



Independent Scientific Advisory Board

*for the Northwest Power and Conservation Council,
Columbia River Basin Indian Tribes,
and National Marine Fisheries Service
851 SW 6th Avenue, Suite 1100
Portland, Oregon 97204*

Review of the Comparative Survival Study Draft 2016 Annual Report

Kurt Fausch
Stan Gregory
William Jaeger
Cynthia Jones
Alec Maule
Katherine Myers
Greg Ruggerone
Laurel Saito
Steve Schroder
Carl Schwarz
Tom Turner

**ISAB 2016-2
October 21, 2016**

Review of the Comparative Survival Study (CSS) 2016 Draft Annual Report

Contents

| | |
|--|----|
| I. Background | 1 |
| II. Summary | 1 |
| III. Suggested Topics for Further Review | 4 |
| IV. Comments on New or Updated Analyses in the draft CSS 2016 Annual Report by Chapter | 6 |
| IV.1. Chapter 1. Overview | 6 |
| IV.2. Chapter 2. Life cycle modeling of alternative spill experiment scenarios..... | 6 |
| IV.3. Chapter 3. Effects of the in-river environment on juvenile travel time, instantaneous mortality rates and survival | 9 |
| IV.4. Chapter 4. Patterns in annual overall SARs | 10 |
| IV.5. Chapter 5. SARs and productivity | 10 |
| IV.6. Chapter 6. Estimation of SARS, TIRS and <i>D</i> for Snake River Subyearling Fall Chinook..... | 11 |
| IV.7. Chapter 7. Effects of juvenile bypass systems on smolt to adult return rates..... | 11 |
| IV.8. Chapter 8. Patterns of variation in age-at-maturity for PIT-tagged spring/summer Chinook salmon in the Columbia River Basin | 11 |
| IV.9. Appendix A: Survivals (SR), SAR, TIR, and <i>D</i> for Snake River Hatchery and Wild Spring/Summer Chinook Salmon, Steelhead, and Sockeye..... | 12 |
| IV.10. Appendix B: Supporting tables on Chapter 4 - Overall SARs | 12 |
| IV.11. Appendix C: Development of a weighted bootstrap for unequally represented hatchery PIT-tag groups | 12 |
| V. Specific editorial comments and requests for clarification | 14 |
| VI. References | 23 |

Review of the Comparative Survival Study (CSS) 2016 Draft Annual Report

I. Background

The Northwest Power and Conservation Council's [2009 amendments](#) to the Columbia River Basin Fish and Wildlife Program called for a regular system of independent and timely science reviews of the [Fish Passage Center's](#) (FPC) analytical products. The [2014 Program's Appendix H](#) maintains this review function. These reviews include evaluations of the Comparative Survival Study's draft annual reports. The ISAB has reviewed these reports annually beginning six years ago with the evaluation of the CSS's draft 2010 Annual Report and most recently the draft 2015 Annual Report ([ISAB 2010-5](#), [ISAB 2011-5](#), [ISAB 2012-7](#), [ISAB 2013-4](#), [ISAB 2014-5](#), and [ISAB 2015-2](#)). This ISAB review of the [draft 2016 CSS Annual Report](#) is the ISAB's seventh review of CSS annual reports in response to the Council's Program language.

II. Summary

This ISAB review begins with an overview of the latest report (this section). It then moves on to suggesting topics for further CSS review (Section III), general comments on each chapter of the 2016 CSS Annual Report (Section IV), and ends with specific queries and suggestions (Section V).

The annual CSS report is a mature product, typically including only updates with the latest year of data and expansion of analyses as more data are acquired. Many of the methods have been reviewed in previous ISAB reports and so now receive only a cursory examination. As more data are acquired, new patterns and questions arise on the interpretation of the results—this is now the primary focus of our reviews. The ISAB appreciates the CSS's detailed response to suggestions provided in previous reviews, and we do not expect the CSS to necessarily respond immediately to new requests for further analyses.

Chapter 1 is similar to previous years with the 2015 results added. Two new fish populations have been added. In the 2016 report, the size of the PIT tags used is reported as being 11 to 12 mm instead of the 9 to 12 mm in previous reports. If this is a real change, the rationale for the change is needed along with a discussion of potential impacts on the fish (e.g., are larger fish now tagged to accommodate the larger tags?).

In Chapter 2, the existing life cycle model was used in a prospective analysis to simulate the relative benefits of flow/spill modifications to habitat. While the approach is generally well implemented, the ISAB has some concerns about specific aspects of the simulation study that suggest the outcomes may not be as clear cut as indicated in the report. For example, what is the justification of choosing particular years as “representative” of low/medium/high flow conditions? Some variables that could vary are held constant—e.g., powerhouse contact rate derived from PIT tag data (PITPH) and water travel time (WTT)—and so the simulation results may underreport the variability in the response.

Chapter 3 is mainly an update with the latest information on in-river effects on juvenile travel time, instantaneous mortality, and survival. A key finding is that there is large variation in the results among years and among cohorts. The variation among years is understandable; the variation within a year less so. Many figures (e.g., Figure 3.2) show a consistent pattern in fish travel time and survival over cohorts as the year progresses. Mortality tends to increase over the migration season and with water temperature (except for sockeye). The report lists four potential mechanisms: (1) declining smolt energy reserves or physiological condition over the migration season and with water temperature, (2) increasing predation rates on smolts over the migration season and with increased water temperature, (3) increases in disease susceptibility or disease-related mortality over the migration season and with increased water temperature, or (4) some combination of these often interrelated mechanisms. Is there an attempt to test these hypotheses using other approaches, either within CSS or by other investigators? Answers to these questions might lead to improvements in survival. We agree that the apparent contradictory response of sockeye warrants further investigation.

Chapter 4 described overall annual SARs and was updated with new data; details are presented in appendices. In addition, an analysis of relationship between the ratio of transport to in-river survival (TIR; transport effect) and in-river survival is now included. It is not surprising that the transport TIR is inversely correlated with in-river survival (Lower Granite Dam [LGR] to Bonneville Dam [BON]). This new analysis identified the value for in-river survival when the benefits of transportation appear to disappear. The CSS also reported on the relatively large absolute difference in SAR based on PIT-tags versus run reconstruction (the values are highly correlated, however). As in previous reports, this report listed various hypotheses. A study is underway to further evaluate PIT-tag effects on salmon survival, but these results will not be ready until after summer 2017 when tagged age-5 Chinook will have returned. Potential bias in survival caused by tagging methodology (or in the run reconstruction methodology) is an important issue to resolve, and the ISAB looks forward to the results of this study.

The material in Chapter 5 was combined with other chapters in previous reports and has now been split out. This is an update with an additional year of data. Chapter 5 continues the examination of the relationships between life cycle productivity and SARs, including the level of SARs needed to reach or exceed population replacement. The findings suggest that pre-harvest SARs of 4%-6% are associated with pre-1970 levels of productivity for Snake River spring/summer Chinook; SARs are much lower in subsequent decades. How might these early SARs (~4-6%) compare with SARs from viable wild Chinook populations in other regions? To what extent might improvements in hydrosystem management, predator control, and estuarine habitat lead to SARs of 4%-6%?

Chapter 6 is mostly an update on Snake River subyearling fall Chinook.

Chapter 7 is a repeat of an analysis done in 2010 with additional years of data. A logistic regression analysis was used to investigate the impact of year, bypass effects, and rearing type on subsequent survival and return as an adult. Estimates (Table 7.4) appear to be on a “per bypass” basis. What is the average number of bypasses encountered by a fish? Wouldn't that be a more accurate reflection of the impact of bypass on an outgoing smolt?

Chapter 8 examined differences in the mean age of maturity among different stocks, years, and fish type (wild or hatchery) using regression methods. While the analysis mostly seems appropriate, the ISAB has numerous suggestions on improving the presentation of the results.

Appendices A and B are updated with an additional year of data. The ISAB is pleased that electronic versions of many of these tables are now available at the FPC website.

Appendix C reports on the development of a weighted bootstrap procedure to deal with a stratified sampling and tagging of smolts that is not proportional to abundance. The authors describe a bootstrap procedure for estimating parameters, but the ISAB suggests that a stratified approach be incorporated directly into the analysis routines currently used to allow for future expansion of stratified-tagging studies.

III. Suggested Topics for Further Review

In 2013, we recommended these topics ([ISAB 2013-4](#), Page 1):

1. Hypotheses on mechanisms regulating smolt-to-adult survival rates (SARs)
2. Life-cycle modeling questions and Fish and Wildlife Program SAR objectives
3. Data gaps
4. Rationalization of CSS's Passive Integrated Transponder (PIT)-tagging, and
5. Publication of a synthesis and critical review of CSS results

In 2014, we recommended these topics ([ISAB 2014-5](#), pages 2-3):

1. Hypotheses on mechanisms regulating smolt-to-adult return rates (SARs) [update from 2013 review]
2. Life-cycle modeling questions and Fish and Wildlife Program SAR objectives [update from 2013 review]
3. New PIT/CWT study to further investigate differential survival among these tag types

In 2015, we recommended these topics ([ISAB 2015-2](#), pages 4-5):

1. Use SAR data to examine both intra- and interspecific density dependence during the smolt out migration and early marine periods
2. Propose actions to improve SARs to pre-1970s levels
3. Explore additional potential relations between SARs and climate and ocean conditions
4. Consider ways to explore the variability of inter-cohort response

The CSS group has incorporated many of our suggestions into the current document. For example, the current report has a substantial discussion of correlations among SARs from different regions or effects of transport on SARs (#1 in 2013; #1 in 2014). The life cycle modeling now allows for variation in stream productivity and hydrosystem survival and simulates the correlative impacts of these changes on predicted future population abundances (#2 in 2013; #2 in 2014; #2, #3 in 2015). The ISAB appreciates the CSS efforts to respond to our queries which in turn lead to further questions.

Some of the recommendations from the ISAB appear to be beyond the scope of the CSS but will become increasingly important in the future. For example, is there evidence of density dependence during the smolt out-migration and early marine periods (2015 #1)? Could the CSS estimate total smolt abundance of each species, say at Bonneville Dam? Is this a potential

mechanism to explain the inter-cohort variation in responses (2015 #4)? This is reflected in our recommendations for future work below.

In 2016, we recommend the following four topics for future reports:

1. Use more realistic and more variable future flow conditions for the study on the impact of flow/spill modifications under future climate change. Simulating only low flows or high flows for decades may not be a realistic scenario. What is the impact of a correlation between Pacific Decadal Oscillations (PDOs) and flows that has not been considered in the simulations presented in the 2016 report?
2. What is the impact of the new restricted tag sizes? Are there fish that were previously marked and are now not marked (e.g. smaller fish) due to the larger PIT tags being used? Similarly, conclusions from studies of compensatory mortality (e.g. in relation to predator control) may be affected by the choice of fish that are tagged. A brief review of the PIT tag procedures should be undertaken so that users of the CSS data are fully aware of any limitations in the conclusions of other studies that are related to types/sizes of fish tagged.
3. There has been a great deal of interest in the impact of predator control programs on salmon returns, especially northern pikeminnow and birds. A life-cycle model is the natural way to study these impacts, but the current version of the life-cycle model appears to incorporate density dependence only at the spawner-to-smolt stage. The ISAB recommends that consideration be given to modifying the life-cycle model to allow a range of compensatory responses ranging from complete additivity (as now is the case) to plausible compensatory mortality effects related to density dependence and predator selectivity (see [ISAB 2016-1](#)). This continues our previous recommendation (#1 in 2015) to investigate impacts of density dependence on subsequent return.
4. Both the CSS and NOAA provide estimates for in-river survival. How do these estimates compare to each other? If there are consistent differences in the estimates, can these be explained?
5. What factors have led to declining proportions of four and five-year olds and increases in three-year olds in spring/summer Chinook? Models that include ocean factors associated with salmon growth and climate change, differences in hatchery practices, or

freshwater environments (tributary temps, or annual differences in migration corridor) may be of interest.

IV. Comments on New or Updated Analyses in the draft CSS 2016 Annual Report by Chapter

IV.1. Chapter 1. Overview

Chapter 1 is similar to previous years, providing a summary of other chapters and what is new in this year's report. Last year's report was updated with 2015 results. In addition, two new fish populations have been added: natural-origin Okanogan River sockeye and natural-origin summer Chinook from above Wells Dam.

According to the 2016 report, the size of PIT tags is now 11 to 12 mm (p.13, l. 26) instead of the 9 to 12 mm reported in previous reports. Is this an actual change in the size of the PIT tags or just a typo? If this is a real change, the rationale for the change needs to be given and a discussion of potential impacts is needed (e.g., are larger fish now tagged to accommodate the larger tags?).

This year's report has three new topics: 1) statistical relationships among total annual flow and salmon population parameters such as survival, smolt-to-adult-return rate (SAR), and other response variables in the life cycle model; 2) impact of the juvenile bypass system on delayed mortality as measured by SARs; and 3) average age of maturity across stocks and years. Additionally, Appendix C presents preliminary results on using a weighted bootstrap procedure to account for a stratified random sampling of fish from certain populations.

IV.2. Chapter 2. Life cycle modeling of alternative spill experiment scenarios

This chapter continues the development of the life-cycle model. No new features were added to the life-cycle model in 2016. However, they used the life-cycle model to evaluate the impact of alternative spill/flow levels on SARs and long-term abundance of spring/summer Chinook to 2050.

The new work uses the 2015 model and investigates the impacts of flow/spill prospectively by applying simulated future environmental conditions to mimic current conditions or preliminarily investigate climate change conditions.

They also investigate the relative benefit of improvement in juvenile passage vs. improvements in spawning productivity and capacity.

They conclude that:

- greatest benefits to SARS occur at highest spill and lowest flow
- relative return abundance appears to be mostly limited by capacity of the habitat to support the fish.

While the approach is generally well implemented, the ISAB has some concerns about specific aspects of the simulation study that suggest the outcomes may not be as clear cut as indicated in the report. For example, in Figure 2.9 (and similar figures), the box plots and whiskers show the variation in the \bar{R} (the average abundance in the last 10 years of a simulated population trajectory over simulated future flow conditions) over the different simulation scenarios. However, when discussing differences in outcomes among scenarios, it is differences in the average of the \bar{R} ($\bar{\bar{R}}$, the average of the average abundances) that is of interest and no information on the uncertainty of $\bar{\bar{R}}$ is shown. Presumably the uncertainty in $\bar{\bar{R}}$ is very small because it is based on 10,000 simulations. These figures would be improved by adding a “typical” measure of uncertainty for $\bar{\bar{R}}$ to the plots. Otherwise the unwary reader may conclude that because the box-plot overlap considerably, there is no evidence of a difference in the $\bar{\bar{R}}$ among scenarios.

Similarly, there are no measures of uncertainty shown in Figure 2.12/2.13. The text implies that such a measure could be inferred (page p.48, l.31) but it was not clear how this is done. Measures of uncertainty should be added to the graphs by the CSS team rather than forcing the reader to try and impute them.

What was the justification for using 2010 as a typical low-flow year, 2009 as an average year, and 2011 as a high-flow year? Did the team do a frequency analysis to choose these years, and if so, what were the exceedance probabilities for these years? If not, was it just by looking at a plot of annual and seasonal flows and deciding that these particular years seemed to appear low, average, or high flow? If the latter, it might be helpful to refer to Figure 1.6, middle panel.

By choosing only one year's conditions to represent low, average, and high flow conditions, the simulation will have reduced future variability (see next point as well).

The range of observed variation in the long-term results may be understated because variation in all variables was not considered. For example, PITPH and WTT were fixed at certain values for each scenario and not allowed to vary (e.g. p.38, l.38). Harvest doesn't appear to have any variation at different levels of abundance. The ISAB assumes that stochastic variability in harvest has been applied; i.e., a 50% harvest probability does not always lead to exactly 50% of fish being harvested, but this should be clarified in the document. Only one flow "year" was considered for "low", "average" and "high" flows (see above), and this single flow year was repeated for the future population projection. This may be unrealistic to assume that there is little future variation in flows. The report was not clear if demographic stochasticity was applied in the forward projections. For example, was there random variation in the number of recruits produced for a given set of productivity and capacity parameter values? Was there random variation in the number of fish surviving given a particular survival parameter value?

The Pacific Decadal Oscillation (PDO) was simulated independently of flow conditions. There are studies that have indicated that PDO is correlated with stream flow—e.g., refer to the NOAA fisheries [website](#) where it states that stream flow is correlated with the PDO. Will the simulation of PDO and flow as independent variables lead to unrealistic outcomes?

There is some confusion in the text that makes it hard to evaluate some of the results. For example, when evaluating impacts of changes in productivity, the text indicates (p. 39, l.9) that evaluations were made for average flow conditions, but it then continues to read that results are only shown at low flows. The text is also confusing on how to evaluate the impacts at the other flow conditions (p. 39, l.16); this needs some reworking and perhaps an example.

There has been a great deal of interest in the impact of predator control programs on salmon returns, especially the northern pikeminnow and bird programs. A life-cycle model is the natural way to study these impacts, but the current version of the life-cycle model appears to incorporate density dependence only at the spawner-to-smolt stage. The ISAB recommends that consideration be given to modifying the life-cycle model to allow a range of compensatory responses ranging from complete additivity (as now is the case) to a range of plausible compensatory mortality effects related to density dependence and predator selectivity. This revised life-cycle model could then serve as a planning tool for the impacts of predator control. Note that we are not suggesting that compensatory response be estimated from any data, but only to modify the code to allow for prospective exploration of a range of impacts.

IV.3. Chapter 3. Effects of the in-river environment on juvenile travel time, instantaneous mortality rates and survival

This is an update based on another year of data. The methodology has not changed from previous years. We have a few questions and comments.

Both the CSS and NOAA provide survival estimates for in-river survival. How do these estimates compare to each other? If there are consistent differences in the estimates, can these be explained?

Do fish tagged and released at LGR have lower survival below LGR than fish tagged and released at hatcheries and from the upstream fish traps, as might be expected from near term tag-related mortality?

Mortality tends to increase over the migration season and with water temperature (except for sockeye). The report lists four potential mechanisms: (1) declining smolt energy reserves or physiological condition over the migration season and with water temperature, (2) increasing predation rates on smolts over the migration season and with water temperature, (3) increases in disease susceptibility or disease-related mortality over the migration season and with water temperature, or (4) some combination of these often interrelated mechanisms. Is there an attempt by CSS or other investigators to test these hypotheses using other approaches? Answers to these questions may lead to additional improvements in salmonid survival. We agree that the apparent contradictory response of sockeye warrants further investigation.

The ISAB recently released a report on density dependence in the Columbia Basin ([ISAB 2015-1](#)) and a report on how to measure the effects of predator control at various points in the life cycle ([ISAB 2016-1](#)). Is it possible to estimate the fraction of hatchery versus wild salmon and total abundances of each group by species and race? Has CSS tested whether total salmonid abundance or abundance of each species within a migration cohort affects survival? Could this type of an approach shed light on potential compensatory mortality caused by predators? To what extent might salmonid size and condition relate to survival? This latter question cannot be investigated with the current cohort model, but there are statistical models based on individual tags that allow for covariates (such as body mass). What are the limitations to doing such an analysis, i.e., is body mass/size at time of tagging collected for all fish? How far back do these data go?

The ISAB appreciates the excellent discussion. Approaches for improving precision of survival estimates seem worthwhile, especially for enhancing detection through spillways. The ISAB

encourages the CSS to continue its investigation to improve precision in the MCN-BON and RIS-MCN reaches via improved spillway detection or by increasing the number of tagged fish.

IV.4. Chapter 4. Patterns in annual overall SARs

This is an update to the previous report with the recent set of data. In addition, an analysis of SARs vs productivity in 2015 was replaced with an analysis of the relationship between TIR (transport effect) and in-river survival. It is not surprising that the transport TIR is inversely correlated with in-river survival (LGR to BON). This new analysis identified the value for in-river survival when the potential benefits of transportation appear to disappear.

Geometric mean survival of juvenile spring Chinook salmon and steelhead is only ~0.6 when migrating from Rock Island Dam (RIS) to McNary Dam (MCN). It would be interesting to know more about the various factors contributing to the high mortality, such as dam passage and predation.

The CSS reported on the relatively large absolute difference in SAR based on PIT-tags versus run reconstruction (the values are highly correlated, however). As in previous reports, this report listed various hypotheses. An email from the FPC's Michele Dehart to the ISAB indicated that a study is underway to further evaluate PIT-tag effects on salmon survival, but that results will not be ready until after summer 2017 when tagged age-5 Chinook will have returned. Potential bias in survival caused by tagging methodology (or in the run reconstruction methodology) is an important issue to resolve.

IV.5. Chapter 5. SARs and productivity

Chapter 5 continues the examination of the relationships between life cycle productivity and SARs, including the level of SARs needed to reach or exceed population replacement. How were hatchery-origin Chinook adults on the spawning grounds identified and excluded from the productivity estimates of the natural-origin population? For steelhead, the text indicates a weir was used to identify natural origin adult returns and was used to exclude hatchery fish.

The findings suggest that pre-harvest SARs of 4%-6% are associated with pre-1970 levels of productivity for Snake River spring/summer Chinook. How do these SARs compare with SARs from viable wild Chinook populations in other regions? To what extent might improvements in hydrosystem management, predator control, and estuarine habitat lead to SARs of 4%-6%?

From the results in Chapter 4, it seems major improvements would be needed to reach the desired SAR range given recent ocean conditions which lead to low SARs.

IV.6. Chapter 6. Estimation of SARS, TIRS and *D* for Snake River Subyearling Fall Chinook

This chapter contains updates from previous years on SARs by route of passage and TIR for 2006 to 2012. There are some minor changes in how results are reported. The ISAB has no major comments other than the editorial notes below.

IV.7. Chapter 7. Effects of juvenile bypass systems on smolt to adult return rates

This is a repeat of an analysis done in 2010 using similar methodology. As before, only smolts that are detected at BON are used in the analysis. A logistic regression analysis was used to investigate the impact of year, bypass effects, and rearing type on subsequent survival and return as an adult.

Estimates (Table 7.4) appear to be on a “per bypass” basis. What is the average number of bypasses encountered by a fish? Wouldn’t that be a more accurate reflection of the impact of bypass on an outgoing smolt?

The results in Figure 7.1 for steelhead appear to show that for all of the hatchery versus wild comparisons by year the 95% confidence intervals overlap. Wouldn’t that indicate that there is no evidence of an effect of rearing type? Wouldn’t a set of models without the effect of rearing type also be of interest?

IV.8. Chapter 8. Patterns of variation in age-at-maturity for PIT-tagged spring/summer Chinook salmon in the Columbia River Basin

This chapter examined the mean age of maturity among different stocks, years, and fish type (wild or hatchery) using regression methods.

We have many editorial comments on this chapter (see comments below) in how the models are described and in reporting the results.

An analysis was conducted on the proportion of age-3, age-4, and age-5 fish using standard regression methods. Usually, proportions are analyzed in regression contexts using variants of logistic regression. Alternatively, standard regression could still be used with the empirical logit of the proportions as the response variable as outlined in Warton (2011).

A large portion of the discussion (e.g., p. 165, p. 166) deals with models that essentially have no weight in the AIC framework with $\Delta AIC > 100$! Why are results from such models even reported? The authors should conduct model averaging and only discuss the model averaged results.

One consideration for the future work is looking for factors that have led to the shift to declining proportions of four and five-year olds and increases in three-year olds such as ocean factors associated with growth and climate change, differences in hatchery practices, or freshwater environments (tributary temps, or annual differences in migration corridor).

IV.9. Appendix A: Survivals (SR), SAR, TIR, and D for Snake River Hatchery and Wild Spring/Summer Chinook Salmon, Steelhead, and Sockeye

The ISAB found the figures to be well done and informative.

IV.10. Appendix B: Supporting tables on Chapter 4 - Overall SARs

The ISAB has no comments on this chapter (other than some editorial notes, see below) as it is an update of previous reports. This is an excellent set of data tables, and the ISAB anticipates these will be updated in future years. These data may be useful in future years to investigate possible density dependent effects on in-river survival.

IV.11. Appendix C: Development of a weighted bootstrap for unequally represented hatchery PIT-tag groups

Appendix C describes a weighted bootstrap procedure to estimate SARS for PIT-tag groups that are not sampled proportional to run size. For example, if the first half of the run has 90% of the

PIT tags applied, the appendix describes how to weight the tags from the early and late parts of the run to get an overall value for the entire run.

This procedure appears to be satisfactory under conditions that the actual sampling weights are known with little or small error (e.g., see Table C.1). If the sampling weights have to be estimated because the actual sampling fractions are estimated, then this will introduce additional steps in the bootstrap procedure.

The half-step method is a fast way to select samples after the cumulative sampling weights are found. This is one of the methods discussed in Brewer and Hanif (1983).

The authors describe a weighted bootstrap method to estimate the parameters. This is likely to be infeasible in more complex model structures such as the life-cycle model. A likelihood approach based on a stratified sample, followed by deriving weighted averages of the parameters from each stratum, will be computationally more feasible rather than using the average of the bootstrap values (Hogg, McKearn, Craig 2005). A stratified bootstrap procedure could then be used to estimate the uncertainty in the estimates in the usual fashion.

One area of concern is a key assumption that every fish has independent fates. This may not be true; i.e., fish travel together and can share fates, and this may result in over-dispersed estimates. Has this been investigated?

V. Specific editorial comments and requests for clarification on each chapter

This section contains editorial suggestions, requests for small clarifications, and the like.

Front matter

p.v, Figures 2.2. and 2.5. Typo in “Corelations”

Chapter 1

p.5, Figure 1.4. Figure is out of place and needs definitions of abbreviations as in Figure 1.1.

p.12, l.31: typo on “reflects” (missing the “r”)

p.12, l.32-33: Suggest wording “but included for review” more clearly. We believe that this means that it’s included in Appendix C for review by ISAB?

p.12, l.42: typo on “implemented”

p.15, l.27-33: There are some grammatical or punctuation errors in this new paragraph.

p.16. Table 1.1. Header needed for 4th column of table (currently says “25”?)

p.17, Table 1.3: The note at the bottom of the table says “** Pre-assigned by NPT” but this notation is not in the table. Should there be a “**” somewhere?

p.19, Figure 1.6. Why was the proportion transported estimated in 2015 rather than known?

p.20, l.11: Typo on “three” (missing the “t”)

p.20, l.15-21: Suggest that the preliminary nature of the results be indicated here.

p.20, l.23: Change “effects” to “relationship”

p.20, l.24: Change “seasonality on” to “seasonality to”

p.20, l.36-38: This sentence is confusing. If the correlation is between SARS and wild and hatchery populations, then it doesn’t seem like common environmental factors were included in the correlations and the second part of the sentence should say something more like “suggesting that common environmental factors could be influencing survival rates.” If common environmental factors were included in the analysis (which it appears it was), then the sentence

should be reworded to make that clear, perhaps "... and common environmental factors were significantly related to survival rates from outmigration to the estuary and ocean environments."

Chapter 2

p. 24, l.19: What does "TAC" stand for?

p.28, l.3-6. Correlations are among the $\log(R/S)$ for each population and the environmental covariates. Reword these sentences to improve readability.

p.28, l.23 "is" should be "are"

p.28, l.24 "know" should be "known"

p.28, l.37 and onwards. Some symbols are not directly defined anywhere?

p.31, l.5 "Bonneville Dam" should be "Bonneville Dam." Check rest of chapter as well for such usage.

p.33. Table 2.1 Move definitions of terms before the equations for the model.

p.34, l.24. Uninformative priors do have an impact on parameter estimates. For example, a $U[0,1]$ commonly used as an uninformative prior for a proportion, pulls the estimated proportions towards 0.5. "Uninformative priors" still provide information.

p.36, l.15: "it's" should be "its"

p.36, l.23. Don't understand why e_t is multiplied by 0.7? The series is normalized later to match the range of the observed variation so this seems to be redundant?

p.37, l.35. "... saving the better combinations (accepting) and not saving (rejecting) worse combinations ..." The MCMC Metropolis-Hastings algorithms accepts/rejects MOVES from the current parameter values but does not say that a MOVE is a better or worse parameter combination in any absolute sense. Similarly, the wording in the whole paragraph needs some tightening up to make it more technically accurate.

p.38, l.19 "the the" needs to be fixed.

p.39, l.15. We didn't understand this explanation. Please provide additional details.

p.43, Figure 2.6. "Triangles" should be defined in upper plot.

p.44, l.26. Reword to “an increase in the AVERAGE total number of returning spawners.” Make similar changes everywhere in this chapter when discussing the results of the simulations.

p.44, l.38. Reword to “Figure 2.10 shows the predicted AVERAGE SARS” and make similar changes throughout this chapter.

p.47, Figure 2.9. The figure indicates that average log-abundance is plotted. But the earlier explanation of the simulation indicated that the average abundance was collected. So perhaps this should be the log(average abundance)? The flow conditions are ordered High, Average, Low; why was this order of flow conditions used rather than Low, Average, and High which would be ordered “numerically.” As the plot now stands, average total returns tend to increase in each cluster as you read from left to right which corresponds to DECREASING flows—this is sure to trap the unwary reader. Y-axis label doesn’t note that results are on log-scale. As noted previously, the ISRP is concerned that not all sources of variation have been captured in the simulation, and so the box plots may underreport the actual variation in average total returns.

p.48, l.14. “You can see it the mouth...” Not clear what is meant here. Do you mean “Bonneville”? Reword.

p.48, l.16. Not clear how the relative average abundance to that of BiOp was computed? Is this the simple ratio of the average of the average abundances? To avoid the influence of outliers, perhaps the ratio of the median of the average abundance should be shown. Measures of variation need to be added to the plots (and discussion).

p.48, l.20. “significant.” How do you know without measures of uncertainty? In general, how do you know if the effects actually vary across treatments without measures of uncertainty?

p.48, l.30. Now to switch to comparing the MEDIAN long-term averages. Why the switch from the average of the averages?

p.48, l.31. Reviewers didn’t understand how the uncertainty corresponded to crossing the shaded area. Further explanation is needed.

p.48, l.39. Reword to “predicted increase in AVERAGE.”

p.49, Figure 2.10. Similar comments about ordering flow levels. Y-axis label should be labeled as average SAR (similar to Figure 2.9 labeled as average return). Similar comment about not capturing all variation in results.

p.50, Figure 2.11. Similar comments about ordering flow levels. Y-axis should be labeled as average SAR. What is meant by “mouth” – Bonneville?

p.51, Figure 2.12. Not clear how these were computed (see earlier comments). Measures of uncertainty need to be presented for each ratio. Legend needs reword to “when compared to expected long-term average abundance at BiOp ...”

p.52, Figure 2.13. Similar comments to Figure 2.12.

p.53, Figure 2.14. Check Y-axis label. Are these really in 1000’s of fish?

p.54, l.27. Is this 28% of 42% or 28 percentage points (i.e. by 0.28)? Best to express survivals as proportions rather than percentages to avoid these confusions, i.e. in-river survival of 0.42 rather than 42%.

p.54, l.32. Similarly, is this 10% or 10 percentage points?

p.55, Figure 2.15. Similar comments to Figure 2.14.

p.56, l.15, Reword to “increasing AVERAGE abundance and AVERAGE SARs” here and elsewhere in the discussion, e.g. p.56, l.25;

p.56, l.8: The table number is missing (should it be Table 2.3?).

p.57, Table 2.3. Without measures of uncertainty how do you know these rankings? Is it possible to show the value of the metrics you used to rank the scenarios in the table (reviewers guess these would be SAR or Rbar)?

p.57, l.2. Reword to “highest AVERAGE SAR”

p.57, l.8. Missing Table number. The wording here is very confusing. It appears to be arguing to remove the highest ranked result (high spill at low flow), and there is a justification for that in the wording. Is it important to have a spill scenario for each flow, if the high spill at low flow is removed?

p. 57, l. 12-27: Be careful of wording about model results. Lines 14-15 say that these precise in-river survivals will happen, but as pointed out in this paragraph, there is uncertainty in the models. The argument about the 20% transportation fraction is a bit confounding as it is being used as a rationale for why model results are giving high predicted SARs, yet the next sentence argues that it’s a reasonable number.

p.57, l.22. “five less powerhouses” than which population?

p. 57, l.28 to p. 58, l.8: The method described here was not clear in the methods section. Over how much time were the fixed levels held? The simulations are a good first step, but just to give preliminary insights into relationships. The variability of flow may not have a predictable effect on model outcomes based on the static conditions simulated.

p. 58, l. 9-10: The model results show that increased abundance can be related to alternative treatments, but they are not showing they are a result of alternative treatments. The results are showing which items have strong relationships, but also because they are so preliminary, the strength of these relationships may change if assumptions are changed (for example, if non-static flows were simulated).

p.58 l.12 Insert AVERAGE before all SARS.

p.58, l.17. Always report in-river survival as a proportion. Don't switch to %.

p. 58, l. 31-40: Reviewers suggest emphasizing the preliminary nature of these results.

The section on the simulation study needed editorial work to make it flow better: remove colloquialisms “e.g. you can see” and be consistent in the presentation (e.g. always express survival in proportions).

Chapter 3

p.66, l.33. We don't particularly like \sqrt{Z} transforms except if the random variable is a count. We would always use $\log(Z)$ based on theoretical considerations that effects are multiplicative. The results should be similar under the two transformations.

Chapter 4

P.98. Table 4.1. AIC column needs 1 decimal place.

p.98, l.7. “Model of the form $\ln(\text{TIR}) = \ln(\text{SR}) + \text{species}$.” Reword, either use a proper model notation with proper coefficients associated with each term or go to a short hand R-like notation, but don't do something in between. Also, traditionally models are written with intercept terms first.

p.98. l.15. “... estimated significance for the species specific intercept at alpha 0.09.” Don't mix paradigms—AIC methods avoid p-values and only use model weights to decide among models.

p.98, l.20. If using AIC, don't report an overall p-value.

p.99, Table 4.2 What does the coefficient associated with "species" represent? This appears to be just the difference in the intercepts for the two species, but why not just present the intercept for the two species instead. Report fewer decimal places for the estimate and SE. Do not report t-values and p-values when using AIC.

p.99. The authors used AIC and found multiple models with similar weight. Yet they didn't model average. Why not? Rather than just using the best model to predict at which point transportation benefits disappear, model average first, and use the model averaged coefficients to make the prediction.

p.100, Figure 4.10. Report survival rates when benefits of transportation cease as a proportion rather than a % to match the graph and text.

p.112, l.36 "survival rates." Not rates, just probabilities. Reword consistently here.

p.113. Label on Y-axis is "survival rate." These are not rates (are they on an annual basis?), but simply probabilities. Reword the axis.

p.115, l.42 "survival rates." Remove "rates" here as they are not an annual basis. Go through entire discussion and fix usage of "rates" where it appears.

Chapter 5

p.121, l.40. Model notation is awkward. The index j is used as brood years, but each brood year also belongs to different periods. The current notation also uses j as an index to year. Perhaps consider a different notation. It is not clear what is a "period." There also doesn't appear to be any results from fitting this model? Was it used in this chapter?

p.121, l.40. Are spawner numbers estimated here? If so, and if the uncertainty in the spawner numbers is appreciable, you have an error-in-variables problem as well which requires a different fitting approach.

p.122, l.19. "survival rate." This is not a rate (per year?) but rather just a survival probability.

p.130, Figure 5.6. Too many decimal places are reported.

p.130. Figure 5.7. Plot all figures on the same scales to make it easier to compare results. Different symbols are used, but these appear to be related to the different locations and not the different brood years as in Figure 5.6. Make the two sets of figures consistent.

p.131. Figure 5.8. Use same symbol color and shape as in Figure 5.6.

Chapter 6

p.148, Table 6.11. Don't bold the confidence interval bounds as this is not sensible. Because these are 90% confidence intervals, the formal test has an alpha of 0.10 rather than the usual 0.05.

p.149, Table 6.12. Similar comments to Table 6.11.

p. 151. Figure 6.5. Because prediction intervals were plotted, it is impossible to get some idea of the uncertainty in the estimated survival at which $\ln(\text{TIR})=0$. Include the confidence interval for the mean response and the end points where the confidence interval also intersects 0 as this will provide a measure of uncertainty on the estimated survival where $\ln(\text{TIR})=0$.

Chapter 7

p.155, Table 7.1 Do these returning adults include harvest?

p.155, l.7. Rather than using a binomial with an index of 1, just use a Bernoulli distribution.

P.155, l.10. Rearing variable never defined.

p.156, Table 7.2. There is no definition of "Bypass Location," which we assume is Snake vs. Columbia river dams.

p.157, l.14. Model averaging ignored models where the bypass effects were forced to be zero (models 1 and 5) why? There are several schools of thought if such models need to be included when estimating the effect of regression variables (see Burnham and Anderson, 2002), but because the interest lies in bypass effects (and a 0 bypass effect is a possibility), these models should be included in the model averaging. This comment is somewhat moot given the near zero model weight placed on these models.

p.157, Table 7.3. Fish are not tagged independently of each other, and so there is a possibility of overdispersion. Model diagnostics need to be done on the best fitting model to see if adjustments for overdispersion are needed.

p.158, Figure 1. But these are model averaged estimates and because the best fitting model for steelhead had no interaction, a consistent difference on the logit scale is enforced by the model

and not by the raw data. A better plot would be of the raw SAR values to see if parallelism is evident.

p.160, l.6. The reported values are per bypass are they not? So, on average how many bypasses are encountered by an outgoing smolt?

p.160, Table 7.4. If you adopt the AIC paradigm, there is no need to report p-values. Just report estimates and confidence intervals. Table legend needs fixing for reference to "Table 2." These appear to be estimates from each model, so what are "adjusted standard errors."

p.161. Table 5. Needs to be re-numbered to Table 7.5? Fix reference to Table 2. Do not report p-values. Check estimates for Chinook as model with different bypass effects by system has important weight.

Chapter 8

p.164, l.8. How was harvest dealt with?

p.164, l.25. The model needs a term for tau to be multiplied against (e.g. an indicator variable for hatchery vs wild). The model assumes equal variance, but the mean age of maturity will be based on different sample sizes and so will have a different variance. It may turn out that the process error is much larger than this sampling error, but this needs to be investigated.

p.164, l.32. If the proportion age 3, age 4 or age 5 is used as the response variable, then this is now a logistic regression exercise. Or following the advice of Warton (2011), use the empirical logit of the proportions as the response variable.

p.164, l.39. This model needs fixing because none of the parameters are multiplied against any design variables!

p.165, l.3. If using AIC, what is the set of models being compared? If using AIC, don't report p-values etc. Later on, we see that Table 8.2 has the model set, but this is never referenced directly.

p.165, l.24. Are these model averaged estimates or model specific estimates? These appear to be specific from model 2, but according to Table 8.2, this model essentially has NO weight with delta AIC more than 100 units away. So why are results from a model with such low weight being reported? Report only the model-averaged estimates.

p.166, l.1. Again, why are models with such low weights being discussed at all? Don't mix AIC with hypothesis testing (i.e. no p-values). The whole discussion about model effects changing delta AIC should be struck and just report model averaged estimates.

p.166. Similar comments about models for proportion of age 3, age 4, or age 5. Only report and interpret model averaged estimates. No p-values if using AIC.

p.167, l.28. "... variability in mean age is a direct result in variation in proportions within each age class." Yes, this is true without needing any statistics to prove it. Again, how was harvest accounted for?

p.169, Figure 8.1 How were the mean ages for a stock averaged across brood years? A simple average? Weighted by the number of adults?

p.172. Figure 8.2, 8.3, 8.4. How was the mean proportion of age 3, 4, 5 averaged across brood years? Was it a simple average? Was it weighted by the number of adults? What is the purpose of the different colors across the three plots? The bars in a different order in each year, so it makes it difficult to compare across plots. Perhaps a combined plot with lines for $p(\text{age } 3)$, $p(\text{age } 4)$, $p(\text{age } 5)$ plotted across the stocks would be more useful. We don't think that stacked bar charts would be very helpful.

p.174. Figure 8.5, 8.6, 8.7, 8.8, and 8.9. Not sure how useful these figure are. There appears to be a high correlation among stocks, but can these be sorted in groups by location in the basin, common experiences at sea? The last paragraph in the discussion on page 169 is getting at this, but perhaps more discussion is needed?

p.176, Figure 8.10. Not clear what the reader is to infer from this plot? Perhaps it can be deleted?

p.178, Table 8.2 It is customary to sort model tables by the delta-AIC value. Add the model weights to the table.

p.179. Tables 8.3, 8.4, 8.5. See comments for Table 8.2. As noted earlier, these should be analyzed on the logit-scale.

Chapter 9

Appendix A

Appendix B

Table B.72. These are not survival “rates”. Capitalize “lgr” in table columns headers. Here SAR are presented as proportions, but elsewhere in the report they are reported as a %? Perhaps also report as % here?

Appendix C

p. C-7. Table c.2 Improve table headers.

VI. References

Burnham, K.P. and D.R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach Springer.

Brewer, K.R.W. and M. Hanif. 1983. Sampling with Unequal Probabilities. New York: Springer-Verlag.

Hogg, R.V., McKean, J.W., and Craig, A.T. 2005. Introduction to Mathematical Statistics. 6th Edition. Pearson, New York.

ISAB (Independent Scientific Advisory Board). 2015-1. Density dependence and its implications for fish management and restoration programs in the Columbia River. Northwest Power and Conservation Council, Portland, Oregon. ISAB Report 2015-1. 246pp. Available at www.nwcouncil.org/fw/isab/isab2015-1.

ISAB (Independent Scientific Advisory Board). 2016-1. Predation Metrics Report: Developing and Assessing Standardized Metrics to Measure the Effects of Predation on Columbia River Basin Salmon and Steelhead. Northwest Power and Conservation Council, Portland, Oregon. ISAB Report 2016-1. Available at www.nwcouncil.org/fw/isab/isab2016-1.

Warton, H. 2011. The arcsine is asinine: the analysis of proportions in ecology. Ecology 92, 3-10.