Department of Energy

Bonneville Power Administration
P.O. Box 3621
Portland, Oregon 97208-3621

Environment, Fish and Wildlife

July 31, 2007

In reply refer to: KEW-4

Michele DeHart
Fish Passage Center
1827 NE 44th Ave., Suite 240
Portland, Oregon 97213

Dear Ms. DeHart:

Thank you for the opportunity to comment on the Comparative Survival Study (CSS), Ten-year Retrospective Analyses Report, May 30, 2007. We have included points below and provide a more detailed analysis as enclosures including a (1) General Technical Comment; (2) Evaluation of the CSS Response to the ISAB Recommendations; and (3) Detailed technical review comments.

The (CSS) 10-year Retrospective Report provides a history of PIT-tagged salmonid fish performance from 1998 through 2006. The length of time, the breadth of geographic coverage, and range of salmon life-history phases investigated in the report have the potential of providing a valuable chronicle of recent Columbia River trends. No other study in the Fish and Wildlife Program has the same scope of effort.

As we have emphasized in past reviews of CSS Annual Reports, and now for this CSS Ten-year Report, an overriding issue for CSS analyses is reproducibility. It is imperative that CSS analyses be capable of accurate reproduction or replication by independent researchers to see if their analyses give similar results to those reported by the original group. The ability to reproduce results is crucial to the scientific review process. Reproducibility requires transparency in terms of sufficient data and detailed methods to allow a third party to reproduce the analyses contained in the Report. As has been noted in the past (e.g., the ISAB 2005 CSS review, in the review by the ISAB on the 2007-2009 CSS Proposal, and BPA's Review of the 2005 Annual Report), CSS analyses have not always been sufficiently transparent. The CSS Ten-year Retrospective Analyses Report continues this pattern, as it does not include sufficient data and detailed methods to allow a third party to reproduce the analyses and conclusions contained in the Report.

- Tagging Results and Reproducibility -- Our attempts to reconstruct final results from intermediate calculations presented in the report have been limited by the absence of necessary information or insufficient technical description.

- Latent Mortality -- When the performance measure for "differential mortality" is corrected for the extra migration of upstream stocks, there is little or no evidence of differential hydrosystem mortality for hatchery Chinook salmon.

- Tagging Results --The CSS Report needs to simply document and display the tagging results for the benefit of most readers and organizations that do not have the resources of the CSS organization. This issue is fundamental to our comment - the need to provide the means to reproduce results.

- Non-standard practices -- The report includes non-standard modeling practices resulting in limited use of the analyses. These practices need to be peer reviewed.

- Missing information -- Basic information and mathematical definitions for equation parameters such as SARs, and also the number of fish actually transported at each dam are absent.

- Upstream and downstream comparisons -- CSS continues to compare upstream and downstream Chinook salmon stocks when the data clearly do not support such comparisons. Previous critique of the upriver-downriver comparison including the 2005 ISAB review has documented this point. The CSS Report does not demonstrate a biological difference given fish size, migration date, marine arrival timing, and year, in addition to upstream/downstream classification.

- Invalid assumptions -- The analyses assumes that no natural mortality occurs once salmon pass the first upstream dam, thus concluding that all mortality between upstream and downstream dams is caused by the hydrosystem. When the performance measure is corrected for the extra migration of the upstream stocks, there is little or no evidence of differential hydrosystem mortality for hatchery Chinook salmon.

- Due to the inability to reproduce these results using accepted modeling and analytical procedures the CSS Report's findings do not demonstrate the scientific rigor and support to authoritatively guide hydrosystem management.

Please let us know if you have any questions or require further clarification on our comments. As we stated in our 2005 and 2006 comments on the CSS Annual Reports, it is critical that the issues raised be addressed because of their importance for the continuing work under the project.

Sincerely,

/S/ Robert J. Austin

Robert J. Austin
Deputy Director of Fish and Wildlife

Enclosures

cc:
Dr. Tom Karier, Northwest Power & Conservation Council
Mr. Bill Booth, Northwest Power & Conservation Council
Mr. Jim Kempton, Northwest Power & Conservation Council
Ms. Joan Dukes, Northwest Power & Conservation Council
Mr. Bruce Measure, Northwest Power & Conservation Council
Ms. Rhonda Whiting, Northwest Power & Conservation Council
Mr. Larry Cassidy, Northwest Power & Conservation Council
Ms. Melinda Eden, Northwest Power & Conservation Council
Mr. Tony Grover, Northwest Power & Conservation Council
Mr. Brian Lipscomb, Columbia Basin Fish & Wildlife Authority
Chairperson Wanda Johnson, Burns Paiute Tribe
Chairman Chief Allen, Coeur d'Alene Tribe
Chairman Michael Marchand, Confederated Tribes of the Colville Reservation
Chairman Glen Nenema, Kalispel Tribe
Chairperson Jennifer Porter, Kootenai Tribe of Idaho
Chairperson Samuel Penney, Nez Perce Tribe
Chairman James Steele Jr., Confederated Salish & Kootenai Tribes
Chairperson Alonzo Coby, Shoshone Bannock Tribes of Fort Hall
Chairman Kyle Prior, Shoshone Paiute Tribes of the Duck Valley Reservation
Chairman Richard Sherwood, Spokane Tribe of Indians
Chairman Antone Minthorn, Confederated Tribes of the Umatilla Indian Reservation
Chairman Ron Suppah, Confederated Tribes of the Warm Springs Reservation
Chairman Lavina Washines, Yakama Indian Nation
Ms. Mary Verner, Upper Columbia United Tribes
Mr. Olney Patt Jr., Columbia River Intertribal Fish Commission

bcc:
S. Wright - A-7
L. Bodi - A-Seattle
S. McNary - A-7
K. Hunt - DKR-7
B. Cobell - DKT-4
K. Johnston - DKT-7
R. Bennett - K-7
S. Stewart - KEC-4
B. Austin - KEW-4
B. Maslen - KEW-4
G. Dondlinger - KEWB-4
P. Lofy - KEWL-4
K. Fisher - KEWR-4
P. Krueger - KEWR-4
A. L'Heureux - KEWR-4
K. Powers - KEWR-4
M. Shaw - KEWU-4
L. Grimm - LC-7
P. Key - LC-7
Official File - KEW (FW-24-11)

Austin:mpr:4748:07/24/2007 (KEW-4-W:\Kew\KEW07\FW-24-11\CSS Comment Letter.doc)

Review of
COMPARATIVE SURVIVAL STUDY (CSS) of PIT-Tagged Spring/Summer Chinook and Steelhead in the Columbia River Basin: Ten-Year Retrospective Analyses Report

by

Bonneville Power Administration
P.O. Box 3621
Portland, Oregon 97208-3621

July 24, 2007

Table of Contents

General Comments 2

Review of Chapter 2 4

Review of Chapter 3 and Appendix D 8

Chapter 4 Review 13

Chapter 5 Review 17

Chapter 6 Review 27

Chapter 7 Review 30

Chapter 8 Review 31

Appendix B Review 33

Literature Cited 35

General Comments

The Comparative Survival Study (CSS) 10-year retrospective analysis provides a useful history of PIT-tagged fish performance from 1998 through 2006. The length and breadth of the tagging data provide a valuable look at the history of salmonid survivals, travel times, transport/inriver ratios (TIRs), smolt adult returns (SARs), etc., in the Columbia Basin. No other study has the same temporal, geographic, or salmonid life-history scope as this project. For this reason, documenting the data collected and the status and trends of the estimate of various performance measures is crucial for the Columbia Basin Fish and Wildlife Program which has supported this work. It is therefore surprising that this important task is limited to a relatively few tables and graphs in Chapters 2 and 3. Appendix D supplements the information in these chapters but never quite reaches the level of showcasing the important trends in the results. In many cases, standard errors or confidence intervals are neither reported nor displayed.

As urged by the Independent Scientific Advisory Board (ISAB), this CSS report now presents some of the methods used for estimating SAR, TIR, and latent mortality (D) for the various groups of interest (i.e., the C0, C1, and T0 groups). It has compiled methods strewn throughout previous reports in one place, and this makes reading the report much easier. Nevertheless, this encouraging start was not continued throughout the report or consistent across chapters. Complicated performance measures such as annual SARs are described verbally but never mathematically defined in equations. Cryptic tables are included, showing values used to estimate annual SARs (Tables D-20 and D-30), but it is not clear what these values are or how they were combined to estimate annual SAR. In a report as important as this 10-year review, the first priority should have been simply presenting the facts (i.e., results). Closely tied to this first objective should have been much more transparency and clarity of methods in this report. Attempts to reconstruct final results from intermediate calculations have usually been difficult due to lack of necessary information or insufficient guidance. The ability to reproduce results is crucial to the scientific peer review process.

A large proportion of the 10-year review focuses on interpreting the PIT-tag results and assessing the influence of environmental and hydrosystem effects on inriver survival and adult returns. The 10-year review includes both approaches previously described in annual reports as well as new analytical methods. These analyses are both the most interesting and often problematic aspects of the report from an analytical perspective.

In Chapter 2, the concept of instantaneous mortality rate (Z) is introduced. However, it is not based on failure times (i.e., death times of PIT-tagged fish) as it properly should be but, rather, on a simple function of the ratio of reach survival estimates and median travel times. The authors then go on to analyze survival, travel times, and Z as if they are three independent pieces of information. Reach survivals within a season are relatively stable while travel times show marked seasonal trends. Using the ratios of this information, Z values are calculated and seasonal trends are (mis)interpreted as survival processes. In fact, the trends in Z are nothing more than inverse trends in travel times misinterpreted or misconstrued as survival effects. Curiously, results of model analyses on reach survivals are not discussed, leaving the impression that results of Z values are applicable to S, which is *not* true. Finally, the authors interpreted the instantaneous mortality rate (Z) as the probability of mortality (i.e., $1-S_t = 1-e^{-Zt}$).

As requested by the ISAB, the CSS has compiled in Chapter 3 and Appendix B many of the methods used to generate the time series of estimates reported. Nevertheless, some methods and definitions are missing here and throughout the report (e.g., annual SAR). Also missing are certain basic results, such as the number of fish actually transported at each dam, which should be documented in this report. It is very helpful to see the figures of trends in the performance measures over time, and to see comparisons between hatchery and wild stocks. Also the 90% confidence intervals included on some figures aid interpretation. However, the CSS Report has based too much inference on whether confidence intervals on two estimates overlap. Non-overlapping confidence intervals is an invalid test of significant differences. Instead, the CSS should find valid methods of testing significance, either within their bootstrap approach or separately with a parametric approach.

Chapter 4 explores the causes and nature of the interannual variation in performance measures such as SAR, TIR, and D. Multiple regression was used to model the responses. Although the summary tables are cryptic, it appears models with partial regression coefficients had signs inconsistent with the investigation philosophy (e.g., negative sign with flow) were consistently omitted. This practice left models that had interaction terms but no main effects or quadratic terms without the linear component inconsistent with general model building practices. Another aspect of the report used parametric models to partition total variance of metrics into natural variation and measurement error. However, the assumption, for example, that SARs are binomially distributed is inconsistent with the mark-recapture models used to estimate the values. Underestimating sampling error will positively bias estimates of natural variation. The report needs to use goodness-of-fit tests to assess the model assumptions and compare their parameter estimates with those of the nonparametric variance component formulas provided. Their inferences concerning natural variation do not take into account their own findings on ambient effects, the historical distribution of those factors, or how influences such as global climate change many affect projections in the future.

The CSS continues in Chapter 5 its comparison of upstream and downstream Chinook salmon stocks. As in the past, multiple upstream hatcheries and collection points are used, while only a single downstream hatchery and collection point (for wild fish) is used, despite the ISAB's recommendation to incorporate more downstream stocks. Given that this is a retrospective report, it is understandable that the CSS could not immediately include additional downstream stocks. While the CSS does perform useful comparisons of biological characteristics of the upstream and downstream stocks, their upstream-downstream analysis is invalid in other critical ways. The CSS uses an invalid performance measure to identify delayed mortality caused by the hydrosystem. This approach assumes no natural mortality for smolt should occur between upstream and downstream sites. When the performance measure is corrected for the longer migration of the upstream stocks, there is little or no evidence of delayed hydrosystem mortality for hatchery Chinook salmon. Similarly, the CSS Report does not consider the longer distance to travel for upstream stocks when comparing travel and arrival times of upstream and downstream stocks. Even if the hydrosystem were not in place, the upstream stocks would still have farther to travel than downstream stocks.

Chapter 6 attempts to partition survival across different portions of the migration, focusing on smolt survival from the hatchery/trap to LGR, perceived adult survival from BON to LGR, and perceived adult survival from LGR back to the hatchery/spawning grounds. Adults are categorized by juvenile migration method. The effect of juvenile migration method—in

particular, transportation—on perceived adult upriver survival is an important question, and the analyses in this chapter relating adult survival to migration method are worthwhile. The CSS should provide the methods used in estimating upriver survival for a given juvenile release group. Reviewing and reproducing their results is difficult without those methods. Additionally, the report misinterprets the odds ratio from their logistic regression when comparing adult survival for LGR-transport fish to other fish; consequently, they overestimate the effect of LGR transportation on upriver adult survival.

Chapter 7 describes simulations done to assess the effect of violations of key Cormack-Jolly-Seber (CJS) assumptions on estimation of C0, T0, C1, SARs, TIR, and D. The assumptions considered were (1) all fish have common survival and detection probabilities, and (2) detection has no effect on subsequent survival. Assumption violations considered were (1) temporal changes in survival and detection probabilities, and (2) differential inriver survival of pre-assigned groups (T=transport group, R=return-to-river group) based on past detections. This is an important exercise, demonstrating the robustness of the estimation methods to all but severe temporal changes in survival and detection probabilities, and the dependence of estimation methods on the assumption of common survival regardless of past detections. The focus on the T and R groups is reasonable, given the ISAB recommendation to pre-assign future transport groups in this way. However, the assessment of assumption violations using the T and R groups does not translate directly to the C0, C1, and T0 groups or to the study design used in the past. The CSS should have performed a third set of simulations assessing the effect of detection-influenced survival directly on estimates of C0, C1, and T0, in order to more correctly assess the robustness of past analyses.

Chapter 8 provides a summary of objectives and findings from the 10-year retrospective report. Because conclusions are at times based on the invalid analysis of the earlier chapters, their inferences are invalid as well. The CSS attributes all differences in survival and travel time between study groups in the upstream-downstream comparison to the hydrosystem, ignoring differences expected because of different migration distances. Additionally, the CSS incorrectly claims to have addressed the question of whether smolt transportation compensates for effects of the FCRPS on survival of Snake River Chinook and steelhead. At most, the comparison of the SARs of transport fish and inriver fish indicates whether transportation is a viable management option; it is not equivalent to comparing transportation to migration through the unimpounded river. The question of the effect of the FCRPS on salmonid migration and survival is important. However, it is not addressed by the analyses presented in this report.

Review of Chapter 2

In this chapter, travel time, survival, and a measure of instantaneous mortality were estimated over two reaches, LGR–MCN and MCN–BON for the years 1998-2006 for hatchery/wild yearling Chinook salmon and steelhead. Multiple regression analysis was used to examine the relationship between these metrics and various environmental covariates. Within season, eight weekly cohorts were formed to monitor trends within the year.

- Page 18, last paragraph

  The report used the exponential decay model

  $$N_t = N_0 e^{-Z_t} \quad (1)$$

  to derive a measure of instantaneous mortality rate Z. Solving for Z in Eq. (1) yields

  $$Z = \frac{-\ln\left(\frac{N_t}{N_0}\right)}{t}$$

  or

  $$Z = \frac{-\ln\left(S_t\right)}{t}. \quad (2)$$

  The report then goes on to estimate Z by

  $$\hat{Z} = \frac{-\ln\left(\hat{S}\right)}{\hat{FTT}} \quad (3)$$

  where

  $\hat{S}$ = reach survival rate,

  $\hat{FTT}$ = median fish travel time for the fish that survived the reach.

  Equation (3) provides a convenient but biased estimate of the instantaneous morality rate. Properly, the maximum likelihood estimate of Z would be based on the likelihood model

  $$L = \prod_{i=1}^{N} Z e^{-Z t_i}$$

  and estimator

  $$\hat{Z} = \frac{1}{\bar{t}} \quad (4)$$

  where t_i = lifetime for the ith fish $\left(i = 1, \ldots, N\right)$. However, PIT-tag data do not provide lifetimes for the fish, only travel times for the survivors. Therefore, the PIT-tag data are incapable of estimating instantaneous mortality rates. Any relationship between the true estimates of Z [Eq. (4)] and that used in the report [Eq. (3)] may be appropriate at best and seriously biased as worst.

- Page 20, second paragraph

  In performing the regression analyses, the response variables were

  a. $\ln\left(\hat{S}\right)$

b. Median ln (FTT)

c. $$Z = \frac{-\ln(\hat{S})}{FTT}$$

or

d. $$\ln \hat{Z} = \ln(-\ln \hat{S}) - \ln FTT$$

Both weighted and unweighted regressions were performed using a variety of weights:

a. Inverse variance

b. Inverse CV

c. Inverse CV^2

The report seemingly takes a shotgun approach to the analysis. In the results section, which weighing scheme and why its selection was not revealed. The weight selection should be objective. Proper weighting should be inversely proportional to the variance except when the variance estimates is correlated with the response variable. In this case, the weight should be inversely proportional to the variance but adjusted to eliminate the correlation.

In the case of $\ln(\hat{S})$

$$\mathrm{Var}(\ln \hat{S}) \mathrm{B} \frac{\mathrm{Var}(\hat{S})}{S^2}.$$

However, $\mathrm{Var}(S)$ in a CJS model is proportional to S, saying $\mathrm{Var}(S) = S \cdot f(n)$ where $f(n)$ is a function of sample size and detection probabilities. Then

$$\mathrm{Var}(\ln \hat{S}) = \frac{Sf(n)}{S^2} = \frac{f(n)}{S}.$$

Consequently, the proper weight should be inversely proportional to that quantity after adjustment for S, where

$$W = \frac{1}{\left(\frac{f(n)}{S}\right)} \cdot \frac{1}{S} = \frac{1}{f(n)},$$

or in other words,

$$W = \frac{\hat{S}}{\mathrm{Var}(\hat{S})},$$

which was *not* one of options considered by the CSS report.

As Z is estimated in the report,

$$\hat{Z} = \frac{-\ln(\hat{S})}{FTT},$$

analyses of $\hat{S}$, FTT, and Z are not independent. For example, by the formulation of Z, if the FTT have a downward seasonal trend and $\hat{S}$ is static, then Z will have an upward seasonal trend (e.g., Fig. 2.4). There is no new information conveyed by the third relationship that is not known for the first two trends. Only if Z was actually estimated by actual fish lifetimes would it provide new information not already captured by $\hat{S}$ and FTT.

- Page 20, last paragraph

The report states, “we examined the sign of the parameter coefficients for plausibility and eliminated models with implausible sign.” This is a dangerous and potentially misguided approach to modeling. First, such an approach eliminates the possibility that new insights might be developed and assumes all preconceptions are correct. Secondly, it is unwise to directly interpret the sign (+ or −) of partial regression coefficients (Neter et al. 1996:290-291). Such signs do not necessarily indicate a positive or negative relationship between dependent and independent variables but, instead, adjustments of the model in the presence of other covariates. This unorthodox model strategy can lead to odd modeling results (see comments below).

- Page 20, last paragraph

The report states, “models were fit and ranked according to their AIC_C and BIC scores.” However, many tables (e.g., Tables 2.7-2.11, 2.13) report AIC scores while other tables (e.g., Tables 2.12, 2.15-2.16) report AIC and AIC_C scores. What was actually done and reported needs to be clarified. For example, are the AIC values in Tables 2.7-2.11 actually AIC_C and “AIC” is a typo?

- Page 21, Section “Comparing survival modeling approaches,” first paragraph

“Integrated models of fish travel time and instantaneous mortality, with each component modeled being a function of environmental covariates” are mentioned but never described. If a multivariate computational model was actually used, it needs to be provided, along with associated assumptions (providing Eq. 2.2 is inadequate).

- Page 22, multiple references on this page

Julian day was found in several instances to help describe regression relationships. The implication of this covariate in the models must be described for it is unlike the other covariates considered (e.g., WTT, percent spill, etc.). Julian date is a surrogate for numerous factors that may have a within-season trend including smoltification, flows, temperature, turbidity, etc. If the purpose of the regression analyses is to describe environmental and hydrosystem factors affecting fish response, inclusion of Julian data obscures the results. In some instances, (e.g., Table 2.15-2.16), it does a very good job all by itself!

- Page 23, fifth paragraph – Comparison of survival modeling approaches

AIC scores cannot be compared across different data sets (Burnham and Anderson 2002:80-81). Comparison of models of FTT and instantaneous mortality versus direct survival is inappropriate and Table 2.2 should be eliminated from the report.

- Page 24, first paragraph

The authors are totally misinterpreting their estimates of instantaneous mortality Z. In this paragraph, they are equating Z to probability of mortality which is *wrong*. For example, the instantaneous rate of 0.112 (steelhead, MCN–BON) is equivalent to a daily survival probability of

$$S = e^{-0.112} = 0.8940$$

or mortality of 0.106, not 0.112 as reported. A half-day has a survival probability of

$$S = e^{-0.112(0.5)} = 0.9455$$

or a mortality probability of 0.0545, not 0.056 as reported. The rest of the paragraph has similar problems and needs to be corrected. The reported values are vaguely close to the actual values only because

$$1 - e^{-x} \text{ B } x \text{ for } x \leq 0.10$$

in a Taylor series expansion.

- Figs. 2.1, 2.2, 2.7

  The symbolism for box and whisker plots is not universally consistent or known. Captions should explain the symbolism.

- Table 2.1

  Caption fails to indicate which models the results refer to.

- Table 2.2

  Omit because AIC are not comparable across different datasets.

- Tables 2.7-2.16

  Captions are inadequately described. Symbols for models are cryptic and need to be explained for clarity of interpretation.
  The selection of models examined is at times eccentric:

  1. Models may include an interaction term without one or both of the main effects included. Purpose of an interaction term is to modify the main effects; it is unclear what the interaction term means in the absence of the main effects.

  2. Higher-order polynomial terms are included in models without corresponding lower-order terms, which is not conventional in linear models; for example, squared term without the linear term.

  Wonder whether this nonconventional approach to modeling is a direct consequence of dropping factors that are perceived to have the wrong sign for the partial regression coefficient (see comment above).

- Fig. 2.17

  The 20-day curve should be eliminated because the model is extrapolated beyond the range of the data. Fig. 2.1 indicates water transit time in LGR-MCN rarely if ever reaches 20 days.

Review of Chapter 3 and Appendix D

Chapter 3 and Appendix D present results on SARs, TIRs, and D for wild and hatchery spring/summer Chinook salmon and steelhead. Point estimates are presented and, in many instances, 90% bootstrap confidence intervals. The point estimates of SAR, TIR, and D are summarized by the geometric mean. Comparisons are made across migration groups, rearing types, and years. The estimated values of SAR, TIR and D are compared to benchmarks (i.e., 2% and 4% for SAR, 1 for TIR and D).

Page 51 (lines 24-26, 33-36); page 63, Table 3.2; page 74, Table 3.4 – Hydrosystem survival and system survival

- "Hydrosystem survival" includes indirect mortality effects of hydrosystem, despite the ISAB's recommendation to stop focusing on latent mortality because of the inability to estimate indirect mortality effects of the hydrosystem.

- "Hydrosystem survival" and "system survival" can be >1, and so are not actual survivals. At the very least, both performance measures are misnamed, and should not be used for management discussions.

- No benchmarks or target values for hydrosystem survival or system survival are given for comparison to estimated values. No expected values are given. Without this information, it is impossible to use the estimated values of these performance measures for management.

- "Hydrosystem survival" is introduced on page 51, defined formally on pages 59-60, and then not used because it cannot be estimated. Instead, "system survival" is reported.

- 2001 has a value of system survival of 2.139 (Table 3.2), which is >1; very high "system survival" in a very low flow year, which generally had poor inriver survival ($\hat{S}_R$ = 0.25 for 2001 [Table D-31]). It is not clear how to interpret this reported result. This result suggests the way the report is estimating system survival is invalid.

- System survival is mostly >1 for wild steelhead (Table 3.4), again inconsistent with general knowledge.

- Values of system survival are not given for hatchery Chinook salmon and steelhead.

Page 51, lines 30-42

The assumption of no density-dependent mortality needs support and should be included here. It has been hypothesized that one way in which hatchery fish negatively impact wild fish is through density-dependent mortality in estuary and nearshore ocean environments, by attracting more predators and competition for resources (food, shelter).

Pages 61-79

- 90% confidence intervals on some (but not all) performance measures were found using bootstrap methods. It is commendable that confidence intervals were computed for the performance measures, because it is impossible to interpret point estimates alone. However, it has been found (Lowther 2002) that bootstrap confidence intervals are not superior to theoretical normal theory confidence intervals arising from mark-recapture data analyzed with the Cormack-Jolly-Seber (CJS) model.

- Report all performance measures with confidence intervals, including:

  - The geometric means of the observed SARs, TIRs, or D values over the years of the study.

- Annual “system survival” estimates.

- Annual extrapolated estimates of inriver survival (S_R) from LGR to BON (Tables D-21 to D-28).

- Bootstrap confidence intervals do not easily yield confidence intervals or standard errors on performance measures that are functions of other parameters. Rather than report measures without some accompanying measure of uncertainty, standard errors or confidence intervals should have been computed in some way:

  - Geometric mean SAR, TIR, or D: A standard error for a geometric mean can be easily derived, assuming $\ln x_i$, nominally distributed, and using the expression for a geometric mean of

    $$\overline{x}_{\mathrm{GM}} = e^{\left\{\frac{1}{n}\sum_{i=1}^{n}\ln x_i\right\}}.$$

    Otherwise, arithmetic means should be reported for they always provide an estimate of expected value.

  - Extrapolated (“expanded”) S_R , inriver survival from LGR to BON: The extrapolated S_R is a function of CJS survival estimates and river km, and a standard error for $\hat{S}_R$ could easily be found using the delta method and CJS-based variances and covariances.

  - System survival, defined in terms of inriver survivals and project-specific D: A bootstrap confidence interval could be found for system survival but would require computing system survival for each bootstrap iterate, as was apparently done for TIR and D.

  Again, standard errors or confidence levels should be computed for all performance measures and included in this report.

- Significant differences in point estimates are incorrectly based on non-overlapping 90% bootstrap confidence intervals.

  - It is possible that two estimates with overlapping confidence intervals are statistically significant.

  - One reason is that confidence intervals ignore the possible correlation between the measures being compared, e.g.,

    - SAR(T0) and SAR(C0) are correlated for a single data set, because both T0 and C0 are estimated using the same CJS parameter estimates

  - Another reason is unequal variances.

  - Even if overlapping confidence intervals were an appropriate gauge of statistical significance for SAR(T0) and SAR(C0), this method focuses on the *difference* between SAR(T0) and SAR(C0), whereas the appropriate measure is their *ratio*, or TIR.

  Therefore, determination of significant differences should be recalculated based on formal statistical tests, not on whether confidence intervals overlap.

Page 54; lines 34-41, 58-59

- S_R , inriver survival from LGR to BON, is extrapolated ("expanded") to BON on a per-km basis in cases where it is not possible to estimate it directly using the CJS model. This is reasonable, but it should be recognized that this is extrapolation past the available data, not simply an "expansion."

- It is unclear if other methods of extrapolation were considered, such as pre-project, and if the goodness-of-fit of the extrapolation was considered. [Should these be considered for this report?]

- Again, confidence intervals or standards errors need to be calculated and included in this report.

  - Confidence intervals are not shown on any estimate of S_R , extrapolated or not, in Figs. 3.8 (p. 68) to 3.18 (p. 77).

  - No measure of uncertainty (e.g., standard error or confidence interval) is provided for the extrapolated S_R point estimates (Tables D-21 to D-28).

Pages 58 and 63, lines 20-22

- Steelhead jacks are included in SARs, but not Chinook jacks, because

  - Steelhead jacks have a fairly stable rate of return.

  - Chinook jacks have a variable return rate.

  Removing jacks from the analysis because of their questionable contribution to spawning is understandable, but not because of a "highly variable jack return rate" (p. 63).

Page 58, lines 16-18

- It appears that conclusions (about D averaging 0.5 for hatchery and wild Chinook "in recent years") are being presented prematurely and inaccurately in the methods section. These statements should be removed from the methods and included, with corrections and justification, in the discussions section.

- Based on CSS estimates of D for hatchery Chinook reported in Tables D-22 through D-26, only 3 of 36 point estimates for D were at 50% or less.

Page 58, lines 26-34

- The measures S_R and S_T are called "hydrosystem survival," but these are *not* the hydrosystem survival described on pages 51, 59, 60. Please explain or clarify.

Page 59, lines 13-21

- Measures SAR_{T1}, SAR_{T2}, SAR_{T3}, SAR_{C1}, SAR_{C2}, SAR_{C3} need to be defined formally using both verbal and mathematical expressions.

- New notation is used and needs to be defined; C1, C2, C3, T1, T2, T3.
    - C1 is previously used a different context, apparently.

Pages 61-78; Appendix D

- Neither the actual numbers of tagged smolts transported from each dam nor the sample sizes used in the analyses are reported. This information needs to be included for a complete and accurate peer review.
- It is unreasonable to assess the effectiveness of transportation based on small transport groups, even if they are augmented by the LGR equivalent approach.
- Present project-specific TIRs; they are used in estimation of annual SAR in the body of the report but are not specified.

Page 61 (lines 26-28); page 68 (lines 14-15)

- 2004 returns are incomplete, so it is unreasonable to say that 2004 SAR is “low” at this point.

Page 61, lines 30-32

- The ISAB review of the 2005 CSS Annual Report referred to the NOAA finding that PIT-tagged survival < untagged survival. If the NOAA finding is true, then comparing SARs from PIT-tagged fish to target values is unreasonable unless we know the size of the bias introduced by tagging or tag loss.

Page 62 (line 13); Table D21

- The geometric mean is used to summarize point estimates of SAR, TIR, and D across years.
    - Use of the geometric mean needs justification, especially considering past criticism and the fact that the geometric mean will always be lower than the arithmetic mean.
    - Standard errors or confidence intervals need to be reported for the geometric mean (see earlier comment and suggestion).
    - Low precision on D and TIR casts doubt on conclusions based on the geometric mean, especially those based only on a point estimate.
    - The geometric mean inherently dampens the effect of extreme values, so the policy of excluding 2001 values from the geometric mean needs further justification.

Page 66, 70, D-17 (Tables D-29, D-30) – Annual SAR

- Annual SAR is discussed often and is described in words, but is never defined formally. An equation is needed to see exactly how the various components are incorporated.
- Annual SAR values should be reported in a table for all species and stocks, with confidence intervals or standard errors.

- It would be useful to compare the annual SAR values to a simple ratio of the number of adults at LGR divided by the number of juveniles at LGR.

- Tables D-29 and D-30

  - These tables should be explained clearly in the text, using precise equations and clear definitions of notation.

  - It is unclear how the values reported here were defined, estimated, and used to compute the annual SARs.

  - It is unclear what the S's mean, and what reaches they apply to.

  - It is unclear where the covariances come from.

  - No comparable tables were provided for hatchery fish.

Page 67, Figure 3.7; Page 70, last paragraph

- Figure 3.7 shows that the trend in SAR for wild fish over two- or three-year time periods mimics the trend in SAR for certain hatchery stocks. However, Fig. 3.7 also shows that SAR for wild fish did not closely track SARs for any single hatchery throughout the entire time period considered. It is therefore uncertain which single hatchery could be used as a surrogate for wild fish in future years.

- No error bars are provided on Fig/ 3.7.

Figures 3.7, 3.8, 3.9, 3.10, 3.13, 3.14, 3.15, 3.18, 3.19, 3.20 (pp. 67-78)

- Need confidence intervals or standard errors on these graphs.

Chapter 4 Review

Chapter 4 attempts to estimate environmental stochasticity in SARs, TIRs, and D by removing variability in estimates due to measurement error. Parametric methods based on beta-binomial random variables and the lognormal distribution are used. Beta and lognormal probability distributions meant to describe variability in SAR, TIR, and D due to environmental stochasticity are presented.

Page 81

- 1st paragraph – Estimates of SARs are also indicators of inriver conditions, fish health, ocean conditions, and harvest survival.

- 2nd paragraph – Opportunistic sampling of fish, more than increasing variance, may result in biased estimates.

- 3rd paragraph – As the SARs are calculated in Chapter 3, they certainly do not have a binomial sampling variance, for both the numerator and denominator (i.e., C_0 fish) are estimated random variables. For a binomial variance to be true, the denominator of the SAR would have be known without error.

- 4th paragraph – The belief that there is a single probability distribution of SAR, TIR, or D over a long time period assumes that there is no temporal trend in the measures, such as may be caused by global climate change. Chapter 3's focus on trends in these measures suggests an assumption that the measures are changing over time, which is inconsistent with the assumption that they arise from a single beta distribution.

Page 82, lines 15-17

- Akcakaya (2002) is cited as a foundation for the method used to remove sampling variance to estimate environmental variance. The method presented in Akcakaya (2002) is appropriate for census data, but not for mark-recapture data, such as the data analyzed in this report. Akcakaya (2002) refers to both Burnham et al. (1987) and Gould and Nichols (1998) for variance-components method of removing sampling variance from mark-recapture data (see below, comment on pages 82-87).

- The methods used in this chapter are not clearly presented, either in the chapter or elsewhere in the report, despite the ISAB request that the report present all methods. They are presented verbally, but not mathematically.

Page 86, lines 19-22

- Equation(4.4) for the variance of a product applies only for independent random variables. This equation cannot be used to calculate the variance of a product of inriver survivals over adjacent reaches (i.e., S_R), because these survival estimates are correlated as based on the CJS model. Instead, the delta method (Seber 1982:7-9) should be used.

Page 82-87

- Kendall's (1998) method is a parametric approach to variance component estimation that makes unnecessarily restrictive assumptions, i.e.,

    a. Measurement error is binomially distributed.

b. SARs are beta-distributed.

Extension of the method to include log-normal distributions is also unnecessarily restrictive.

- Using the conditional variance formula

$$\mathrm{Var}\left(\hat{\theta}_i\right) = \mathrm{Var}_2\left[E_1\left(\hat{\theta}_i \middle| 2\right)\right] + E_2\left[\mathrm{Var}_1\left(\hat{\theta}_i \middle| 2\right)\right]$$

where

1 = sampling stage where $\hat{\theta}_i$ estimates θ_i,

2 = sampling stage where θ_i is a random sampling from the population values of θ,

then

$$\mathrm{Var}\left(\hat{\theta}_i\right) = \mathrm{Var}_2\left[\theta_i\right] + E_2\left[\mathrm{Var}\left(\hat{\theta}_i \middle| \theta_i\right)\right]$$

$$= \sigma_{\theta_i}^2 + \overline{\mathrm{Var}\left(\hat{\theta}_i \middle| \theta_i\right)} \qquad (1)$$

and where

$\sigma_{\theta_i}^2$ = natural variance in θ_i,

$\overline{\mathrm{Var}\left(\hat{\theta}_i \middle| \theta_i\right)}$ = average measurement error.

Using Eq. (1) and the method-of-moments, where

$$E\left(s_{\hat{\theta}_i}^2\right) = \sigma_{\theta_i}^2 + \overline{\mathrm{Var}\left(\hat{\theta}_i \middle| \theta_i\right)},$$

then an estimate of natural variability can be calculated as follows:

$$\hat{\sigma}_{\theta_i}^2 = s_{\hat{\theta}_i}^2 - \overline{\hat{\mathrm{V}}\mathrm{ar}\left(\hat{\theta}_i \middle| \theta_i\right)} \qquad (2)$$

where

$$s_{\hat{\theta}_i}^2 = \frac{\sum_{i=1}^{n}\left(\hat{\theta}_i - \hat{\bar{\theta}}\right)^2}{n-1}, \qquad (3)$$

$$\overline{\hat{\mathrm{V}}\mathrm{ar}\left(\hat{\theta}_i \middle| \theta_i\right)} = \frac{\sum_{i=1}^{n}\hat{\mathrm{V}}\mathrm{ar}\left(\hat{\theta}_i \middle| \theta_i\right)}{n}, \qquad (4)$$

$$\hat{\bar{\theta}} = \frac{\sum_{i=1}^{n}\hat{\theta}_i}{n}.$$

In other words, you can estimate the natural variance in responses $\left(\sigma_{\theta_i}^2\right)$ such as SARs, TIRs, or D based on the empirical variance among the replicate values [Eq. (3)] and average measurement error [Eq. (4)] without any distributional assumptions whatsoever. The only assumptions are:

1. $\hat{\theta}_i$ is an unbiased estimator of θ_i.

2. $\hat{\mathrm{V}}\mathrm{ar}\left(\hat{\theta}_i \middle| \theta_i\right)$ is an unbiased estimator of sampling error.

3. A random sample of the population of inference.

In the case where seasonal trends exist as indicated in travel times (Figs. 2,3-2.8), regression can be used to describe the pattern, leaving the error mean square (MSE) as an estimate of total variability [Eq. (3)]. This MSE can then be partitioned into natural variation about the trend and measurement error.

- There are several implications of the parametric approach taken to variance component estimation used in the CSS report, including the following:

1. Incorrectly using a binomial variance for the measurement error of the SARs will underestimate that component and overestimate natural variation $\left(\sigma^2\right)$.

2. The CSS report neglects to present the exact form of the beta distribution used, and there is an entire family of beta distributions to choose from. In the typical beta distribution, the means and variances are as follows:

$$\mu = \frac{\alpha}{(\alpha + \beta)} \qquad (5)$$

with a variance of

$$\sigma^2 = \frac{\alpha\beta(\alpha + \beta + 1)}{(a + \beta)^2}. \qquad (6)$$

If the CSS approach is correct, then the values $\alpha/(\alpha + \beta)$ for the SARs in Table 4.1 should be very close to the average SAR values across years. Unfortunately, the exact parameter estimates used in their calculations is not provided in the report. Such critical information and evaluation of assumptions need to be included in this report.

3. Similarly, if the fitted beta distributions are adequate, the beta variance (6) should reasonably approximate the nonparametric estimates of Eq. (2) and should be compared. Again, this critical information and analysis are not presented in this report.

4. The assumptions that SARs are beta-distributed are critical to the inference concerning the frequency of events. A goodness-of-fit test to the beta distribution needs to be performed using, for example, a Kolmogorov-Smirnov test, to verify the assumptions.

5. The use of the beta distribution to describe the frequency of SARs assumes the observed data are independent and identically distributed. However, this contradicts the results in Chapter 2, where the inriver survival, which contributes to the overall SARs, was found to be significantly correlated with environmental factors (e.g., Table 2.1). In other words, annual conditions influence the values of SARs for different stocks. The beta distribution ignores that previous set of findings and ignores the expected distribution of environmental conditions in the past or possible future. This should include projecting the possible consequences of global warming on inriver conditions and subsequent SARs.

Page 88, Table 4.1

- The estimated demographic variance is greater than total variance, suggesting something is wrong and thus casting doubt on all methods and results in this chapter.

- Observed correlations between point estimates of SAR for transport and C_0 groups for wild steelhead are explained by small transport groups and so are not used. However, such small transport groups (we are not told the actual sizes) produce unreliable parameter estimates that can seriously distort interpretation of results.

Chapter 5 Review

Chapter 5 compares estimates of annual SARs to target values indicated by the Northwest Power and Conservation Council (NPCC) (2003), and historical SARs based on run reconstruction methods. Multiple regressions are reported, relating Chinook salmon SAR to environmental variables. Upstream-downstream comparisons are made between Snake River Chinook and Chinook salmon from the John Day River. Biological comparisons between Snake River and John Day River Chinook are reported.

<u>Introduction (pp. 105-106)</u>

Page 106, lines 11-22

- Critiques of the single release-recapture (SR) analysis and PATH have demonstrated the reliance of latent mortality results on untestable assumptions, e.g., stock-specific Ricker *a*'s versus a common Ricker *a*. Additionally, climate effects have been shown to account for the majority of latent mortality. These criticisms should be addressed in this chapter.

Page 106, lines 19-20

- It is not explained and it is unclear how direct mortality, differential delayed mortality of transported smolts, and the common year effect were accounted for in the SR comparisons.

<u>Methods: General (pp. 107-109)</u>
Page 107, line 26

- "Overall SAR" is never defined, either here or elsewhere in the report. Presumably it is equal to "annual SAR," which is also never defined analytically.

<u>Methods: Relationships between Chinook SAR and environmental covariates (pp. 110 - 115)</u>

Pages 110-111

- Run Reconstruction SARs: Include jacks and adults; measure returns to mouth of Columbia River.

- CSS SARs: Include only adults (Chinook), no jacks; Measure returns to LGR

- Are run-reconstruction SARs and CSS SARs really comparable? It has not been justified that direct comparison of the measures is appropriate.

Page 112, lines 15-19: How is WTT defined?

Page 114, lines 3, 9; Figures 5.5, 5.6

- What does "frequently incorporated in multiple regression models" mean?

Page 115, Multiple Factor Model, lines 5-31

- How were candidates for independent environmental covariates selected? What were they? Only WTT, PDO, and an upwelling index are named, and it is unclear whether other covariates were considered.

- Harvest and temperature are known to affect SARs and do not appear to have been considered.

- Were any other “inriver” predictors than WTT considered?

- Were interaction terms considered in the multiple regression models?

- Typo in SAS version (presumably 9.1, not 91).

<u>Methods: Snake River and Downriver SAR Comparison (pp. 115-119)</u>

- There has been much previous criticism of the upriver-downriver comparisons made by the CSS and of the spawner-recruit model used to justify the upriver-downriver comparisons. Insufficient response has been made to these criticisms.

Page 116

- Lines 7-8

  - How is μ_t defined and estimated? Provide an equation showing how value is calculated. Is this the same μ as in Eq. 5.3, or is it the differential mortality defined verbally based on Eq. 5.2?

  - The “delta model” should be defined.

- Equation 5.3: If there is no delayed mortality from hydrosystem, then we expect $\exp\left(-\mu_{SAR,t}\right) = S_{J(\text{LGR-JD})}$. This important point is omitted from the report.

Page 117

- Line 18: Only a single hatchery (Carson) is used for the downstream hatchery Chinook salmon.

Page 119, Table 5.9

- This table is very difficult to understand. The caption does not agree with the notation used in the table. Values reported in the table are not sufficiently explained. It appears that the formula used to estimate BOA detection efficiency (*p*) is wrong.

- What are the values reported in the row “GRA, MCA, IHA?”

- MCN and IHA are not mentioned in table caption.

- The estimate of *p* based on detections at BON and upstream is INVALID if it is based on detections from different years, unless upriver adult survival to GRA is constant across return years, and detection probabilities at MCN, IH, and GRA are constant across return years. This is not true, so estimates of BOA detection efficiency presented here are invalid.

- On a yearly basis, *p* should be estimated as,

$$\hat{\rho} = \frac{\text{No. detected at BON and upstream}}{\text{No. detected upstream of BON}}$$

(from Manly-Parr) or equivalently as

$$\hat{\rho} = \frac{\text{No. detected at BON}}{\left(\text{No. det. at BON}\right) + \left(\text{No. det. upstream, not at BON}\right) \times \left(\text{Survival from BON to upstream}\right)}$$

(from CJS model), but NOT as is estimated here:

$$\rho_{CSS} = \frac{\text{No. detected at BON}}{\left(\text{No. det. at BON}\right) + \left(\text{No. det. upstream, not at BON}\right)}$$

The estimates of *p* as reported by the CSS will be positively biased, i.e., too large.

<u>Methods: Comparison of biological characteristics of Snake River and downriver smolts (pp. 119-121)</u>

Page 120

- In general for upstream/downstream comparisons, was goodness-of-fit considered or examination of residuals performed? Show results.

- With only 6 years of data, this is not a long time series, which limits the amount of useful information that can be gleaned from it.

Page 121

- Lines 13-14: No migration distance is given for JDAR1 fish. Comparison of survival and travel time between upstream and downstream fish should incorporate migration distance for the two groups of fish.

- Lines 40-41: Basing analyses on (Number of BON detects/Number released at trap) assumes that all groups have the same conditional detection probability at BON. This is likely to change with arrival timing.

<u>Results: Overall SARs (pp. 122-127)</u>

Page 122, lines 32-34

- "Removing sampling variability" resulted in lower mean SAR. Does this always occur?

Page 126, lines 17-19

- The CSS has been using a geometric mean previously, but here does not identify the type of mean used for mean SAR.

- It is unclear what the reference to the *t*-distribution means. If a formal *t*-test is being performed, this should be stated simply. Note that while these arithmetic means may be compared using a *t*-distribution, the geometric mean should not.

<u>Results: Relationship between SAR and environmental covariates (pp. 128-131)</u>

Page 128

- Lines 4-8: The data for the PIT-tag SARs and environmental factors are not presented in this report.

- In general, references to figures should be proofread. There are mistakes in figure references throughout the chapter, making it difficult to follow the narrative.

- Line 11: What is meant by "bi-variate results?" Is this regression of a single response variable on a single predictor variable? A vector response variable on one or more predictor variables? A single response variable on two or more predictor variables?

- Table 5.4: Did the CSS consider correlation between PDO and UP45n? Both types of measures are used in the same regressions, apparently.

Page 129, Table 5.5

- It should be explained why SepPDO is used rather than JulyPDO as a covariate, when JulyPDO looks better than SepPDO for both the long and current time series.

Page 130

- In general for regression with environmental variables:
  - What was the set of candidate predictor variables? Was it only PDOs, UP45ns, and WTT?
  - How model selection was performed needs to be specified?

- Lines 13-14 – The report says that WTT was "less significant for the shorter time series," but Table 5.7 indicates that WTT was not at all significant if the model includes upwelling index (Table 5.7, Current Time Series).

Page 131, lines 1-2, and Table 5.7, Current time series

- What model selection criterion were used to identify the "best" model?

- The "best" model shows no predictor variable significant at the 10% level when upwelling index is included.

- Without upwelling index (NovUP45n), SNWTT and MayPDO become significant. Was multicollinearity between these parameters and NovUP45n considered? And how?

<u>Results: Snake River and Downriver SAR Comparisons (pp. 131-136)</u>

- The CSS upstream-downstream comparison of SARs is based on the performance measure $U/D = \frac{S_{LGR-BON}}{S_{JD-BON}}$. If there is no differential post-JD mortality for upstream fish, then we expect U/D to equal S_{LGR-JD}, inriver smolt survival from LGR to JD.

- The CSS also reports values of $\mu_{SAR} = -\ln(U/D)$ for wild Chinook, although not for hatchery Chinook salmon. There is no benchmark for μ_{SAR}, however, because it compares the SAR from LGR to BON for upriver stocks to the SAR from JD to BON for downriver stocks.

- Interpretation of both U/D and μ_{SAR} estimates depends on the inriver survival of upriver stocks from LGR to JD, which is never considered by the author.

- Tables 5.8 (p. 132), 5.9 (p. 133), and 5.10 (p. 135), and Figure 5.16 (p. 136) cannot be usefully interpreted as they are, because they do not compare the reported measures to S_{LGR-JD}. Figure 5.16, showing the pattern of μ_{SAR} across years for wild and hatchery Chinook salmon, demonstrates the variation in μ_{SAR} across stock. Without also showing S_{LGR-JD} across stock, however, it is impossible to reach any conclusions.

- Table 1 (below) shows CSS estimates of U/D and μ_{SAR} taken from Tables 5.8-5.10, and compares them to estimates of S_{LGR-JD} calculated from Tables D-31 through D-36. Using the criterion of $S_{LGR-JD} > U/D$, or equivalently, $\exp[-\mu_{SAR}] < S_{LGR-JD}$, it is determined whether upstream stocks had lower SARs from JD to BON (SAR_{JD-BON}) than downstream stocks (referred to as "differential mortality" by the CSS).

  - In 4 of 5 years, wild Chinook upstream stocks showed lower SAR_{JD-BON} than wild Chinook stocks from John Day River (i.e., differential mortality).
  - Rapid River Hatchery spring Chinook salmon showed no differential mortality.
  - Dworshak Hatchery spring Chinook salmon showed differential mortality in 3 of 5 years.
  - Catherine Creek Acclimation Pond spring Chinook salmon showed differential mortality in only 1 of 4 years.
  - McCall Hatchery summer Chinook salmon showed no differential mortality in 5 years.
  - Imnaha Acclimation Pond summer Chinook salmon showed differential mortality in only 1 of 5 years.
  - In some years, the U/D measure is considerably greater than S_{LGR-JD}, such as Rapid River spring Chinook salmon for 2003, when U/D was estimated at 1.21 and S_{LGR-JD} was estimated at 0.502. There are similar

examples from most hatchery Chinook stocks, in which U/D is estimated to be greater than 1, and S_{LGR-JD} is estimated to be less than 1. The CSS report does not address this situation, and gives no indication how U/D > 1 should be interpreted. In these cases, upstream stocks had higher SARs to BON, whether from LGR or from JD. If we were to follow the CSS's example, we must conclude that passage through the hydrosystem *improves* survival for many upstream hatchery stocks.

– It is obvious from Table 1 that the value of U/D (and by extension, μ_{SAR}) alone does not indicate whether or not "differential mortality" has occurred.

Table 1. Comparison of $U/D = (SAR_U)/(SAR_D)$ from Tables 5.8 (p. 132) and 5.10 in the CSS report to estimated inriver smolt survival from LGR to JD, calculated from Tables D-31 to D-36. If S_{LGR-JD} is greater than U/D, then upstream fish had a lower SAR than downstream fish from JD to BON.

Stock	Year	$\frac{SAR_U}{SAR_D}$	S_{LGR-JD}	SAR (JD-BON) lower for upstream stock?
Wild Chinook	2000	0.24	0.622	Yes
	2001	0.47	0.377	No
	2002	0.31	0.704	Yes
	2003	0.12	0.693	Yes
	2004	0.15	0.542	Yes
RAPH Sp Chinook	2000	0.79	0.741	No
	2001	0.76	0.529	No
	2002	0.83	0.745	No
	2003	1.21	0.502	No
	2004	0.50	0.508	No
DWOR Sp Chinook	2000	0.46	0.658	Yes
	2001	0.24	0.380	Yes
	2002	0.59	0.676	Yes
	2003	1.11	0.683	No
	2004	0.63	0.583	No
CATH Sp Chinook	2001	0.20	0.389	Yes
	2002	0.87	0.721	No
	2003	1.25	0.694	No
	2004	0.66	0.570	No
MCCA Su Chinook	2000	1.09	1.07	No
	2001	0.81	0.399	No
	2002	1.35	0.840	No
	2003	2.85	0.749	No
	2004	0.69	0.627	No
IMNA Su Chinook	2000	1.05	0.655	No
	2001	0.45	0.547	Yes
	2002	0.73	0.640	No
	2003	2.50	0.765	No
	2004	0.78	0.642	No

Page 132, Table 5.8

- How are the SARs for downriver wild Chinook salmon estimated? If simple return ratio, why not use same method for Snake River fish?

Page 134

- Lines 2-5: The CSS claims that the SAR to BON is *always* higher for the downriver (hatchery) fish, but that is not true for 2003.

- Lines 13-16: The reason given for not providing a confidence interval on SAR for downriver fish in 2004 is because an average survival to BON from previous years is used. However, that survival is not known without error, so a measure of uncertainty should be reported on survival to BON for 2004, and that error could be propagated to produce a CI on SARs.

- In general, the CSS addresses uncertainties incorporated by using a single downstream hatchery stock when the upstream/downstream results show no effect of the hydrosystem (i.e., for hatchery Chinook salmon), but not when the upstream/downstream results *do* imply hydrosystem effects (i.e., for wild Chinook salmon). This sounds like an inconsistent approach.

Page 135, Table 5.10

- In some years, upriver SAR > downriver SAR for hatchery Chinook salmon, despite additional inriver migration for upriver fish. Presumably, this result is unexpected and should be addressed. Such results may be due to large measurement error that obscures the relationship or the upstream/downstream pairing is a mismatch.

<u>Results: Comparison of Biological Characteristics of Snake River and downriver smolts (pp. 136-143)</u>

Page 136, Figure 5.16 - Needs confidence intervals.

Page 139

- Lines 2-3 It says that there is a significant ($P < 0.001$) difference in density-adjusted mean fork lengths of 106 and 106 mm (for IMNTRP and JDAR1), and separately of 100 and 100 mm (for SALTRP and SNKTRP).

- Lines 6-7**:** The report is inconsistent when it says 74 mm vs. 121 mm in fork length is not significant, especially considering that they previously defined any differences >5 mm to be biologically significant.

Page 141, lines 11-13

- "Smolts from upriver populations and downriver-origin smolts migrated at a similar rate, once their different migration distances were accounted for." What does this mean? Their migration "rate" (i.e., distance traveled per unit time) already accounts for differing migration distances.

Page 142

- Lines 15-17 – The observation that upriver smolts took longer to travel to BON than downriver smolts is not surprising since they leave at the same time and travel at the same rate, given that upriver smolts have farther to travel.

Results: SARs by Bonneville Arrival Timing (pp. 144-146)

Page 143, lines 1-2

- The "pattern of delayed arrival" was *not* consistent across years, as is stated – See years 2000 and 2003.

Page 144

- Lines 11-14 – What reference point is used to determine that upstream smolts experience delayed migration?

- Lines 16-18 – What does "significantly experienced lower SARs" mean? Does this mean that the difference or ratio between the SAR for wild upstream Chinook and wild downstream Chinook was *statistically* significant? *Biologically* significant?

Page 145, Figure 5.23

- Binomial confidence intervals are shown, but error is not binomial for C0, C1, and T0. Recalculate appropriately.

- In some years, large numbers of upriver migrants are omitted from the analysis by restricting attention to 16 April – 31 May window.

PIT-tag SARs versus SAR of run-at-large (p. 147)

- Lines 3-5: Are run-reconstruction SARs and CSS SARs mathematically comparable? Justify.

- Lines 12-19: Assumptions necessary for the run-reconstruction SARs are discussed, but not assumptions for the CSS SARs.

Discussion (pp. 148-151)

Page 148

- Lines 21-22: The limitations of small sample size cannot be avoided by using multi-year methods, as indicated here. Multi-year methods result in conclusions that are based on many uncertain estimates (due to small yearly sample sizes), instead of based on only a single uncertain estimate. This simply expands the problem of small sample size.

- Lines 29-31: WTT is named the "best" predictor variable for SARs, but it is not clear that the CSS considered other inriver covariates.

- Lines 37-38: It was found here that WTT influences the smolt migration rate. But JDAR1 and Snake fish have similar migration rates. Did they have different WTT? This needs to be addressed.

- Lines 42-43: SARs of downriver fish are compared to SARs from upriver fish, but these SARs are estimated over DIFFERENT reaches and distances, so we *expect* them to be different. The CSS needs to investigate whether the differences are more than expected.

Page 149

- Lines 39-40: "Hydrosystem migration rates did not differ between groups but were strongly influenced by water travel time." It is not clear how to interpret this statement. Did groups have different water travel times but the same migration rate? Or did they have the same WTT? Or was migration rate and travel time examined on an individual fish basis, instead of a group basis? It is not clear.

- Lines 41-46: Distance to travel is not considered as a factor of travel time.

Page 150

- lines 1-6 – It is claimed that the "potential confounding effects due to life history differences are probably negligible," but the CSS does not attempt to model SAR using both the upstream/downstream designation and the life history differences. Additionally, the effect of distance to travel was ignored. A model that includes all possible factors affecting SAR should be considered, in order to claim that it is the hydrosystem rather than other factors that cause the difference in return rates.

<u>In general for Chapter 5</u>

- In order to determine if there is a biological difference that explains any differences in SAR between upriver and downriver stocks, model SAR using fork length, migration date, arrival timing, year, in addition to upstream/downstream classification. Is upstream/downstream effect significant, given presence of all others?

- Looking at population differences in fork length, migration date, etc., one at a time, is reasonable for initial data exploration, but insufficient for conclusions about the significance of the upstream/downstream effect.

<u>Throughout Chapter 5</u>

- Typos are made in references to tables and figures throughout the entirety of Chapter 5.

- Pages 139–144: The reader is referred to a nonexistent figure for release site abbreviations.

Chapter 6 Review

Page 154, line 9

- The notation RY has not been defined. The context suggests Return Year, but Release Year is also a possibility.

Page 154, Tables 6.1 and 6.2

- Pooling migration success data across migration year and return year is valid only if those factors are nonsignificant. Perform test of homogeneity.

- Also applies to Page 155 (lines 17-23); Page 156 (Table 6.3).

Page 155, line 41

- Was return year modeled as a fixed or random effect? Most blocking factors are modeled as random effects, although there are times when a fixed effect is more appropriate.

Page 156, Table 6.3

- Chi-squared tests indicate whether there is a difference in perceived upriver adult survival across juvenile migration groups, but they do not indicate the nature of the difference. The p-values reported do not indicate that the actual ranking in the Success Rate Ranking column is significant, simply that at least one of the juvenile migration groups had a significantly different success (survival) rate than the others. One-sided tests should be performed comparing pairs of juvenile migration groups in order to test the significance of the ranking.

Page 157, Figure 6.1

- Needs error bars or confidence intervals.

Page 158, Figure 6.2

- Needs error bars or confidence intervals.

- The interpretation of Fig. 6.2, showing the proportion of LGR-detected adults and jacks detected at hatcheries, depends on the detection effort at each hatchery in each year. Without that information, it is useful only for comparing transported to inriver fish. It appears that transported fish had slightly better survival from LGR to the hatcheries, but without error bars and without information about detection effort (and harvest pressures, etc.), no real conclusion can be reached from Fig. 6.2.

Page 159, lines 11-15

- The overall average perceived BON-LGR adult survivals for the three migration groups are not very useful without standard errors or confidence intervals.

- It is not clear how these average survivals were computed. Were yearly estimates weighted by the number of fish returning in each year? Or were migration year estimates averaged?

- Given the finding that return year is a significant factor in perceived upriver adult survival (from the logistic regressions presented later in this chapter), pooling data over return year is not warranted.

Page 160

- Lines 16-19: The model evidence ratio does not indicate that one model is "more likely" than another, in a Bayesian sense. Rather, it means that there is more evidence for one model compared to the others.
  - Also applies to results for wild Chinook salmon (p. 161, lines 22-23).
  - The highest evidence ratio for the best model for wild Chinook salmon (p. 161, lines 21-25; p. 162, Table 6.6) is at most 4, thus there is *not* clear evidence that transportation is an important factor in determining adult migration success when compared to environmental factors.
- Lines 29, 32: It is not clear how the confidence intervals on the odds ratios are computed. Provide explanation. Asymptotic normal-theory confidence intervals are considerably narrower than those reported, and do not include 1 for either LGR-transport fish or LGSdown fish. If the confidence intervals were based on a *t*-distribution, the degrees of freedom should be reported (Table 6.5).

Page 161

- Table 6.5
  - Degrees of freedom should be reported for each parameter estimate.
  - Surprisingly, warmer temperatures were associated with higher perceived adult survival. Perhaps temperature is confounded with run (spring versus summer).
- Lines 29-30: The odds ratio is misinterpreted here. An odds ratio of 0.5 does *not* mean that the probability of success of LGR-transport fish is half that of inriver fish. If the probability of success (i.e., perceived adult survival from BON to LGR) is P_{LGR} for LGR-transport fish, and is $P_{inriver}$ for inriver fish, then:

$$\text{Odds ratio} = \frac{\left(\dfrac{P_{LGR}}{1-P_{LGR}}\right)}{\left(\dfrac{P_{inriver}}{1-P_{inriver}}\right)} = \frac{1}{2}$$

$$\Rightarrow P_{LGR} = \frac{P_{inriver}}{2-P_{inriver}}$$

This means that the probability of success of LGR-transport fish depends on the value of the success probability for inriver fish, as demonstrated in Table 2 below. Table 2 indicates that for an odds ratio of 0.5, the probability of success of LGR-transport fish is generally greater than half that of inriver fish, except for very small inriver success probabilities, which are not applicable here.

Table 2. The probability of adult migration success (BON to LGR) for inriver fish and LGR-transport fish for an odds ratio of 0.5.

$P_{inriver}$	P_{LGR}	$\frac{P_{LGR}}{P_{inriver}}$
0.1	0.05	0.53
0.25	0.14	0.57
0.33	0.2	0.6
0.5	0.33	0.67
0.75	0.6	0.8
0.9	0.82	0.91
1	1	1

Pages 162-163, Hatchery Chinook arrival and travel time ANOVAs

- For both arrival time and travel time, the interaction term between return year and juvenile migration (outmigration) method was significant. This affects interpretation of the main effects of both return year and outmigration method, so conclusions based on the main effects alone are invalid.

- The ANOVAs should be included in the report.

Page 164, lines 38-40

- How much of TIR or D is explained by observed differences in perceived upriver adult survival between inriver and LGR-transport fish?

Chapter 7 Review

Page 168; lines 31, 46

- Survival from release to LGR = 0.95 seems high, and does not correspond to year 2000 data used as basis for default values

- SAR = 0.03 seems high, given that observed SAR has been lower than the target value of 2% in most years according to this report.

Page 169, line 9

- Joint probability of survival from BON to TWX and detection at TWX = 0.10 is high, based on past years.

Page 170 (lines 9-11, 21-23, 39-41), Page 171 (lines 8-10)

- The survival probabilities used in simulation scenarios #5, 7, 10, and 12 include inriver survival probabilities > 1, when the variable *day* is 0 or very low. Inriver survival should be parameterized using only admissible parameter values (i.e., ≤ 1) and included in this report.

Page 171, last paragraph, and Page 174

- The emphasis on the T and R groups is confusing. The underlying cause of the assumption violation is not that R fish have higher or lower inriver survival than T fish, but that detected fish have higher or lower inriver survival than non-detected fish. While understanding the effect on the T and R groups will be useful in the future, it is not clear how they apply to estimation of C0, C1, and T0 for previous years' data, in which T and R groups were not used. A simpler method of assessing the effect of detection-influenced survival would be to simulate data under the scenario described (Post-turbine survival < Post-bypass survival < Post-spill survival, with varying proportions of undetected fish passing via turbine or spill) and examine estimates of C0, C1, and T0.

Page 175, last paragraph

- It is not clear if the "true" survival parameters used to compute LGR equivalents are averages of seasonal survival parameters, or if LGR equivalents are computed on a daily basis and then summed over the season. Given the temporal variation in survival parameters introduced in these simulations, the latter approach should give a better representation of the "true" C0, C1, and T0 groups. Clarify approach and, if necessary, rerun simulations.

Page 194

- The CSS uses results of the second set of simulations to address how to best analyze data using the NPT approach, in which tagged fish are pre-assigned into migration groups: T (transport) fish are transported upon their first detection at a transport dam; R (river) fish are returned to river upon all detections. Using the C0/C1/T0 approach to analyze data with pre-assigned migration groups is not intuitive. It would be simpler and more defensible to simply compare the SAR of the T group to the SAR of the R group. All "R" fish will have migrated wholly inriver, while some "T" fish will have been transported and others (undetected) will have migrated inriver. The comparison of SAR(T) to SAR(R) is more easily interpreted for management, because the alternative to transportation is to return detected fish to the river, whereas the transportation alternative being tested in the SAR(T0) vs. SAR(C0) comparison is not clear.

Chapter 8 Review

Page 198

- Lines 35-38: The trend of performance measures for wild fish mimicked the overall trend of performance measures for the collection of hatcheries, but did not agree well with the trend from any single hatchery across all years. It is not clear which single hatchery could be used to make inference to wild fish. Also applies to Pages 199-200, bullet (b) of Chapter 5 summary.

- Report confidence intervals for results (e.g., geometric means).

Page 199

- Lines 32-35: The inference made from declining SAR(C1) over the season to hydrosystem-caused post-Bonneville mortality is unfounded. There are alternative possible causes of post-Bonneville mortality, including temperature, pollution, disease, and seasonal changes in estuary conditions. No conclusions about the relative importance of the various potential sources of mortality can be reached here.

- Lines 40-42: The CSS claims that Snake River wild steelhead SARs averaged less than 2%. It is difficult to confirm this statement, because the annual SARs are not presented in tabular form in this report. However, Fig. 3.12 suggests that average annual SAR for wild steelhead may be greater than 2%. Document annual SARs in the table and explain apparent inconsistency.

Page 200

- Lines 8-14: The CSS did not compare the ratio of upstream and downstream SARs to inriver survival between Lower Granite and John Day, so the conclusion that upstream fish experience extra mortality caused by the hydrosystem is unjustified. Also applies to Page 200 (lines 33-34).

- Lines 23-24: The claim is made that that the CSS shows clear evidence of delayed estuary entry of Snake River inriver smolts, caused by passage through the hydrosystem, on the basis of comparisons with John Day smolts. This is not true. The CSS found that Snake River and John Day smolts (1) initiate migration at the same times, and (2) migrate at similar rates through the first three dams passed. Given the extra distance traveled by the Snake River smolts, it is not surprising that Snake River smolts enter the estuary later than John Day smolts. The CSS analysis would be more useful if it had compared the observed and expected arrival dates of the Snake River fish, given their migration initiation date, migration rate (through the first three dams), and distance to travel.

- Lines 26-30: The conclusion that differing seasonal SARs for upstream versus downstream smolts is evidence of delayed mortality ignores possible alternative explanations, including potentially different ocean residencies.

- Lines 37-42: It appears here that wild and hatchery Chinook salmon transported from LGR always had 10% lower SAR than fish passing through the hydrosystem by alternative routes. It should be noted that the effect for hatchery fish (4% to 7%) was considerably less than the effect for wild fish (15%), so the 10% effect reported is somewhat misleading.

Page 202

- Lines 11-16; lines 39-41: The claim is made that the CSS addresses the question of whether smolt transportation compensates for effects of the Federal Columbia Power System (FCRPS) on survival of Snake River chinook and steelhead. This claim extrapolates past the available data. The CSS compares the SAR of transport fish to the SAR of fish migrating inriver. While the inriver fish experience effects of migrating through the FCRPS, available data do not indicate the magnitude of those effects; this would require comparing the SAR of fish migrating through the FCRPS to the SAR of fish migrating through the same reaches but not through the FCRPS. That is not possible. At most, the comparison of the SAR of transport fish to the SAR of inriver fish indicates whether transportation is a viable management option; it is *not* equivalent to comparing transportation to migration through the unimpounded river. It is worth noting that the SAR from BON to BOA for hatchery Chinook salmon from the John Day river was less than 2% for 2001 through 2004 (Table 5.10). Regardless of the validity of upstream-downstream comparisons, these low SARs for John Day fish suggest that the hydrosystem is not the only factor in below-target SARs.

- 3rd paragraph:

  - The CSS reports “mean” values for TIR for steelhead, failing to mention that these are geometric means. Typically, “mean” implies the arithmetic mean. Geometric means produce lower values than arithmetic means. Omitting 2001, the arithmetic mean of TIR for wild steelhead was 2.4, versus a geometric mean of 1.7; the arithmetic mean TIR for hatchery steelhead was 1.7, versus a geometric mean of 1.5.

  - The CSS says that D < 1.0 for steelhead. However, wild steelhead showed D values >1.0 in 5 of 7 years, with an arithmetic mean of 1.12 (including 2001). Thus, it appears that in most cases, D > 1.0 for steelhead. This inconsistency should be explained.

  - When TIR values are at 1.0 or greater, the CSS points out that D values are nevertheless less than 1.0. They do not discuss the implications of this. Even if D < 1, the decision to transport should be based on TIR values, not on D.

- Last paragraph: The CSS compares observed SAR estimates from PIT-tagged fish to the NPCC objectives for SAR (2% minimum, 4% average), without addressing the NOAA finding that PIT-tagged fish have lower survival than untagged fish (as requested by the ISAB). Without knowing the size of the PIT-tag bias, comparisons of PIT-tag SAR to target values are not completely useful.

Page 203, 3rd paragraph

- The CSS mentions that the decision of when to initiate transportation is an important management decision, and implies that this study fully addresses that question. While some estimation and analysis of seasonal TIR was done, it was hardly a complete analysis, and provides little management guidance.

- The CSS claims that seasonal TIRs “may contain some positive bias” because they are based on the C1 group (detected, not transported) rather than the C0 group (undetected). However, because the management alternative to transportation is to return bypassed fish to the river, the C1 group is more appropriate than the C0 group for comparison to transport SARs.

Appendix B Review

Page B-3, Figure 1

- The estimators of ϕ_1, ϕ_2, and ϕ_3 are correct.

- The figure is somewhat cryptic. The parameters ϕ_i are not defined, nor are the statistics R_i, $R'_{1\bullet 2}$, $R'_{12\bullet 3}$, ..., r_i, m_i. The reduced *m*-matrix is not so standard that the CSS should expect all readers to recognize and understand it without further explanation. Provide more detail.

Page B-4

- The CSS explains that they allow individual reach survival estimates exceeding 100% when computing an overall multi-reach survival estimate. Why, then, do they not allow $\hat{S}_{JD-BON} > 1$ for 2004 for Carson NFH Chinook, in Chapter 5?

- The verbal description of the weighted average of survival estimates provided in the second full paragraph is insufficient. An equation demonstrating precisely how the overall survival estimate was estimated is required.

- The CSS used weights equal to the inverse relative variance of the reach- and cohort-specific survival estimates. How were the variances of those reach and cohort survival estimates computed? How was the standard error on the weighted average survival computed? Provide details.

- In the final partial paragraph, the CSS discusses using a "per-mile" expansion of juvenile survival in cases where it was impossible to estimate survival to BON directly. Previously (Chapter 3), they used a per-km method of extrapolation. Either there or here, did they consider any other basis for extrapolation? Did they consider the goodness-of-fit of the extrapolation method used? Did they estimate the standard error on the survival estimate to BON, either with or without the extrapolation?

Page B-5

- The CSS lists the three ways in which fish can pass an individual transport dam, and indicates that these three passage routes describe the passage routes through the hydrosystem. However, their three passage routes must be combined over multiple dams to describe the possible passage routes through the entire hydrosystem. For example, there are seven possible passage routes through LGR, LGS, and LMN that result in transportation from one of those dams.:

  - Transportation at LGR (route 1)

  - Transportation at LGS following either detection at LGR, or non-detection at LGR (routes 2 and 3)

  - Transportation at LMN following either detection at both LGR and LGS, detection at only one of LGR and LGS, or non-detection at both LGR and LGS (routes 4-7)

  Thus, the CSS "partition" of PIT-tagged smolts arriving at LGR is, at best, unclear from their description and, at worst, potentially omitting considerable numbers of fish. Clarification in this report is required.

Page B-6

- #5. Is “run-at-large” equal to “untagged” here, or does it also include tagged fish? How is P_{t2} estimated? Is P_{t2} really the proportion of the entire run-at-large that were transported at LGR, or only the proportion of the run-at-large collected at LGR that were transported? Similar comments pertain to #7 and #9.

Page B-7

- #13 - #15: It is essential for the CSS to actually write out the expected values of the statistics T0, T0*, and C1 in terms of the underlying model (i.e., survival, detection, transportation, and removal parameters), rather than leaving them partially defined. This level of technical detail is essential for all readers to know exactly what is being estimated by the parameters in the report.

- #15, #16: The statistics d0 and d1 are never defined. The 50% survival probability is not explained—50% survival to where? On what basis is 50% chosen? Why not use the actual estimated survival probability to whatever site or sites are used?

Page B-8

- Finally, an attempt is made at an explanation for the 50% survival probability used to deal with downstream removals. It is not sufficient, however. Why not use a dam-specific adjustment, rather than pooling all downstream removals and assuming a common survival to every downstream dam? Have the effects of violations of this 50% survival assumption been examined? It is known that violations of this assumption occur, because survival between MCN, JD, and BON is not 100%, so survival to one dam (e.g., at 50%) is not equivalent to survival to the other dams, as is implicitly assumed by using a single survival probability to all downstream dams. Additionally, if using a single survival rate is warranted and if survival to BON is to be used each year, it should be possible to use the estimated survival to BON for the year, rather than assuming 50% survival each year.

- Show the number of removals on a dam-specific basis that you contribute to d0 and d1?

- “Estimation of SARs for study categories:” SAR1(T0) and SAR2(T0) have been discussed but not defined in this report. Define all measures.

Page B-9

- A “common annual routing rate to the raceways” was used—what is this? Is it known or estimated? What value was used?

- The notation used to define AC0 and AC1 is insufficient. It does not preclude using adults that were removed at downstream dams for any reason. Because many removed fish are not sacrificed, it is conceivable that some of these “removed” fish may return as adults. Are these adults included in AC0 and AC1? One assumes not, because this would positively bias the SAR for the C0 and C1 groups; however, the notation used implies that these removed fish are included in AC0 and AC1.

- It looks like SAR2(T0) is used in this report for overall SAR of transported fish, rather than SAR1(T0), unless otherwise specified. Is this correct? Clarify.

Page B-11

- The expected value of the size of the C1* group should be presented. At the least, the definition of the C1* group should be explained. It does not make intuitive sense to define it in terms of the T0, C1, and T0* statistics, because the T0 and T0* statistics are based on different groups of fish.

Page B-12

- The CSS states that "the rate of harvest is assumed independent of whether fish had been transported as smolts. [These] assumptions … apply to both TIR and *D*." Where does the CSS actually make use of this assumption? Is it only in their interpretation of results about TIR and D?

Literature Cited

Akcakaya, H.R. 2002. Estimating the variance of the survival rates and fecundities. Animal Conservation **5**: 333-336.

Burnham, K.P., and Anderson, D.R. 2002. Model selection and multimodel inference. Springer-Verlag, New York, New York, USA.

Burnham, K.P., Anderson, D.R., White, G.C., Brownie, C., and Pollock, K.H. 1987. Design and analysis methods for fish survival experiments based on release-recapture. American Fisheries Society Monographs **5**.

Gould, W.R., and Nichols, J.D. 1998. Estimation of temporal variability of survival in animal populations. Ecology **79**(7): 2531-2538.

Lowther, A.B. 2002. Development, expansion, and evaluation of release-recapture survival models for Snake River juvenile salmonids, with new algorithms allowing time-dependent individual covariates Dissertation, Quantitative Ecology and Resource Management, University of Washington, Seattle, Washington.

Neter, J., Kutner, M.H., Nachtsheim, C.J., and Wasserman, W. 1996. Applied linear statistical models. Irwin, Chicago, Illinois, USA.

Seber, G.A.F. 1982. The estimation of animal abundance. MacMillan, New York, New York.

CSS Response to ISAB Recommendations

1. **Describe methods clearly.**—Methods used to define and estimate the C0, C1, and T0 study groups and SAR, TIR, and D are presented. Methods used to define and estimate annual SAR are not presented clearly. Other methods (e.g., to remove sampling variability) are not presented fully or clearly.

2. **Report size at tagging to survival and relate to survival.**—Size at tagging is analyzed for the upstream-downstream comparison but is not reported for releases in general or related to survival.

3. **Address validity of inference from tagged fish to untagged fish.**—This point was addressed briefly, with criticisms of the methods used to determine that untagged fish have different survival than tagged fish. In general, results from tagged fish are compared to target values with no mention of any possible bias based on tagged fish.

4. **Use more downstream hatcheries in the upstream-downstream comparison.**—This was not done for the retrospective analysis.

5. **Do not limit analyses to an annual time scale; consider environmental and operational factors.**—Within-year patterns of SAR, TIR, and D are addressed briefly. The main focus of the analysis is on the annual time scale, due both to sample size and to the use of the C0 group, which cannot be analyzed on a smaller time scale. SAR, TIR, and D are related to several environmental factors. Operational factors are not considered.

6. **Perform a 10-year summary report.**—This is it.

7. **Test assumptions.**—Estimation results are analyzed for robustness to CJS-assumption violations. Little attention is given to whether or not those assumptions are violated.

8. **Pre-assign routes of passage to simplify analytical methods.**—This could not be done for the retrospective report. The simulations testing the robustness of estimation methods to CJS assumption violations incorporated pre-assigned routes of passage for future analysis. However, it appears that the analysis methods to be used with pre-assigned passage routes will remain unchanged, so the pre-assigned routes will not simplify analytical methods.

9. **Use more diverse metrics of differential survival (not only TIR and D).**—"Hydrosystem survival" was defined, but (1) was not used because it cannot be estimated, and (2) does not appear to be an improvement over TIR and D. "System survival" was used, but no expected or target values were given, and there was no

guidance for interpreting results. Conclusions continue to be based on TIR and D. Distance from ocean and hatchery practices were not considered.

General Technical Comments on CSS 10-Year Report

The Comparative Survival Study (CSS) 10-Year Retrospective Analyses Report provides a history of PIT-tagged salmonid fish performance from 1998 through 2006. The length of time, the breadth of geographic coverage, and range of salmon life-history phases investigated in the report have the potential of providing a valuable chronicle of recent Columbia River trends. No other study in the Fish and Wildlife Program has the same scope of effort.

Given the unique range of this project and the importance of this 10-year review, it is therefore unfortunate that the report does not document the tagging results more thoroughly.

- Reporting the empirical results of the tagging study is largely limited in this report to relatively few tables and graphs in Chapters 2 and 3. Appendix D supplements this information in these chapters but does not provide a showcase for the important trends and comparisons one might expect from a 10-year summary. For example, comparison of trends among the many hatchery stocks tagged is completely absent. Furthermore, in many cases, standard errors or confidence intervals for performance measures are neither reported nor displayed. The CSS Report needs to simply document and display the tagging results for the benefit of most readers and organizations that do not have the resources of the CSS organization. This issue is fundamental to our comment - the need to provide the means to reproduce results. (See also the closing comment on the last page.)

- Again we suggest that the CSS Report provide a straightforward presentation of tagging results. The Retrospective Report instead focuses on interpreting estimates of survival (S), smolt-to-adult ratios (SARs), transport-inriver ratios (TIR), and delayed mortality (D), using both previous as well as new approaches. This is unfortunate because it is in these analyses (as discussed in subsequent bullets) where the Retrospective Report most often falters in providing basic data and analyses useful for fish research and management.

- By definition, these PIT-tag studies are observational, thereby precluding direct causal inferences to any natural or anthropogenic factors. Replicated, randomized, and manipulative studies beyond the scope of the CSS study would be required for such inferences. Consequently, any attempt to identify environmental driving variables or differentiate ambient from hydrosystem effects is very difficult. The methods CSS uses in the report are not exempt from these problems, and contain several technical errors, as summarized below by chapter. Again, a direct causal inference to any natural or anthropogenic factor is precluded.

- Beyond that, there are conceptual issues, e.g., the approach of basing transportation analysis on C0 (undetected) fish. Not only does the C0 group not represent a real management alternative to transportation, but that group also migrates through the hydrosystem, not through an unimpounded river. Consequently, using the C0 group as a surrogate for a non-hydrosystem alternative is invalid. The CSS approach to estimating differential mortality using upstream-downstream comparisons is equally invalid. The CSS methods in many cases have not been peer-reviewed in the scientific literature, as might be expected for a 10-year-old program.

Below are summarized some of the more important technical concerns by chapter and also attached is a list of recommendations by the ISAB and our assessment of how well the CSS 10-year review complied.

- In Chapter 2, the concept of instantaneous mortality rate (Z) is introduced. However, it is not based on failure times (i.e., death times of PIT-tagged fish) as it properly should be, but rather on a simple function of the ratio of reach survival estimates and median travel times. The report inappropriately analyzes survivals, travel times (FTT), and Z as if they are three independent pieces of information. Reach survivals throughout the season are relatively stable while travel times show marked seasonal trends. Using the ratios of this information, Z values are calculated and seasonal trends are interpreted as survival processes. In fact, the trends in Z are essentially nothing more than the inverse trends in travel times construed as survival effects. Finally, the Report misinterpreted the instantaneous mortality rate, Z, as the probability of mortality (i.e., $1-S_t = 1-e^{-Zt}$), which it is not.

- Results of the modeling exercises on reach survivals in Chapter 2 are not discussed, leaving the impression that results for Z values are applicable to S, which is not true as discussed above

- The summary tables in Chapter 2 for the modeling exercise are difficult to interpret. Nevertheless, it appears as though models with partial regression coefficients (e.g., negative sign with flow) were routinely omitted. This practice left models that sometimes had an interaction term but no main effect, or a quadratic term without the linear component, which is inconsistent with general modeling-building practices. These nonstandard practices, as well as using Julian date as a surrogate for any number of unspecified environmental factors, greatly limited the interpretation and efficacy of the analyses.

- As requested by the Independent Scientific Advisory Board (ISAB), the CSS has compiled in Chapter 3 and Appendix B many of the methods used to generate the time series of estimates reported. Nevertheless, important definitions such as annual SARs, upriver adult survival, and project-specific TIR, are never mathematically defined in equations. Also missing are basic results, such as the numbers of fish actually

transported at each dam, which should be documented in a 10-year review such as this report.

- The CSS report inferences are often based on whether confidence intervals overlap. Non-overlapping confidence intervals do not provide a valid test of significant differences. Instead, the CSS should use valid methods of testing significance, either within their bootstrap approach or separately with a parametric approach.

- Chapter 4 used parametric models to partition the total variance of SARs, TIR, and D into natural variation and measurement error. However, an underlying assumption, that the SARs are binomially distributed, is inconsistent with the mark-recapture models used to estimate the values. This invalid assumption results in underestimating the sampling error, which will inflate estimates of natural variation. The report needs to use goodness-of-fit tests to assess their parametric model assumptions and compare their parameter estimates with nonparametric variance components. Additionally, their inferences concerning natural variation do not take into account their own findings on ambient effects, the historical distribution of those ambient factors, or how influences such as global climate change may affect projections into the future.

- In Chapter 5, the CSS Report continues a practice of comparing upstream and downstream Chinook salmon stocks. As in the past, multiple upstream hatcheries and collection points are compared to only a single downstream hatchery and collection point (for wild fish), despite the ISAB's recommendation to incorporate more downstream stocks. Given that this is a retrospective report, it is understandable that the CSS report could not immediately include additional downstream stocks. However, the Report should have included the early data from downstream hatcheries that were originally used in the CSS. These hatcheries were removed from the study design by CSS management, contrary to the urging by some who viewed their inclusion as providing the exact perspective that the ISAB called for later in their 2005 review of the CSS.

- While the CSS does perform some useful comparisons of biological characteristics of the upstream and downstream stocks, their upstream-downstream analysis is invalid in other critical ways.

- The CSS uses an incorrectly conceived and constructed measure of "differential mortality"—just another name for latent mortality. The approach assumes that no natural mortality should occur for smolts between upstream and downstream sites. When the performance measure is corrected for the extra migration of the upstream stocks, there is little or no evidence of differential hydrosystem mortality for hatchery Chinook salmon.

- Additionally, in comparing travel times between upstream and downstream stocks in Chapter 5, the CSS report ignores the longer distance upstream stocks have to travel and then attributes their later estuary entry on the Federal Columbia River Power System (FCRPS).

- Finally, all these efforts in Chapter 5 to estimate latent mortality are contrary to the recent ISAB 2007 recommendations that such attempts be abandoned because this task is impossible with existing data. The report does not explain why it continues to pursue this rationale.

- Chapter 6 attempts to partition survival across different portions of the migration, focusing on smolt survival from the hatchery/trap to Lower Granite Dam (LGR), perceived adult survival from Bonneville Dam to LGR, and perceived adult survival from LGR back to the hatchery/spawning grounds. Adults are categorized by juvenile migration method. The effect of juvenile migration method—in particular, transportation—on adult upriver survival is an important question, and the analyses in this chapter relating adult survival to migration method are worthwhile. However, the CSS Report does not provide the methods used in estimating upriver survival for a given juvenile release group, so reviewing and reproducing their results is impossible. Additionally, the Report misinterprets the odds ratio from its own logistic regression when comparing adult survival for LGR-transport fish to other fish. Consequently, the Report overestimates the effect of LGR transportation on upriver adult survival.

- Chapter 7 describes the results of useful computer simulations to assess the effects of violations of key assumptions of the Cormack-Jolly-Seber (CJS) release-recapture model. The study demonstrated the robustness of the estimation methods to all but the most severe temporal changes in survival and detection probabilities. The results also showed the release-recapture model to be reasonably robust to changes in survival due to prior detection history. The focus of the simulation study was on the preassigned transport (T) and return-to-river (R) groups. Unfortunately, these two groups of fish do not directly translate into the C0, C1, and T0 groups used in the CSS analysis. Consequently, more focused simulations are still needed to assure the CSS methods are robust enough to model violations.

- The question of the effect of the FCRPS on salmonid migration and survival is important. However, it is not addressed by the analyses presented in this Report. Because the overall conclusions provided in the report are based on the invalid analysis of the previous chapters, the final inferences are unreliable.

- The CSS Report attributes all differences in survival and travel time between study groups in the upstream-downstream comparison to the hydrosystem, ignoring differences expected because of different migration distances and resulting natural mortality. Additionally, the Report incorrectly claims to have addressed the question of whether smolt transportation compensates for effects of the FCRPS on survival of Snake River Chinook salmon and steelhead. The comparison of the SARs of transport fish and inriver fish is not equivalent to comparing transportation to migration through the unimpounded river.

As urged by the ISAB, this CSS Report now presents some of the methods used in estimating SAR, TIR, and D for the various groups of interest (i.e., the C0, C1, and T0 groups). It has compiled methods strewn through previous reports in one place, and this makes the reading much easier. However, this encouraging start was not consistent across chapters. Our attempts to reconstruct final results from intermediate calculations presented in the report have been frustrated by a lack of necessary information and insufficient technical descriptions. The ability to reproduce results is crucial to the scientific peer review process. The Retrospective Analyses Report would have benefited from greater documentation of basic tagging results and from far less exploratory efforts to assign effects to the hydrosystem when causation really cannot be partitioned or identified using the CSS data. Due to the inability to reproduce these results using accepted modeling and analytical procedures the CSS Report's findings do not demonstrate the scientific rigor and support to authoritatively guide hydrosystem management.