basin using Program RealTime. Technical Report (DOE/BP-35885-9) to BPA, Project 91-051-00, Contract 87-BI-35885.
- Townsend, R. L., P. Westhagen, D. Yasuda, and J. R. Skalski. 1995. Evaluation of the 1994 predictions of run-timing capabilities using PIT-tag databases. Technical Report (DOE/BP-35885-8) to BPA, Project 91-051-00, Contract 87-BI-35885.
- Skalski, J. R., G. Tartakovsky, S. G. Smith, P. Westhagen, and A. E. Giorgi. 1994. Pre-1994 season projection of run-timing capabilities using PIT-tag databases. Technical Report (DOE/BP-35885-7) to BPA, Project 91-051-00, Contract 87-BI-35885.
- Skalski, J. R., and A. E. Giorgi. 1993. A plan for estimating smolt travel time and survival in the Snake and Columbia Rivers. Technical Report (DOE/BP-35885-3) to BPA, Project 91-051-00, Contract 87-BI-35885.
- Smith, S. G., J. R. Skalski, and A. E. Giorgi. 1993. Statistical evaluation of travel time estimation based on data from freeze-branded chinook salmon on the Snake River, 1982-1990. Technical Report (DOE/BP-35885-4) to BPA, Project 91-051-00, Contract 87-BI-35885.

# DRAFT

## PREFACE

Project 91-051 was initiated in 1991 in response to the Endangered Species Act (ESA) listings in the Snake River Basin of the Columbia River Basin. Primary objectives and management implications of this project include: (1) to address the need for further synthesis of historical tagging and other biological information to improve understanding and identify future research and analysis needs; (2) to assist in the development of improved monitoring capabilities, statistical methodologies and software tools to aid management in optimizing operational and fish passage strategies to maximize the protection and survival of listed threatened and endangered Snake River salmon populations and other listed and nonlisted stocks in the Columbia River Basin; (3) to design better analysis tools for evaluation programs; and (4) to provide statistical support to the Bonneville Power Administration and the Northwest fisheries community.

The following report addresses measure 4.3C of the 1994 Northwest Power Planning Council's Fish and Wildlife Program with emphasis on improved monitoring and evaluation of smolt migration in the Columbia River Basin. In this report, statistical models are used to evaluate the framework for compliance testing of the RPA improvements using the information provided in the Federal Columbia River Power System (FCRPS) 2000 Biological Opinion (BO). The main concern is to evaluate the anticipated performance of two statistical hypothesis tests proposed in the 2000 FCRPS BO. It is hoped that assessing the compliance rules before actual data are collected will help avoid any unpleasant surprises concerning the statistical behavior of the proposed decision rules for compliance evaluation in 2005 and 2008. Having the capability to correctly assess the outcome of the RPAs should improve the public confidence in the recovery process and should also contribute to the regional goal of increasing juvenile passage survival through the Columbia River System.

## DRAFT

### ABSTRACT

Using the pre-2000 reach survival probabilities reported in the 2000 FCRPS Biological Opinion (BO) for three selected stocks: yearling and sub-yearling chinook and steelhead, power curves were constructed for each of the two statistical hypothesis tests suggested in the BO. These power calculation results were interpreted in terms of the ability of the statistical tests to correctly identify the true states of recovery (i.e., fail or succeed in fulfilling RPA expectations). The proposed one-sided tests have a moderate to low probability of correctly assessing the true status of the recovery by the years 2005 and 2008. The relatively poor odds of making the correct decision with the BO proposed Tests 1 and 2 suggest alternative decision rules need to be investigated and developed for assessing RPA compliance. Therefore, we propose to immediately examine alternative decision rules that might maximize the likelihood of correct decisions while minimizing the prospect of incorrect decisions. The Bayesian analysis will incorporate scientific/biological knowledge/expertise.

## DRAFT

### EXECUTIVE SUMMARY

The Federal Columbia River Power System (FCRPS) completed its 2000 Biological Opinion (BO) in compliance with Section 7(a)(2) of the Endangered Species Act (ESA), which requires Federal agencies to consult with the US Fish and Wildlife Services (USFWS) and the National Marine Fisheries Services (NMFS) to ensure that their actions are not likely to jeopardize the continued existence of endangered or threatened species or to adversely modify or destroy their designated critical habitats.

The 2000 FCRPS BO recommended performance measures and system goals to help recover listed salmonid species. Under the BO, NMFS will evaluate the RPA performance in 2005 and 2008 by comparing post-2000 average survival with pre-2000 average survival estimates, published in the 2000 BO, plus the expected RPA survival improvements.

#### **Objective**

The objective of this preliminary report was to evaluate the framework for compliance testing of the RPA improvements using the information provided in the BO. The main goal was to evaluate the anticipated performance of two statistical hypothesis tests proposed in the 2000 FCRPS BO. Using power calculations, performance of the RPA decision rules was assessed. As such, we seek to determine the statistical behavior of the proposed decision rules well before the schedule compliance evaluations in 2005 and 2008. Should problems in the compliance rules be identified, follow-up research will help to identify alternative testing procedures in time for formal RPA evaluation by NMFS.

#### **Results**

Using the pre-2000 reach survival probabilities reported in the 2000 FCRPS BO for three selected stocks: yearling and sub-yearling chinook salmon and steelhead, power curves were constructed for each of the two statistical hypothesis tests suggested in the BO. The results of the power calculations were interpreted in terms of the ability of the statistical tests to correctly identify the true states of recovery (i.e., fail or succeed in fulfilling RPA expectations).

Tables 6, 9, and 12 summarize the probabilities of making the correct decisions with the proposed statistical hypothesis tests. The proposed one-sided tests have only a low to moderate probability of correctly assessing the true status of the recovery by the year 2005 and 2008.

## **Recommendations**

The relatively poor chance of making the correct decisions with the proposed BO Tests 1 and 2 suggest alternative decision rules need to be investigated and developed for assessing RPA compliance. The development and selection of decision rules should proceed immediately. The credibility of the scientific process begun by the BO could be seriously jeopardized if the public perceives the rules will be established only after the results are known. Lack of scientific objectivity could undermine public confidence in not only the ESA process but also in the agencies involved

The next phase of this project is to examine alternative decision rules that might maximize the likelihood of correct decisions while minimizing the prospect of incorrect decisions. In particular, special attention will be given to the novel “two one-sided tests” (TOST), spline regression techniques, as well as Bayesian decision rules.

# DRAFT

## ACKNOWLEDGMENTS

This report was funded by the Bonneville Power Administration (BPA), US Department of Energy, under Contract No. 96BI91572, as part of the BPA's program to protect, mitigate, and enhance fish and wildlife affected by the development and operation of hydroelectric facilities on the Columbia River and its tributaries.

We wish to express thanks to the many federal agencies that have expended considerable resources in completing the Federal Columbia River Power System (FCRPS) 2000 Biological Opinion.

DRAFT

**TABLE OF CONTENTS**

1.0 INTRODUCTION .....	1
2.0 DESCRIPTION OF DATA .....	2
3.0 STATISTICAL HYPOTHESIS TESTING .....	4
4.0 RESULTS .....	8
4.1 YEARLING CHINOOK SALMON .....	8
4.2 SUBYEARLING CHINOOK SALMON .....	14
4.3 STEELHEAD .....	20
5.0 MONTE CARLO MODELING APPROACH .....	26
5.1 METHODS .....	26
5.2 RESULTS .....	28
6.0 DISCUSSION .....	30
7.0 LITERATURE CITED .....	32

DRAFT

LIST OF FIGURES

Figure 1. Power of Test #1 for yearling chinook salmon based on inriver smolt survival between Lower Granite and Bonneville dams for (a)  $n_2 = 5$  years of RPA (2001-2005) and (b)  $n_2 = 8$  years of RPA (2001-2008). .....9

Figure 2. Power of Test #2 for yearling chinook salmon based on inriver survival between Lower Granite and Bonneville dams for (1) (a)  $n_2 = 5$  years of RPA (2001-2005) and (b)  $n_2 = 8$  years of RPA (2001-2008). .....10

Figure 3. Power of Test #1 for sub-yearling chinook salmon based on inriver smolt survival between Lower Granite and Bonneville dams for (a)  $n_2 = 5$  years of RPA (2001-2005) and (b)  $n_2 = 8$  years of RPA (2001-2008).....15

Figure 4. Power of Test #2 for sub-yearling chinook salmon based on inriver survival between Lower Granite and Bonneville dams for (1) (a)  $n_2 = 5$  years of RPA (2001-2005) and (b)  $n_2 = 8$  years of RPA (2001-2008). .....16

Figure 5. Power of Test #1 for steelhead based on inriver smolt survival between Lower Granite and Bonneville dams for (a)  $n_2 = 5$  years of RPA (2001-2005) and (b)  $n_2 = 8$  years of RPA (2001-2008). .....21

Figure 6. Power of Test #2 for steelhead based on inriver survival between Lower Granite and Bonneville dams for (1) (a)  $n_2 = 5$  years of RPA (2001-2005) and (b)  $n_2 = 8$  years of RPA (2001-2008). .....22

Figure 7. Immediate and sustained improvement of size  $\Delta = 0.09$  beginning in 2000. ....27

Figure 8. Gradual improvement in survival of size  $\Delta = 0.09$  by the end of year 2003, 2004, 2005, 2006, or 2007.....27

DRAFT

LIST OF TABLES

Table 1. Reach survival rates of juvenile yearling (spring/summer) chinook salmon through FCRPS: 1994-1999, Lower Granite (LGR) to Bonneville (BON), McNary (MCN) to Bonneville, and LGR to MCN.....2

Table 2. Reach survival rates of juvenile subyearling (fall) chinook salmon through FCRPS: 1995-1999, Lower Granite (LGR) to Bonneville (BON), McNary (MCN) to Bonneville, LGR to MCN.....3

Table 3. Each survival rates of juvenile Steelhead through FCRPS: 1994-1999, Lower Granite (LGR) to Bonneville (BON), McNary (MCN) to Bonneville, LGR to MCN. ....3

Table 4. Summary of estimated effects of the RPA in the action areas by fish stock .....4

Table 5. Probabilities of marking correct (shaded) and incorrect (unshaded) decisions using Test #1 at  $\alpha = 0.05$  for yearling chinook salmon for the survival estimates from Lower Granite to Bonneville dams under different states of nature by (a) 2005 and (b) 2008.....11

Table 6. Probabilities of making correct (shaded) and incorrect (unshaded) decisions using Test #2 at  $\alpha = 0.05$  for yearling chinook salmon for the survival estimates from Lower Granite to Bonneville dams under different states of nature by (a) 2005 and (b) 2008. ....12

Table 7. Probabilities Tests #1 and #2 will make the correct decisions, individually and jointly, under alternative states of nature at  $\alpha = 0.05$  for the yearling chinook salmon for (a) 2005 and (b) 2008. ....13

Table 8. Probabilities of marking correct (shaded) and incorrect (unshaded) decisions using Test #1 at  $\alpha = 0.05$  for subyearling chinook salmon for the survival estimates from Lower Granite to Bonneville dams under different states of nature by (a) 2005 and (b) 2008.....17

Table 9. Probabilities of making correct (shaded) and incorrect (unshaded) decisions using Test #2 at  $\alpha = 0.05$  for subyearling chinook salmon for the survival estimates from Lower Granite to Bonneville dams under different states of nature by (a) 2005 and (b) 2008.....18

Table 10. Probabilities Tests #1 and #2 will make the correct decisions, individually and jointly, under alternative states of nature at  $\alpha = 0.05$  for the subyearling chinook salmon for (a) 2005 and (b) 2008.....19

Table 11. Probabilities of Probabilities of marking correct (shaded) and incorrect (unshaded) decisions using Test #1 at  $\alpha = 0.05$  for steelhead for the survival estimates from Lower Granite to Bonneville dams under different states of nature by (a) 2005 and (b) 2008.....23

Table 12. Probabilities of making correct (shaded) and incorrect (unshaded) decisions using Test #2 at  $\alpha = 0.05$  for steelhead for the survival estimates from Lower Granite to Bonneville dams under different states of nature by (a) 2005 and (b) 2008.....24

DRAFT

LIST OF TABLES (Continued)

Table 13. Probabilities Tests #1 and #2 will make the correct decisions, individually and jointly, under alternative states of nature at  $\alpha = 0.05$  for the steelhead for (a) 2005 and (b) 2008. ....25

Table 14. Statistical power for the three stocks (yearling chinook, subyearling chinook, and steelhead) based on inriver smolt survival between Lower Granite and Bonneville dams for eight years of RPA using Tests #1 and #2 at  $\alpha = 0.05$  under the scenario of gradual recovery. Statistical probabilities are the probabilities of correctly identifying the intermediate state of some but not all recovery.....29

Table 15. Statistical power for the three stocks (yearling chinook, subyearling chinook, and steelhead) based on inriver smolt survival between Lower Granite and Bonneville dams for eight years of RPA using Tests #1 and #2 at  $\alpha = 0.05$  under the scenario of gradual recovery. Statistical probabilities are the probabilities of correctly identifying the state of full recovery with a change of size  $\Delta$  or greater. ....29

Table 16. Required improvement in survival between Lower Granite and Bonneville dams for the three fish stocks (yearling chinook, subyearling chinook, and steelhead)in order for Tests #1 and #2 to have the same statistical power under the scenario of gradual improvement as in the case of immediate and substantial improvement beginning in 2000. ....30

## 1.0 Introduction

The Federal Columbia River Power System (FCRPS) completed its 2000 Biological Opinion (BO) in compliance with Section 7(a)(2) of the Endangered Species Act (ESA), which requires Federal agencies to consult with the US Fish and Wildlife Services (USFWS) and the National Marine Fisheries Services (NMFS) to ensure that their actions are not likely to jeopardize the continued existence of endangered or threatened species or to adversely modify or destroy their designated critical habitats.

The ESA requires that the mortality of listed salmonids in the different Evolutionary Significant Units (ESUs) that can be attributed to the actions must be below the following:

“A level that, when combined with mortality occurring in other life stages, results in a high likelihood of survival and a moderate to high likelihood of recovery.”

To achieve this goal, the 2000 FCRPS BO recommended performance measures and system goals to help recover listed salmonid species. Under the BO, NMFS will evaluate the RPA performance in 2005 and 2008 by comparing post-2000 average survival with pre-2000 average survival estimates, published in the 2000 BO, plus the expected RPA survival improvements. For this, the BO suggested the following:

“The progress check might consist of series of two-samples statistical tests on one-sided hypotheses about juvenile survival levels. The tests would take into account uncertainty in both the 1994-to-1999 and the more recent average. A first test could check whether the post-2000 estimate of survival was significantly lower than the 1994-to-1999 average, plus RPA improvements. The second test could check whether post-2000 survival was significantly higher than the 1994-to-1999 average.”

In this report, we evaluated the anticipated performance of these two proposed statistical hypothesis tests to analyze the RPA before any post-2000 data are collected. Testing the compliance criteria while the data are still being collected is the safest approach to avoid any unpleasant surprises concerning the statistical behavior of the proposed decision rules. The challenge is to define decision rules that minimize the risk of making an error in assessing BO standards compliance, in other words, concluding compliance when it has not been achieved, or concluding noncompliance when compliance has actually been achieved.

## 2.0 Description of Data

Survival probabilities at each FCRPS project were estimated by NMFS with the Simulated Passage (SIMPAS) spreadsheet model. NMFS used the most recent empirical passive integrated transponder (PIT)-tag reach information collected from 1994 through 1999 to estimate survival probabilities between successive dams (i.e., detection sites) for yearling and sub-yearling chinook and steelhead salmon. This study concentrates on the overall inriver survival rates of juvenile chinook and steelhead throughout the system (i.e., between Lower Granite and Bonneville dams). Given the inriver survival rates for each dam of the FCRPS network, the overall reach survival through the FCRPS projects for a specified year is the product of the estimates for each of the shorter reaches.

The data used in this analysis came from tables showing project survival rates of juvenile salmonids in Appendix D of the 2000 FCRPS biological opinion. These tables recorded for a given year three types of survival rates: reach, pool and dam. One table was presented for each of the three ESUs: yearling chinook, subyearling chinook, and steelhead. Tables 1-3 summarize the survival rates of juvenile salmonids used to investigate the RPA improvements. For the yearling chinook and steelhead salmonids, data are available from 1994 to 1999. The subyearling chinook data is available only from 1995 to 1999. The parameters of interest in our study are the number of years and the mean and variance of the annual survival estimates. For example, from Table 1, the test of RPA compliance for the survival from McNary to Bonneville will use 6 years of baseline estimates, the mean survival 0.575, and the variance 0.0044.

Table 1. Reach survival rates of juvenile yearling (spring/summer) chinook salmon through FCRPS: 1994-1999, Lower Granite (LGR) to Bonneville (BON), McNary (MCN) to Bonneville, and LGR to MCN.

Year	LGR to BON	LGR to MCN	MCN to BON
1994	0.272	0.586	0.465
1995	0.418	0.692	0.604
1996	0.407	0.733	0.555
1997	0.385	0.687	0.560
1998	0.451	0.743	0.607
1999	0.518	0.786	0.660
Mean	0.409	0.704	0.575

DRAFT

Table 2. Reach survival rates of juvenile subyearling (fall) chinook salmon through FCRPS: 1995-1999, Lower Granite (LGR) to Bonneville (BON), McNary (MCN) to Bonneville, LGR to MCN.

Year	LGR to BON	LGR to MCN	MCN to BON
1995	0.164	0.415	0.396
1996	0.113	0.294	0.386
1997	0.005	0.082	0.059
1998	0.139	0.348	0.399
1999	0.086	0.364	0.237
Mean	0.102	0.301	0.296

Table 3. Each survival rates of juvenile Steelhead through FCRPS: 1994-1999, Lower Granite (LGR) to Bonneville (BON), McNary (MCN) to Bonneville, LGR to MCN.

Year	LGR to BON	LGR to MCN	MCN to BON
1994	0.322	0.615	0.523
1995	0.478	0.747	0.640
1996	0.428	0.730	0.586
1997	0.456	0.766	0.595
1998	0.417	0.683	0.611
1999	0.402	0.702	0.573
Mean	0.417	0.707	0.588

The broad ranges of the survival rates on Table 4 reflect variable hydraulic/environmental condition and uncertainty about delayed mortality.

Table 4. Summary of estimated effects of the RPA in the action areas by fish stock

Location / Stock	Survival Rates		
	Juvenile Standard*	Juvenile Current*	$\Delta^{**}$
<u>Snake River Spring/Summer Chinook</u>	35-62%	27-52%	~9%
Snake River Fall Chinook	1-22%	0.5-15%	~5%
Upper Columbia Spring Chinook	55-76%	46-66%	~9%
Lower Columbia Spring Chinook	87-95%	83-91%	~5%
Lower Columbia Fall Chinook	57-85%	50-80%	~5%
Snake River Steelhead	42-58%	32-46%	~9%
Upper Columbia Steelhead	61-74%	57-64%	~9%
Middle Columbia Steelhead	61-74%	57-64%	~9%
Lower Columbia Steelhead	86-96%	85-92%	~4%

\* From Table 9.7-5 (2000 FCRPS BO)

\*\* From Table 9.7-18 (2000 FCRPS BO)

### 3.0 Statistical Hypothesis Testing

Hypothesis testing is the process of making rational decisions concerning the choice between alternative and mutually exclusive states of nature. The null hypothesis ( $H_o$ ) describes one state of nature; the alternative hypothesis ( $H_a$ ), an opposing state of nature. Only one of these states of nature can be true at any one time. The goal of hypothesis testing is to try to make the correct decision regarding the true state of nature as often as possible.

Under each state of nature (i.e.,  $H_o$  or  $H_a$ ), a correct or incorrect decision could be made. The Type I error is the probability of rejecting the null hypothesis ( $H_o$ ) when it is indeed true, denoted by  $\alpha$ . The value of  $\alpha$  is the significance level of the hypothesis test, often set *a priori* at 0.10, 0.05, or 0.01. The probability  $1-\alpha$  is the choice of not rejecting  $H_o$  when it is true. The Type II error is the probability of not rejecting  $H_o$  and concluding  $H_a$  when the alternative hypothesis ( $H_a$ ) is indeed true, denoted by  $\beta$ . The complement,  $1-\beta$ , is the probability of rejecting  $H_o$  when  $H_a$  is true, and called the power of the test.

## DRAFT

In order to evaluate the compliance criteria, the tests of hypotheses recommended by the BO

“A first test could check whether the post-2000 estimate of survival was significantly lower than the 1994-to-1999 average, plus RPA improvements”

**and**

“The second test could check whether post-2000 survival was significantly higher than the 1994-to-1999 average”

will be evaluated using the smolt survival data from Lower Granite Dam to Bonneville Dam (Tables 1-3). The RPA anticipates a 0.09 (Table 4) improvement in survival over that reach in subsequent years. In which case, the two tests can be more formally stated as follows.

### **Test #1:**

$H_o$ : The post-2000 estimate of survival is greater than the 1994-to-1999 (pre-2000) average plus 9% RPA expected survival improvements.

$H_a$ : The post-2000 estimate of survival is lower than the 1994-to-1999 (pre-2000) average plus 9% RPA expected survival improvements.

### **Test #2:**

$H_o$ : The post-2000 survival is lower than the 1994-to-1999 (pre-2000) survival rate in average.

$H_a$ : The post-2000 survival is greater than the 1994-to-1999 (pre-2000) survival rate in average.

These statistical hypotheses can be further rewritten in terms of population (post-2000 and pre-2000) parameters as follows.

### **Test #1:**

$$\begin{aligned} H_o &: \mu_{\text{post-2000}} - \mu_{\text{pre-2000}} \geq 0.09 \\ H_a &: \mu_{\text{post-2000}} - \mu_{\text{pre-2000}} < 0.09 \end{aligned} \tag{1}$$

**Test #2:**

$$\begin{aligned} H_o &: \mu_{\text{post-2000}} - \mu_{\text{pre-2000}} \leq 0 \\ H_a &: \mu_{\text{post-2000}} - \mu_{\text{pre-2000}} \geq 0 \end{aligned} \quad (2)$$

where

$\mu_{\text{pre-2000}}$  = pre-2000 mean survival,

$\mu_{\text{post-2000}}$  = post-2000 mean survival.

From the perspective of the BO, the null hypothesis of the first test assumes the RPA has been satisfied, unless there is evidence to the contrary. The null hypothesis of the second test assumes no improvement whatsoever, unless there is evidence to the contrary. As such, the two proposed tests of hypotheses juxtapose the nature of the statistical test. The apparent motivation of the two tests is to provide equal opportunity to conclude or reject the premise of recovery.

The tests of hypotheses (1) and (2) can be based on sample means using the annual survival estimate pre- and post-2000. Let

$\bar{x}_{\text{pre-2000}}$  = mean survival estimate for the years 1994 or 1995 to 1999,

$\bar{x}_{\text{post-2000}}$  = mean survival estimate for the years 2000 to 2005 or 2008.

The tests of hypotheses can be based on two-sample t-tests of the form

$$t = \frac{(\bar{x}_{\text{post-2000}} - \bar{x}_{\text{pre-2000}}) - (\mu_{\text{post-2000}} - \mu_{\text{pre-2000}})}{\sqrt{\widehat{Var}(\bar{x}_{\text{post-2000}} - \bar{x}_{\text{pre-2000}})}}$$

that follows a t-distribution with  $n_1 + n_2 - 2$  degrees of freedom.

Because the sample  $\bar{x}_{\text{post-2000}}$  has not yet been collected, we shall assume for this analysis an equal inter-annual variance (i.e.  $s^2$ ) during pre- and post-2000 years. Therefore,

$$\widehat{Var}(\bar{x}_{\text{post-2000}} - \bar{x}_{\text{pre-2000}}) = s^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)$$

for sample sizes  $n_1$  and  $n_2$  for the pre- and post-2000 periods, respectively.

Hence, the tests of hypotheses will be based on the t-statistic

DRAFT

$$t = \frac{(\bar{x}_{\text{post-2000}} - \bar{x}_{\text{pre-2000}}) - \Delta}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (3)$$

where  $\Delta = \mu_{\text{post-2000}} - \mu_{\text{pre-2000}}$  under the null hypotheses. Test #1 specifies  $\Delta = 0.09$  while Test #2 specifies  $\Delta = 0$  for yearling chinook salmon between Lower Granite and Bonneville dams. The power of Test #1 is given by the probability that the test statistic falls in the rejection region:

$$P \left[ t \leq \frac{\Delta}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} - t_{\alpha, \mu} \right]$$

The power of Test #2 is given by:

$$P \left[ t \leq -\frac{\Delta - 0.09}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} - t_{\alpha, \mu} \right].$$

Power calculations were performed for the one-tailed hypotheses (1) and (2) at  $\alpha = 0.05$  or  $\alpha = 0.10$  using test statistic (3) for different values of  $\Delta$ .

## 4.0 Results

For yearling and subyearling chinook salmon and steelhead, power curves were constructed for tests (1) and (2) as a function of the size of  $\Delta = \mu_{\text{post-2000}} - \mu_{\text{pre-2000}}$ . The results of the power calculations were then interpreted in terms of the ability of the statistical tests to correctly identify the states of recovery (i.e.,  $\Delta \geq .09$  or  $\Delta < .09$ )

### 4.1 Yearling Chinook Salmon

Figure 1 presents the power of the two-sample t-test to reject the null hypothesis of 0.09 improvement or greater in survival (1) between Lower Granite and Bonneville dams. When  $\Delta \approx 0$ , the first test has about a 50% chance of rejecting (1) at  $\alpha = 0.05$ .

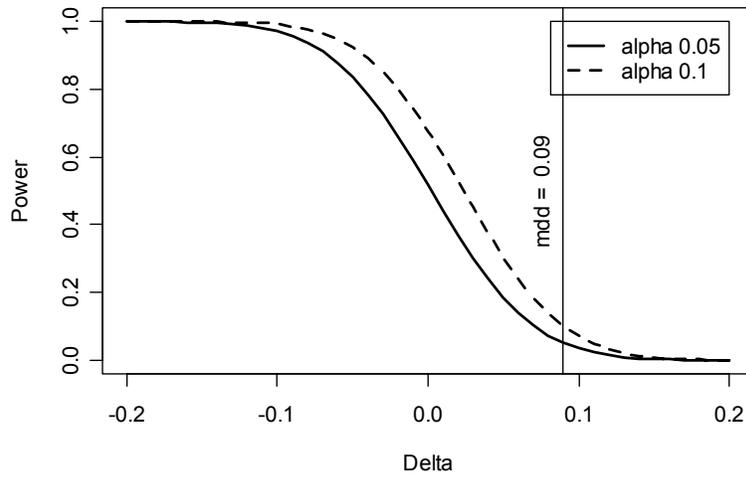
Figure 2 presents power of the two-sample t-test to reject the null hypothesis of no improvement in survival (2) between Lower Granite and Bonneville dams for yearling chinook salmon. At  $\Delta \approx 0.10$ , the t-test has approximately a 58% chance of rejecting (2) at  $\alpha = 0.05$ . A 0.20 improvement in survival between periods is needed before the t-test is almost certain to reject the null hypothesis of no improvement (2).

The results of the power curves in Figures 1-2 are summarized in Tables 5 and 6, respectively. By design, Test #1 will make an incorrect decision  $\alpha \cdot 100\%$  of the time and conclude  $\Delta < 0.09$  when impact recovery has occurred with  $\Delta \geq 0.09$ . However, Test #1 will make an incorrect decision between 49%-95% of the time and conclude  $\Delta > 0.09$  when in fact  $0 \leq \Delta < 0.09$ . The mean survival during post-2000 years can be less than pre-2000 years, and Test #1 has a  $\leq 48\%$  chance of concluding  $\Delta > 0.09$  (Table 5). By design, Test #2 will make an incorrect decision  $\alpha \cdot 100\%$  of the time and conclude  $\Delta > 0$  when, in fact,  $\Delta \leq 0$ . However, Test #2 will make an incorrect decision between 48%-95% of the time and conclude  $\Delta < 0$  when, in fact,  $0 \leq \Delta < 0.09$ . Test #2 has a  $\geq 52\%$  chance of making the correct decision when the improvement in survival  $\geq 0.09$ .

The results of the power calculations for Tests 1 and 2 are summarized under alternative scenarios for recovery by the years 2005 and 2008 in Tables 5 and 6. The ideal results for Tests #1 and #2 would be to have probabilities of correct decisions near 1 in the shaded boxes and probabilities of incorrect decision near 0 in the unshaded boxes in Tables 5 and 6, respectively. Deviations from these expectations are a measure of the lack of performance of the proposed tests of compliance. Table 7 summarizes the probabilities of making the correct decisions with Tests 1-2 under alternative states of nature. The chance of both Tests 1 and 2 making the correct decision for the yearling chinook stock when  $\Delta < 0$  is  $\geq 59\%$  of the time. The chance is  $\geq 49\%$  that Tests 1 and 2 will both make the correct decision when  $\Delta > 0.09$ . There is only 0.25%-26% chance of the correct decision for both tests when  $0 < \Delta < 0.09$  by the year 2008 (Table 7b).

Figure 1. Power of Test #1 for yearling chinook salmon based on inriver smolt survival between Lower Granite and Bonneville dams for (a)  $n_2 = 5$  years of RPA (2001-2005) and (b)  $n_2 = 8$  years of RPA (2001-2008).

a.  $n_2 = 5$



b.  $n_2 = 8$

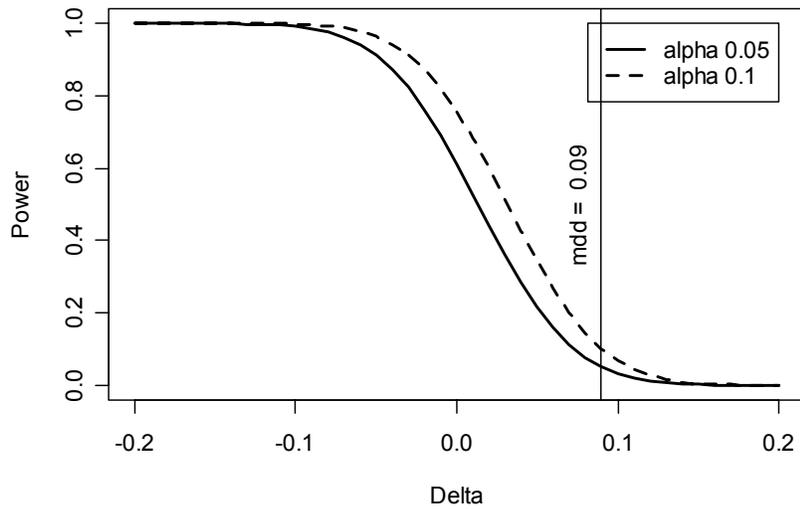
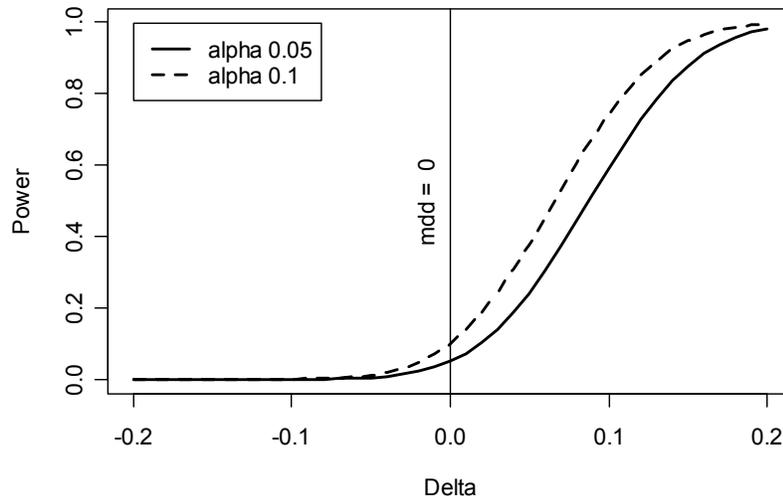
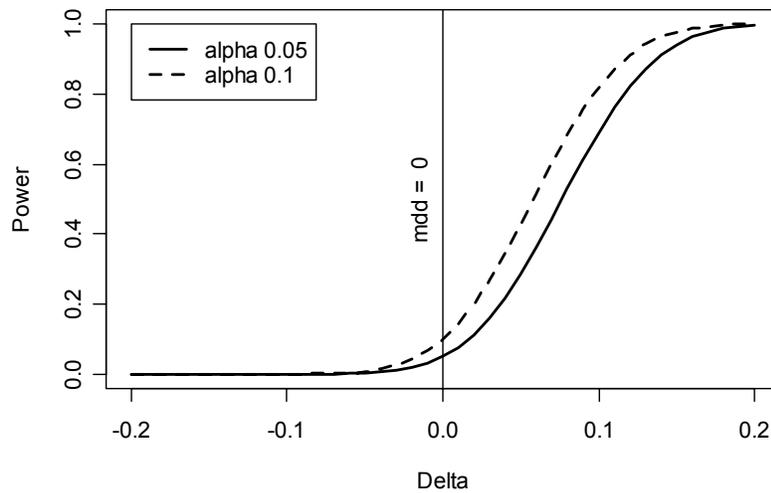


Figure 2. Power of Test #2 for yearling chinook salmon based on inriver survival between Lower Granite and Bonneville dams for (1) (a)  $n_2 = 5$  years of RPA (2001-2005) and (b)  $n_2 = 8$  years of RPA (2001-2008).

a.  $n_2 = 5$



b.  $n_2 = 8$



DRAFT

Table 5. Probabilities of marking correct (shaded) and incorrect (unshaded) decisions using Test #1 at  $\alpha = 0.05$  for yearling chinook salmon for the survival estimates from Lower Granite to Bonneville dams under different states of nature by (a) 2005 and (b) 2008.

a. 2005

		Alternative States of Nature		
		No Improvement $\Delta \leq 0$	Some Improvement $0 < \Delta < 0.09$	Recovery $\Delta \geq 0.09$
Conclude $\Delta \geq 0.09$		$0 < \beta < 0.49$	$0.49 < \beta < 0.95$	$1 - \alpha = 95$
Conclude $\Delta < 0.09$		$0.51 < 1 - \beta < 1.0$	$0.05 < 1 - \beta < 0.51$	$\alpha = 0.05$

b. 2008

		Alternative States of Nature		
		No Improvement $\Delta \leq 0$	Some Improvement $0 < \Delta < 0.09$	Recovery $\Delta \geq 0.09$
Conclude $\Delta \geq 0.09$		$0 < \beta < 0.38$	$0.38 < \beta < 0.95$	$1 - \alpha = 95$
Conclude $\Delta < 0.09$		$0.62 < 1 - \beta < 1.0$	$0.05 < 1 - \beta < 0.62$	$\alpha = 0.05$

DRAFT

Table 6. Probabilities of making correct (shaded) and incorrect (unshaded) decisions using Test #2 at  $\alpha = 0.05$  for yearling chinook salmon for the survival estimates from Lower Granite to Bonneville dams under different states of nature by (a) 2005 and (b) 2008.

a. 2005

		Alternative States of Nature		
		No Improvement $\Delta \leq 0$	Some Improvement $0 < \Delta < 0.09$	Recovery $\Delta \geq 0.09$
Conclude $\Delta \leq 0$		$1 - \alpha = 0.95$	$0.48 < \beta < 0.95$	$\beta < 0.48$
Conclude $\Delta > 0$		$\alpha = 0.05$	$0.05 < 1 - \beta < 0.52$	$1 - \beta \geq 0.52$

b. 2008

		Alternative States of Nature		
		No Improvement $\Delta \leq 0$	Some Improvement $0 < \Delta < 0.09$	Recovery $\Delta \geq 0.09$
Conclude $\Delta \leq 0$		$1 - \alpha = 0.95$	$0.38 < \beta < 0.95$	$\beta < 0.38$
Conclude $\Delta > 0$		$\alpha = 0.05$	$0.05 < 1 - \beta < 0.62$	$1 - \beta \geq 0.62$

DRAFT

Table 7. Probabilities Tests #1 and #2 will make the correct decisions, individually and jointly, under alternative states of nature at  $\alpha = 0.05$  for the yearling chinook salmon for (a) 2005 and (b) 2008.

a. 2005

		Alternative States of Nature		
		$\Delta \leq 0$	$0 < \Delta < 0.09$	$\Delta \geq 0.09$
Test #1	Reject $H_0$ $0.52 \leq 1 - \beta \leq 1.0$	Reject $H_0$ 0.05 1 .51	Do not reject $H_0$ $1 - \alpha = 0$	
Test #2	Do not reject $H_0$ $1 - \alpha = 0.95$	Reject $H_0$ 0.05 1 .52	Reject $H_0$ 0.52 1 1.0	
Joint Tests #1 and #2	0.49 – 0.95	0.0025 – 0.26	0.49 – 0.95	

b. 2008

		Alternative States of Nature		
		$\Delta \leq 0$	$0 < \Delta < 0.09$	$\Delta \geq 0.09$
Test #1	Reject $H_0$ $0.62 \leq 1 - \beta \leq 1.0$	Reject $H_0$ 0.05 1 .62	Do not reject $H_0$ $1 - \alpha = 0$	
Test #2	Do not reject $H_0$ $1 - \alpha = 0.95$	Reject $H_0$ 0.05 1 .62	Reject $H_0$ 0.62 1 1.0	
Joint Tests #1 and #2	0.59 – 0.95	0.0025 – 0.38	0.59 – 0.95	

## 4.2 Subyearling Chinook Salmon

Figure 3 presents the power of the two-sample t-test to reject the null hypothesis of 0.09 improvement or greater in survival (1) between Lower Granite and Bonneville dams. When  $\Delta \approx 0$ , the first test has about a 68% chance of rejecting (1) at  $\alpha = 0.05$ .

Figure 4 presents power of the two-sample t-test to reject the null hypothesis of no improvement in survival (2) between Lower Granite and Bonneville dams for sub-yearling chinook salmon. At  $\Delta \approx 0.10$ , the t-test has approximately a 75% chance of rejecting (2) at  $\alpha = 0.05$ . A 0.20 improvement in survival between periods is needed before the t-test is almost certain to reject the null hypothesis of no improvement (2).

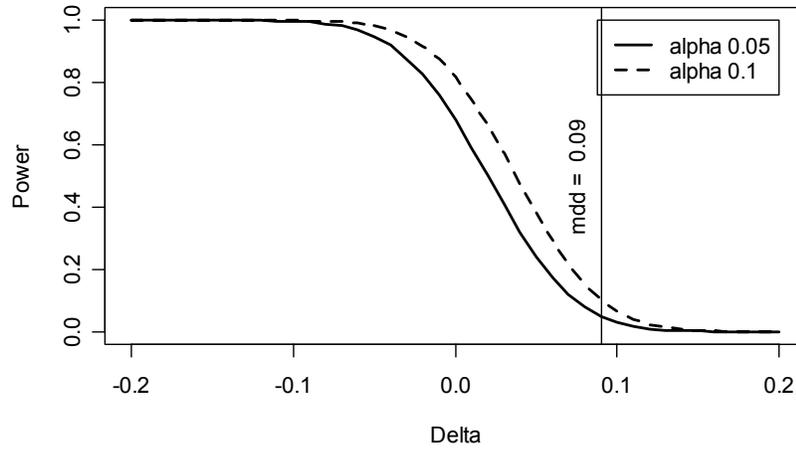
The results of the power curves in Figures 3-4 are summarized in Tables 8 and 9, respectively. By design, Test #1 will make an incorrect decision  $\alpha \cdot 100\%$  of the time and conclude  $\Delta < 0.09$  when impact recovery has occurred with  $\Delta \geq 0.09$ . However, Test #1 will make an incorrect decision between 32%-95% of the time and conclude  $\Delta > 0.09$  when in fact  $0 \leq \Delta < 0.09$ . The mean survival during post-2000 years can be less than pre-2000 years, and Test #1 has a  $\leq 32\%$  chance of concluding  $\Delta > 0.09$  (Table 8). By design, Test #2 will make an incorrect decision  $\alpha \cdot 100\%$  of the time and

conclude  $\Delta > 0$  when, in fact,  $\Delta \leq 0$ . However, Test #2 will make an incorrect decision between 32%-95% of the time and conclude  $\Delta < 0$  when, in fact,  $0 < \Delta < 0.09$ . Test #2 has a  $\geq 68\%$  chance of making the correct decision when the improvement in survival  $\geq 0.09$ .

The result of the power calculations for Tests #1 and #2 are summarized under alternative scenarios for recovery by the year 2005 and 2008 in Tables 8 and 9. The ideal results for Tests #1 and #2 would be to have probabilities of correct decisions near 1 in the shaded boxes and probabilities of incorrect decision near 0 in the unshaded boxes in Tables 8 and 9, respectively. Deviations from these expectations are a measure of the lack of performance of the proposed tests of compliance. Table 10 summarizes the probabilities of making the correct decisions with Test 1-2 under alternative states of nature for improvement in survival of the sub-yearling chinook salmon. The chance of both Tests 1 and 2 making the correct decision for the sub-yearling chinook stock when  $\Delta < 0$  is  $\geq 65\%$  of the time. The chance is  $\geq 65\%$  that Tests 1 and 2 will both make the correct decision when  $\Delta > 0.09$ . There is only 0.25%-61% chance of the correct decision for both tests when  $0 < \Delta < 0.09$  by the year 2008 (Table 10b).

Figure 3. Power of Test #1 for sub-yearling chinook salmon based on inriver smolt survival between Lower Granite and Bonneville dams for (a)  $n_2 = 5$  years of RPA (2001-2005) and (b)  $n_2 = 8$  years of RPA (2001-2008).

a.  $n_2 = 5$



b.  $n_2 = 8$

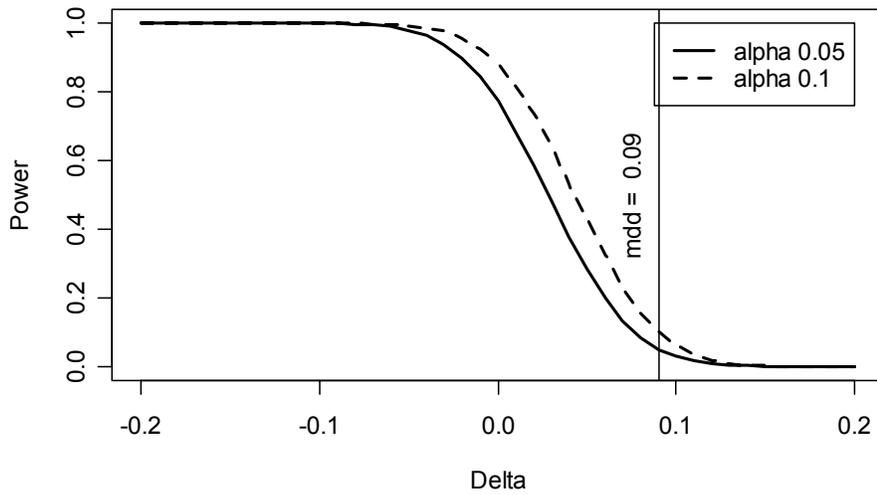
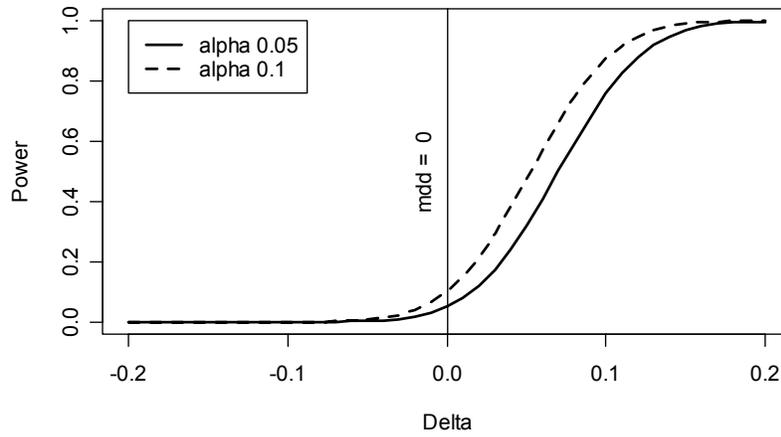
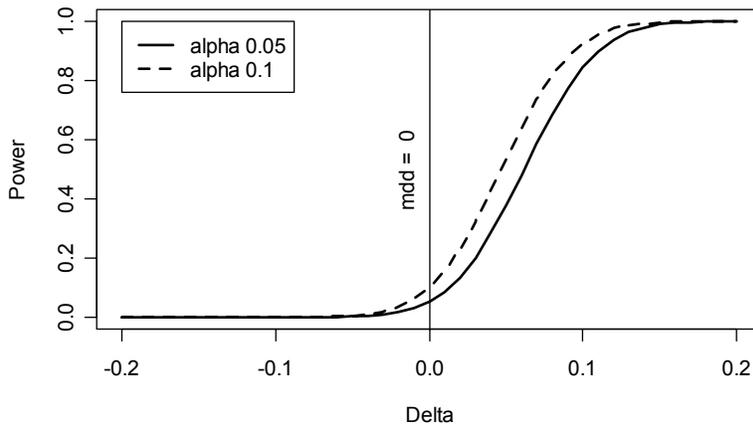


Figure 4. Power of Test #2 for sub-yearling chinook salmon based on inriver survival between Lower Granite and Bonneville dams for (1) (a)  $n_2 = 5$  years of RPA (2001-2005) and (b)  $n_2 = 8$  years of RPA (2001-2008).

a.  $n_2 = 5$



b.  $n_2 = 8$



DRAFT

Table 8. Probabilities of marking correct (shaded) and incorrect (unshaded) decisions using Test #1 at  $\alpha = 0.05$  for subyearling chinook salmon for the survival estimates from Lower Granite to Bonneville dams under different states of nature by (a) 2005 and (b) 2008.

a. 2005

		Alternative States of Nature		
		No Improvement $\Delta \leq 0$	Some Improvement $0 < \Delta < 0.09$	Recovery $\Delta \geq 0.09$
Conclude $\Delta \geq 0.09$		$0 < \beta < 0.32$	$0.32 < \beta < 0.95$	$1 - \alpha = 95$
Conclude $\Delta < 0.09$		$0.68 \leq 1 - \beta \leq 1.0$	$0.05 < 1 - \beta < 0.68$	$\alpha = 0.05$

b. 2008

		Alternative States of Nature		
		No Improvement $\Delta \leq 0$	Some Improvement $0 < \Delta < 0.09$	Recovery $\Delta \geq 0.09$
Conclude $\Delta \geq 0.09$		$0 < \beta < 0.22$	$0.22 < \beta < 0.95$	$1 - \alpha = 95$
Conclude $\Delta < 0.09$		$0.78 \leq 1 - \beta \leq 1.0$	$0.05 < 1 - \beta < 0.78$	$\alpha = 0.05$

DRAFT

Table 9. Probabilities of making correct (shaded) and incorrect (unshaded) decisions using Test #2 at  $\alpha = 0.05$  for subyearling chinook salmon for the survival estimates from Lower Granite to Bonneville dams under different states of nature by (a) 2005 and (b) 2008.

a. 2005

		Alternative States of Nature		
		No Improvement $\Delta \leq 0$	Some Improvement $0 < \Delta < 0.09$	Recovery $\Delta \geq 0.09$
Conclude $\Delta \leq 0$		$1 - \alpha = 0.95$	$0.32 < \beta < 0.95$	$\beta < 0.32$
Conclude $\Delta > 0$		$\alpha = 0.05$	$0.05 < 1 - \beta < 0.68$	$1 - \beta \geq 0.78$

b. 2008

		Alternative States of Nature		
		No Improvement $\Delta \leq 0$	Some Improvement $0 < \Delta < 0.09$	Recovery $\Delta \geq 0.09$
Conclude $\Delta \leq 0$		$1 - \alpha = 0.95$	$0.22 < \beta < 0.95$	$\beta < 0.22$
Conclude $\Delta > 0$		$\alpha = 0.05$	$0.05 < 1 - \beta < 0.78$	$1 - \beta \geq 0.78$

DRAFT

Table 10. Probabilities Tests #1 and #2 will make the correct decisions, individually and jointly, under alternative states of nature at  $\alpha = 0.05$  for the subyearling chinook salmon for (a) 2005 and (b) 2008.

a. 2005

		Alternative States of Nature		
		$\Delta \leq 0$	$0 < \Delta < 0.09$	$\Delta \geq 0.09$
Test #1		Reject $H_0$ $0.68 \leq 1 - \beta \leq 1.0$	Reject $H_0$ 0.05 1 .68	Do not reject $H_0$ $1 - \alpha = 0$
Test #2		Do not reject $H_0$ $1 - \alpha = 0.95$	Reject $H_0$ 0.05 1 .68	Reject $H_0$ 0.68 1 1.0
Joint Tests #1 and #2		0.65 – 0.95	0.0025 – 0.46	0.65 – 0.95

b. 2008

		Alternative States of Nature		
		$\Delta \leq 0$	$0 < \Delta < 0.09$	$\Delta \geq 0.09$
Test #1		Reject $H_0$ $0.78 \leq 1 - \beta \leq 1.0$	Reject $H_0$ 0.05 1 .78	Do not reject $H_0$ $1 - \alpha = 0$
Test #2		Do not reject $H_0$ $1 - \alpha = 0.95$	Reject $H_0$ 0.05 1 .78	Reject $H_0$ 0.78 1 1.0
Joint Tests #1 and #2		0.94 – 0.95	0.0025 – 0.61	0.74 – 0.95

### 4.3 Steelhead

Figure 5 presents the power of the two-sample t-test to reject the null hypothesis of 0.09 improvement or greater in survival (1) between Lower Granite and Bonneville dams. When  $\Delta \approx 0$ , the first test has about a 80% chance of rejecting (1) at  $\alpha = 0.05$ .

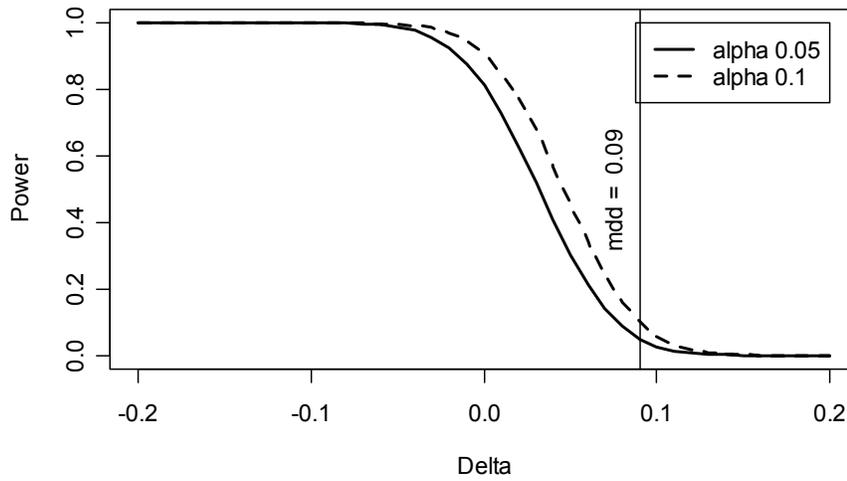
Figure 6 presents power of the two-sample t-test to reject the null hypothesis of no improvement in survival (2) between Lower Granite and Bonneville dams for steelhead. At  $\Delta \approx 0.10$ , the t-test has approximately a 85% chance of rejecting (2) at  $\alpha = 0.05$ . A 0.20 improvement in survival between periods is needed before the t-test is almost certain to reject the null hypothesis of no improvement (2).

The results of the power curves in Figures 5-6 are summarized in Tables 11 and 12, respectively. By design, Test #1 will make an incorrect decision  $\alpha \cdot 100\%$  of the time and conclude  $\Delta < 0.09$  when impact recovery has occurred with  $\Delta \geq 0.09$ . However, Test #1 will make an incorrect decision between 21%-95% of the time and conclude  $\Delta > 0.09$  when in fact  $0 \leq \Delta < 0.09$ . The mean survival during post-2000 years can be less than pre-2000 years, and Test #1 has a  $\leq 79\%$  chance of concluding  $\Delta > 0.09$  (Table 11). By design, Test #2 will make an incorrect decision  $\alpha \cdot 100\%$  of the time and conclude  $\Delta > 0$  when, in fact,  $\Delta \leq 0$ . However, Test #2 will make an incorrect decision between 21%-95% of the time and conclude  $\Delta < 0$  when, in fact,  $0 < \Delta < 0.09$ . Test #2 has a  $\geq 79\%$  chance of making the correct decision when the improvement in survival  $\geq 0.09$ .

The result of the power calculations for Tests #1 and #2 are summarized under alternative scenarios for recovery by the year 2005 and 2008 in Tables 11 and 12. The ideal results for Tests #1 and #2 would be to have probabilities of correct decisions near 1 in the shaded boxes and probabilities of incorrect decision near 0 in the unshaded boxes in Tables 11 and 12, respectively. Deviation from these expectations are a measure of the lack of performance of the proposed tests of compliance. Table 13 summarizes the probabilities of making the correct decisions with Test 1-2 under alternative states of nature for improvement in survival of the juvenile steelhead. The chance of both Tests 1 and 2 making the correct decision for the steelhead stock when  $\Delta < 0$  is  $\geq 75\%$  of the time. The chance is  $\geq 75\%$  that Tests 1 and 2 will both make the correct decision when  $\Delta > 0.09$ . There is only 0.25%-79% chance of the correct decision for both tests when  $0 < \Delta < 0.09$  for the year 2008 (Table 13b).

Figure 5. Power of Test #1 for steelhead based on inriver smolt survival between Lower Granite and Bonneville dams for (a)  $n_2 = 5$  years of RPA (2001-2005) and (b)  $n_2 = 8$  years of RPA (2001-2008).

a.  $n_2 = 5$



b.  $n_2 = 8$

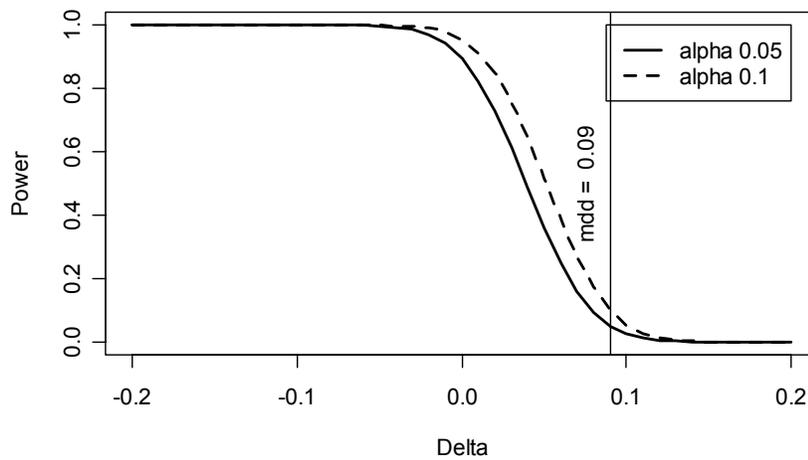
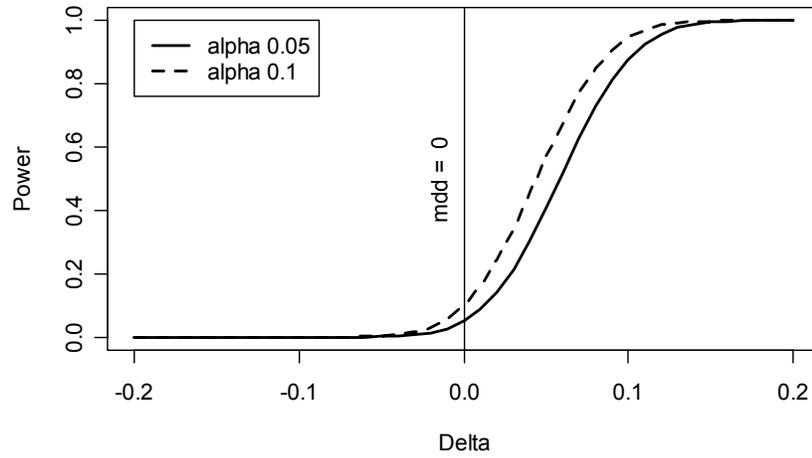


Figure 6. Power of Test #2 for steelhead based on inriver survival between Lower Granite and Bonneville dams for (1) (a)  $n_2 = 5$  years of RPA (2001-2005) and (b)  $n_2 = 8$  years of RPA (2001-2008).

a.  $n_2 = 5$



b.  $n_2 = 8$

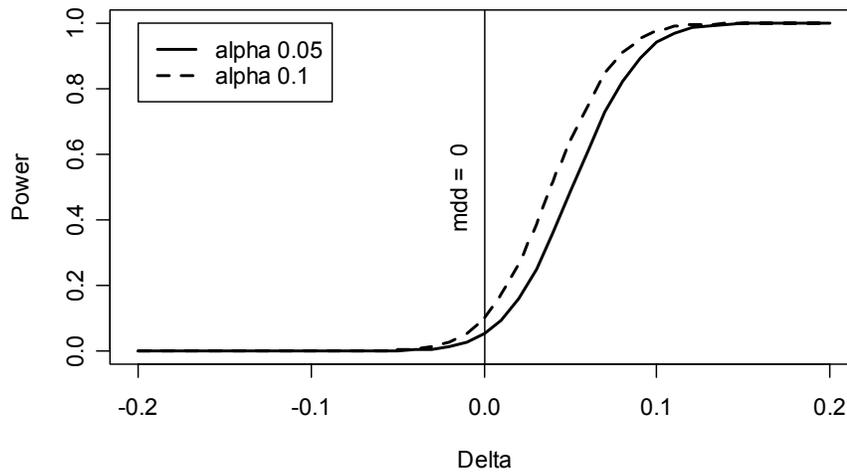


Table 11. Probabilities of Probabilities of marking correct (shaded) and incorrect (unshaded) decisions using Test #1 at  $\alpha = 0.05$  for steelhead for the survival estimates from Lower Granite to Bonneville dams under different states of nature by (a) 2005 and (b) 2008.

a. 2005

		Alternative States of Nature		
		No Improvement $\Delta \leq 0$	Some Improvement $0 < \Delta < 0.09$	Recovery $\Delta \geq 0.09$
Conclude $\Delta \geq 0.09$		$0 < \beta < 0.21$	$0.21 < \beta < 0.95$	$1 - \alpha = 95$
	Conclude $\Delta < 0.09$	$0.79 \leq 1 - \beta \leq 1.0$	$0.05 < 1 - \beta < 0.79$	$\alpha = 0.05$

b. 2008

		Alternative States of Nature		
		No Improvement $\Delta \leq 0$	Some Improvement $0 < \Delta < 0.09$	Recovery $\Delta \geq 0.09$
Conclude $\Delta \geq 0.09$		$0 < \beta < 0.11$	$0.11 < \beta < 0.95$	$1 - \alpha = 95$
	Conclude $\Delta < 0.09$	$0.89 \leq 1 - \beta \leq 1.0$	$0.05 < 1 - \beta < 0.89$	$\alpha = 0.05$

DRAFT

Table 12. Probabilities of making correct (shaded) and incorrect (unshaded) decisions using Test #2 at  $\alpha = 0.05$  for steelhead for the survival estimates from Lower Granite to Bonneville dams under different states of nature by (a) 2005 and (b) 2008.

a. 2005

		Alternative States of Nature		
		No Improvement $\Delta \leq 0$	Some Improvement $0 < \Delta < 0.09$	Recovery $\Delta \geq 0.09$
Conclude $\Delta \leq 0$		$1 - \alpha = 0.95$	$0.21 < \beta < 0.95$	$\beta < 0.21$
Conclude $\Delta > 0$		$\alpha = 0.05$	$0.05 < 1 - \beta < 0.79$	$1 - \beta \geq 0.79$

b. 2008

		Alternative States of Nature		
		No Improvement $\Delta \leq 0$	Some Improvement $0 < \Delta < 0.09$	Recovery $\Delta \geq 0.09$
Conclude $\Delta \leq 0$		$1 - \alpha = 0.95$	$0.11 < \beta < 0.95$	$\beta < 0.11$
Conclude $\Delta > 0$		$\alpha = 0.05$	$0.05 < 1 - \beta < 0.89$	$1 - \beta \geq 0.89$

Table 13. Probabilities Tests #1 and #2 will make the correct decisions, individually and jointly, under alternative states of nature at  $\alpha = 0.05$  for the steelhead for (a) 2005 and (b) 2008.

a. 2005

		Alternative States of Nature		
		$\Delta \leq 0$	$0 < \Delta < 0.09$	$\Delta \geq 0.09$
Test #1		Reject $H_0$ $0.79 \leq 1 - \beta \leq 1.0$	Reject $H_0$ 0.05 1 .79	Do not reject $H_0$ $1 - \alpha = 0$
Test #2		Do not reject $H_0$ $1 - \alpha = 0.95$	Reject $H_0$ 0.05 1 .79	Reject $H_0$ 0.79 1 1.0
Joint Tests #1 and #2		0.75 – 0.95	0.0025 – 0.62	0.75 – 0.95

b. 2008

		Alternative States of Nature		
		$\Delta \leq 0$	$0 < \Delta < 0.09$	$\Delta \geq 0.09$
Test #1		Reject $H_0$ $0.89 \leq 1 - \beta \leq 1.0$	Reject $H_0$ 0.05 1 .89	Do not reject $H_0$ $1 - \alpha = 0$
Test #2		Do not reject $H_0$ $1 - \alpha = 0.95$	Reject $H_0$ 0.05 1 .89	Reject $H_0$ 0.89 1 1.0
Joint Tests #1 and #2		0.84 – 0.95	0.0025 – 0.79	0.84 – 0.95

## 5.0 Monte Carlo Modeling Approach

The Tests #1 and #2 power calculations results in Sections 3 and 4 of this report are based on an immediate and sustained improvement of size  $\Delta = 0.09$  during all of the post-2000 monitoring years. Although this overly optimistic scheme for recovery already has been shown to yield relatively poor odds of making the correct decisions with the proposed tests, a more realistic assumption on RPA performance also needs to be evaluated. In reality, RPA improvement might be gradual, reaching the size  $\Delta$  only by the end of the post-2000 period, in which case, the statistical power of Tests #1 and #2 may be substantially less than presented in Sections 3 and 4. To test this contention, we investigated the power of Tests #1 and #2 using Monte Carlo simulations

Using a known set of parameters, we generated a time series of simulated survival estimates. The gradual improvement in survival was simulated using a regression model that reproduces the natural changes in survival over time. The simulated time series was then subjected to the power calculations exactly as if it were data from field observations.

### 5.1 Methods

We started with a baseline of survival probabilities that followed an upward linear trend with the improvement reaching the RPA target in a specified year. The slope is estimated with a linear regression model fit on pre-2000 survival data. Then these initial conditions are used to randomly generate sets of future survivals for  $N$  years, where  $N$  is equal to 5 or 8 and represents the time span set by the FCRPS 2000 BO for RPA compliance testing. We simulated different scenarios with the improvement reaching  $\Delta = 0.09$  by the end of the post-2000 period or during the subsequent years.

Figure 7 presents the optimistic scenario, used in Sections 3 and 4, where the improvement target is suddenly attained in 2000 and is sustained through years 2005 and 2008. Figure 8 presents different scenarios of gradual improvement with the RPA target reached by the end of a given period. For example, the worst-case scenario is on the far right of Figure 8, where the continuous line reaches the RPA target in 2007. The far left, dashed line assumes that the RPA target was reached in 2003, while the middle, dotted line assumes that the RPA target was reached in 2005.

In order to simulate a gradual recovery process, we assumed that the improvement in survival followed a linear trend and computed the expected survival series using the linear regression model as follows:

Figure 7. Immediate and sustained improvement of size  $\Delta = 0.09$  beginning in 2000.

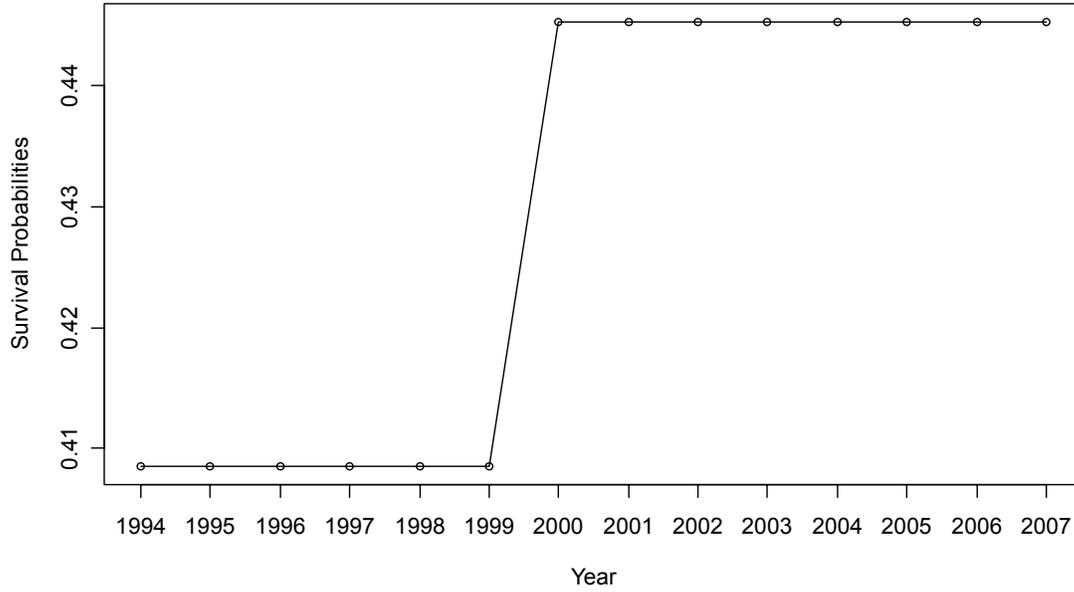
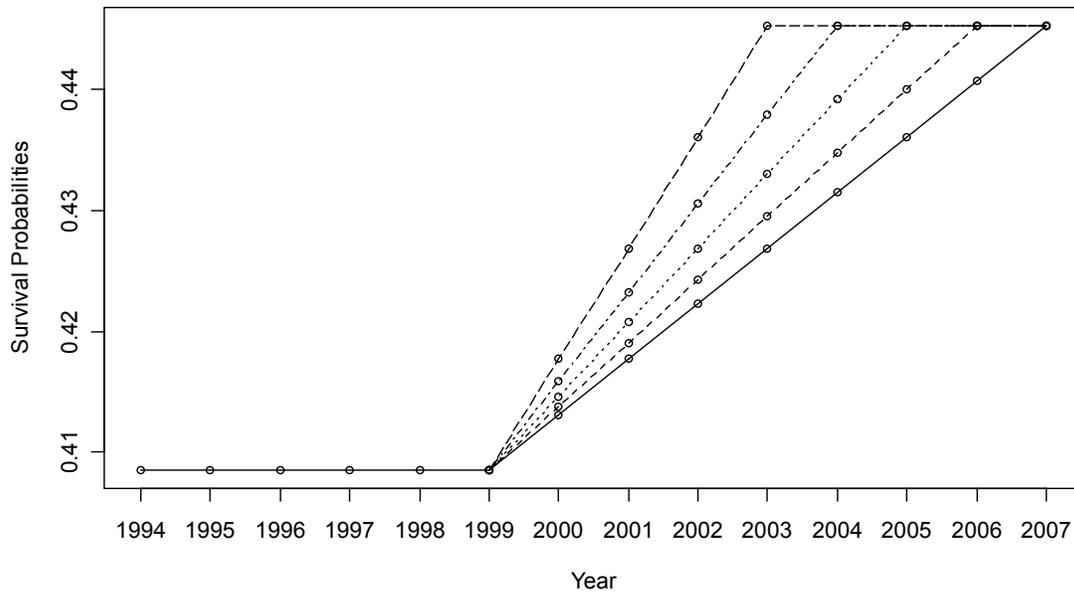


Figure 8. Gradual improvement in survival of size  $\Delta = 0.09$  by the end of year 2003, 2004, 2005, 2006, or 2007.



$$E_b(S_{\text{post-2000},i}) = \left(1 + \frac{0.09 \times i}{n}\right) \times \bar{S}_{\text{pre-2000}} \quad (4)$$

where

$n$  = number of post-2000 years involved in the study (8 or 5),

= year index,

$\bar{S}_{\text{pre-2000}}$  = mean survival estimate for the years 1994 or 1995 to 1999,

0.09 = anticipated 9% RPA improvement by the end of year 2005 (2008).

To create the post-2000 series, Monte Carlo simulations were performed and annual survival probabilities were generated using the expected values in Equation (4) plus a random noise term  $\varepsilon$  where

$$S_{\text{post-2000},i} = E_b(S_{\text{post-2000},i}) + \varepsilon_i \quad (5)$$

where  $\varepsilon \sim N(0, s^2)$  is a normal random noise with the variance  $s^2$  equal to the pre-2000 survival inter-annual variance  $Var(x_{\text{pre-2000}})$ .

## 5.2 Results

Monte Carlo simulations were run based on the several scenarios of gradual improvement in survival depicted in Figure 8. We compared statistical power of the tests of compliance under gradual recovery with the scenario examined in Section 4 of an immediate and sustained improvement of 9% attained in 2000, as shown in Figure 7.

Results of the power analyses under gradual recovery are summarized in Table 14 for all the three fish stocks. Table 14 gives the probability of correctly identifying the intermediate state of some but not complete recovery. Table 15 gives the probability of correctly identifying the recovery at a value of  $\Delta$  or greater. These tables show that as the RPA target is reached later in the eight-year period, it becomes more and more difficult to provide it with Tests #1 and #2. As expected, the power of Test #1 and #2 decreases substantially under the scenario of gradual recovery. Table 16 shows how large of an RPA improvement is needed under gradual recovery to equal the statistical power of the tests when recovery was immediate and sustained. These Monte Carlo simulation results indicate the proposed statistical tests of compliance have extremely low power to demonstrate RPA compliance when the recovery process gradually reaches its target by 2007 (sic 2008). Statistical tests based on assumptions other than immediate recovery need to be examined to identify better tests of RPA compliance.

DRAFT

Table 14. Statistical power for the three stocks (yearling chinook, subyearling chinook, and steelhead) based on inriver smolt survival between Lower Granite and Bonneville dams for eight years of RPA using Tests #1 and #2 at  $\alpha = 0.05$  under the scenario of gradual recovery. Statistical probabilities are the probabilities of correctly identifying the intermediate state of some but not all recovery.

Scenario No. - Year of Compliance	Joint Tests #1 and #2 (Reject $H_0$ )		
	Yearling Chinook	Subyearling Chinook	Steelhead
1 - 2000 (Baseline)	0.0025-0.38	0.0025-0.61	0.0025-0.79
2 - 2001	0.0025-0.30	0.0025-0.30	0.0025-0.55
3 - 2002	0.0025-0.27	0.0025-0.28	0.0025-0.49
4 - 2003	0.0025-0.23	0.0025-0.27	0.0025-0.44
5 - 2004	0.0025-0.21	0.0025-0.25	0.0025-0.38
6 - 2005	0.0025-0.19	0.0025-0.24	0.0025-0.34
7 - 2006	0.0025-0.16	0.0025-0.23	0.0025-0.30
8 - 2007	0.0025-0.13	0.0025-0.22	0.0025-0.26

Table 15. Statistical power for the three stocks (yearling chinook, subyearling chinook, and steelhead) based on inriver smolt survival between Lower Granite and Bonneville dams for eight years of RPA using Tests #1 and #2 at  $\alpha = 0.05$  under the scenario of gradual recovery. Statistical probabilities are the probabilities of correctly identifying the state of full recovery with a change of size  $\Delta$  or greater.

Scenario No. - Year of Compliance	Joint Tests #1 and #2 (Reject $H_0$ )		
	Yearling Chinook	Subyearling Chinook	Steelhead
1 - 2000 (Baseline)	$\geq 0.59$	$\geq 0.74$	$\geq 0.84$
2 - 2001	$\geq 0.52$	$\geq 0.52$	$\geq 0.70$
3 - 2002	$\geq 0.49$	$\geq 0.50$	$\geq 0.66$
4 - 2003	$\geq 0.46$	$\geq 0.49$	$\geq 0.63$
5 - 2004	$\geq 0.44$	$\geq 0.48$	$\geq 0.59$
6 - 2005	$\geq 0.41$	$\geq 0.47$	$\geq 0.55$
7 - 2006	$\geq 0.38$	$\geq 0.46$	$\geq 0.52$
8 - 2007	$\geq 0.33$	$\geq 0.45$	$\geq 0.48$

Table 16. Required improvement in survival between Lower Granite and Bonneville dams for the three fish stocks (yearling chinook, subyearling chinook, and steelhead) in order for Tests #1 and #2 to have the same statistical power under the scenario of gradual improvement as in the case of immediate and substantial improvement beginning in 2000.

Scenario No. - Year of Compliance	Improvement Needed to Reach the Baseline Power		
	Yearling Chinook	Subyearling Chinook	Steelhead
1 - 2000 (Baseline Power)	0.0025-0.38	0.0025-0.61	0.0025-0.79
2 - 2001	11%	21%	12%
3 - 2002	12%	22%	14%
4 - 2003	13%	23%	15%
5 - 2004	14%	24%	16%
6 - 2005	15%	26%	17%
7 - 2006	17%	28%	18%
8 - 2007	19%	30%	20%

## 6.0 Discussion

Based on Monte-Carlo simulations and an underlying regression model, the power calculations strongly suggest the inability of the two novel statistical hypothesis tests recommended by the Biological Opinion (BO) to show the anticipated compliance of the RPA scheduled in 2005 and 2008 using pre-2000 historical survival data.

The statistical power calculation results were interpreted in terms of the ability of the two tests to correctly identify the true states of recovery (i.e., fail or succeed in fulfilling RPA expectations). Both tests used the same monitoring and evaluation data on different scenarios:

- Realistic schemes for recovery based on gradual improvement in survival of size  $\Delta = 0.09$  by the end of year 2003, 2004, 2005, 2006, or 2007 (Figure 8).
- Overly optimistic scheme for recovery based on immediate and sustained improvement of size  $\Delta = 0.09$  beginning in 2000 (Figure 7).

Tables 14-15 summarize the simulated statistical power of jointly making the correct decisions with Tests #1 and #2 for all three stocks under a recovery process based on gradual

improvement. These simulated scenarios indicate there are very poor odds of making the correct decision with the proposed tests in the BO and as the 9% target is reached late in the 8-year RPA period. The statistical power of the proposed tests to identify recovery is only 0.25%-26% for the steelhead stock, 0.25%-22% for the sub-yearling chinook salmon, and 0.25%-13% for the yearling chinook. Thus, the simulations show the longer it takes for the RPA improvement to be attained the more difficult it will be to statistically demonstrate improvement.

In the case where the recovery scenario is overly optimistic and based on immediate and sustained 9% improvement beginning in 2000, the probabilities of jointly making the correct decision with the proposed tests are higher with values 0.25%-79% for the steelhead stock, 0.25%-61% for the sub-yearling chinook salmon, and 0.25%-38% for the yearling chinook. However, these probabilities of correctly identify recovery remain relatively poor and indicate the inability of the two proposed BO Tests #1 and #2 to correctly identify the true state of survival recovery.

The next phase of this project is to examine alternative decision rules that might maximize the likelihood of correct decisions while minimizing the prospect of incorrect decisions. In particular, special attention will be given to the novel “two one-sided tests” (TOST), which test the “interval hypotheses” (Brown et al. 1995, Berger and Hsu 1996) as stated statistically as follows:

$$H_0 : \Delta_1 \leq \mu_{\text{post-2000}} - \mu_{\text{pre-2000}} \leq \Delta_2$$

$$H_a : \mu_{\text{post-2000}} - \mu_{\text{pre-2000}} < \Delta_1 \text{ or } \mu_{\text{post-2000}} - \mu_{\text{pre-2000}} > \Delta_2$$

where  $\Delta_1$  and  $\Delta_2$  are known constants,  $\Delta_1 < \Delta_2$ .

Unlike the proposed Tests #1 and #2, the TOST will test if there are some improvements within a specific range of values. The current 9% RPA expected survival improvements for Snake River smolts suggests a  $\Delta_1 = 0$  and  $\Delta_2 = 0.09$  for the TOST. However, a value of  $\Delta_1 = 0$  may be regarded as simplistic. Therefore, investigations should look at a range of values for  $\Delta_1$  and  $\Delta_2$ . The TOST will be implemented in both classical and unbiased versions. Unbiased versions of the TOST are generally uniformly more powerful than the classical version but often require a good deal of computing (Martin 1990).

In addition to the TOST, Bayesian methods will be addressed. Bayesian decision analysis incorporates prior probability distribution and likelihoods of observed data to determine a posterior probability distribution of events. As such, they helps achieve a precautionary

## DRAFT

approach to compliance evaluation. Given the relatively small size of our data, Markov Chain Monte Carlo (MCMC) simulations will play a crucial computational role in these evaluations.

The relatively poor odds of making the correct decisions with proposed Tests #1 and #2 suggest alternative decision rules need to be investigated and developed for assessing RPA compliance. The development and selection of decision rules should proceed immediately. The credibility of the scientific process begun by the BO could be seriously jeopardized if the public perceives the rules will be established only after the results are known. Lack of scientific objectivity could undermine public confidence in not only the ESA process but also in the agencies involved.

### **7.0 Literature Cited**

- National Marine Fisheries Service. 2000. 2000 FCRPS Biological Opinion. National Marine Fisheries Service (NMFS), etc.
- Berger, R. L., and J. C. Hsu. 1996. Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Science* 11:283-319.
- Brown, L. D., J. T. Gene Hwang, and A. Munk. 1995. An unbiased test for the bioequivalence problem. Technical report. Cornell University. Ithaca, NY.
- Martin, A. A. 1990. On testing for bioequivalence. *Biometrical Journal* 32:125-126.
- Searle, S. R. 1971. Linear models. John Wiley & Sons. New York, NY.