CSMEP

Collaborative Systemwide Monitoring and Evaluation Project (CSMEP)

Collaborative Systemwide Monitoring and Evaluation Project (CSMEP) – Year 3

Project No. 2003-036-00

Annual Report for FY 2006

Collaborative Systemwide Monitoring and Evaluation Project (CSMEP) – Year 3

Project No. 2003-036-00

Annual Report for FY 2006

Submitted to

Tracy Yerxa
KEWB-4
Bonneville Power Administration
P. O. Box 3621
Portland, OR 97208-3621

Compiled by

David R. Marmorek, Marc Porter, and Darcy Pickard
ESSA Technologies Ltd.
1765 West 8th Avenue, Suite 300
Vancouver, BC V6J 5C6 Canada

On behalf of the

Columbia Basin Fish and Wildlife Authority
851 SW Sixth Avenue, Suite 260
Portland, OR 97204

October 31, 2006

Contributors:

Robert Al-Chokhachy (EcoLogical)
Chris Beasley (QC-NP)
Tom Berggren (FPC)
Bill Bosch (YN)
Nick Bouwes (EcoLogical)
Evan Brown (IDFG)
Cedric Cooney (ODFW)
Tim Copeland (IDFG)
Tim Dalton (ODFW)
Dave Fast (YN)
Jeff Fryer (CRITFC)
Peter Galbraith (CRITFC)
Jay Hesse (NP)
Peter Hahn (WDFW)
Annette Hoffman (WDFW)
Saang-Yoon Hyun (CRITFiC)
Chris Jordan (NOAA)
Steve Katz (NOAA)
Phil Larsen (EPA)
Dave Marmorek (ESSA)
Claire McGrath (IDFG)
Dick OConnor (WDFW)
Charlie Paulsen (PES-BPA)
Charlie Petrosky (IDFG)
Darcy Pickard (ESSA)
Marc Porter (ESSA)
Tom Rien (ODFW)
Phil Roger (CRITFC)
Kris Ryding (WDFW)
Bruce Schmidt (StreamNet)
Jesse Schwartz (CTUIR)
Sam Sharr (IDFG)
Dick Stone (WDFW)
Eric Tinus (ODFW)
Chris Toole (NOAA)
Earl Weber (CRITFC)
Paul Wilson (USFWS)
Keith Wolf (KWA-CTCR)
Frank Young (CBFWA)

Citation: **Marmorek, D.R., M. Porter and D. Pickard (eds).** 2006. Collaborative Systemwide Monitoring and Evaluation Project (CSMEP) – Year 3, Project No. 2003-036-00, Annual Report for FY 2006. Prepared by ESSA Technologies Ltd., Vancouver, B.C. on behalf of the Columbia Basin Fish and Wildlife Authority, Portland, OR. 126 pp. + appendices.

Executive Summary

The Collaborative Systemwide Monitoring and Evaluation Project (CSMEP) is a coordinated effort to improve the quality, consistency, and focus of fish population and habitat data to answer key monitoring and evaluation questions relevant to major decisions in the Columbia River Basin. CSMEP was initiated in October 2003 and is administered by the Columbia Basin Fish and Wildlife Authority (CBFWA), with the participation of several federal, state and tribal fish and wildlife agencies.[1] CSMEP is a major commitment of the Council towards regionally integrated monitoring and evaluation (M&E) across the Columbia River Basin, and is a critical element of the Pacific Northwest Aquatic Monitoring Partnership (PNAMP). CSMEP's specific goals are to: 1) interact with federal, state and tribal programmatic and technical entities responsible for M&E of fish and wildlife, to ensure that work plans developed and executed under this project are well integrated with ongoing work by these entities; 2) document, integrate, and make available existing monitoring data on listed salmon, steelhead, bull trout and other fish species of concern; 3) critically assess strengths and weaknesses of these data for answering key monitoring questions; and 4) collaboratively design, implement and evaluate improved M&E methods with other programmatic entities in the Pacific Northwest.

Progress in FY2006

During FY2006 CSMEP made considerable progress on its inventory and assessment goals. CSMEP and StreamNet jointly completed inventories of salmon and steelhead data for a second set of selected subbasins in Washington—Okanagan, Methow, Kalama; Oregon—Deschutes, Grande Ronde; and Idaho—Upper Fork Salmon, Middle Fork Salmon. Inventory efforts undertaken by StreamNet in FY2006 also began to focus more intensively on collection of metadata for resident fish species (e.g., bull trout, cutthroat trout). CSMEP biologists continued with their reviews of the strengths and weaknesses of these subbasin data for addressing a structured set of monitoring questions about fish population status and trends at different spatial and temporal scales. The CSMEP web database originally developed in FY2004 to store inventory metadata in a readily accessible format and location was further developed and populated with a growing body of metadata from the pilot watersheds. The CSMEP public website developed for communication and coordination amongst CSMEP members and interested parties was extensively restructured in FY2006 for greater ease of use.

Significant progress was also made in FY2006 on CSMEP's goals of collaborative design of improved M&E methods. Three multi-agency monitoring design workshops (one in collaboration with PNAMP) were held to explore how best to integrate the most robust features of existing monitoring programs with new approaches (e.g., Federal RME pilot studies, AREMP, EPA EMAP, etc.). CSMEP continued to build on this information to develop general 'design templates' or 'design processes' for monitoring the status and trends of fish populations and the effectiveness of habitat, harvest, hatchery and hydrosystem recovery actions within the Columbia River Basin. As a pilot exercise, information from the CSMEP metadata inventories as well as from ongoing regional RME studies is being used to develop design

[1] **Agencies**: NOAA Fisheries, US Fish and Wildlife Service (USFWS), Columbia Fish and Wildlife Authority (CBFWA), Columbia River Intertribal Fish Council (CRITFC), Bonneville Power Administration (BPA), Oregon Department of Fish and Wildlife (ODFW), Washington Department of Fish and Wildlife (WDFW), Idaho Department of Fish and Game (IDGF), Fish Passage Center (FPC), StreamNet, Nez Perce Tribe, Umatilla Indian Reservation, Confederated Tribes of the Colville Reservation, Yakama Indian Nation

Consultants: ESSA Technologies Ltd. (Facilitators), Eco Logical Research, Quantitative Consultants, Paulsen Environmental Research, KWA Ecological Sciences templates at the spatial scale of the Snake River Basin ESUs. Elements of CSMEP's work on the Snake River Basin Pilot Project have fed into design considerations for the NOAA-F /BPA Salmon River Basin Pilot Study and the Lemhi Basin Habitat Conservation Plan (HCP). Further information on CSMEP metadata inventories, strengths and weaknesses assessments and monitoring design products for FY2006 are presented in the main text of this Annual Report as well as on the CSMEP public website.

CSMEP M&E Design Subgroups:

1) Status and trends

In FY2006, the CSMEP Status and Trends Subgroup focused on continued development of a simulation model for evaluating alternative designs (low, medium, high) for monitoring status and trends of Snake River spring Chinook at the population, MPG and ESU scales. The model incorporates the four data elements required for informing TRT decisions on species delisting (i.e., abundance, productivity, spatial structure and diversity). The Subgroup also began a comparative analysis of alternative survey protocols for evaluating TRT criteria in relation to steelhead monitoring in the Mid Columbia.

Simulation model

Overview and key insights:

- Provides a framework for assessing the variability in the data used to measure abundance, productivity, spatial structure, and diversity
- Employs misclassification rates to describe error in SS/D risk levels, resulting in a simple but realistic framework for assessing variability in spatial structure and diversity data
- Developing test datasets is not easy, and we may need to create data sets with extreme behavior initially, to help us learn how the TRT rules function.
- The model provides a useful tool for evaluating the sensitivity of the TRT viability criteria to changes in the quality of data
- Realistically, it is difficult to test all possible M&E designs that could potentially be used in a particular ESU. The Snake River Sp/Su ESU contains 32 populations and we need to collect multiple types of information in each population. Rather than try to enumerate each possible data collection method in each population we consider very basic designs initially in order to bound the problem:

  Design 1: all 32 populations are monitored using Low monitoring methods

  Design 2: all 32 populations are monitored using Medium monitoring methods

  Design 3: all 32 populations are monitored using High monitoring methods
- The next step is to determine what the status quo M&E design is and compare this to these three simple designs in order to provide direction as to which other more complex designs might be worth testing.
- This information can help the manager to determine where it is feasible to improve monitoring activities. The model can then be used to test how much value would be gained by making those improvements.
- Due to the lack of consistent information regarding accuracy of different monitoring methods, the model doesn't specify the method used but rather a particular level of variability i.e., the coefficient of variation (CV). The model can tell us, for example, 'the correct viability assessment is made 80% of the time if you use a CV of *X*', and then the individual scientist in the field can

determine the best way to achieve that CV dependent on the conditions of a specific monitoring site.

Interpretation of results for the different M&E designs

Population level results:

- How often each population is correctly assessed for either A/P risk or SS/D risk?
- In which direction did the error occur? (Over or under estimated the risk?)
- Was the viability of the population correctly assessed?

MPG results:

- How often was each MPG assessed as viable?
- How often did a particular MPG fail each of the 7 requirements?

ESU results:

- How often was the ESU assessed as viable?
- Which MPGs caused the ESU to fail?

Mid Columbia (Oregon) steelhead monitoring, evaluation and population viability analysis

Current practices:

- Index redd surveys have been commonly used to monitor steelhead spawner numbers in Oregon Mid-Columbia streams, as well as in many other areas, for several decades. They are likely to be continued to ensure the benefits of such long-term databases and the maintenance of existing funding sources and personnel positions.
- Methods used to estimate spawner abundance and spawner-to-spawner productivity for population viability assessments for ICTRT-defined Mid-Columbia steelhead populations in Oregon include trap counts and stratified random sampling, but most estimates employ expansions from index survey average redd densities in some fashion.
- In most subbasins, redd density estimations and expansions employ a GIS-based spawning habitat quality/weighting model. For John Day subbasin populations, EMAP protocol redd density surveys over the past three years are used to calibrate index survey redd density expansions.
- Annual estimates of redd numbers for a population are multiplied by a standard average fish/redd ratio to estimate spawner abundance. This ratio is currently based on only four years of data from Deer Creek in the Grande Ronde subbasin.
- Empirical data necessary to assess spatial structure and diversity criteria for population viability assessments are often infrequent, as well as being difficult and expensive to obtain.

Next steps — scope of work summary:

- Use the EMAP and index redd density data from the John Day basin, encompassing a wide range of densities and habitat qualities, to calibrate the intrinsic habitat quality rating model. Also iteratively test and recalibrate the model using Umatilla and Warm Springs River (Deschutes subbasin) datasets, which include index survey data and complete adult counts at downstream locations.

- Use John Day EMAP and index survey data to assess, for each protocol, sampling requirements to measure spatial and temporal variations in abundance, particularly minimal changes required to reach population viability threshold levels (i.e., accuracy, precision and power analysis). Usefulness of each protocol for assessing spatial structure and diversity criteria will also be evaluated. This analysis will be performed on a whole subbasin as well as an individual population basis, and associated costs will be assessed.
- Attempt to obtain additional fish/redd ratio data from streams within the Mid-Columbia.
- Identify Pacific Northwest data sets for comparison of levels of variability between index survey based spawner estimates or average redd densities to “true” abundances or average densities measured by other means (e.g., weirs, dam counts, aerial counts, probabilistic survey protocols).
- Employ the above assessments to parameterize the simulation model.
- Based upon the above analyses, for each Mid-Columbia steelhead population in Oregon, make recommendations on which monitoring methods currently in place to maintain and build upon, methodologies to adopt, and, likely, hybrid monitoring systems (e.g., EMAP/index) that would best meet population viability assessment needs.

2) Hydrosystem

The CSMEP pilot effort in the Snake Basin involves a number of steps:

1. characterizing the current monitoring effort in the Snake Basin (“Status Quo”);
2. developing a set of Low, Medium, and High designs that *integrate* across the various monitoring objectives (i.e., Status & Trends; Action Effectiveness monitoring for hydro, hatchery, harvest and habitat), and provide increasing reliability of responses to decision-focused questions; and
3. evaluating the cost-precision and other tradeoffs in these designs, so as to demonstrate a logical process of making decisions on M&E.

To help move CSMEP through these steps, the Hydro subgroup began in FY2006 with two elements: a) the description of Status Quo, Low, Medium and High alternatives prepared for *Status and Trends* monitoring in the CSMEP FY2005 report; and b) various Status Quo, Low, Medium and High alternatives developed in FY2005 for hydro action effectiveness monitoring directed at a long list of different questions (CSMEP 2005). In FY2006, the CSMEP Hydro Subgroup first narrowed their scope, focusing on just three major sets of decisions and four questions related to those decisions, as shown in the following table.

Decisions / Alternative Actions	Hydro Action Effectiveness Questions
Are SARs, and important SAR ratios relating to effectiveness of transportation, meeting NPCC and BiOp targets? If targets are not met, (*by how much?*), then decision makers may need to consider changes in FCRPS operations (e.g., when, how much to transport and spill) or FCRPS configuration.	1. Is SAR sufficient for 1) NPCC goal[2] & 2) recovery goals? 2. Is transportation more effective than in-river passage?
Has hydrosystem complied with performance standards set out in 2000 FCRPS BiOp? If not, what changes are required?	3. How does annual in-river survival of spring summer Chinook and steelhead (Lower Granite to Bonneville) compare to 2000 FCRPS BiOp performance standards?
Should FCRPS change the timing of transportation of some species within the season to improve survival?	4. How does effectiveness of transportation change over the course of the season?

The CSMEP Hydro Subgroup then built upon the SQ, L, M and H design options generated by the CSMEP Status and Trends group, recognizing that Status & Trends monitoring is the long term foundation for all other M&E. The preliminary Hydro L, M and H options supplement the Status and Trend group options with additional monitoring (e.g., PIT-tagging) required to answer the set of Hydro questions in the above table. The Hydro Subgroup also developed a preliminary estimate of the costs associated with PIT-tagging hatchery and wild fish, which can be applied to each of these options. As the L, M, and H options were evaluated, we made iterative improvements to achieve higher levels of cost-effectiveness. This process is continuing.

We looked at the ability to answer questions 1 & 2 over several years, assuming a period of relatively stable management. Here are the subgroup's major conclusions:

- Combining data from multiple years of PIT-tag data from outmigrating smolts allows a better overall picture of whether SARs and ratios of SARs are, in general, meeting survival targets. Getting the best possible estimates of SARs and *TIR*s in individual years (by marking large numbers of fish) is useful for other purposes (e.g., questions 3 and 4), but not necessary for estimating long-term mean values under questions 1 and 2.
- The power to distinguish between alternative hypotheses about the values of SARs and ratios of SARs is generally much more sensitive to the number of migration years for which data are collected, rather than to the number of PIT-tagged smolts each year (at least over the range of number of tags examined). This is likely due to the fact that, at the tagging rates simulated, sampling error is dwarfed by process error (true environmental variation) in SARs.
- One caveat to the above conclusion is that as the number of tags increases up to 5000, the chances that estimated confidence intervals for SARs will include the true mean (known as 'Coverage') also increase. Coverage also improves with time, as well (presumably from the sampling effect of drawing yearly values at random from beta distributions). More confidence can be invested in results from monitoring using more tags and covering more years.
- When a value of interest is very close to the target (e.g., SAR ≈ 2%, TIR ≈ 1), it's very difficult to decide if the true value is actually higher or lower. The results suggest that, assuming observed variation is proportional to mean values (i.e., that CV is constant over a range of mean SAR values), true SARs will need to exceed target values by 25% or more to have a reasonable chance of correctly concluding that the target values have been exceeded in a reasonable amount of time.

[2] Pg. 13 of NPCC mainstem amendments of 2003-2004. www.nwcouncil.org/library/2003/2003-11.pdf ; interim goals of 2-6% SAR

- For *TIRs*, the benefit of more tags is evident in the decreasing width of confidence intervals and decisions about hypotheses, though the benefit is relatively small and declines with time. The decision rule used to draw conclusions about the relative efficacy of transportation is influential, but a relatively high probability of reaching the correct conclusion would be achieved even after only 5 years, using a "neutral" decision rule[3], <u>if the true *TIR* differs from 1.0 by at least 20%</u>. If the true *TIR* differs from 1.0 by 50%, high probabilities of reaching the correct conclusion are achieved quickly even under the "wrong" decision rule (e.g., having a transportation averse decision rule when TIR > 1, or having a transportation tolerant decision rule when TIR < 1).
- The simulations used in the analyses presented here assumed the number of PIT-tagged fish in each group in each year in the time series was constant. In reality, the numbers of PIT-tagged wild fish vary substantially between years. However, because the methods explicitly account for the numbers of tags each year in estimating sampling variance and in weighting estimates among years, the benefits of this approach over estimating simple means and associated confidence intervals for a time series of annual estimates of SAR or *TIR* would likely be large.
- Further, application of the methods used here may be especially useful for making inferences from even smaller numbers of marked fish, such as might be the case in estimating within-season trends in Snake River ESU SAR or *TIR*, or in estimating values of these parameters at a population level finer than ESU (e.g., major population group).

Questions 3 and 4 have a different time scale than those for questions 1 and 2. Questions 1 and 2 focus on long-term, multi-year averages, while questions 3 and 4 focus on individual annual estimates of survival and TIR. For question 3, we have drawn the following conclusions:

- In years past (1998-2005), estimates of in-river survival for spring Chinook, Lower Granite (LGR) to Bonneville, have been highly variable, due to a mixture of sampling variation (caused by the limited number of tagged fish) and process variability, associated with varying in-river conditions and fish condition. While process variation is largely beyond our control, sampling variation can be reduced by increasing the number of in-river smolts tagged and, perhaps, by increasing the proportion of survivors detected at dams and at the trawl below Bonneville.
- None of the preliminary strategies developed by the Hydro Subgroup (low, medium or high), greatly increases the number of in-river migrants. To reduce measurement variation by a factor of two, one would need to increase tagged in-river migrants from about 70,000 at present to about four times that number. While this is possible to do in most years, the estimated cost of the increased sampling effort would be substantial—on the order of $700K per year. On the other hand, these tagged fish could be used for other purposes, including more precise estimates of annual SARs, upstream survival, and in-river harvest, assuming harvested adults were sampled for PIT-tags.

For question 4, we have drawn the following conclusions:

- Wild spring Chinook TIRs have generally shown an increasing trend over time—fish transported late in the season have higher SARs than in-river migrants, while the opposite is true early on. As with in-river survival, none of the preliminary strategies (L, M, H) initially developed by the Hydro Subgroup greatly increases the number of tagged, transported wild smolts, so there is little change from the base case in the precision of annual within-season TIR estimates. The

[3] The three decision rules used were:
1. "Transportation averse": reject conclusion that *TIR* > 1 unless Pr[*TIR* > 1] ≥ .8.
2. "Transportation neutral": accept conclusion that *TIR* > 1 if Pr[*TIR* > 1] ≥ .5
3. "Transportation tolerant": accept conclusion that *TIR* > 1 unless Pr[*TIR* >1] < .2

measurements require fish tagged or detected at LGR, so it may make sense to continue or increase tagging efforts at LGR for all scenarios—this is surely cheaper per fish than tagging above LGR, especially for wild smolts.

- While fish tagged at LGR cannot, of course, be used to estimate survival from release above LGR to the dam, the numbers required for the latter are much smaller than those needed for precise estimates of LGR-BON survival or SARs. A stratified sampling system might be useful, depending on the questions being asked. Further, if consistent relationships between wild and hatchery fish can be established, substitution of hatchery fish for wild fish may be possible, reducing costs substantially.

Further work is required to assess alternative monitoring designs and multi-year evaluation approaches during periods of significant changes in management actions (e.g., installation of Removal Spillway Weirs, changes in the timing of transportation). Similarly, these analyses could change if there are increases in the natural variability of flows (Jain et al. 2005) or in freshwater and ocean survival, as may be occurring with climate change.

3) Harvest

In FY2006, the CSMEP Harvest Subgroup focused on evaluating potential improvements in harvest monitoring through alternative tagging approaches, as well as exploring existing harvest impact models to suggest means for incorporating variance estimates.

Collection and use of harvest data

- The mortality rate of federal ESA-listed fish in Columbia River fisheries (incidental in selective fisheries and direct in targeted fisheries) is referred to as an impact rate, or an impact. Conceptually, in selective fisheries, an impact rate can be thought of as the product of the probability of being captured and the probability of mortality after release.
- The variables needed to assess incidental mortality in selective fisheries are: run size, harvest number, stock composition of the catch, marked fish release rate, and post release mortality rate. In assessments of upriver spring Chinook take, TAC views the estimate of harvested fish to be strong and accurate. The preseason forecast is reasonably accurate and is adjusted in season and post season using passage and harvest estimates. Stock composition may be improved by PIT-tag monitoring and genetic stock identification (GSI) technology. Onboard observations are used to monitor marked fish release rate. Standard gear-specific values are applied to estimate post release mortality.
- Stock composition in mainstem fisheries is estimated within season by applying juvenile mark rates to pre-season adult forecasts and assumed proportions of wild fish in the juvenile runs. This could be improved or corroborated by real-time Genetic Stock Index sampling or PIT-tag sampling.

Challenges of harvest management - Uncertainties in precision of estimates

- In general fisheries managers do not provide precision bounds on estimates of harvest and incidental take.
- Models to assess impacts of Columbia River fisheries on listed fish species are frequently revised (of necessity) and poorly documented (due to lack of staff time). The *US v Oregon* Technical Advisory Committee (TAC) agrees that better documentation is needed and asks if CSMEP staff will be able to take on this task.

- Onboard observation of catch and release numbers in selective commercial fisheries have a potential bias in estimates of steelhead stock composition that might be addressed through stratification or randomizing sampling design.
- Post release mortality estimates used in selective fisheries are based on just a few field studies and a technical consensus of TAC membership. Small-mesh fishery rate is based on 3 yrs of study – but each year the experiment was conducted differently. Sensitivity analysis may help describe the potential magnitude of bias and how much variance might affect estimates of incidental mortality. The TAC, by and large, greets additional field studies of release mortality skeptically, given the inherent expense and difficulty of obtaining definitive results.

Next Steps – Could alternative monitoring approaches improve estimates?

- The CSMEP harvest group is examining existing models to suggest means for incorporating variance estimates.
- The CSMEP harvest subgroup is currently conducting power analyses to describe the effect a range of variation in marked-fish release rate has on estimates of incidental mortality.
- CSMEP auditing of the lower Columbia spring fisheries impact model has identified a probable calculation error that results in small-magnitude errors in estimates of incidental mortality. We will work with TAC membership to confirm and correct the error.
- The results of these analyses will provide a context to develop potential alternative harvest M&E approaches.
- *Beyond the CSMEP Harvest Subgroup:* TAC envisions a need for full ESU/population-based run reconstructions for steelhead.

4) Habitat

CSMEP has recognized that there are serious challenges to the development of consistent or uniform habitat effectiveness monitoring. These include:

1. Habitat conditions vary greatly across subbasins in terms of their natural biogeoclimatic regimes, the status of their fish populations, the degree of human impact and management, and the number and nature of restoration actions that have been implemented, or are being considered for implementation within them.
2. Habitat effectiveness questions encompass different scales of inquiry, which imply different scales of monitoring. Widely different scales of study impose quite different demands on monitoring design.
3. Management actions are usually planned and implemented on local scales within frameworks provided by regional funding processes, but management questions and objectives are set on a larger scale. As such, the specific designs applied to individual or small groups of habitat management actions rarely match up with regional questions in a manner that allows monitoring that is easily evaluated quantitatively.
4. The mechanistic linkages between habitat change and fish response that empower quantitative predictions are often poorly understood. Therefore, monitoring of habitat actions requires explicit experimental design to be incorporated into the monitoring design. Unfortunately, neither regional nor local habitat action planning commonly includes experimental design in their processes.

CSMEP's Habitat Subgroup has therefore not pursued development of a generic template for habitat effectiveness monitoring, but has instead been attempting to develop a consistent "process" that can be applied to development of individual monitoring designs dependent on the particular situation in different subbasins. They have been piloting this approach within the Lemhi Subbasin. In FY2006 the Habitat Subgroup:

- Compared CSMEP's M&E design development process and their final design recommendations for the Lemhi Subbasin with a parallel ISMEP design process being undertaken concurrently for the Lemhi HCP. This "side-by-side comparison" undertaken by CSMEP has provided insights into commonalities for design of habitat effectiveness monitoring that may occur across subbasins. It has also identified some of the elements that are likely unique to individual subbasins and will not readily lend themselves to standardized design templates.
- Explored the pros and cons of "top-down' (i.e., management objectives) vs. "bottom-up" (i.e., scientific questions) design approaches for habitat effectiveness monitoring in the Lemhi Subbasin; an issue that will likely re-emerge in development of an acceptable "process" for improving habitat effectiveness M&E designs within Columbia subbasins.
- Finalized their Lemhi Subbasin design work and sought to engage regional managers to determine the management objectives for the subbasin; began a "closing the loop" exercise to evaluate how well CSMEP analysts have matched up their design questions with the actual management objectives for the Lemhi Subbasin.
- Extended their exploration of the experimental designs and associated statistical analyses that could be undertaken for testing each of the Lemhi habitat hypotheses formulated by the subgroup.
- Incorporated M&E designs for bull trout in the Lemhi Subbasin and evaluated how these might be integrated with the originally proposed low, medium, high Lemhi designs that were focused on spring Chinook.
- Began to explore whether other subbasins could benefit from similar CSMEP design efforts for habitat effectiveness monitoring. Such proactive design efforts for individual subbasins are likely to be of real benefit only when directed to subbasins with major habitat projects planned for the near future and that are supported by a robust, well-funded management and monitoring program.

5) Hatchery

Efforts of the Hatchery Subgroup in 2006 were focused on the further development of study designs at the spatial scale of the Snake River subbasin. However, all designs continued to be developed in a manner that: 1) enables them to act as "replicates" upon expansion of the project to the entire Columbia River Basin; or 2) as a small-scale test of a larger Columbia River Basin scale design. Due to a substantial decrease in participation relative to 2005, work in 2006 focused primarily on:

- Initial development of a stratified study design to estimate the proportion of hatchery origin strays in target and non-target populations across the Snake River subbasin.
- Initial development of a stratified design to representatively allocate research using genetic parentage analysis to address the relative reproductive success of hatchery origin adults.

Thus far, design development has focused on identification of appropriate strata, distribution of current sampling effort, and identification of data gaps. It is anticipated that strata will be populated in early FY2007, and that sample "draws" using EMAP will enable cost estimation and statistical power analyses later in FY2007.

6) Design Integration

In FY2006 CSMEP began to explore the integration of the individual M&E component parts within a larger monitoring framework (i.e., generate improved efficiencies through integrated designs) for the Snake Basin pilot design. This integration effort across scales and subgroups is a challenge faced by all subbasins; hence the results will be of general benefit basin wide. The group has begun to develop a comprehensive matrix of shared performance measures and data interdependencies across the different CSMEP subgroups. The matrix is providing a starting foundation for identifying the priority performance measures for monitoring and the relevant spatial scale(s) of these data for varied subgroup monitoring needs. CSMEP has also begun to explore how to integrate monitoring costs for the shared performance measures to achieve greater efficiencies across monitoring programs. To ensure that analyses and monitoring designs explored as part of the project are consistent with the overarching objectives of Columbia River Basin monitoring agencies CSMEP has been working closely with PNAMP in FY2006 (e.g., shared workshops, etc.) to solicit appropriate direction and feedback from the key monitoring groups in the Basin.

Ultimately, all M&E decisions involve tradeoffs and a balancing of risks. The PrOACT approach is being employed by CSMEP as a simplified multi-objective decision analysis that provides a suitable framework for dealing with the large number of objectives associated with the Columbia Basin M&E issues. PrOACT is an iterative process that involves cycling over the development of M&E alternatives, evaluating them, assessing tradeoffs, revising alternatives and then starting again, starting from a broad set of alternatives that gradually narrows to an acceptable choice or set of choices. CSMEP has been attempting to apply the PrOACT approach for the generation and filtering of their alternative M&E designs across the subgroups based on a suite of criteria which includes: 1) high inferential ability; 2) strong statistical performance; 3) reasonable cost; 4) practical application; and 5) environmental impact CSMEP is following an approach where the base requirements for status and trends M&E will provide the foundation for low, medium and high design alternatives, while monitoring requirements for the various action effectiveness issues are to be built incrementally onto this foundation (as feasible).

Table of Contents

List of Tables

List of Figures

1. Introduction

1.1 Background

The Collaborative Systemwide Monitoring and Evaluation Project (CSMEP) is a collaborative effort led by the Columbia Basin Fish and Wildlife Authority (CBFWA). Project participants include NOAA Fisheries (NOAAF), the U.S. Fish and Wildlife Service (USFWS), three state fish and wildlife agencies (WDFW, ODFW, IDFG), StreamNet, the Columbia River Inter-Tribal Fish Commission (CRITFC), The Nez Perce Tribe (NPT), Yakama Indian Nation (YIN), Confederated Tribes of the Umatilla Indian Reservation (CTUIR) and the Confederated Tribes of the Colville Reservation (CTCR). Close coordination occurs with the Pacific Northwest Aquatic Monitoring Partnership (PNAMP).

This three-year project (now renewed for an additional two years) focuses on the issue of systemwide monitoring and evaluation of fish status, addressing requirements of NOAAF and USFWS biological Opinions and recovery plans as well as the NPCC Fish and Wildlife Program. CSMEP's goal is to demonstrate the benefits of systematic development and evaluation of alternative M&E designs on a regional scale, for answering key questions related to fish and watershed management decisions in the Columbia Basin. It involves an integrated, collaborative effort by fisheries scientists and biometricians to fulfill seven objectives:

1. *Interact with federal, state and tribal programmatic and technical entities* responsible for monitoring and evaluation of fish and wildlife, to ensure that quarterly work plans developed and executed under this project are well integrated with ongoing work by these entities.
2. *Collaboratively inventory existing monitoring data* that bear on the problem of evaluating the status and trend of salmon, steelhead, bull trout and other species of regional importance across the U.S. portion of the Columbia Basin, and for selected parts of the Columbia River Basin in Canada which affect the status of key fish stocks in the U.S. portion of the Columbia River Basin (e.g., Okanagan sockeye).
3. *Work with existing entities (e.g., StreamNet, NOAAF) to make a subset of existing monitoring data available through the Internet,* recognizing the continuing evolution of data management in the Columbia Basin.
4. *Critically assess the strengths and weaknesses* of existing monitoring data and associated evaluation methods for answering key questions at various spatial scales concerning the state of ecosystems and fish habitat, as well as fish distributions, stock status and responses to management actions.
5. *Collaboratively design* improved monitoring and evaluation methods that will fill information gaps and provide better answers to these questions in the future, by providing state and tribal fish agency participation and work products for multi-agency development of regionally coordinated monitoring programs.
6. *Coordinate state and tribal participation and work products* for regionally coordinated, multi-agency *implementation* of pilot projects or large scale monitoring programs.
7. *Participate in regional forums* to evaluate new monitoring program results, assess new ability to answer key questions, propose revisions to monitoring approaches, and coordinate proposed changes with regional monitoring programs.

Since project initiation in October 2003, CSMEP participants have collaboratively developed work plans in close consultation with other programmatic and technical entities (Objective 1). For Objective 2 (data inventory), CSMEP began with a set of 16 specific M&E questions adapted from Jordan et al. (2002), and a set of 45 performance measures for viable salmonid populations, adapted from McElhany et al. (2000). This original set of questions has been expanded by CSMEP workgroups to more comprehensively cover the key M&E questions perceived of relevance to decision makers in Columbia River Basin fish and wildlife managers (Appendix A). To evaluate the range of data quality that exists within the Columbia River Basin, CSMEP selected pilot subbasins that included both data rich and data poor areas and were located across a range of Basin Ecoregions. For each of these pilot subbasins, StreamNet staff and CSMEP biologists jointly completed an inventory of the information available for each of the key performance measures for each of the target fish species. An Internet-based CSMEP database (Objective 3) has been developed by StreamNet which allows access to the metadata recorded from these CSMEP inventories. For Objective 4, CSMEP biologists have reviewed the strengths and weaknesses of these data for addressing Tier 2 status and trend questions, and considered opportunities for using these data to answer Tier 3 action effectiveness questions (see Appendix A for definition of tiers). CSMEP workshops have provided continuing opportunities for biologists and biometricians from across the region to meet and discuss recent advances in M&E approaches (e.g., EMAP sampling frames, results from pilot projects, IMW strategies). CSMEP thus represents a unique forum for the cross-fertilization of M&E ideas among federal, state and tribal fish agency staff (Objective 7). Ideas expressed at these workshops have been incorporated into CSMEP's pilot M&E study designs. CSMEP's preliminary designs for their Snake Basin Pilot Project have provided input to the NOAAF/BPA Salmon River Subbasin Pilot Study, and have assisted in the continuing evolution of potential M&E designs to assess effectiveness of the Lemhi River Subbasin Habitat Conservation Plan (Objective 6).

CSMEP made considerable progress in FY2006 in the creation and evaluation of alternative monitoring designs for both status & trends and action effectiveness monitoring (Objective 5). To keep the scope of their design work manageable for initial evaluation purposes, CSMEP constrained their area of application of approaches to the Snake Basin for FY2006, focusing principally on the spring/summer Chinook ESU. While initial analyses have been focused on the Snake Basin, the intended area of relevance of our methodological findings is the entire Columbia Basin and beyond to the larger Pacific Region. CSMEP's continuing design work is intended to fulfill the following overall objectives:

- Collaboratively develop Tier 1, 2 and 3 designs in an integrated, cohesive manner to ensure that experimental designs and monitoring protocols integrate across tiers, spatial hierarchy levels and life cycles in a cost-effective manner, to address the information needs of decision makers.
- Apply the EPA's Data Quality Objectives process to work systematically from decisions to M&E designs.
- Consider multiple objectives, observation error, natural spatial and temporal variability, future trends, and types of analytical methods to estimate parameters of interest, building upon existing work of the FCRPS RME Plan and other regional federal, state and tribal M&E efforts to date.
- Review the designs of existing pilot studies, and assess their applicability to other regions.
- Evaluate the success of existing pilot projects as results become available.
- Recommend the most cost effective M&E designs within available budget constraints for each sub-basin with well integrated M&E methods.

The conceptual approach for the collaborative design and evaluation process is captured in Figure 1.1. Specific work tasks and products associated with these objectives through FY2005 and 2006 and beyond were to:

- Develop a draft design template and the general structure of a decision analysis to guide the evaluation of monitoring designs appropriate for different performance measures at various spatial and temporal scales.
- Review and revise the design process, design template and evaluative framework.
- Adapt / build tools and perform quantitative evaluations of alternative monitoring designs, taking into consideration their statistical and cost properties.
- Present a preliminary evaluation of alternative monitoring designs for the Snake River Basin to client agencies and reviewers.
- Revise the M&E designs and present the revised plans to clients and reviewers.

CSMEP intends to build on their initial work undertaken in the Snake Basin Pilot Study in FY2006 and during FY2007-08 will begin to expand their design processes to a larger set of regional pilot projects in the Columbia Basin. CSMEP's design efforts during this period will increasingly focus on simultaneous assessment of and optimizing of monitoring programs for multiple species, including salmon, steelhead, bull trout and other resident species of concern.

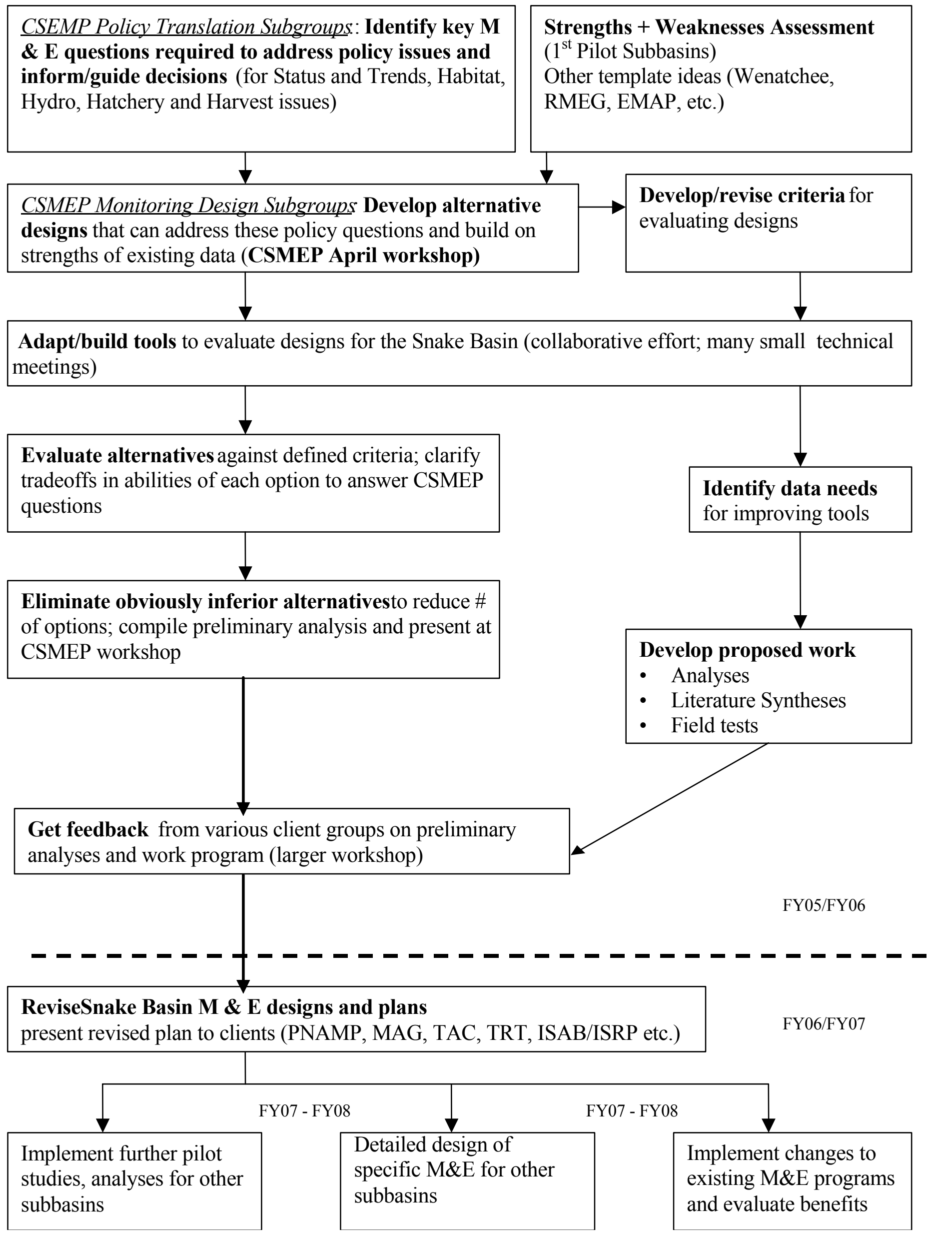


Figure 1.1. Process for CSMEP development of basin-wide M&E designs—FY 2005/2006 and beyond.

1.2 Challenges

Development of an effective, integrated systemwide M&E program for the Basin requires long term commitment and cooperation by all participating state, federal and tribal entities. The challenges are many and include:

- getting buy-in from all M&E entities;
- creating M&E alternatives to address multiple agency data needs at multiple spatial / temporal scales;

- determining the trade-offs that make most sense; and
- getting sufficient replication of consistent, high quality monitoring data to better inform Basin decisions.

A more complete summary of the range of policy/programmatic and technical challenges that face development of integrated M&E for status and trends and effectiveness monitoring are presented in Table 1.1.

Table 1.1. Policy, technical and field challenges to development of an effective **systemwide** status and action effectiveness monitoring program (adapted from Jordan et al. 2002, CSMEP 2002). Challenges which CSMEP began to address in CSMEP's Phase 1 (FY2004-06) are indicated by a "*". These challenges, plus those indicated with a "+", are intended to be addressed in Phase 2 (FY2007-08).

Policy / Programmatic Challenges	Technical and On-the Ground Challenges
Unspecified level of acceptable uncertainty for decision making, and lack of clear decision criteria	Existing monitoring efforts are not catalogued*
Cooperation of necessary private, local, state, tribal, and federal jurisdictions is difficult to achieve*	Quantitative information on data quality (accuracy and precision) often is unavailable.
Entities have different scopes of responsibility and authority, and different priorities for monitoring information*	No concise, clearly described basin-wide monitoring program presently exists
Entities often have no mandate for supporting regional programs	Non-random index sites for trend monitoring precludes inferences to larger scales*
Different entities and programs operate at different spatial and temporal scales, from project-scale evaluations to high level indicators at provincial scales*	Specific monitoring responsibilities need to be assigned to, and accepted by, multiple entities*
Perceived high cost *	Data management technology is evolving rapidly; various entities have different levels of ability and available resources.
Insufficient technical feedback to policy makers*	Lack of integration of monitoring designs across spatial scales, life history stages, and M&E domains (i.e., status and trend, action effectiveness of habitat, hatchery, harvest and hydro actions)*
Inaccurate perception that effects of management actions are well understood by scientists and project implementers*	No systematic process for evaluating the tradeoffs among different monitoring designs for meeting competing M&E objectives*
Lack of coordination in regional implementation of management actions and highly constrained management regimes results in low contrasts in actions, and poor ability to evaluate effectiveness +	Coordinating field crews from multiple agencies is operationally difficult
Tension of bottom-up, local implementation of restoration actions vs. more top-down regional implementation to maximize rate of learning +	No common protocols / manuals for collecting field data
	Field crews often do not have time for data entry and QA/QC activities
	Lack of documentation of actual implementation of habitat restoration actions
	Poor inferences on action effectiveness due to inadequately framed hypotheses, insufficient spatial/temporal contrast in management actions (effect sizes), insufficient duration of monitoring and inability to account for confounding covariates.*
	Non-random allocation of management actions in space and time limits inferences on action effectiveness; +

During the past four years considerable progress has been made by a variety of entities on understanding, and beginning to address, many of the technical challenges in Table 1.1. Over this period various groups both within and outside the Basin have provided useful guidance for both status and trend monitoring, and action effectiveness monitoring (e.g., Jordan et al. 2003; CSMEP 2004, 2005; NWPCC 2005; ISAB 2003, 2005; ISRP 2005; ISRP/ISAB 2005a, 2005b; Marmorek et al. 2004a; Paulsen and Fisher 2003; Porter and Marmorek 2004, 2005; Roni et al. 2005; Bradford et al. 2005). Pilot projects in the Upper Columbia, John Day, and Salmon subbasins (Jordan et al. 2003, Hillman 2004), as well as the Oregon Plan coastal coho monitoring (Stevens 2002), are exploring the effectiveness of alternative sampling designs and monitoring protocols. The State of Washington's Governor's Forum on Monitoring has developed guidelines for monitoring (WA Dec 2005) and implemented a project to assess the effectiveness of a representative subset of different types of habitat restoration projects at a reach scale (SFRB 2003b). PNAMP is currently reviewing fish monitoring protocols (Johnson and O'Neal, in prep) building on similar work for habitat protocols (Johnson et al. 2001), and is facilitating agency field comparisons of protocols for assessing habitat attributes.

Each of the above efforts has developed various tools and insights valuable to the long term development of effective M&E programs. With such a rapidly evolving landscape and a patchwork quilt of entities, coordination and co-evolution is critical. CSMEP work plans, tasks and products have been carefully designed to build on, complement and interact with other ongoing M&E efforts. CSMEP seeks opportunities to engage with existing efforts and established forums to maximize both our learning from their experience, and the catalytic benefits of our work to others. Together with these other efforts, CSMEP work products have made major contributions towards addressing many of the technical challenges in Table 1.1.

On the policy side of the table, the development of the PNAMP Charter has begun a process of collaboration on M&E issues at the policy and programmatic level amongst some of the key entities in the Pacific Northwest. The development of subbasin plans, led by the NWPCC, has catalyzed cooperation *within* many subbasins, though much stronger levels of coordination both within and *among* subbasins will be required for effective design and implementation of systemwide M&E programs.

Despite the considerable progress made over the last several years by both CSMEP and other entities, there remain considerable challenges to be overcome, on both the technical and policy sides of Table 1.1, before the Basin can implement an effective and affordable M&E program. The Columbia Basin is a huge area with a very complex set of jurisdictions and entities. The most feasible strategy for making progress on these challenges is to incrementally learn from successive pilot projects on sub-basin or ESU scales, while at the same time addressing M&E issues that operate on larger scales (e.g., tracking survival through successive life stages for understanding both limiting factors and action effectiveness). This has been called a "tile the basin" strategy and is the strategy that CSMEP has been pursuing and hopes to continue in to FY2006-08. Sub-basin and ESU scales are large enough to force consideration of various integration issues, yet small enough to develop a good working effort among participating agencies. Each pilot project provides insights that build knowledge for subsequent efforts.

2. Summary of Progress on M&E Designs in FY2006

Overview

CSMEP has developed a set of strategies and general principles to meet the challenge of integrating multiple M&E objectives for the Basin:

1. involve federal, state, tribal and local entities in the collaborative development of M&E designs for multiple scales, questions and species, closely coordinating to ensure no duplication of effort;
2. survey managers and policy people to ascertain their relative priorities for different questions, scales, and species;
3. use *decisions* as the starting point for developing sampling, response and evaluation designs[4], rather than *questions*, which permits a more rigorous assessment of the exact inputs and level of precision required in monitoring data, and the risks of making different types of decision errors (Marmorek et al. 2005); and
4. recognize that M&E designs inevitably involve tradeoffs across a number of design objectives and evaluation criteria, and attempt to address these tradeoffs explicitly.

Three multi-agency monitoring design workshops were undertaken in FY2006 to further explore how best to integrate the strengths of existing monitoring, together with novel approaches that help to deal with their weaknesses. CSMEP has continued to evaluate the ability of varied approaches to answer the questions in Appendix A, and is attempting to lay out a structured approach to evaluating the costs, benefits and tradeoffs of different M&E strategies (i.e., ProAct Process—see Section 2.6). The CSMEP design process is fully outlined in the Proposed Evaluation and Design of Preliminary Design Templates (Parnell et al. 2005) document available on the CSMEP website. As a pilot example of this design process CSMEP has focused their efforts to date principally on the Snake River Basin spring/summer/fall Chinook ESUs (see map of the Columbia River subbasin areas encompassed by the CSMEP pilot); this pilot exercise is however intended to illustrate the steps that will be required for development of an integrated monitoring program across the Columbia River Basin.

CSMEP has been using the 7-step EPA Data Quality Objectives (DQO) process to rigorously connect policy decisions and the M&E designs that provide the input for these decisions (Table 2.1). The DQO process forces rigor: clarification of the critical management decisions to be made in the Columbia River Basin, the alternative evaluation approaches to those decisions, the performance measures required to feed those evaluation approaches, and the sampling options available to generate data for the key performance measures. In FY2006, the five CSMEP subgroups (Status and Trends, Habitat, Harvest, Hydro and Hatcheries—see Table 2.2 for participants in each subgroup) have continued their work on applying the DQO process to develop a set of robust M&E designs for evaluating both the status and trends of fish populations and the effectiveness of habitat, harvest, hatchery and hydrosystem recovery actions. That is, what are the M&E alternatives for answering the questions laid out in Appendix A, how well can each option answer those questions, and at what cost? What are the risks of not answering certain questions well? The draft results of steps 1-5 of the DQO process for the spring/summer Chinook

[4] *Sampling designs* refer to the selection of locations and times to sample, *response designs* to what is monitored (and how) at those locations and times, and *evaluation designs* to the analytical methods used on the data to make a decision or answer a question which feeds into a decision.

ESU pilot exercise for each CSMEP subgroup are available on the CSMEP Website (Marmorek et al. 2005). Participants in each of these CSMEP subgroups on FY2006 are listed in Table 2.2.

Table 2.1. EPA Data Quality Objectives process for developing monitoring and evaluation designs. (Source: **United States Environmental Protection Agency**. 2000. Guidance for the Data Quality Objectives Process. EPA QA/G-4. www.epa.gov/swerust1/cat/epaqag4.pdf

1. State the problem
2. Identify the decision
3. Identify inputs to the decision
4. Define the study boundaries
5. Develop an "if-then" decision rule
6. Specify limits on decision errors (both directions)
7. Optimize the design for obtaining data

CSMEP's focus on developing its M&E designs employing EPA's DQO process is intended to emphasize iterative learning within an adaptive management loop. CSMEP's overall design process should involve the following steps:

1. initial problem *assessment* to make explicit our current understanding of the system, clarify our understanding of management goals in the Columbia Basin, and identify the key uncertainties in evaluating agency management actions;
2. careful *design* of monitoring to evaluate management actions and reduce the key uncertainties;
3. *monitoring* of key performance measures to test key management hypotheses and assess progress towards management goals;
4. *evaluation* of monitoring results against the goals defined in the assessment phase; and
5. *adjustments* in our understanding of the system and the effects of management actions; and proceed back to step 1.

In some cases (i.e., the Habitat subgroup) CSMEP's attempts to work through the DQO process have identified where this adaptive management loop can break down for Basin M&E efforts (i.e., an inability to translate non-quantitative management goals into the quantitative performance measures that are required to inform this process, and/or a limited understanding of the mechanistic linkages required to anticipate how the system might respond to actions). For example, riparian areas might be replanted to decrease water temperature, but rarely do we see the goal identified quantitatively (e.g., riparian plantings will decrease stream temperatures by three degrees Celsius within a period of five years). Within CSMEP we are still grappling with how best to incorporate non-quantitative goals into the adaptive management loop.

The preliminary M&E alternatives developed for the Snake River Basin pilot have been presented to fish and wildlife analysts/managers in FY2006 by the various CSMEP subgroups (i.e., Habitat subgroup to the Lemhi HCP, Harvest subgroup to the TAC, Status & Trends subgroup to regional TRT, Hydro subgroup to the BiOP Remand groups, Hatchery subgroup to the Adhoc Hatchery Supplementation Subcommittee). Efforts have also been made to obtain additional feedback from regional policy representatives on the key management decisions (and associated questions) through continuing regional workshops, and surveys undertaken by both CSMEP (key findings of the CSMEP survey are presented in Appendix B of this report) and PNAMP . In FY2007-08, CSMEP intends to work with agency partners to expand our pilot

DQO design efforts in the Snake Basin and develop integrated M&E guidance applicable to other subbasins, both within the Columbia River Basin and potentially to other areas served by PNAMP.

Table 2.2. Participants in each of the CSMEP design subgroups in FY2006. Individuals with bold italicized names are the designated subgroup leaders.

I) **Status and Trends** of Listed Species/Stocks for Extinction Risks and Recovery Evaluations:

Chris Jordan (NOAA), Sam Sharr (IDFG), Claire McGrath (IDFG), Frank Young (CBFWA), Charlie Petrosky (IDFG), Eric Tinus (ODFW), Pete Hahn (WDFW), Paul Wilson (USFWS), Charlie Paulsen (Paulsen Environmental Research), Nick Bouwes (Eco Logical Research), Phil Larsen (EPA), Dave Marmorek (ESSA), Tim Dalton (ODFW), Marc Porter (ESSA), Darcy Pickard (ESSA)

II) Effects of **Habitat** Restoration Actions:

Steve Katz (NOAA), Keith Wolf (KWA-Colville.), Chris Beasley (NP-Quantitative Consultants), Charlie Paulsen (Paulsen Environmental Research), Tim Copeland (IDFG), Nick Bouwes (Eco Logical Research), Robert Al_Chokhachy (Eco Logical Research), Marc Porter (ESSA)

III) Effects of **Hydrosystem** Operations:

Charlie Petrosky (IDFG), Earl Weber (CRITFC), Paul Wilson (USFWS), Charlie Paulsen (Paulsen Environmental Research), Nick Bouwes (EcoLogical Research), Frank Young (CBFWA), Tom Berggren (FPC), David Marmorek (ESSA)

IV) Effects of **Hatchery** Operations:

Chris Beasley (NP-Quantitative Consultants), Peter Galbraith (CRITFC), Dave Fast (YN), Bill Bosch (YN), Jay Hesse (NP), Tim Dalton (ODFW), Pete Hahn (WDFW), Marc Porter (ESSA)

V) Effects of **Harvest** Management Decisions

Tom Rien (ODFW), Eric Tinus (ODFW), Jeff Fryer (CRITFC), Sam Sharr (IDFG), Kris Ryding (WDFW), Stuart Ellis (CRITFC), Saang-Yun Hyun (CRITFC), Marc Porter (ESSA)

2.1 Status & trends

CSMEP's Status and Trends subgroup has focused its DQO efforts on identifying monitoring design elements necessary to address one of the most important management decisions in the Snake Basin: has there been sufficient improvement in status of Snake River S/S Chinook to justify delisting the ESU and allow removal of ESA restrictions? This decision is based on the abundance, productivity, spatial structure and diversity of SRSS Chinook salmon over the prior 10 years (IC-TRT 2005). A full description of the subgroup's work on DQO steps 1-5 for the Snake pilot is presented as a chapter in Marmorek et al. 2005. A brief PowerPoint presentation describing the subgroup's DQO steps 1-5 is also provided on the CSMEP website.

In FY2006 as part of its work on DQO steps 6 and 7, the S &T subgroup continued development of a simulation model that can be used for evaluating alternative designs for monitoring fish at the population, major population group and ESU scales. Monitoring designs are described as high, medium, and low templates, in reference to the levels of accuracy and precision in data that are collected using each template. Alternative design templates will be compared in terms of cost ($/yr) and probability of error in decisions that are associated with individual templates. A presentation of the subgroup's progress on this model in FY2006 was given recently to the Interior TRT and is available on the CSMEP website.

2.1.1 Status & trends simulation model

The immediate objective of this model is to evaluate alternative design templates for determining the status of Snake River Spring/Summer (SRSS) Chinook salmon, while limiting risk in the decision to acceptable levels. The ultimate objective is to develop a tool that can be adapted for monitoring designs in other basins and for other species.

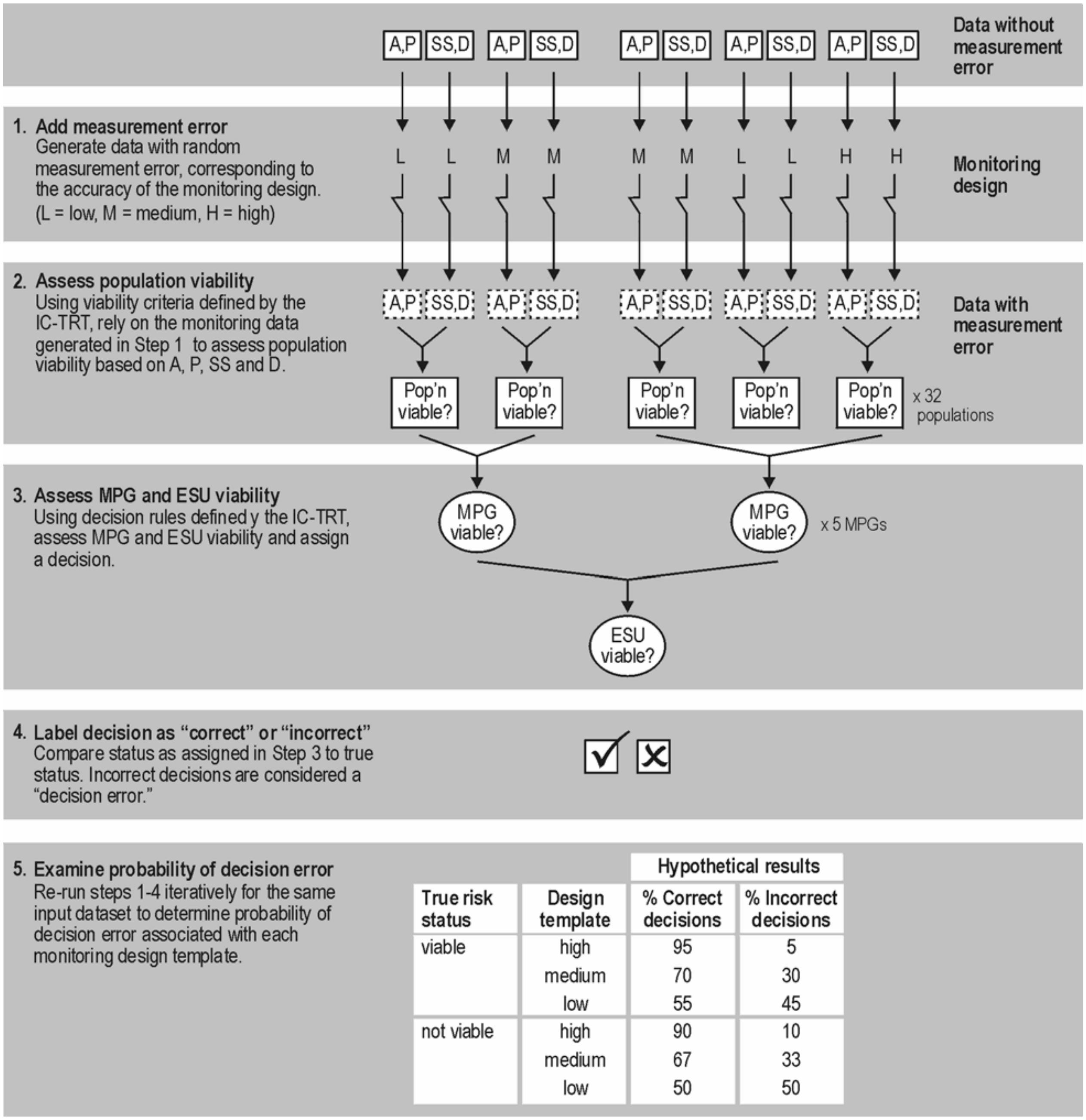
The Interior Columbia Technical Recovery Team (TRT) developed viability criteria for application to the Interior Columbia Basin salmonid ESUs (July 2005). The viability criteria are based on four types of information: abundance, productivity, spatial structure and diversity (McElhany et al. 2000). The TRT criteria define rules (Appendix C) for taking this information at the population scale and assessing the viability at the population, major population group (MPG) and ESU scale.

Methods

Model Overview:

The simulation model can be used to evaluate the sensitivity of the TRT viability assessment process to alternative designs for monitoring fish populations (Figure 2.1.1). Sensitivity analysis will be used to examine how the rate and direction (type I versus type II) of decision error responds to uncertainty in monitoring data.

The simulation model uses data for abundance, productivity, spatial structure and diversity for populations with known viability status and then adds variability to these data on each run of the simulation to produce data with measurement error corresponding to alternative monitoring designs. The TRT rules are then applied to the data with measurement error and, subsequently, the ability to correctly determine the viability status is assessed (Figure 2.1.1).



True risk status	Design template	% Correct decisions	% Incorrect decisions
viable	high	95	5
	medium	70	30
	low	55	45
not viable	high	90	10
	medium	67	33
	low	50	50

Figure 2.1.1. Flow chart of the simulation process.

Test Datasets:

Simulated data that resemble currently collected monitoring indicators were generated to allow testing of the viability decision criteria processes. The viability decision process requires the input of 14 VSP metrics for each population for each year. The required VSP metrics are Abundance, Productivity, the eight Spatial Structure and Diversity Factors with unique Metrics, and the four Metrics of Factor B.2 (ICTRT Viability Criteria document, Table 12).

A time series of Abundance was generated to represent each population in the test case ESU (Snake River Spring/Summer Chinook). The abundance time series were 60 years in length, and were generated as independent instances of a population process engine. The population process engine was a two stage population model designed by the ICTRT to mimic time series data that are representative of current stage specific population processes. The two stage model uses an input spawner abundance and a Beverton-Holt smolt production function based on a SAR or 1%, a smolt capacity of 100,000 and a B-H productivity term of 216, to generate an estimate of brood year smolt production. The smolt production estimate is multiplied by a random variate (lognormal, median = 0, variance = 0.443; parameters estimated from Chiwawa River population data), and a year specific SAR that mimics the variance and autocorrelation of the SAR time series from Lower Granite Dam. This process is repeated iteratively to generate time series of spawners that mimic natural population processes.

Productivity time series were generated for each of the Abundance time series in the manner of simple Recruits per Spawner run reconstructions. Assuming an age structure of 50:50 for 4 and 5 year olds, for each run year the corresponding recruits were accumulated by brood year. When the run year abundance was <10% of the time series average, a productivity was not calculated as these productivities tended to be artificially high, and thus biased the distribution of the metric. The ICTRT also recommended not calculating productivities for Abundance values >75% of the population threshold, since these productivities would be reduced due to the density dependent feedback inherent in the Beverton-Holt production function.

All 12 Spatial Structure and Diversity Metrics / Factors generated for the viability decision simulations were done at the "risk" level, that is each Metric/Factor was scored [-1, 0, 1, 2] for each population for each year, representing a risk score for that Metric/Factor x Population x Year of High, Moderate, Low or Very low, respectively. Since the risk scores for each Metric/Factor x Population x Year is a somewhat abstract categorization of actual monitoring data based on the ICTRT Spatial Structure and Diversity criteria, there are no existing time series of these values for Snake River Spring/Summer chinook populations as there would be for Abundance and Productivity. Therefore, there aren't any examples of what the temporal or spatial pattern of variability might be for these metrics. As a result, the 60 year simulated data set had to be generated by hand based on the current Status Assessments for each populations (e.g., worked examples), constraints due to the rule set (e.g., some populations can never reach Very Low risk levels due to the geographic distribution of m/MSA) and the following guiding principles: populations within MPGs were more like each other in risk score than between MPGs; risk scores could vary annually, but generally were held constant for 5–10 year runs; risk scores changed only one level at a time, e.g. 0 to -1 or 1, not 2; each of the five MPGs were given an over all expected Viability score, and Metric/Factor x Population x Year risk scores were generated to preserve this expectation through time.

Design Overview:

We consider three possible monitoring methods at the population scale: L = low quality data, M = moderate quality data, and H = high quality data. The user specifies the level of accuracy that corresponds to each of the L, M, and H monitoring methods. The monitoring designs describe what method (L, M, H) will be applied to each of the populations in the ESU. The initial simulation analysis compares three very simple alternative monitoring designs:

Design 1: All 32 populations are monitored using L monitoring methods.

Design 2: All 32 populations are monitored using M monitoring methods.

Design 3: All 32 populations are monitored using H monitoring methods.

In the present report we provide preliminary results from Designs 1-3 described above. In future work, more complicated designs can be compared (see maps of possible alternative designs suggested in FY2005).

Noise Overview:

The monitoring design defines the measurement error or noise that will be added to the data at the population level. The structure of the noise added to the input data depends on the type of data. For abundance data, noise with a log-normal distribution is generated and multiplied by the true abundance data in order to simulate abundance data with measurement error (Figure 2.1.2a). For spatial structure and diversity data, a probability transition matrix is used to define the conditional probability of classifying the data in each of the 4 categories given the truth. These multinomial probabilities are applied to the true SSD risk data in order to simulate categorical data with occasional misclassification due to measurement error (Figure 2.1.2b).

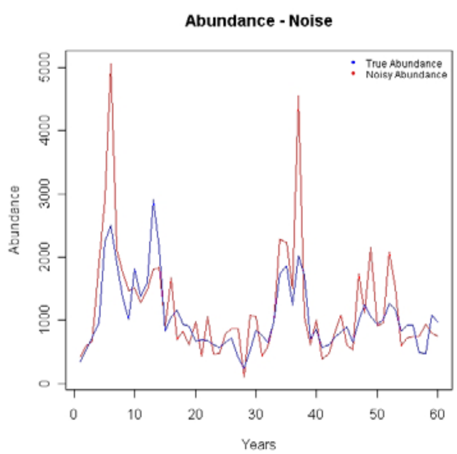


Figure 2.1.2a. Example of random noise added to abundance data.

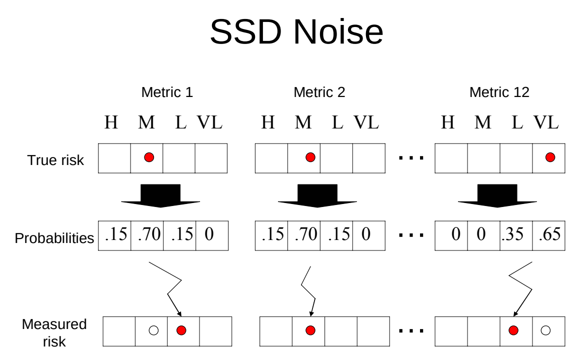


Figure 2.1.2b. Example of random noise added to Spatial Structure and Diversity data.

Abundance is often well approximated by a log-normal distribution, as abundance data must be non-negative and will often have a long right tail. We assumed that the measurement or observation error was also log-normally distributed. Simulating data with random observation error or noise included was completed by generating noise that was log-normally distributed with an expected value of 1, and then multiplying the true data by this noise (eq. 2.1.1). This results in noisy abundance data whose expected value is the same as that of the true abundance data (this is what we would expect with unbiased monitoring methods). In future work, we could also consider allowing for biased monitoring methods or other error distributions. The variability of the noise was determined by a user input coefficient of variation (CV). The observed abundance, N_{obs}, was generated as per equation 2.1.1 (after Hilborn and Mangel 1997, eq. 7.33).

$$N_{obs,t} = N_t \exp(Z\sigma_{V,t} - \sigma^2_{V,t} / 2) \qquad \text{equation 2.1.1}$$

where Z is a normally distributed random variable with mean 0 and standard deviation 1, and $\sigma_{V,t}$ = standard deviation of observation error in year t. $\sigma_{V,t}$ was calculated based on the user input CV for year t and the mean abundance over the previous ten years.

Productivity is calculated from abundance and age-structure information. We calculate the noisy productivity from the noisy abundance data and population specific age-structure information. In future, we would like to test the effect of adding noise to the age-structure information as well. Occasionally, the abundance temporarily falls to zero or near zero. In these years it isn't appropriate to calculate the productivity since any small errors in the age-structure information may lead to the productivity being inflated unreasonably. We chose not to calculate the productivity in years where the abundance was less than 10% of the mean abundance for the population.

The **spatial structure and diversity** risk level assessment differs from the abundance/productivity risk level assessment. Twelve different metrics are evaluated for each population. It is difficult to define the exact data required to assess risk for each of the 12 metrics. Each metric has a series of complex questions requiring a range of data along with expert opinion to evaluate them. Since the data themselves are difficult to define, adding noise to the raw data is not straightforward. Instead, we consider the 12 metrics as the input data. Each of the 12 metrics can belong to one of 4 possible risk categories: very low (VL), low (L), moderate (M) or high (H). The input dataset with known viability status is a time-series of risk categories for each of the 12 metrics in each of the 32 populations. Since there are only 4 possible outcomes for each metric, the data for each metric can be described as ~multinomial (n, p_1, p_2, p_3, p_4). Depending on the monitoring methods used, the probability (p_i) of choosing category i changes. More precise monitoring methods will result in a greater probability of choosing the correct category and less precise methods will spread the probability among the other categories. A probability transition matrix is used to define the probability of classifying the data in each of the 4 categories given the truth. Figure 2.1.3 illustrates an example of a probability transition matrix. If the true risk level is low, then according to this probability transition matrix 70% of the time you would assess the correct risk level, but 15% of the time you would overestimate the risk and 15% of the time you would underestimate the risk. A different probability transition matrix was used for each of the three monitoring designs and for each of the metrics in each of the populations.

Probability Transition Matrix

SSD True Risk \ SSD Measured Risk	H	M	L	VL
H	.85	0.15	0	0
M	0.15	0.7	0.15	0
L	0	0.15	0.7	0.15
VL	0	0	0.15	0.85

Figure 2.1.3. Example of a possible probability transition matrix for defining risk classifications.

For each population, we noted when a specific risk category was not possible and ensured that adding error to the risk category for SS/D would not place the population into an "impossible" risk category. To determine these impossible risk categories, we referred to information on attributes of individual populations including: 1) historical population characteristics (e.g., population size, configuration, number of major spawning areas and other areas with spawners); 2) location of hatcheries; and 3) number of historic and current ecoregions occupied .

Next, for each individual metric we assigned miscategorization rates for each population. To quantify the distribution of random and directional error, we used several sources of information for guidance, including:

1. published data and grey literature that describe accuracy and precision of monitoring data for Chinook salmon;
2. professional judgment from Idaho Department of Fish and Game (IDFG) fishery biologists on the amount of error in SS/D data using status quo monitoring methods for Chamberlain Creek as an example population (C. Petrosky, P. Hassemer, S. Sharr, and J. White, personal communications).
3. population characteristics including presence of hatcheries, human influence, and habitat alteration (e.g., error in hatchery stray rates is low with little or no hatchery influence).

For each metric, SS/D input data sets never assigned a population to an impossible risk category and subsequent addition of error to the input data also never assigned a population to an impossible risk category. The magnitude and direction of error added to input data was greatest for the “low” design and was minimized in the “high” monitoring design. The error structures were deemed as realistic estimates for monitoring using currently available methods of monitoring for SS and D metrics.

Results

Population level results:

There are multiple ways to achieve a particular viability status at the population level. Table 2.1.1 determines the viability status resulting from each of the 16 possible combinations of AP and SSD risk levels (Table 13 of the TRT July 05 viability draft, reprinted with revisions described by Pete Hassemer, personal communication).

Table 2.1.1. Shows the viability status for all 16 possible A/P and SS/D risk combinations. There are four possible outcomes: HV= highly viable, V= viable, M= maintained, and NV= not viable.

		SS/D risk			
		Very Low (VL)	Low (L)	Moderate (M)	High (H)
A/P risk	Very Low (VL) <1%	HV (1)	HV (2)	V (3)	M (4)
A/P risk	Low (L) 5%	V (5)	V (6)	V (7)	M (8)
A/P risk	Moderate (M) 25%	M (9)	M (10)	M (11)	NV (12)
A/P risk	High (H) >25%	NV (13)	NV (14)	NV (15)	NV (16)

Every run of the simulation assigns a viability status to each population in each decision year. The results for a single population can be plotted over all runs and all decision years to provide information about how the A/P and SS/D risk levels affect the viability status. An example is shown for a single population in Figure 2.1.4 for each of the three monitoring designs discussed in Section 2.1. In this example, the population’s true viability status is ‘Maintained’ for 3 decision years and ‘Not Viable’ for 2 decision years (Table 2.1.2).

Table 2.1.2. True risk levels for example population.

Decision Year	A/P risk level	SS/D risk level	Table 2.1.1 cell # *	Viability status
1	High	Low	14	Not Viable
2	Moderate	Low	10	Maintained
3	Moderate	Low	10	Maintained
4	High	Low	14	Not Viable
5	Moderate	Very Low	9	Maintained

* Table 2.2.1 cell #'s are assigned from left to right, top to bottom

When the data are collected using the 'Low' monitoring design the 'Not Viable' status is assigned more often than expected. Figure 2.1.4a suggests that the SS/D risk level is being overestimated and causing the error in viability assessment. Both the 'Moderate' and the 'High' monitoring designs result in roughly the correct overall viability assessment (60% Maintained : 40% Not Viable). Even though the 'High' design results in the correct Table 2.1.1 cell more frequently than the 'Moderate' design, the overall viability assessment is approximately the same. In this case, the 'High' design provides no obvious advantage in terms of overall viability assessment over the 'Moderate' design. The 'Low' design is substantially worse at correctly assigning viability status than the 'Moderate' and 'High' designs, but the error is almost always in the conservative direction. Figure 2.1.4 also indicates that this population is limited to 'Maintained' or 'Not Viable' status due to the high A/P risk level, not the SS/D risk level.

We are also interested in summarizing how often we make the correct viability assessment using each monitoring design. In order to do this, we compare the measured viability status to the viability status with no measurement error, for each population and each decision year in each simulation. We summarize the results by plotting how often the measured viability status was the same as the true viability status over all populations, decision years and simulation runs. The initial results are shown in Figure 2.1.5, the on-diagonal bars represent correct viability assessments and the off-diagonal bars are incorrect. For example, in Figure 2.1.5a) there were ~3500 instances where the true viability was 'Not Viable', but ~500 of those times the measured viability status was incorrectly found to be 'Maintained' instead.

As expected, the overall % of correct viability assessments increase with increased monitoring levels (L=75%, M=87%, H=94%). This plot also indicates where the errors tend to occur. For this particular test dataset and monitoring design definition, we see that in general and particularly for the 'Low' monitoring design the TRT rules tend to err on the conservative side.

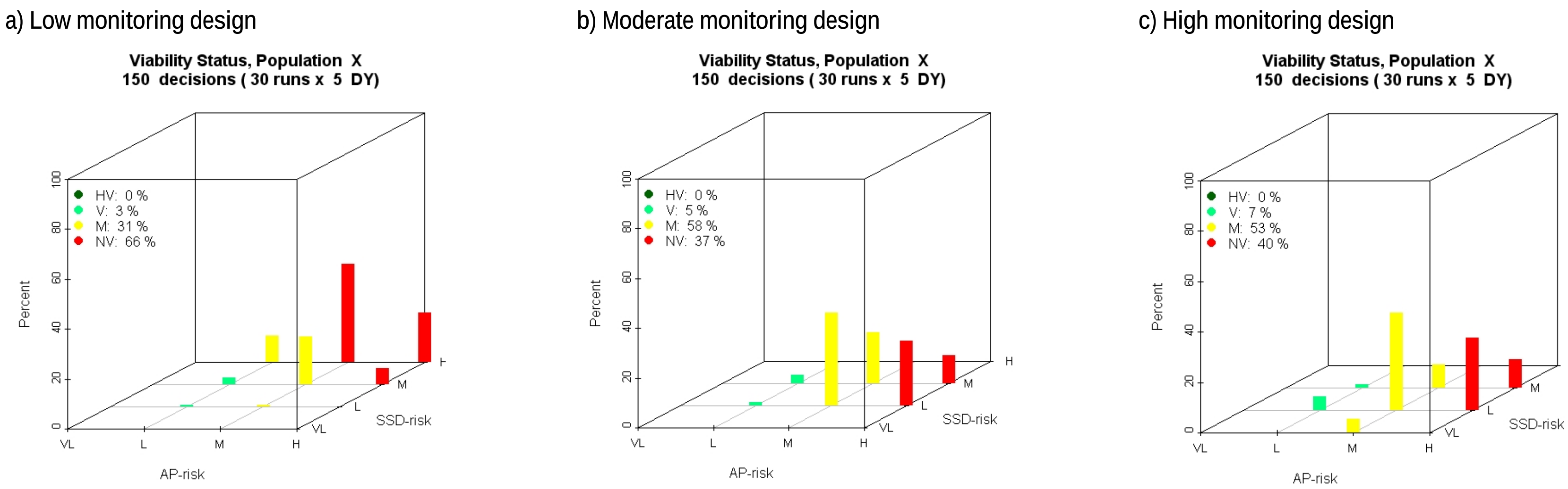


Figure 2.1.4. Example of how the viability status might change with increased monitoring effort.

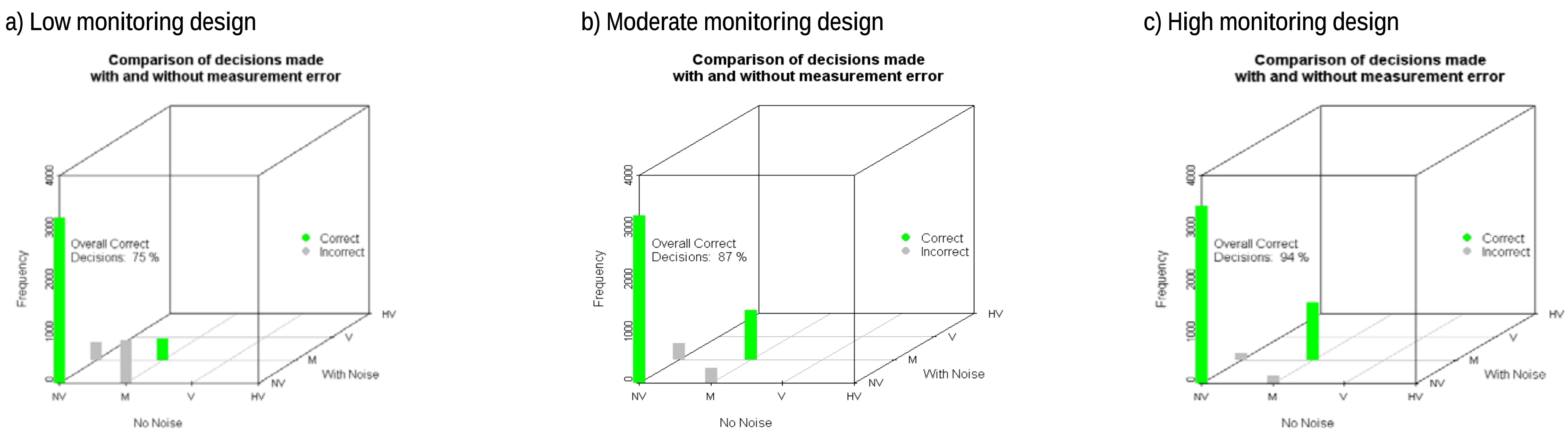


Figure 2.1.5. Example of how often we correctly identify the viability status with alternative monitoring designs. The “No Noise” axis represents the true viability category, while the “With Noise” axis represents the measured viability category. The “Frequency” axis represents the frequency of decisions that fall into each of the 16 possible combinations. For example, the height of the bottom left bar, indicates how many times the measured viability category was “Not Viable” given that the true viability category was “Not Viable”. Recall: Not Viable = NV, Maintained = M, Viable = V, Highly Viable = HV.

MPG level results:

The TRT defines seven MPG specific rules that must all be met for an MPG to be viable (Table 2.1.3). The frequency that each of the rules is violated for each of the monitoring designs can be plotted to describe which MPGs fail to meet viability standards, and which of the seven criteria are limiting the MPG. An example of possible results for a single design over many runs is shown in Figure 2.1.6.

Table 2.1.3. TRT viability criteria for an MPG.

Criteria	Description*
1	At least one population in the MPG must be highly viable (HV)
2	At least half the populations in the MPG must be viable (V)
3	At least x Intermediate – V. Large populations must be viable (V)
4	At least x Large-V. Large populations must be viable (V)
5	At least x spring life history populations must be viable (V)
6	At least x summer life history populations must be viable (V)
7	There can be no populations that are not viable (NV)

* The number of populations required to meet criteria 3 through 6 (represented by an x) is MPG specific.

In this example, Figure 2.1.6a indicates that MPG 2 is never viable, MPG 3 & 5 are rarely found to be viable, but MPG 1 is viable in 60% of the runs and MPG 4 is viable in 80% of the runs. Figure 2.1.6 b-f show how often each of the seven criteria failed for each of the MPGs. The results for this example are summarized in Table 2.1.4.

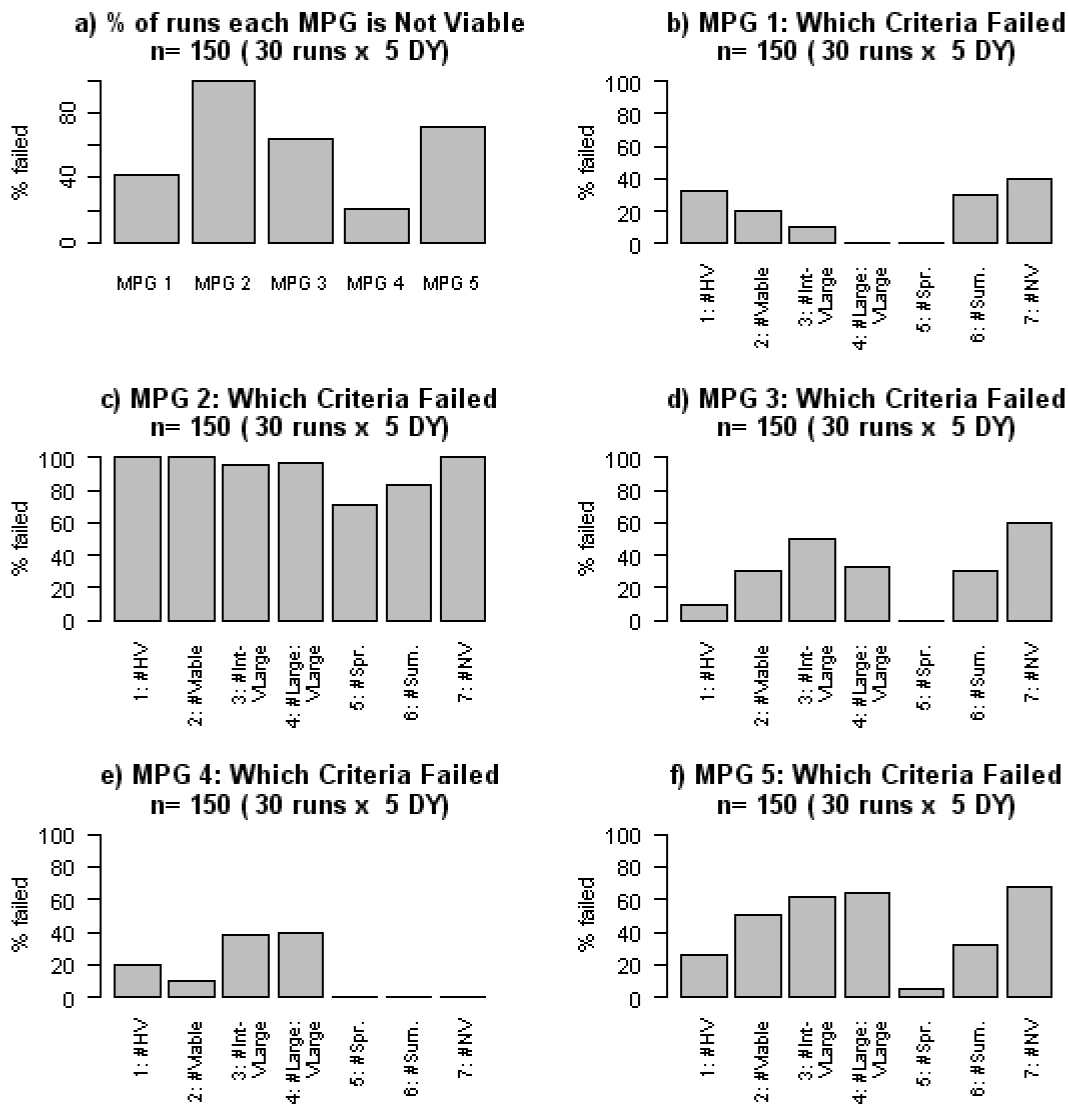


Figure 2.1.6. Example of possible MPG level results.

Table 2.1.4. Summary of MPG level results from Figure 2.1.6.

MPG	Strengths	Weaknesses
1	Enough viable (V) spring and large populations	Often has at least one not viable (NV) population and doesn't have enough viable (V) summer or highly viable (HV) populations
2		Generally fails all seven criteria.
3	Enough viable (V) spring populations, has enough highly viable (HV) populations	Often has at least one not viable (NV) population
4	Always has enough viable (V) spring and summer populations, never has any not viable (NV) populations	Doesn't always have enough viable (V) large or very large populations.
5	Enough viable (V) spring populations, generally has enough highly viable (HV) populations.	Often has at least one not viable (NV) population. Generally doesn't have enough viable (V) large populations.

ESU level results:

For the Snake River Spring/Summer Chinook Salmon ESU, the TRT has decided that all five of the MPGs must be viable, for the ESU to be viable (Pete Hassemer, personal communication). In the example above, MPG 2 failed during 100% of the simulation runs and therefore the ESU would never be considered viable for this example. In future work, we will consider test data sets where we would expect the ESU to be viable and will report how often we correctly assess the ESU as viable given alternative monitoring designs.

Discussion

The model provides a framework to help managers understand the variability in the information used to make decisions about viability status. The process of assessing the effectiveness of current monitoring methods is in itself a very useful tool. This information can help the manager to determine where it is feasible to improve monitoring methods, and the model can be used to test how much value would be gained by making those improvements. The model can evaluate the sensitivity of the TRT viability criteria to changes in the quality of data. The tool is flexible and information specific to the ESU can be used.

A literature review of ~50 papers found that there was very little information available specifically addressing the problem of assessing measurement error due to various spawner abundance assessment methods, and even where studies had been completed there was little consistency in the results. It is likely that the error in methods varies dramatically by site, e.g., weirs may be unreliable in streams with high or fluctuating flow, inter-observer variability may be high (Dunham et al 2001, Jones et al 1998). We decided it was not sensible to try to relate the specific method (e.g., redd count vs. snorkel count) to data quality in the model. Instead, the model can tell us 'the correct viability assessment is made 80% of the time, given a specified CV. Subsequently, individual scientists in the field can determine how to achieve that CV depending on the conditions of a specific site.

In future work, we will consider test data sets where we would expect the ESU to be viable and will report how often we correctly assess the ESU as viable given alternative monitoring designs. We would also like to consider more complicated monitoring designs, such as the designs proposed in the CSMEP FY05 annual report or comparing three alternative designs with the same cost of implementation. We plan to compare the status quo monitoring designs in the Snake River Spring/Summer Chinook ESU to our Low, Moderate and High designs. We can look at biased monitoring methods and/or different error structures based on feedback from the TRT and other CSMEP scientists. Finally, we'd like to adapt this model as necessary in order to apply it to other ESUs.

2.1.2 Assessment of monitoring approaches for evaluating TRT population viability assessment criteria: Oregon mid-Columbia steelhead monitoring—proposed scope of work in FY2007

In assessing viability of Interior Columbia Technical Recovery Team (ICTRT) defined populations of Mid-Columbia steelhead in Oregon, Carmichael (2005) estimated annual spawner abundance, and thus spawner-to-spawner productivity. Occasionally, complete passage block counting facilities were in place, or mark/recapture estimates were made using partial passage block facilities and weirs. For one population, total known spawning habitat was divided into five mile sections, from each of which one randomly selected mile was surveyed and these observed redd densities were expanded accordingly. Otherwise, abundance estimates employed redd counts from index surveys, which have been commonly used to monitor steelhead (as well as Chinook salmon) spawner abundance in the Mid-Columbia, as well as other areas, for several decades. These surveys are likely to be continued in order to take advantage of the long-term databases. Thus, they can supply redd density data, and occasionally information on live adult numbers and hatchery/wild proportions, usually employing current personnel positions and funding sources.

Carmichael (2005) expanded average annual redd densities from index surveys over the various populations' historic spawning areas, delineated by expert opinion, GIS analysis and identified blocks to migration. In most cases, all reaches were rated for intrinsic spawning habitat quality, yielding a weighted area of spawning habitat for each reach, and these were summed for each population. However, an Environmental Monitoring and Assessment Program (EMAP) protocol has been implemented for the past three years to estimate steelhead redd density and numbers for the John Day subbasin. The EMAP results were also used in estimating adult numbers for the five subbasin populations of this basin. For each population, the average index survey redd density was adjusted downward by multiplying by the ratio between the average density of all EMAP and all index surveys conducted throughout the basin.

The EMAP program in the John Day subbasin provides an excellent opportunity to examine the accuracy of the Mid-Columbia index survey expansion estimation methods. We will be able to compare the efficacy of EMAP and index surveying protocols for Status and Trends assessments, and determine whether the two protocols can perhaps complement each other. Each protocol has its own strengths and weaknesses. EMAP surveys are chosen randomly from among all known surveyable spawning areas, but a good degree of spatial balance among the populations (subbasins) is guaranteed by stratification in the selection process. Index surveys were selected many years ago, also for dispersal throughout the basin, as well as for ease of access and relatively high densities of spawners. There is a small amount of variation among index surveys performed each year.

The intrinsic spawning habitat quality model employs GIS analysis of stream width, gradient and valley width, and depth is currently being added as a factor in the weighting. Using this model, the areas (square meters) of all reaches in which steelhead (or Chinook) are distributed within a subbasin are multiplied by a weighting factor (High = 1.0; Medium = 0.5; Low = 0.25; None =0.0) and the resultant weighted-reach areas are summed for the population. Using index surveys, estimated spawner abundance is the product of weighted average index survey redd density in square meters (i.e., from application of the intrinsic spawning habitat quality model), total weighted spawning area for the population, and a standard average fish/redd ratio conversion factor. Thus, the accuracy of the spawner-abundance estimates depends on the strength and consistency of the spawning habitat weighting model.

The combined EMAP and index-survey sampling in the John Day subbasin also affords an excellent opportunity to test and to better calibrate the intrinsic spawning habitat model, and this should improve redd and spawner estimates. This analysis might also provide for an increase in the number of intrinsic quality levels, further improving accuracy and precision of the estimates. Also, we will be able to test the performance of the calibrated model for streams for which index survey datasets as well as adult estimates are available (complete adult counts at passage facilities are available for the Warm Springs River in the Deschutes subbasin as well as for the Umatilla subbasin). Model calibration employing the John Day data coupled with these performance tests will allow for an iterative process to optimize both calibration and performance.

Some pertinent comparisons between the index surveys and EMAP surveys are shown in Table 2.1.5. In the 2004 EMAP sampling, spatial balance among the five populations was considerably better than for the 15 year average of index surveys, whereas the index surveys averaged almost twice as much stream length coverage. In 2005, however, there was little more spatial balance among the EMAP surveys than among the 15 year index survey average; additionally only one EMAP survey was performed in the South Fork population area. Again, the index surveys covered much greater stream length.

Here, for the purpose of contrasting overall precision of redd density estimates between index and EMAP surveys in the two years, confidence intervals of subbasin-wide mean redd densities for both protocols were calculated using a local neighborhood variance estimator. Index surveys indicated a difference in mean redd densities between the two years at >90% significance, but EMAP surveys indicated a difference equal to only 65% significance. Perhaps more importantly, in both years, the proportion of surveys that were zeroes (i.e., no redds were observed) in each population area was much higher for EMAP surveys than for index surveys. Additionally, in both years there was a population area in which all EMAP surveys were zeroes (besides the South Fork in 2005 which had only one EMAP survey, also a zero). It is perhaps not surprising, considering the current conditions of the populations, that redds are seldom encountered on randomly chosen surveys, and that redd abundances in both years followed negative binomial distributions. This would likely also be the case under probable recovery scenarios. One planned line of inquiry is to conduct analyses of the relationships between sample sizes and precision, as well as power to detect annual differences in mean redd densities, for both protocols and for the entire subbasin as well as the population areas. Of particular interest will be the minimal power and sample sizes associated with detecting the differences between the estimated 10-year average annual population abundances and the threshold abundances necessary to achieve viability for each population (Carmichael 2005). In relation to differing recovery goals for populations, this analysis could involve assessing variations in sampling needs associated with different confidence bound requirements for population viability performance measures. Costs of sampling levels indicated in the above analyses also will be evaluated.

As noted above, the average of the annual EMAP/index survey redd density ratios was used to adjust the spawner abundance expansions from the index survey mean redd densities for each population (i.e., estimated spawners equals the product of average index redds/mile, total miles of spawning reaches, the standard average fish/redd ratio, and the two-year average EMAP/index survey redd density ratio). As part of our analysis, we are also identifying datasets for other populations, basins and species in the Pacific Northwest for which we can compare index survey based fish or redd densities, or spawner abundance estimates, to some presumably unbiased version of "true" density or abundance such as EMAP estimates or weir counts. Currently, we have assessed all Streamnet datasets for Chinook salmon in Washington for utility in this analysis, and among these were found six possibly useful datasets for further review.

Besides the annual EMAP/index survey average redd density ratios for the John Day basin (0.35 and 0.34 for 2004 and 2005, respectively; 2006 data is not complete but will be considerably higher), there are two further examples we have reviewed. For coho salmon along the coast of Oregon from 1990 to 1997, Jacobs and Nickelson (1999) reported that the EMAP/index survey spawner density ratios ranged from about 0.20-0.31, averaging 0.27. For the Warm Springs River (Deschutes subbasin, Mid-Columbia), complete counts of steelhead passed at the Warm Springs Hatchery weir can be compared to spawner abundance estimates made by expanding average index survey redd densities, adjusted for intrinsic habitat quality, over total estimated spawning habitat, also adjusted for habitat quality, and then multiplying by the standard average fish/redd ratio. From 1993-2004, these ”truth”/index ratios ranged from 0.13-0.62 and averaged 0.38. Thus, among these few datasets the averages of these ratios were surprisingly consistent, and variability was similar to or perhaps somewhat less than that often observed spatially and temporally among annual fish/redds ratios.

Empirical fish/redd ratio data for steelhead is infrequent in the Interior Columbia; Carmichael (2005) based his expansions solely on data from Deer Creek in the Wallowas (Grande Ronde subbasin; four years of data currently available). Therefore, we are assessing other streams where such data could be obtained. Within the Mid-Columbia, streams that currently have adult trapping facilities, and that therefore may be useful for this purpose include, Shitike Creek and the Warm Springs River in the Deschutes subbasin, the Umatilla River, and a watershed of the Walla-Walla River.

The ICTRT population viability criteria (ICTRT 2005) also include measurements and indices of spatial structure and diversity. Most of these criteria can be assessed or verified to a lesser or greater extent by spawning surveys. These can demonstrate how many, where, when and which species and races are using certain spawning reaches such as TRT defined Major Spawning Areas (MSAs). Proportions of hatchery fish present can be assessed when live fish or carcasses are observed, and in the latter case biological samples might determine sex, age and origin and can certainly yield genetic information. More information of this nature is likely to be obtained when surveying areas where fish densities are expected to be relatively high, such as in index reaches or reaches rated highly for intrinsic spawning habitat quality by the habitat quality model.

Finally, the above assessments will provide useful information for parameterizing the Status and Trends subgroup simulation model. In our assessment of monitoring methods for Oregon populations of Mid-Columbia steelhead, we shall assess and consider the factors discussed above with the various datasets we currently have and those that we are reviewing. Our goal will be to investigate sampling robustness, as well as cost effectiveness, and to recommend, monitoring methods and plans that will meet various management and, particularly, population viability assessment needs. These will sometimes be similar between or among populations, but in some cases they will be specific to a population, such as when a weir or dam is in place. Few or perhaps no changes may be required in some instances, whereas in others much or complete alteration in methods may be recommended. It is probable that combinations of methods will sometimes be advisable. We plan to pay particular attention to the possibilities of hybrid survey designs (e.g., combinations of index and EMAP surveys) to capitalize on the advantages and strengths of individual monitoring systems.

Table 2.1.5. Comparisons of John Day basin steelhead redd survey results, indexes (15 year average) to EMAP (2004, 2005).

Spatial Balance

Populations (sub-basins)	index 15 yr avg% surveyed	EMAP 05% surveyed	EMAP 04% surveyed
Lower Mainstem (MS)	2.142	1.984	1.745
Upper Mainstem (MS)	2.559	5.471	2.917
North Fork	3.209	2.389	3.149
Middle Fork	10.476	3.635	2.454
South Fork	7.894	0.844	3.438
AVERAGE	5.256	2.865	2.741
VAR	13.843	3.118	0.439
STD DEV	3.721	1.766	0.663
CV	0.708	0.616	0.242

2004

Index	Redds/ mile	Estimates with CI's	Estimate EMAP/ Index	EMAP	Estimates with CI's
	95%	2.298<3.246<4.194	0.351	95%	0.382<1.140<1.898
	90%	2.458<3.246<4.034		65%	0.767<1.140<1.513
MEAN	3.246			1.140	
VAR	7.381			6.813	
STD DEV	2.717			2.610	
STD ERR	0.466			0.377	
CV	0.144			0.330	

	Upper MS	South Fork	Middle Fork	North Fork	Lower MS	Total
Index: zero redds proportion	1/9	1/6	0/4	1/8	1/7	4/34
EMAP: zero redds proportion	7/7	3/4	2/7	14/19	6/11	32/48

2005

Index	Redds/mile	Estimates with CI's	Estimate EMAP/Index	EMAP	Estimates with CI's
	95%	0.969<1.693<2.417	0.342	95%	0.200<0.579<0.958
	90%	1.092<1.693<2.294		65%	0.392<0.579<0.766
MEAN	1.693			0.579	
VAR	3.892			1.776	
STD DEV	1.973			1.333	
STD ERR	0.354			0.189	
CV	0.209			0.326	

	Upper MS	South Fork	Middle Fork	North Fork	Lower MS	Total
Index: zero redds proportion	4/10	2/4	2/4	0/8	2/5	10/31
EMAP: zero redds proportion	9/13	1/1	10/10	9/14	7/12	36/50

Combined 2004 and 2005

	Upper MS	South Fork	Middle Fork	North Fork	Lower MS	Total
index: zero redds proportion	5/18	3/10	2/8	1/16	3/12	14/65
EMAP: zero redds proportion	16/20	4/5	12/17	23/33	13/23	68/98

2.1.3 Monitoring design work to sample multiple fish species

CSMEP's Status and Trends subgroup has begun to draw on lessons learned from existing monitoring activities that address status and trends of fish species in the Snake River Basin in addition to SRB spring/summer Chinook salmon. We reviewed the CSMEP "Table C1" metadata inventories prepared for selected Columbia River Basins, as well as the "Table B2" strengths and weaknesses assessments of existing monitoring activities that provide information on status and trends of Columbia Basin steelhead populations. Additionally, we reviewed previous and ongoing status assessments conducted by NOAA-Fisheries Biological Review Teams and the Interior Columbia Basin Technical Recovery Team (Good et al. 2005; ICBTRT 2005).

To begin to examine how monitoring designs might provide information on Viable Salmonid Population attributes we drafted an overview of monitoring activities for Snake River steelhead populations by data type, Snake River population, and Major Population Groups (MPG) as defined by the ICBTRT (Table 2.1.6). We used the same approach used previously by CSMEP for the SRB spring/summer Chinook ESU (CSMEP 2005). To address the potential of monitoring designs to provide information on additional anadromous species and resident fish species (bull trout), we have begun to assemble information on the distribution by habitat use of the three runs of Snake Basin Chinook salmon, sockeye salmon, steelhead trout, and bull trout (Table 2).

While Table 2.1.6 is a working draft, a very preliminary assessment of the current monitoring of Snake River steelhead populations reveals that not all populations and MPGs are represented with similar data types and collection methods. For example, spawning ground surveys are conducted in streams in the Lower Snake, Grande Ronde and Imnaha MPGs, but not in the Clearwater and Salmon river MPGs. Parr monitoring has occurred in the Imnaha, Clearwater, and Salmon River MPGs, but not in the Lower Snake and Grande Ronde MPGs. Approaches to monitor adult abundance may necessarily vary across MPGs for logistical reasons. Winter and spring freshet conditions may allow spawning ground surveys in some MPGs but not in others.

Designing monitoring approaches that provide information on multiple species would need to address life history differences, spatial distributions, as well as the temporal distribution of species by life stage. For example, while age information can be collected from Chinook salmon carcasses during spawning ground surveys, few steelhead carcasses are recovered from spawning ground surveys because they do not die immediately after spawning. Adult weirs can provide an opportunity to collect age information, but the time of year to trap returning adults of different species may vary depending on the sampling location.

While GIS coverages of the SRB fish distribution by species (see Figure 2.1.7) have not yet been analyzed by the CSMEP Status and Trends subgroup, Table 2.1.7 suggests a challenge to cover spatially multiple species. For example the streams occupied by bull trout are relatively numerous, while those occupied by fall Chinook salmon are relatively few. An EMAP approach to monitoring both fish distribution and abundance might provide differing levels of resolution depending on the species. The effectiveness of juvenile traps and adult weirs to sample multiple species would likely vary depending on stream order – sampling in higher order streams might provide information on multiple migratory species but not for some species with resident life histories. Sampling in lower order streams, would likely yield less information on some species such as fall and summer Chinook salmon.

The population attributes might vary by sample method for different species. For example, a spring Chinook salmon spawning ground survey might provide an opportunity for live counts of bull trout, whereas a bull trout spawning ground survey might provide information on spring Chinook salmon spawning distribution.

Table 2.1.6. ***DRAFT*** overview of monitoring activities for Snake River steelhead populations.

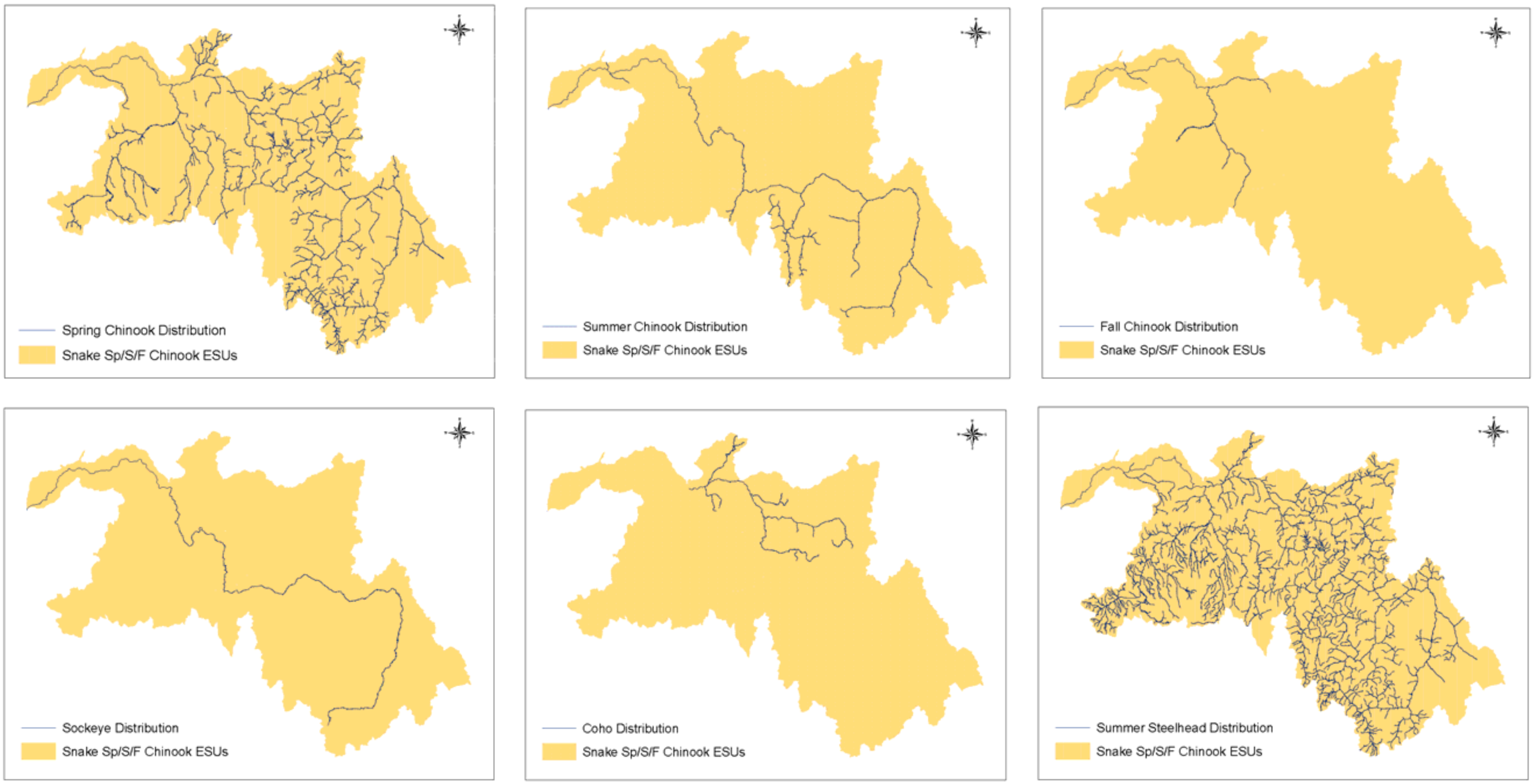
"Current" Design Template:

			Snake River Steelhead ESU			Lower Snake MPG		Grande Ronde MPG				Imnaha MPG	Clearwater River MPG					Salmon River MPG											
Data need	Method/Description		Snake River Steelhead ESU	Snake River "A type"	Snake River "B type"	Tucannon	Asotin	Upper Grande Ronde	Wallowa	Lower Grande Ronde	Joseph Creek	Imnaha River	Lower Mainstem	Lochsa River	Selway River	South Fork	Lolo Creek	Upper Middle Fork	Lower Middle Fork	Lemhi	Upper Salmon East Fork	Upper Salmon Mainstem	Chamberlain Creek	Pahsimeroi	Panther Creek	Little Salmon River	South Fork	Secesh River	North Fork
Abundance of adult fish	A1	Census weir	?	?	?	?																							
	A2	Weir with MR	?	?	?																								
	A3	Weir without MR	?	?	?							1		1												1			
	A4	MR survey, no weir																											
Abundance/spatial distribution of redds	B1	Cen-cen																											
	B2	Cen-multi																											
	B3	Cen-once																											
	B4	Stat-cen																											
	B5	Stat-multi																											
	B6	Stat-once																											
	B7	Fixed-cen																											
	B8	Fixed-multi																											
	B9	Fixed-once						1	1	1	1	1																	
Age structure of spawners	C1	Tags																											
	C2	Hard parts	?	?	?				1			1														1			
	C3	Length at age	?	?	?				1			1														1			
	C4	Basinwide estimate					?																						
Origin of spawners ***(hatchery fraction)***	D1	Hatchery marks, handle fish at weirs	?	?	?				1			1														1			
	D2	Hatchery marks, remotely sense	?	?	?																								
	D3	Hatchery marks on carcasses																											
Sex ratio of spawners	E1	?? Handle fish at weirs	?	?	?				1			1														?			
	E2	??																											
Abundance/spatial distribution of smolts/parr	F1	Juvenile trap	?	?	?							1																	
	F2	Electrofishing										1	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***
	F3	Snorkel survey										1	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***	***1***
Survival of juveniles	G1	Mark recapture	?	?	?							1																	

Table 2.1.7. Overview of fish distribution reported through StreamNet by Snake River subbasin and species.

	Number of streams with distribution information					
Subbasin	Bulltrout	Steelhead	Spring Chinook	Summer Chinook	Fall Chinook	Sockeye
Snake lower	1	8	1	1	1	1
Tucannon	8	10	1	1	1	
Clearwater	376	225	138		3	
Asotin	4	10				
Grande Ronde	61	297	66		1	
Salmon	605	400	226	51	1	4
Imnaha	19	61	14		1	
Snake Hells Canyon	11	34	5	1	1	1
Snake Basin	1,085	1,045	451	54	9	6

1/ Source: StreamNet, 3/13/2006, URL: query.streamnet.org/Request.cfm?cmd=BuildCriteria&NewQuery=BuildCriteria&Required=Run&DataCategory=23


Spring Chinook Distribution
Snake Sp/S/F Chinook ESUs
Summer Chinook Distribution
Snake Sp/S/F Chinook ESUs
Fall Chinook Distribution
Snake Sp/S/F Chinook ESUs
Sockeye Distribution
Snake Sp/S/F Chinook ESUs
Coho Distribution
Snake Sp/S/F Chinook ESUs
Summer Steelhead Distribution
Snake Sp/S/F Chinook ESUs

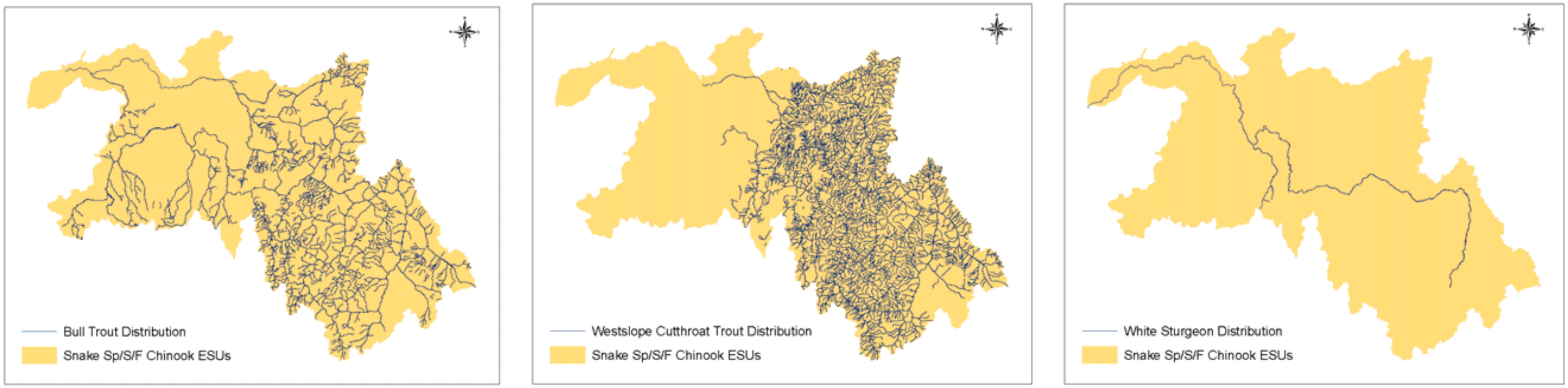


Figure 2.1.7. Distributions of key fish species in the Snake River Basin Sp/S/F Chinook ESUs.

2.2 Hydrosystem

In FY2005, CSMEP's Hydro Subgroup tackled a set of ten hydro management questions across several scales: individual projects, survival by different passage routes through the hydrosystem, and overall life cycle survival (Marmorek et al. 2005; CSMEP Hydro subgroup 2005*)*. These different scales relate to a variety of decisions: operations at individual projects (e.g., spill, bypass, removable spillway weirs); overall operations (e.g., when to transport fish within season, compliance with hydrosystem biological opinions), longer term hydrosystem decisions (e.g., flow management, effectiveness of transportation over multiple years, system configuration); and adequacy of hydrosystem operations for stock recovery. Moving along these scales, the performance measures of interest change. Performance measures range from direct survival at and between dams, to smolt-to-adult survival rates (e.g., smolts leaving Lower Granite Dam to adults returning there 2-3 years later) to inferences about delayed mortality from contrasts in mortality patterns (contrasts in recruits/spawner or smolt-to-adult survival rates).

In FY2006, the CSMEP Hydro Subgroup narrowed our focus to three major sets of decisions and four questions related to those decisions (Table 2.2.1). The subgroup worked on advancing through steps 6 and 7 of the DQO process, that is: specifying limits on decision errors in both directions (step 6); and optimizing the design for obtaining data (step 7).

Table 2.2.1. Hydrosystem decisions and associated questions tackled by the CSMEP hydro group in FY2006.

Decisions / Alternative Actions	Hydro Action Effectiveness Questions
Are SARs, and important SAR ratios relating to effectiveness of transportation, meeting NPCC and BiOp targets? If targets are not met, (*by how much?*), then decision makers may need to consider changes in FCRPS operations (e.g., when, how much to transport and spill) or FCRPS configuration.	1. Is SAR sufficient for 1) NPCC goal[5] & 2) recovery goals? 2. Is transportation more effective than in-river passage?
Has hydrosystem complied with performance standards set out in 2000 FCRPS BiOp? If not, what changes are required?	3. How does annual in-river survival of spring summer Chinook and steelhead (Lower Granite to Bonneville) compare to 2000 FCRPS BiOp performance standards?
Should FCRPS change the timing of transportation of some species within the season to improve survival?	4. How does effectiveness of transportation change over the course of the season?

To date, neither NOAA and USFWS Biological Opinions nor NPCC documents have specified limits on decision errors related to the decisions in Table 2.2.1. Therefore the Subgroup used various models and statistical methods to examine *hypothetical* decision rules and the potential decision errors associated with these hypothetical rules under different monitoring and evaluation designs. Decision errors were measured by a variety of metrics, as described below. Alternative designs included varying the number of PIT-tagged fish, the location of that tagging, and the duration of monitoring.

This work was conducted in close coordination with the BiOp Remand Passage Modeling Group and BiOp Remand RME group, as five members of the CSMEP Hydro Subgroup participate in these BiOp Remand efforts. Some of the CSMEP Hydro Subgroup's work products have already been shared with the BiOp Remand groups, and others will be presented later this year.

[5] Pg. 13 of NPCC mainstem amendments of 2003-2004. www.nwcouncil.org/library/2003/2003-11.pdf ; interim goals of 2-6% SAR

Our previous work on the DQO process tackled nine hydrosystem related decisions, the questions associated with these decisions, and for a subset of the questions, a range of Low, Medium and High preliminary monitoring strategies specific to each question (CSMEP Hydro subgroup 2005). The subgroup began their efforts by consolidating their previous preliminary strategies for multiple hydrosystem questions into a single set of Status Quo (SQ), Low (L), Medium (M) and High (H) strategies that spanned multiple questions (Appendix D). To encourage integration across domains, this set of hydro monitoring strategies built upon the preliminary strategies developed by the Status and Trend group. Ultimately it is CSMEP's intent to have an integrated set of L, M, H designs across all five subgroups (i.e., status and trend, hatchery, habitat, hydro, harvest) to illustrate various dimensions of M&E tradeoffs (i.e., cost, precision, monitoring objective). However, it's recognized that these designs will need to iteratively evolve through the DQO process of evaluating alternatives.

2.2.1 Labor costs of PIT-tagging fish

PIT-tags themselves cost $2.10 per tag. The labor costs of tagging both hatchery and wild fish were estimated from existing and proposed CSS studies, and opportunistic use of tagging data from ongoing studies in Idaho to evaluate the effectiveness of supplementation, and monitor the status and trend of populations (Appendix E). Some of the cost data and estimates in Appendix E are preliminary values still under discussion, but do provide give a general idea of relative costs (Table 2.2.2). There's a huge variation in the number of spring/summer Chinook caught in the 29 rotary screw traps in Idaho tributaries, as different streams have different communities of fish. We therefore looked at the average number of fish caught (2,150) and estimated labor cost per year ($65,000), to derive an estimate of the cost / tagged fish over all traps ($30.23).

Table 2.2.2. Summary of labor cost information in Appendix E.

Type of Fish & Tagging	Average Labor Cost of Tagging / PIT-tagged fish (min – max; *n*)	Key Assumptions
Hatchery fish	$1.16 ($0.92 - $1.51; *5 data points*)	Labor costs are comprehensive
Wild fish at tributary-population level	$30.23 (*average across 29 traps)* This is a maximum cost, as more fish could be trapped at many of these traps.	Estimated cost of $65,000 / trap / yr. Total of 29 traps, capturing 62,357 fish.
Wild fish at Major Population Group (MPG) level	$6.29 (estimated, not from actual cost data)	Estimated cost of $65,000 / trap / yr.; assumed to be able to catch 186,000 fish with 18 traps (about 10,000 fish / trap)
Wild fish at sub-basin / MPG level (Salmon, Clearwater, Grande Ronde)	$12.36 (*3 traps)*	Based on total labor costs for Salmon, Snake and Clearwater traps operated under the Smolt Monitoring Program ($359,074) in spring, which in 2006 trapped 29,050 fish (trapping all that they can at these 3 traps). This works out to roughly 10,000 fish / trap.

The main caveats on these cost estimates are that: 1) CSS programs are often piggybacked on top of other funded programs, so the incremental costs of improving existing monitoring depends on whether or not these programs have base funding; 2) labor cost estimates for trapping at the population level have some uncertainty, and MPG level trapping costs are highly uncertain; and 3) trapping at an MPG rather than population level will mean losing potentially valuable population information such as parr to smolt survival which is valuable for both status & trend and effectiveness monitoring for habitat and hatchery effectiveness (e.g., Idaho Supplementation Study). In addition, adult arrival times to Bonneville of different populations can vary considerably within an MPG (e.g., Lemhi is early, Pahsimeroi is late); similar variation in migration timing exists for juveniles at a population scale.

The following two sections of this report are new work completed in FY2006, which will be placed into the CSMEP Hydro Subgroup DQO report. Readers requiring more background should consult CSMEP Hydro subgroup 2005.

2.2.2 Questions 1 and 2: Are annual SARs and T/C ratios meeting management targets over multiple years of observations?

Ultimately, inferences about the relative value of different long term hydrosystem management strategies will be based in part on inferences about expected, mean overall SARs, *T/C* values, and *D* values. Uncertainty in SARs, *T/C*s, and *D*s, obscuring the mean underlying value (for a given time period), is due to process error (inter-annual variation in survival rates) and sampling error. Previous work suggests environmental variation in these measures is large and can be expected to influence population viability. For parameter estimates for wild (ESA listed) fish in particular, sampling variance may also be substantial, since these fish are opportunistically sampled and tend to be available for capture and tagging in much lower numbers than hatchery fish.

Combining data from multiple years may allow us to better estimate the long-term distributions and expected values of these indicators of survival during and subsequent to the hydrosystem migration, thereby facilitating relevant inferences. A previous analysis explored how the power of hypothesis tests and confidence intervals about the mean value of *D* increased with the number of years included in the study (PATH 2000, Appendix F). However, that analysis did not attempt to separate sampling from process errors in estimating the true distribution of *D*, nor did it produce probability distributions of the parameter.

These analyses present distributions of *TIR*s (*T/C*s) reflecting inter-annual variability due to environmental conditions. These can be used in conjunction with the passage and life cycle models to explore the effects of different strategies involving transportation of smolts. The distributions can also be used for statistical inference in answering questions such as “Does transportation of species X from dam Y provide a survival benefit compared to leaving fish in-river under a particular hydrosystem management strategy?” An obvious test value for an if-then decision related to this kind of question is *TIR* = 1. Levels of acceptable Type I and II errors appropriate to the framing of the research question could be chosen, or the question could be framed in terms of the degree of confidence (credibility) to invest in the hypothesis that over the long term *TIR* is greater than 1.

When survival rates are estimated from counts of individuals (from a census or from marking a sample of the population) at the start and end of the interval, the sampling error is binomial (assuming minimal error in enumerating individuals) and can be removed from the variance estimated from a time series of such survival rate estimates. One method is to use a beta-binomial likelihood function to estimate the underling parameters of a beta distribution representing distribution of actual survival rates. Kendall (1998) used census data and a likelihood function that assumed binomial demographic error and underlying, beta-distributed environmental stochasticity. Morris and Doak (2002) also note the flexibility of the beta distribution and recommend it as ideal for modeling variability in survival rates. Mean and variance can be calculated directly from a and b (the resulting parameters of a beta distribution), though Kendall’s method results in a negative bias in the estimate of environmental variance, which a correction factor he proposed reduces, but does not eliminate.

The sampling variance can also be estimated and removed with a simpler method (Akçakaya 2002). Akçakaya's paper was intended to present a simpler and lower-bias alternative to Kendall's (1998) approach. The method for estimating beta distributions representing environmental variation in SARs is described in Berggren et al. (2005), Chapter 3. The analysis presented here differs from that in Berggren et al. in that: 1) the SARs and *TIR*s are assumed to be estimated for each transport project separately; 2) the method of producing parameters for distributions of *TIR*s that include covariance between transport and control SARs is modified (since the earlier analytical method was strictly correct only for ratios of binomial, rather than beta, random variables, and led to underestimates of variance).

We used Akçakaya's method to estimate the variance in PIT-tag SAR estimates from sampling error, and remove it from the total variance in the time series. The mean and total variance can be estimated in different ways: unweighted (i.e., each annual estimate gets the same weight in calculating mean and variance); or weighted in some manner, where the influence of each year's estimate reflects some measure of precision and/or relevance of that estimate. Akçakaya (2002) cites Kendall (1998) as pointing out that different ways of calculating variance reflect different assumptions about the reliability of individual estimates. Akçakaya recommends that in general, weighted methods should be used when the variation in sample size results from variation in sampling effort. For our purposes, the number of PIT-tagged smolts in a category can be considered an index of sampling effort. However, independent of considerations of sample size, individual year estimates for PIT-tagged fish in a particular category may be more or less representative, depending on how well they reflect the experience of the relevant untagged population, and how large a portion of the total population of smolts that category represented in that year. Although this simulation analysis assumes annual SAR estimates are made, the methods can also be used to explore seasonal patterns in SARs across years. The migration season could be broken into segments based on arrival timing at a collector project, and the method applied to each of the segments, to test for differences in SARs and *TIR*s among them.

We use the total weighted variance method used by Akçakaya (2002) and Kendall (1998: equation 1) to estimate the multi-year mean and variance of both transport and in-river SARs:

$$\mathrm{var}(p) = \frac{\sum_{t=1}^{Y} N_t (p_t - \overline{p})^2}{\sum_{t=1}^{Y} N_t}, \qquad [1]$$

where $\overline{p} = \sum_{t=1}^{Y} m_t / \sum_{t=1}^{Y} N_t$ and Y = number of years of data, m_t = number of survivors remaining (i.e., returning adults) from N_t individuals in year *t*. This is equivalent to weighting the estimates from each year by inverse variance. The weighting methods for both transport and in-river SARs ensure that the contribution of each year to demographic variance is proportional to the year's contribution to total variance.

The number of transported PIT-tagged fish from a particular project is known from summing fish with the appropriate capture history code. The number of smolts falling into the in-river category at Lower Granite Dam can be taken directly from capture histories if C1 fish are used (Berggren et al. 2005), or estimated if C0 fish are used, according to the methods of Berggren et al. (2005). For the lower projects, C0 smolts alive at those projects can be estimated by multiplying the estimate of C0 smolts at LGR from Berggren et al. by the point estimate of survival rate for the appropriate reach(es).

The values for the mean and remaining variance (after subtraction of sampling variance) of the time series for a given SAR are then converted into the parameters of a beta distribution, using

$$a = \mu\left(\frac{\mu(1-\mu)}{\sigma^2} - 1\right) \qquad [2]$$

and

$$b = (1-\mu)\left(\frac{\mu(1-\mu)}{\sigma^2} - 1\right) \qquad [3]$$

where μ is the estimate of the mean and σ is the square root of the estimate of the variance, after Kendall (1998) equations 7 and 8. The resulting distributions reflect an estimate of variance due only to environmental stochasticity in SARs over time. The resulting distributions of each particular measure under environmental stochasticity can also be used to estimate the standard error of the mean value, based on the number of years of data used.

Simulations of the ratio of independent beta random variables (using the parameters estimated for SARs as described above) indicated that the distribution of a large number of realizations of the ratio appeared to closely approximate the lognormal distribution. This assumption can also be examined analytically, since the exact distribution of the ratio of beta random variables has been worked out. The exact form of the ratio of two standard, independently distributed beta random variables was derived by Pham-Gia (2000). The probability density function is a complex expression of beta functions and the Gauss hypergeometric function in three parameters, but can be calculated using appropriate software. The parameters of the lognormal distribution describing the ratio of the SARs are derived from statistics of the simulated *TIR*s. These parameters (μ and σ) are derived by estimating the mean and variance, respectively, of the natural logarithm of the simulated ratios.

To test the goodness of the lognormal assumption, 5000 realizations of the ratio of two beta random variables were simulated and recorded, using the parameters derived from the data for Chinook LGR transport and in-river SAR beta distributions. From the simulated values, the parameters of a lognormal were estimated as described above. The exact distribution was computed per Pham-Gia (2000) from the same SAR beta distribution parameters and plotted along with the lognormal distribution. The lognormal distribution is easier to implement in modeling than the exact PDF, and appears to provide an excellent approximation to the exact distribution (Figure 2.2.1).

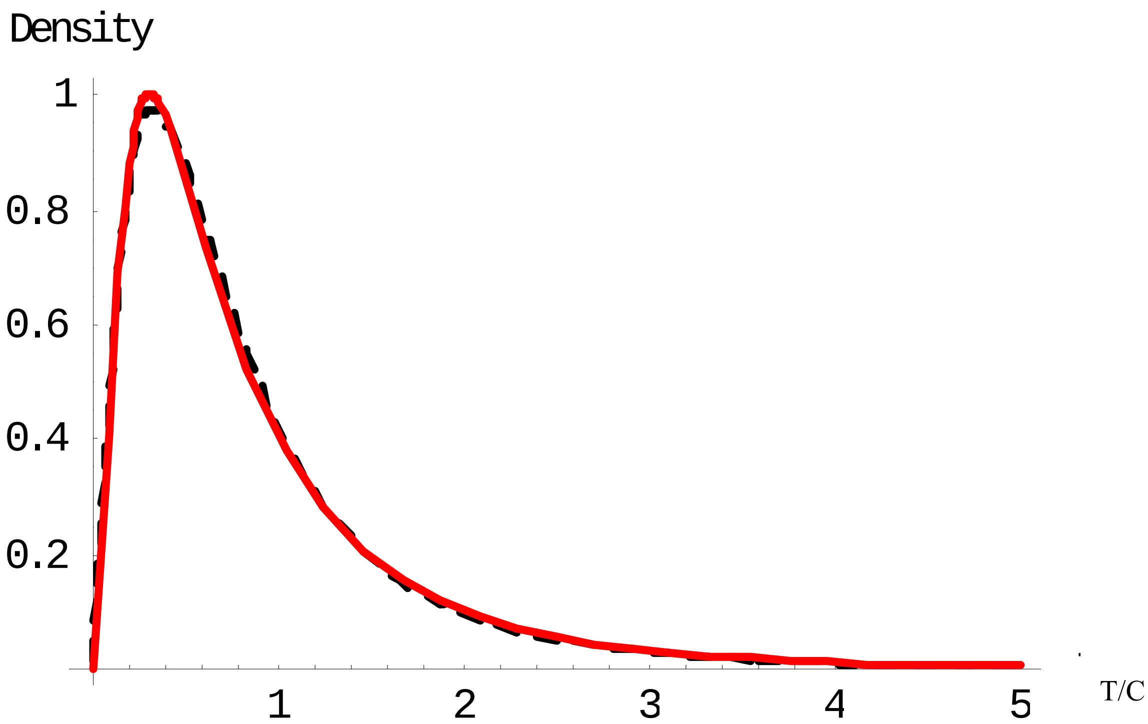


Figure 2.2.1. Density function of exact ratio of beta random variables, based on parameters of Chinook SARs from LGR (dashed line); lognormal approximation using values for μ and σ derived from 5000 values of simulated TIR (solid red line). (μ = -.402, σ = .866).

The ratio of correlated beta random variables, reflecting observed correlation between annual in-river and transport SARs, is simulated using the CORAND array function from the Excel add-in SimTools (home.uchicago.edu/~rmyerson/addins.htm) and the BETAINV function of Excel. For the correlation coefficients observed, this method provides two beta random variables with the intended distributions, with a median correlation approximately equal to the nominal correlation. The resulting distributions of simulated *TIR*s with positive correlations between the SARs are approximately lognormal, with smaller variances than simulations using the same beta parameters with assumed independence of SARs. Table 2.2.3 shows the estimated parameters of the beta distributions representing transport and in-river SAR from each transport project, and the observed correlation between them.

Table 2.2.3. Parameters of SAR distributions for wild spring/summer Chinook and Steelhead, and observed correlation coefficient between point estimates of annual T and C0 SARs. Migration years 1994-2002 for Chinook; 1997-2001 for Steelhead.

	Transport		Control		
Species / Project	Alpha	Beta	Alpha	Beta	Corr Coeff
Chinook LGR	2.63	259	3.59	241	0.61
Chinook LGS	5.27	429	3.68	227	0.74
Chinook LMN	2.13	267	3.38	190	0.59
Steelhead LGR	10.9	508	38.5	2502	0.63
Steelhead LGS	2.47	116	10.5	608	0.02
Steelhead LMN[1]	1.69	92	9.38	487	0.002

[1] For transport SARs, demographic variance estimate was higher than total variance, so total variance was used in calculating beta distribution parameters.

[2] Observed correlation was -.08; assumed to be 0 in deriving *TIR* distribution

The resulting distributions of *TIR* for each species and project for available data can be calculated by the lognormal parameters in Table 2.2.4. The median of the distribution is equal to e^{μ}, the mean of the distribution is $e^{\left(\mu+\frac{\sigma^2}{2}\right)}$, and the variance is equal to $e^{2\mu}e^{\sigma^2}\left(e^{\sigma^2}-1\right)$ (Johnson et al. 1994).

Table 2.2.4. Estimated species- and project-specific parameters of lognormal *TIR* distributions estimated from parameters in Table 1.

Species	Project		
Chinook	LGR	-0.432	0.538
Chinook	LGS	-0.246	0.412
Chinook	LMN	-0.840	0.576
Steelhead	LGR	0.310	0.233
Steelhead	LGS	0.091	0.651
Steelhead	LMN	-0.211	0.749

Simulation of monitoring

A model has been developed to simulate random variation in SARs of the two groups, and the process of estimating SAR from returning PIT-tagged fish for a fixed number of tagged fish at LGR dam. The number of fish that would need to be marked in order to result in a desired number in a particular group, e.g., tagged true control fish at LGR, would have to be estimated elsewhere. It is assumed that all fish are transported from one project (LGR), or alternatively fish are transported at multiple projects, with the *D* value the same for each project, and the LGR equivalent transport smolt number is estimated with little error (i.e., survival of T0 group represents a binomial process).

Correlation between transport and in-river SARs is modeled using a value (.65) close estimated from the point estimates from 1994-2002 migration years for wild Chinook at LGR. Data from LGR and the other two projects also informed the choice of the assumed coefficient of variation (CV) of the underlying beta distributions. Estimated CVs from beta distributions for Chinook SARs for both transport and in-river groups ranged from 0.4 to 0.7, with no clear difference between transport and in-river values. A value of 0.6 was used in the simulations as a conservative (i.e., slightly pessimistic) choice. Estimated CVs for steelhead were generally lower than for Chinook.

SARs are modeled as beta random variables, with underlying environmental coefficient of variation and correlation coefficient as described above. The number of smolts surviving to adult in a group in a particular cohort is determined in a beta-binomial process—i.e., the probability of survival is drawn randomly from the relevant beta distribution, and that survival probability is used in a binomial draw with N = number of smolts to determine adults actually surviving to and being detected at LGR. For exploration of SAR distributions, each simulation is run independently of others. For TIR distributions, with each combination of parameters in the simulation describing the expected outcome, the same sequence of “actual” realized SARs for both groups is used as the seed from which estimated SARs are derived through survival of PIT-tagged fish, for each level of PIT-tagging. This is due to the relatively low number of simulations used to explore TIR, compared to SAR. Unlike simulation of the monitoring of SARs themselves, the simulation of monitoring TIR involves simulation of the ratio of SARs, and hence is particularly computer-intensive.

The model ignores correlation structure *within* a group. Adults returning in a given year can contribute to SARs and *TIR*s of different migration years (because adjacent cohorts overlap in the ocean). Actual survival rates probably exhibit some serial autocorrelation as well. Within-group correlation structure is assumed not to exist, both in the simulation model and the estimation procedure. This assumption is tenuous, and it may affect results, more for shorter time series. As an alternative to this, a more realistic simulation mode could be created, based on a stochastic, age-structured projection matrix with a correlation matrix of parameters. This would likely necessitate an alternative approach to removing sampling variance that considers covariance between survival rates, such as the variance-components approach used by Gould and Nichols (1998).

The measures used to evaluate the influence of number of fish tagged and number of years on inferential ability are 1) width of the 95% confidence interval on the standard error of the estimated mean (expected) value of the parameters; 2) (for SARs) probability of the alternative hypothesis, given that it is true (or false) at a certain effect size (i.e., true expected value of SAR); 3) (For TIRs) probability of making correct conclusion about the hypothesis, given different decision rules.

A range of assumptions can be modeled, about:

1. true value of $E[SAR_C]$;
2. true value of *TIR* (in combination with SAR_c and V_c assumption, determines true value of $E[SAR_T]$ and $E[D]$;
3. number of smolts PIT-tagged at LGR or PIT-tagged smolts alive on reaching LGR; and
4. ratio of number of tags between T and C groups.

For initial simulations, the number of marks in each group was fixed. In other words, for a given target number of PIT-tagged fish in the two groups, there was no inter-annual variation in the numbers of marked smolts in either transport or control groups. In practice, PIT-tagged smolts falling into the different groups vary widely over years, especially for wild fish. Generally, the control group is larger than the transport group, due in part to the need to return PIT-tagged fish to the river for use in reach survival estimation.

The planned values from which the simulation will be designed, are listed in Table 2.2.5.

Table 2.2.5. Range of assumptions used in generating data for TIR monitoring simulation.

	Assumption No.		
	1	2	3
E[SAR_C]	1.0%	2.0%	
E[*TIR*]	0.8	1.2	1.5
Tagged T fish @ LGR	1000	2500	5000
C:T tag ratio	1:1	2:1	

Twenty years of monitoring are simulated for both SAR and TIR estimation, with performance measures estimated every 5 years.

Simulation Results

SARs

Results showed that the width of the 95% confidence interval (CI) on the mean of SAR declined with time, and after a given interval of time, was directly proportional to the true underlying mean of the distribution. The exact dependence on assumed mean is a consequence of the assumption of constant CV over different mean SARs. Consequently, Figure 2.2.2 shows the relative width of the CI as a function of time for the two mean SAR values examined. Also, the width of the estimated interval was almost independent of the number of PIT-tagged fish used to estimate the value (Figure 2.2.2). After 5 years of data, the width of the interval is close to the value of the mean of the distribution. After 20 years, the CI width has declined to about half this value (Figure 2.2.2).

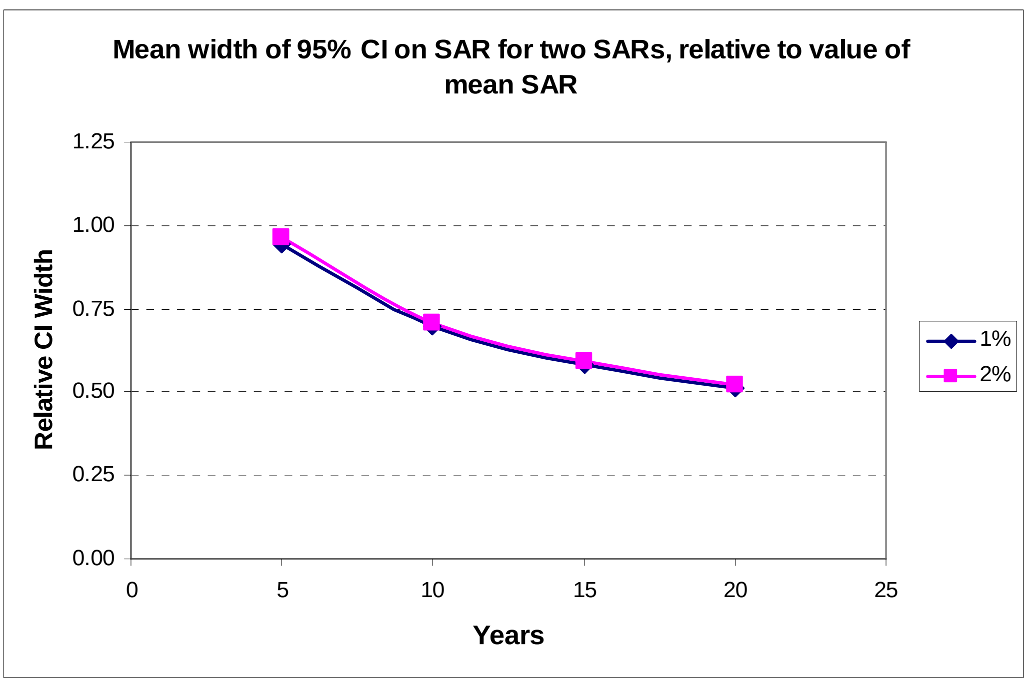


Figure 2.2.2. Relative width of confidence interval of mean SAR (CI width / mean SAR), from method described earlier to remove sampling variance and estimate beta distribution of environmental variance alone. CV = 0.60 (from environmental variance). Width insensitive to annual number of tagged fish examined (1000, 2500, 5000, 10000).

The influence of time and tag numbers on the estimated probability of the hypothesis that SAR is > 2.0% was explored using two different values of true mean SAR: 2.2% and 2.5%. The probability increased with time and again was largely independent of the number of tagged fish used to estimate SARs (Figure 2.2.3). Because of the non-symmetrical distribution of probability estimates around the mean (since it's greater than .5), the median of the distribution of probability for both SAR values is greater than the means shown in Figure 2.2.3. This means that more than 50% of the time, we would expect the estimated probability to exceed the mean values shown for any of the time periods. Nevertheless, the results indicate that the true average SAR will probably need to be close to 2.5% to be highly confident that true SAR is greater than 2.0 within 20 years.

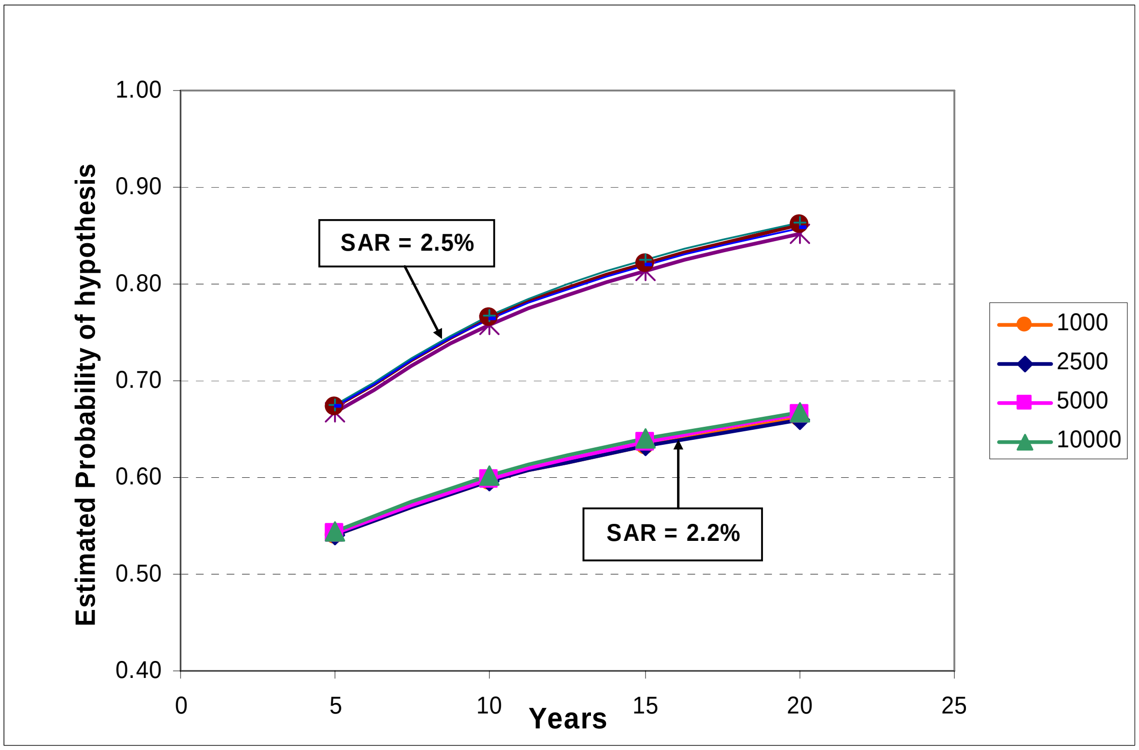


Figure 2.2.3. Expected (average) estimated probability, from beta distribution of mean of SAR, that mean SAR is greater than 2.0%, for two values of true SAR and for four values for annual PIT-tagged smolts. CV = 0.60 (from environmental variance), 10,000 simulations.

The lack of sensitivity of CI width or estimated probability of the hypothesis that SAR > 2.0% to number of PIT-tagged fish used in estimation results in part from poorer coverage of nominal CIs with low tag numbers. The percentage of 10,000 simulations where the estimated 95% CI contained the true, underlying mean value of SAR from the beta distribution is shown in Figures 2.2.4 and 2.2.5, for true mean SARs of 1% and 2%, respectively. The figures show that for all tag numbers, there is a significant improvement in coverage with time, to 90% or above, due to the diminishing of the environmental variance sampling effect as years accumulate. Also evident is that coverage increases with tag number, at least up to 5000 annual tags. The results suggest that the lack of sensitivity to tag number in Figures 2.2.2 and 2.2.3 is a result of the diminished coverage, resulting in "overconfidence" intervals, at low tag numbers. Adjusting of the nominal confidence of estimated confidence intervals might be warranted when using these methods for the inferences of interest.

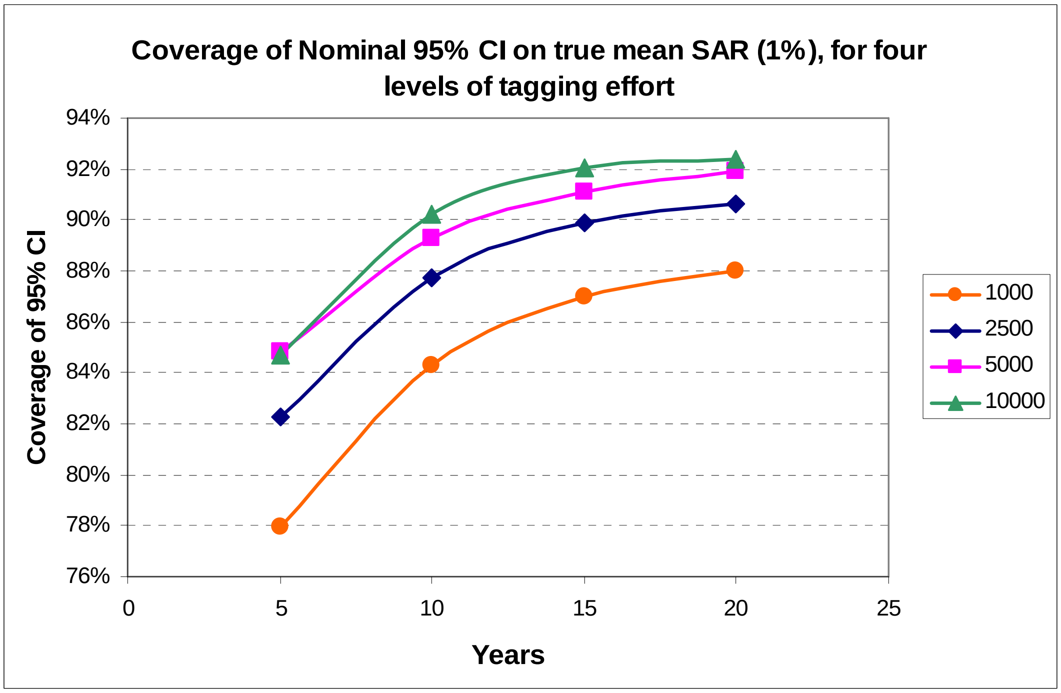


Figure 2.2.4. Percentage of 10,000 simulations in which estimated 95% CI of mean contained "true" expected value of SAR (i.e., mean of beta distribution used to generate simulated SARs). SAR = 1%, CV = 0.60.

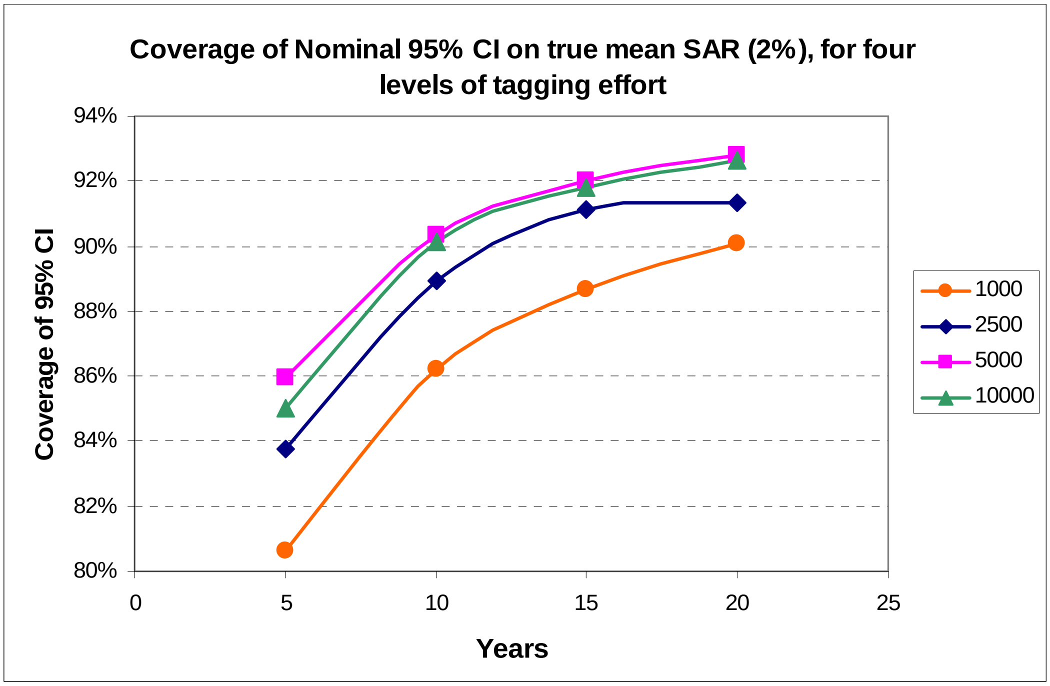


Figure 2.2.5. Percentage of 10,000 simulations in which estimated 95% CI of mean contained "true" expected value of SAR (i.e., mean of beta distribution used to generate simulated SARs). SAR = 2%, CV = 0.60.

TIRs

Because of time limitations, it was not possible to perform simulations on all permutations of the assumptions for generating simulated data, shown in Table 2.2.5. The in-river to transport tag ratio of 2 was dropped, cutting the number of simulations in half. Table 2.2.6 shows the complete simulation design, with skipped runs shaded. For each run, a 20 year time series of transport and control SARs were simulated. For each set of three runs with the same value of parameters in columns 2-4, "true" SARs were simulated, and the estimation of SAR simulated by using each of the tag numbers with the same SARs. From the appropriate number of PIT-tagged smolts adult recoveries, "monitoring" was performed by estimating alpha and beta for transport and control SARs, along with correlation coefficient between the two, at 5, 10, 15 and 20 years. This was done 100 times for each simulator run. For each of the 100 sets of parameter estimates, 5000 simulated estimations of the *TIR*s are done for each of the four time periods.

Table 2.2.6. Design for TIR estimation simulations. Shaded rows have not yet been run.

Simulator Run #	Expected In-river SAR	Expected TIR	I/T tag ratio	Annual # T tags
1	1%	0.8	1	1000
2	1%	0.8	1	2500
3	1%	0.8	1	5000
4	1%	0.8	2	1000
5	1%	0.8	2	2500
6	1%	0.8	2	5000
7	1%	1.2	1	1000
8	1%	1.2	1	2500
9	1%	1.2	1	5000
10	1%	1.2	2	1000
11	1%	1.2	2	2500
12	1%	1.2	2	5000
13	1%	1.5	1	1000
14	1%	1.5	1	2500
15	1%	1.5	1	5000
16	1%	1.5	2	1000
17	1%	1.5	2	2500
18	1%	1.5	2	5000
19	2%	0.8	1	1000
20	2%	0.8	1	2500
21	2%	0.8	1	5000
22	2%	0.8	2	1000
23	2%	0.8	2	2500
24	2%	0.8	2	5000
25	2%	1.2	1	1000
26	2%	1.2	1	2500
27	2%	1.2	1	5000
28	2%	1.2	2	1000
29	2%	1.2	2	2500
30	2%	1.2	2	5000
31	2%	1.5	1	1000
32	2%	1.5	1	2500
33	2%	1.5	1	5000
34	2%	1.5	2	1000
35	2%	1.5	2	2500
36	2%	1.5	2	5000

The estimated mean 95% CIs of the geometric mean TIR is shown as a function of annual tag numbers (for each group) and number of years monitored, for expected TIRs of 0.8, 1.2, and 1.5, for expected in-river SARs of either 1% or 2%, in Figures 2.2.6 to 2.2.11.

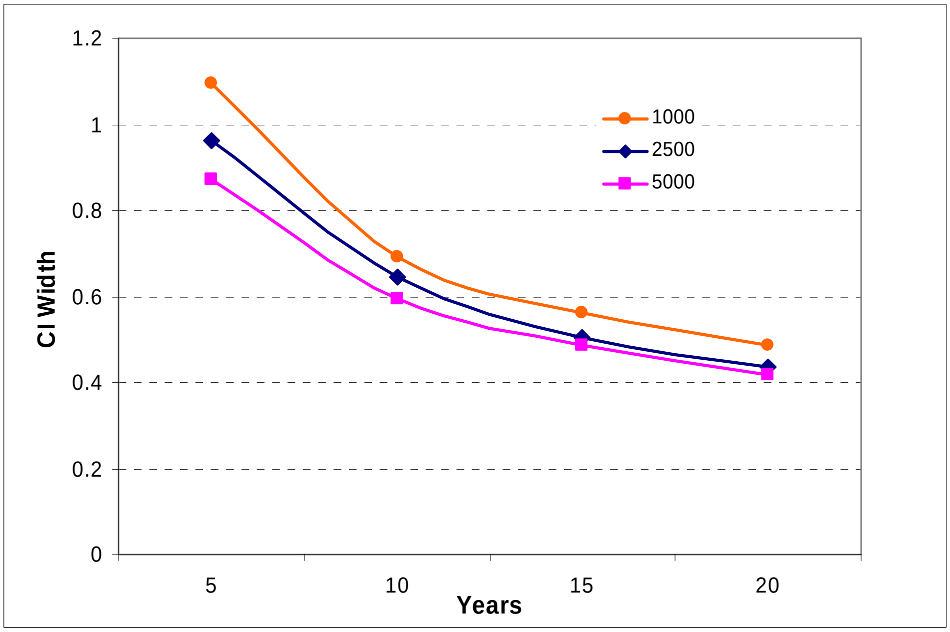


Figure 2.2.6. Average width of confidence interval of geometric mean TIR for three levels of annual Transport tags (100 simulations, 5000 replicates each). Runs 1-3: Mean In-river SAR = 1.0%; True expected TIR = 0.8; Ratio of number of In-river tags to Transport tags = 1:1.

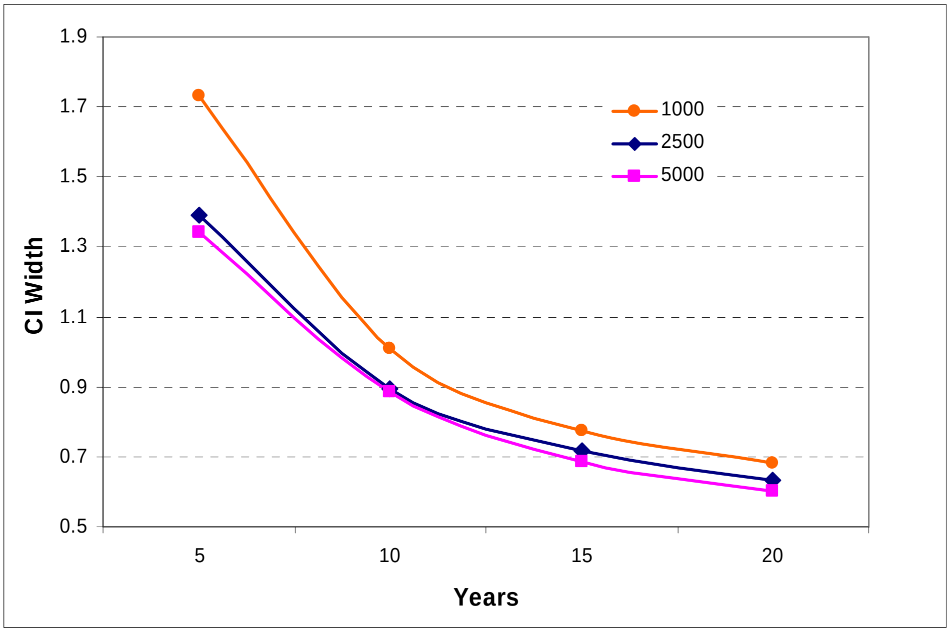


Figure 2.2.7. Average width of confidence interval of geometric mean TIR for three levels of Transport tags (100 simulations, 5000 replicates each). Runs 7-9: Mean In-river SAR = 1.0%; True expected TIR = 1.2; Ratio of number of In-river tags to Transport tags = 1:1.

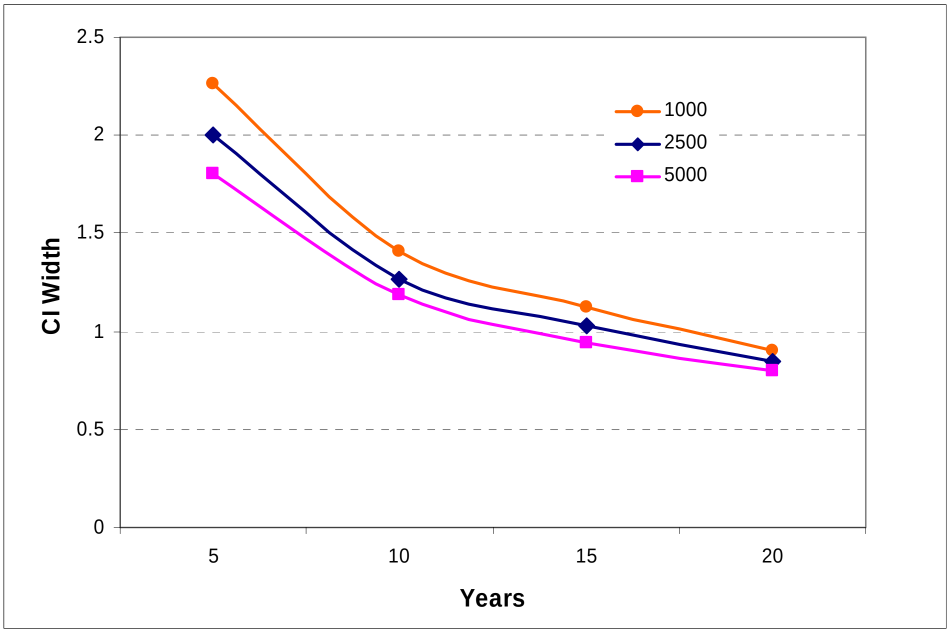


Figure 2.2.8. Average width of confidence interval of geometric mean TIR for three levels of Transport tags (100 simulations, 5000 replicates each). Runs 13-15: Mean In-river SAR = 1.0%; True expected TIR = 1.5; Ratio of number of In-river tags to Transport tags = 1:1.

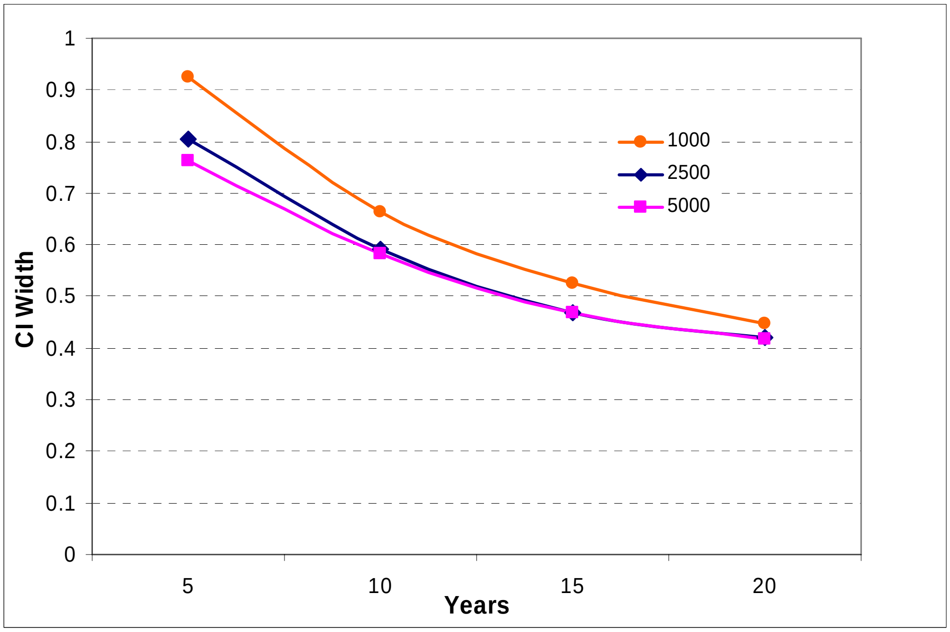


Figure 2.2.9. Average width of confidence interval of geometric mean TIR for three levels of Transport tags (100 simulations, 5000 replicates each). Runs 19-21: Mean In-river SAR = 2.0%; True expected TIR = 0.8; Ratio of number of In-river tags to Transport tags = 1:1.

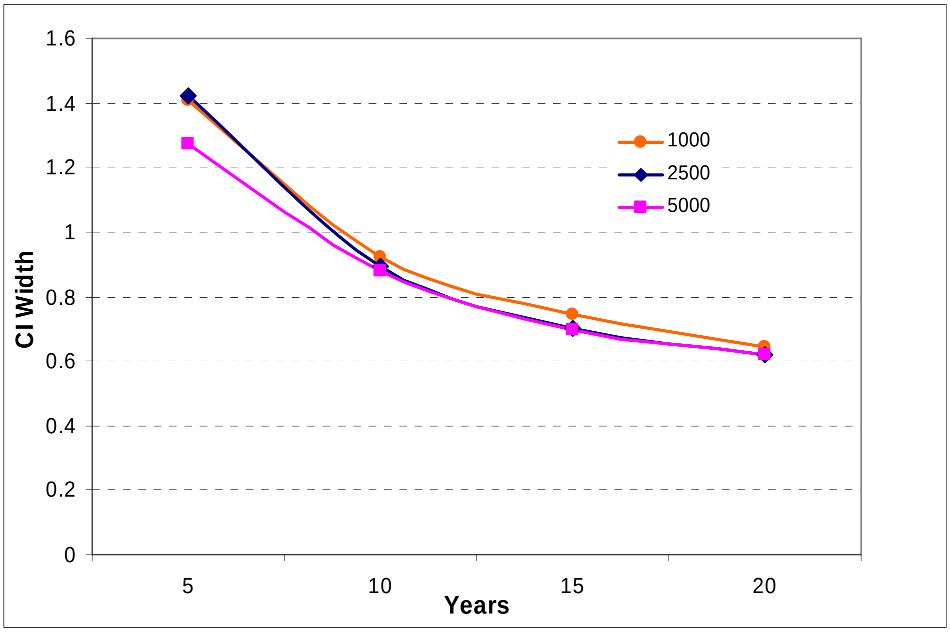


Figure 2.2.10. Average width of confidence interval of geometric mean TIR for three levels of Transport tags (100 simulations, 5000 replicates each). Runs 25-27: Mean In-river SAR = 2.0%; True expected TIR = 1.2; Ratio of number of In-river tags to Transport tags = 1:1.

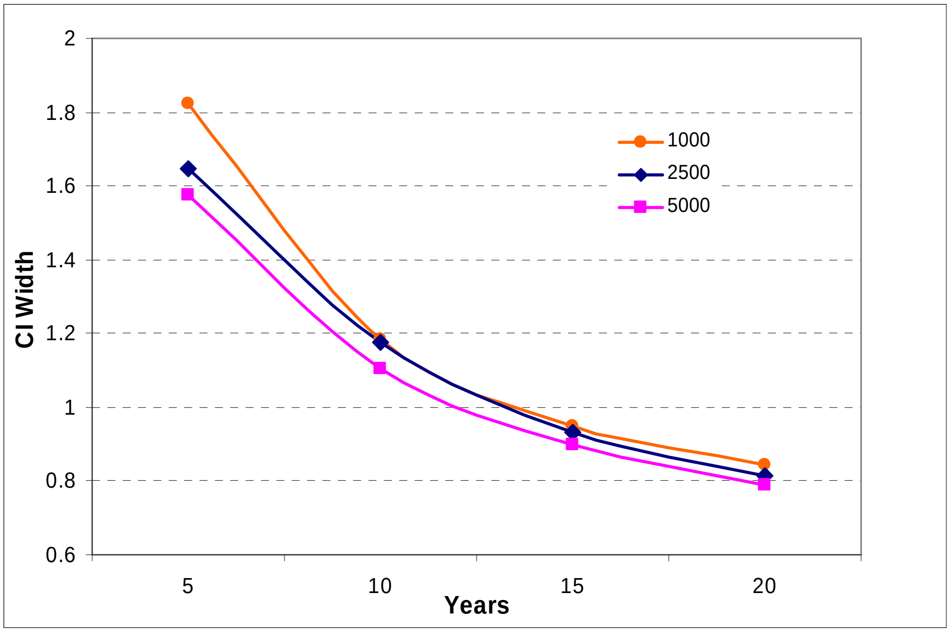


Figure 2.2.11. Average width of confidence interval of geometric mean TIR for three levels of Transport tags (100 simulations, 5000 replicates each). Runs 31-33. Mean In-river SAR = 2.0%; True expected TIR = 1.5; Ratio of number of In-river tags to Transport tags = 1:1

Unlike with SAR CIs, increasing tag numbers results in narrowed TIR CIs at a given true TIR. This effect is seen also by comparing figures with the same TIR but different in-river SAR (e.g., Figures 2.2.6 and 2.2.9): at a given number of tagged smolts, the greater number of adult returns resulting from an assumed greater SAR has the same effect on CI width as tagging more fish. The increase in precision about the mean TIR with increased number of tagged smolts in the two groups is likely due, at least in part, to the improvement in estimation of correlation coefficient between transport and in-river SARs, due to more reliable point estimates of these SARs.

The figures also show that the absolute width of CIs is proportionally related to the underlying *TIR*—i.e., the greater the *TIR*, the greater the CI width. Overall, the relative benefit to CI of accumulating years is similar to the case with SARs: at 20 years, the CI for a given annual number of tags is approximately half that at 5 years, for the same annual number of tags.

The probability of making the correct conclusion about the hypothesis that $TIR > 1$ was estimated using the estimated probability from the lognormal distribution that the geometric mean *TIR* was greater than 1, applying three different decision rules. The rules used were:

1. "Transportation averse": reject conclusion that $TIR > 1$ unless $\Pr[TIR > 1] \geq .8$.
2. "Transportation neutral": accept conclusion that $TIR > 1$ if $\Pr[TIR > 1] \geq .5$
3. "Transportation tolerant": accept conclusion that $TIR > 1$ unless $\Pr[TIR >1] < .2$

Applying each rule to the estimated probabilities of $TIR > 1$ at 5 year intervals, the number of the 100 simulations where the correct conclusion is made is recorded and presented in Figures 2.2.12 to 2.2.17, for the same runs in Figures 2.2.6 to 2.2.11.

Expectedly, accumulating more years of data leads to better decisions. From the figures, it's clear that choosing the "correct" decision rule (i.e., averse if true $TIR < 1$, tolerant if true $TIR > 1$) allows the correct conclusion to be made sooner. Predictably, if the wrong decision rule is chosen, outcomes are much less favorable. However, if the wrong rule is chosen, the correct decision is still made at least 50% of the time even after only 5 years. With an agnostic "neutral" decision rule, in all cases the expected frequency of the correct conclusion is at least 70% or so after only 5 years. More tagged fish generally result in higher probability of the correct decision, although the benefit appears relatively small (and small differences in frequency aren't reliably estimated with only 100 simulations). A higher in-river mean SAR results in improved decision ability; but if the true *TIR* is very different from 1, a higher number of returning tagged adults as a function of either more tags or higher SAR is less important, and the consequences of choosing the wrong decision rule are less severe (Figures 2.2.14 and 2.2.17).

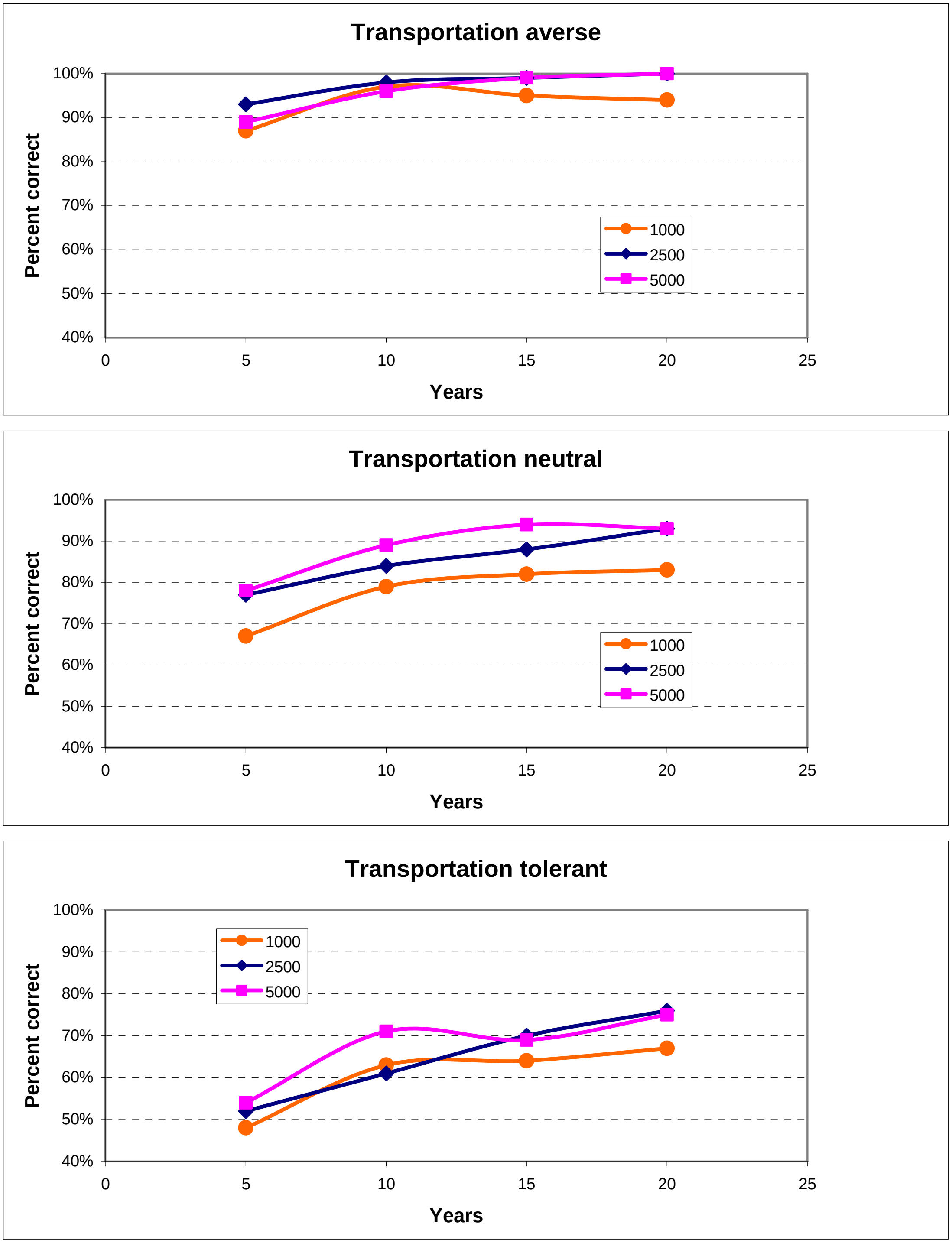


Figure 2.2.12. Frequency of correct decision about hypothesis that TIR > 1 for three decision rules, each with three levels of annual Transport tags (100 simulations, 5000 replicates each). Runs 1-3: Mean In-river SAR = 1.0%; True expected TIR = 0.8; Ratio of number of In-river tags to Transport tags = 1:1.

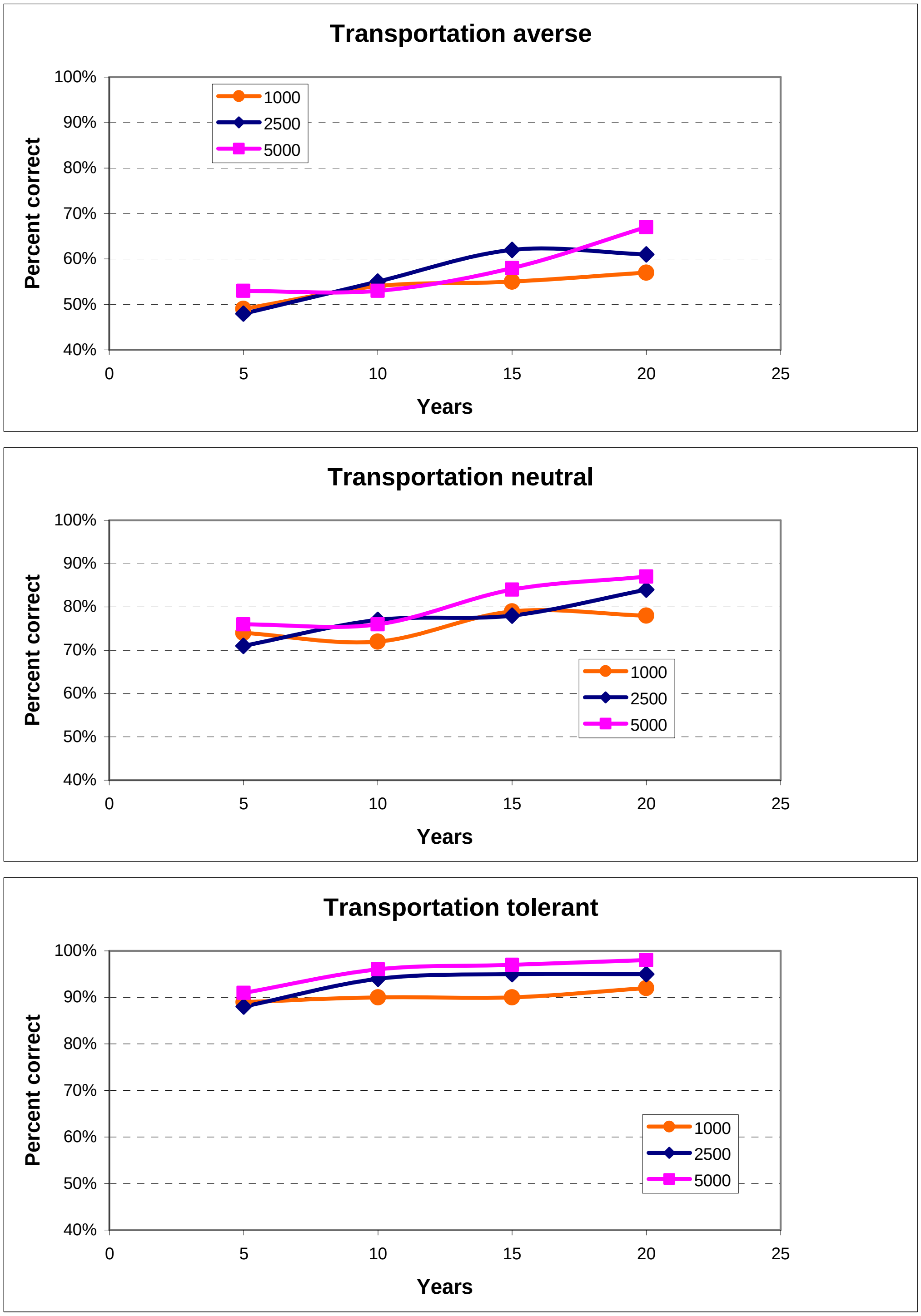


Figure 2.2.13. Frequency of correct decision about hypothesis that TIR > 1 for three decision rules, each with three levels of annual Transport tags (100 simulations, 5000 replicates each). Runs 7-9: Mean In-river SAR = 1.0%; True expected TIR = 1.2; Ratio of number of In-river tags to Transport tags = 1:1.

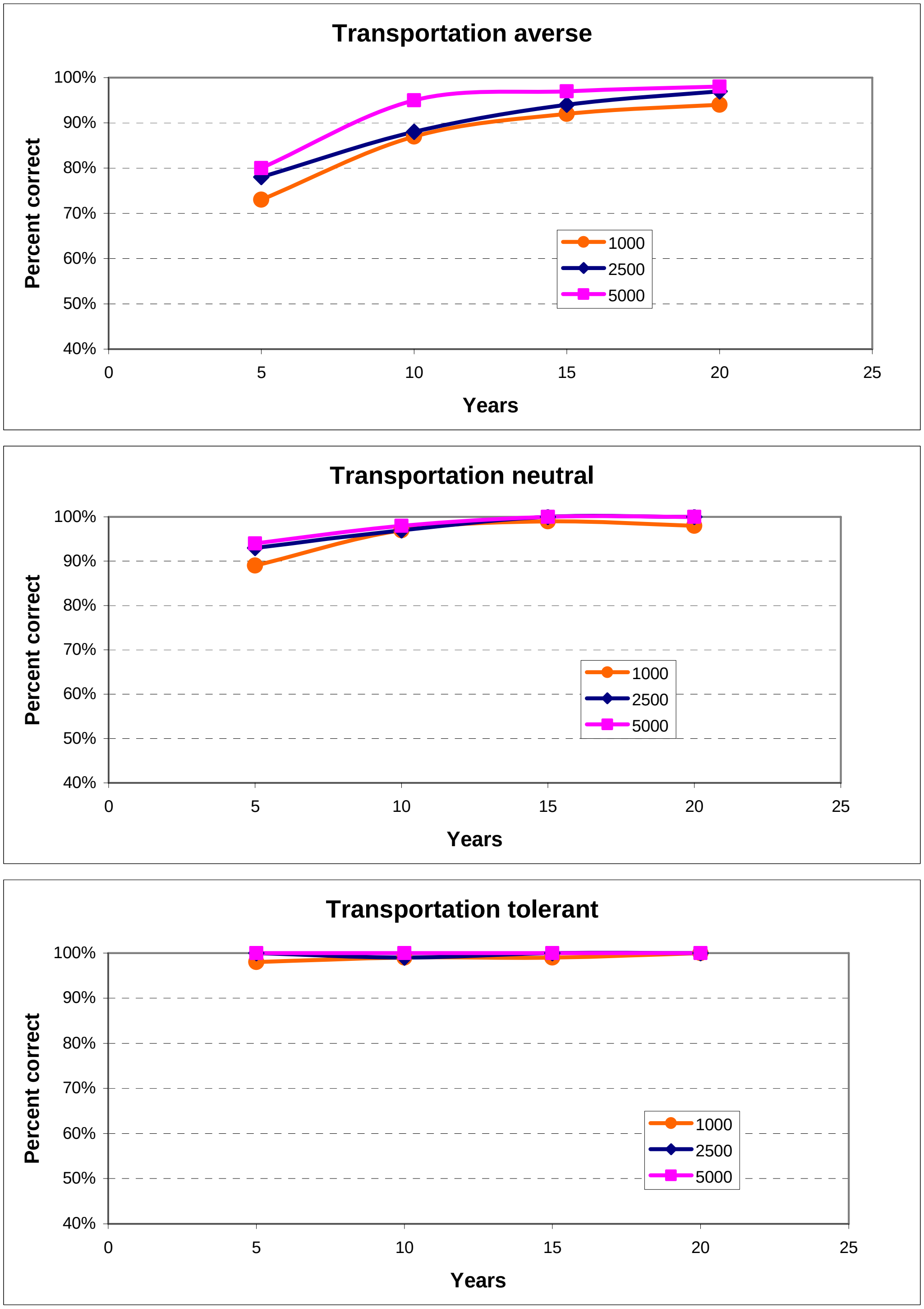


Figure 2.2.14. Frequency of correct decision about hypothesis that TIR > 1 for three decision rules, each with three levels of annual Transport tags (100 simulations, 5000 replicates each). Runs 13-15: Mean In-river SAR = 1.0%; True expected TIR = 1.5; Ratio of number of In-river tags to Transport tags = 1:1.

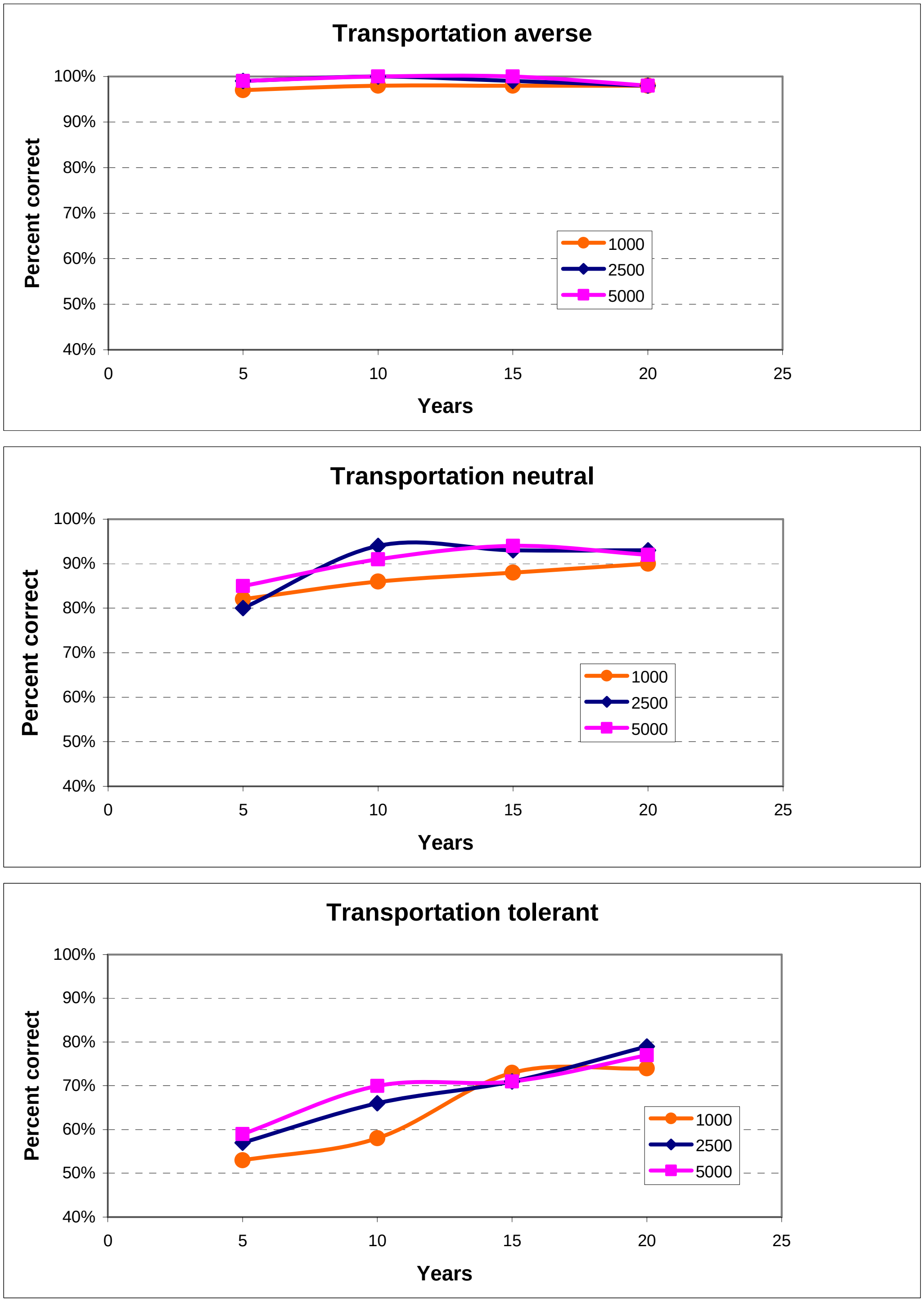


Figure 2.2.15. Frequency of correct decision about hypothesis that TIR > 1 for three decision rules, each with three levels of annual Transport tags (100 simulations, 5000 replicates each). Runs 19-21: Mean In-river SAR = 2.0%; True expected TIR = 0.8; Ratio of number of In-river tags to Transport tags = 1:1.

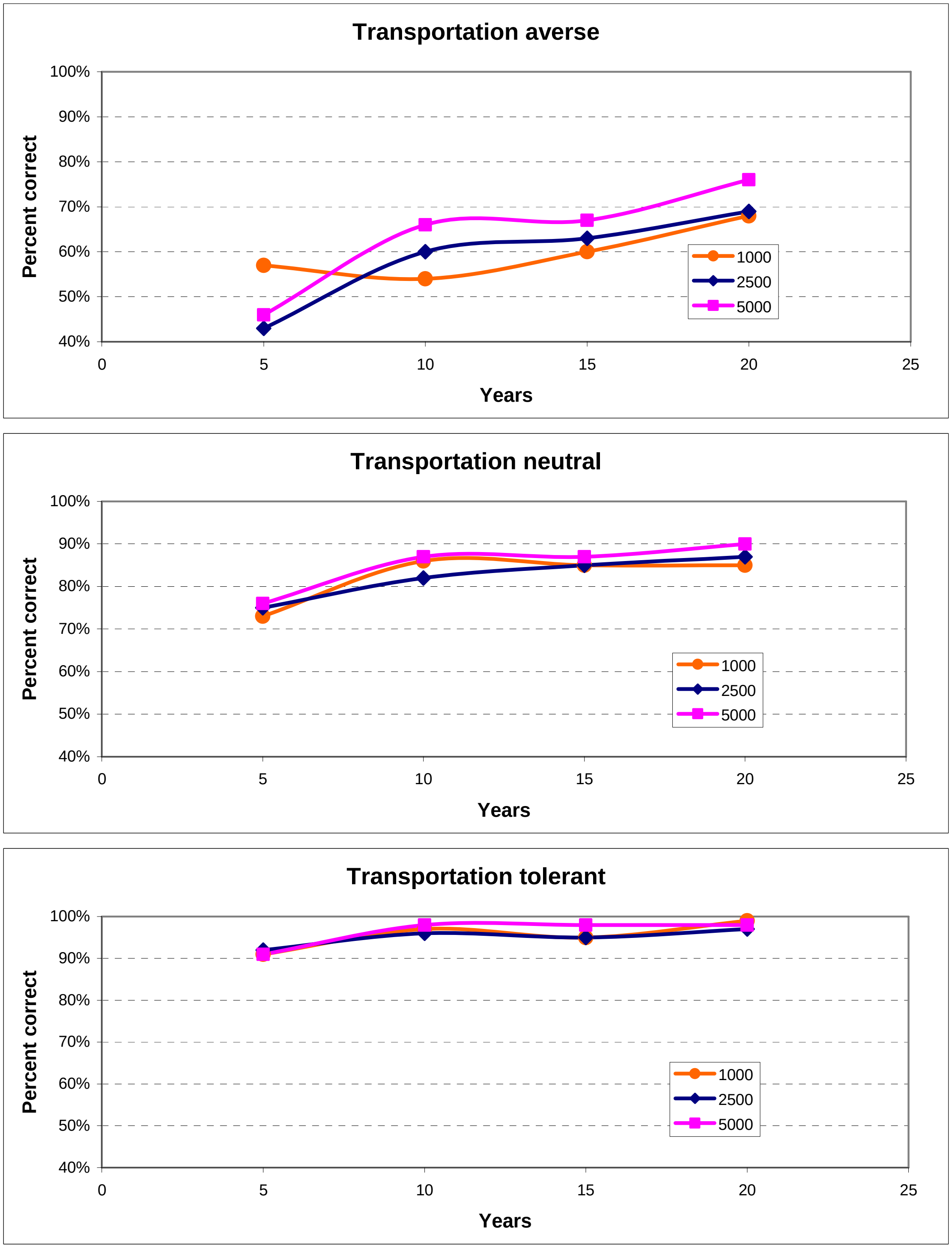


Figure 2.2.16. Frequency of correct decision about hypothesis that TIR > 1 for three decision rules, each with three levels of annual Transport tags (100 simulations, 5000 replicates each). Runs 25-27: Mean In-river SAR = 2.0%; True expected TIR = 1.2; Ratio of number of In-river tags to Transport tags = 1:1

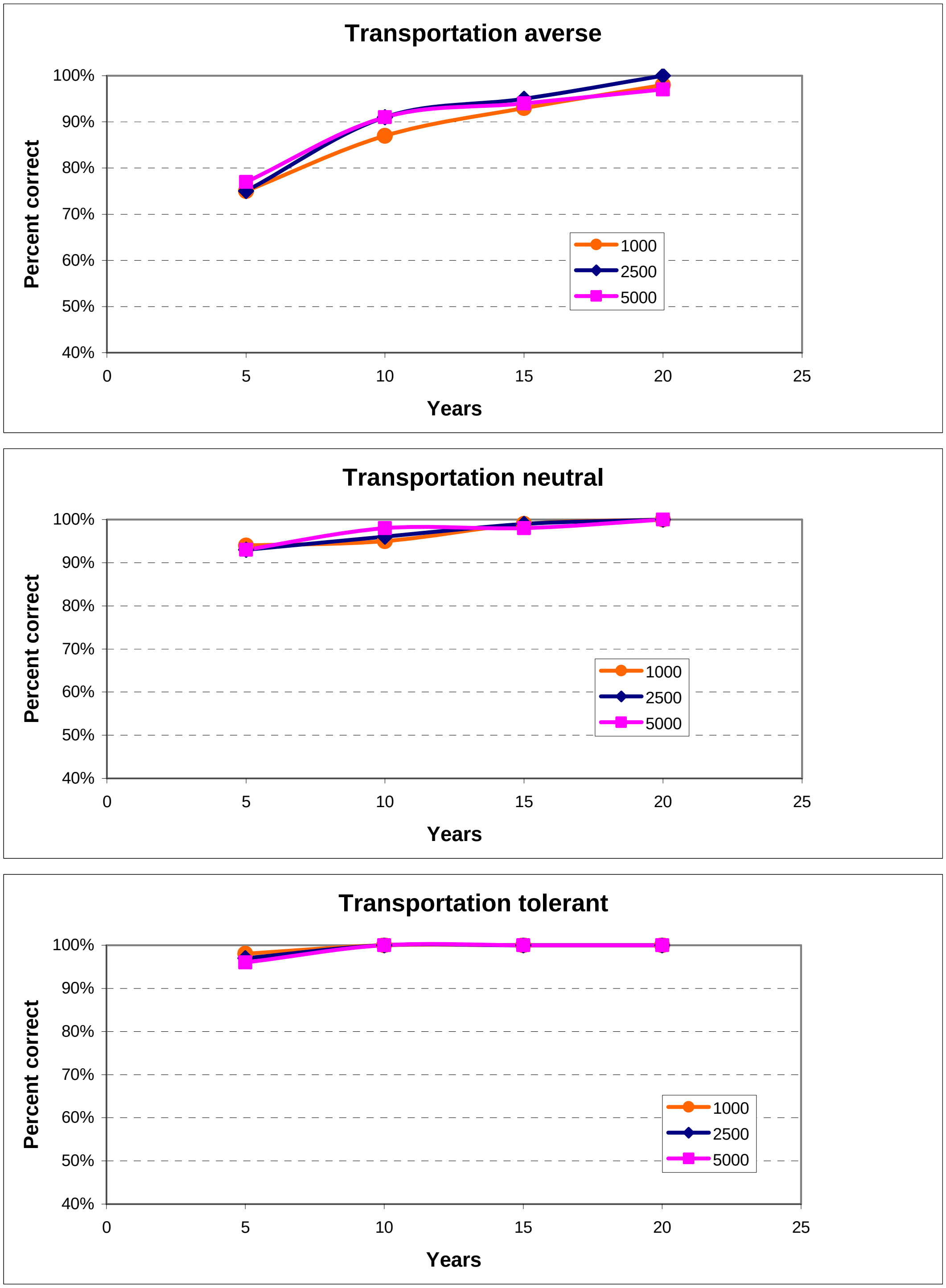


Figure 2.2.17. Frequency of correct decision about hypothesis that TIR > 1 for three decision rules, each with three levels of annual Transport tags (100 simulations, 5000 replicates each). Runs 31-33: Mean In-river SAR = 2.0%; True expected TIR = 1.5; Ratio of number of In-river tags to Transport tags = 1:1

Realized coverage of the estimated 95% confidence interval for *TIR*s ranged generally from the low to mid-80s percent at 5 years and 1000 Transport PIT-tags to around 95% at 20 years and 5000 tags; however many more simulations per run would have to be performed to arrive at statistically reliable conclusions about the coverage.

Conclusions

Combining data from multiple years of PIT-tag data from outmigrating smolts allows a better overall picture of whether SARs and ratios of SARs are, in general, meeting survival targets. Getting the best possible estimates of SARs and *TIR*s in individual years (by marking large numbers of fish) is useful for other purposes, but not necessary for estimating long-term mean values. With the methods used here, the power to distinguish between alternative hypotheses about the values of SARs and ratios of SARs is generally much more sensitive to the number of migration years for which data is collected than to the number of PIT-tagged smolts each year (at least over the range of number of tags examined). This is likely due to the fact that, at the tagging rates simulated, sampling error is dwarfed by process error (true environmental variation) in SARs. One caveat is that the true coverage the underlying mean value of SAR of estimated confidence intervals is somewhat sensitive to tag numbers, with increasing annual number of tags improving coverage (up to 5000 tags, at least). Coverage also improves with time, as well (presumably from the sampling effect of drawing yearly values at random from beta distributions). More confidence can be invested in results from monitoring using more tags and covering more years.

The results suggest that, assuming observed variation is proportional to mean values (i.e., that CV is constant over a range of mean SAR values), true SARs will need to exceed target values by 25% or more to have a reasonable chance of correctly concluding that the target values have been exceeded in a reasonable amount of time. This is seen in Figure 2.2.3, in the upper group of lines, where the true underlying mean SAR (2.5%) is 25% higher than the target (2.0%).

For *TIRs*, the benefit of more tags is evident in CI width and decisions about hypotheses, though the benefit is relatively small and declines with time. The decision rule used to draw conclusions about the relative efficacy of transportation is influential, but a relatively high probability of reaching the correct conclusion would be achieved even after only 5 years, using a “neutral” decision rule, if true *TIR* differs from 1.0 by at least 20%. If true *TIR* differs from 1.0 by 50%, high probability of reaching the correct conclusion are achieved quickly even under the “wrong” decision rule.

The simulations used in the analyses presented here assumed the number of PIT-tagged fish in each group in each year in the time series was constant. In reality, the numbers of PIT-tagged wild fish vary substantially between years. However, because the methods explicitly account for the numbers of tags each year in estimating sampling variance and in weighting estimates among years, the benefits of this approach over estimating simple means and associated confidence intervals for a time series of annual estimates of SAR or *TIR* would likely be large. Further, application of the methods used here may be especially useful for making inferences from even smaller numbers of marked fish, such as might be the case in estimating within-season trends in Snake River ESU SAR or *TIR*, or in estimating values of these parameters at a population level finer than ESU (e.g., major population group).

2.2.3 Question 3: Compliance of Hydrosystem with Performance Standards in 2000 BiOp

Management questions of interest

Does in-river survival of Snake spring/summer Chinook and steelhead smolts comply with the 2000 BiOp? Here, we consider the effects of the status quo (SQ), low, medium, and high (L, M, and H) designs (Appendix D) on our ability to answer this question in future.

Sampling design and response design options

Table 2.2.7 displays the number of fish PIT-tagged under the SQ, L, M, and H options. Table 2.2.8 displays the number of tagged smolts at LGR, a subset of total fish tagged. Numbers are divided into hatchery and wild in-river and transported smolts. For this section, where the metric of interest is in-river survival, the numbers of most interest are the fish left to migrate in-river following detection or tagging at LGR.

As can be seen from the table, the number of in-river migrants does not vary substantially from the status quo, with the exception of hatchery fish for the Low option. Past work (Hinrichsen and Paulsen, unpublished manuscript) has not uncovered any systematic differences in in-river survival for hatchery and wild Chinook, and so we assume here that the two groups would be combined for any in-river survival analysis. In fact, then numbers on average are sufficiently close to the status quo that we do not expect substantial changes in the in-river survival sampling variance with any of the options (Figure 2.2.18, upper panel). Only if numbers were to decrease dramatically (Figure 2.2.18, lower panel) would the sampling variance of annual estimates of in-river survival increase substantially.

To decrease in the sampling variance will require substantial increases in the number of tagged in-river migrants at LGR beyond any of the scenarios in Tables 2.2.7 and 2.2.8. In general, the confidence bounds displayed in Figure 2.2.18 (upper panel) will decrease by $1/\sqrt{N}$, where N is the factor by which sample size increases. For example, for the SQ, roughly 70,000 in-river migrants leave LGR each year (43,000 hatchery and 27,000 wild). To halve to width of the confidence bounds, one would need to increase the sample size by a factor of 4, to about 280,000 fish per year.

Evaluation of tradeoffs among the above designs

Describe tradeoffs: what do you lose in accuracy, precision, ability to answer question as you move from H to M to L; what do you gain in $ saved?

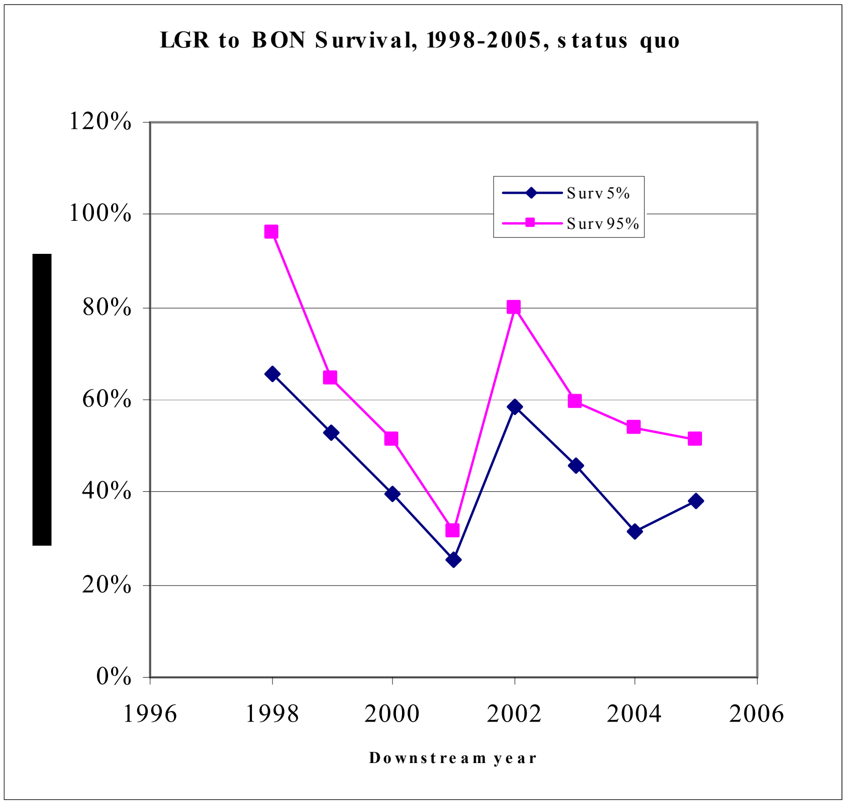
Continuing the above example of increasing the number of tagged in-river migrants, if tags cost roughly $2 and labor costs roughly $1.50 per tagged fish, at the high end of hatchery labor costs (from Table 2.2.2 and Appendix E), increasing sample size by 210,000 (280,000–70,000), assuming fish are tagged at LGR, would require an additional costs of $735K per annum. Note that these costs are very much approximate, and that the tagged fish might well be useful in other monitoring programs, including TIRs, harvest rate estimation using PIT-tags and upstream survival rates as adults pass through the hydrosystem. These and related trade-offs will be analyzed further in FY2007.

Additional work required to improve designs

Propose limited field and analytical investigations to acquire additional information relating to data variability and improve designs (e.g., existing/upcoming Federal RME pilots, novel research undertakings, etc.) Question 3 can be broken down into two different questions:

- *Question 3A : Has in-river survival increased by 9% between 1995-2000 and 2001-2010?*
- *Question 3B: Is in-river survival above the standard for each stock (i.e., 49.6% from LGR to BON for spring-summer Chinook, and 50.6% for steelhead)?*

It would be worthwhile to develop nomograms which explain how much of a change in survival rates (e.g., 9%, 15%, 20%) WOULD actually be detectable at 80% or 90% statistical power over different time frames (i.e., 1, 5, 10, 20 years), using the Skalski et al. (2003) multi-dimensional test for question 3A. It would be good to have an analogous nomogram for 3B (i.e., how much above 50% can be detected over different time frames at 80% statistical power). We also speculate that it might be useful to relate survival rates to in-river environmental conditions (flow, temperature, etc.) to “control” for some of the variability in survival rates over time.



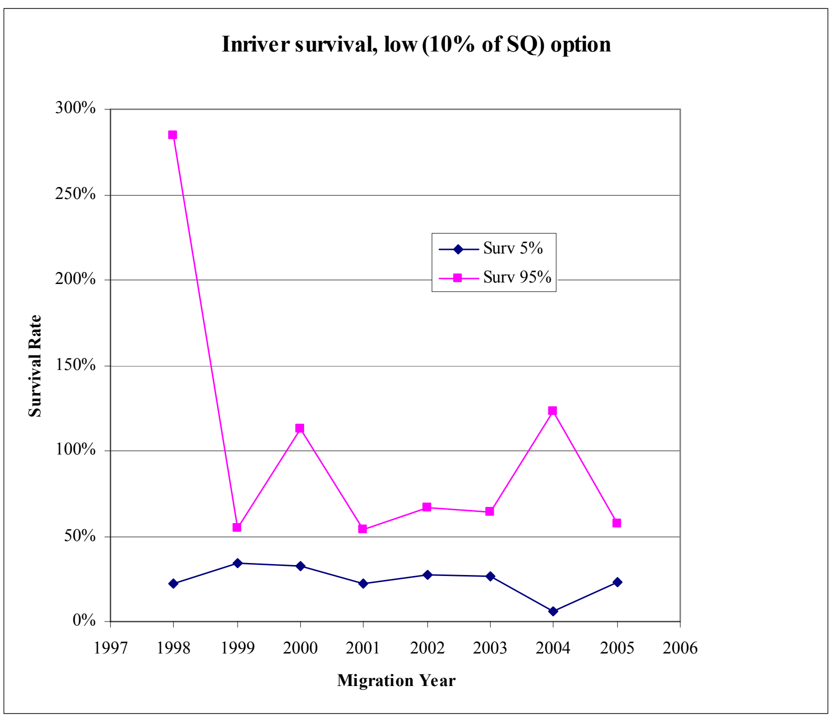


Figure 2.2.18. In-river survival estimates, SQ and low (10% of SQ) scenarios

2.2.4 Question 4: What's the effect of different within-season transportation management actions on post-Bonneville survival of transported fish?

Management questions of interest

Past work by CSS and NOAA suggests that transport-in-river ratios may vary in a roughly predictable way within the course of the (spring) migration season. (See Figure C2 in Marmorek et al. 2004b—CSS Workshop Report), pg. 98). Further work by Paulsen (CSMEP 2005, CSMEP Hydro subgroup 2005) showed that TIR increased over the season for both hatchery and wild Chinook. For hatchery fish, TIR exceeded one across the entire season, while for wild Chinook, it increased from roughly 0.5 to 1.5 as the season progressed.

Sampling design and response design options

Table 2.2.7 displays the number of fish PIT-tagged under the SQ, L, M, and H options. Table 2.2.8 displays the number of tagged smolts at LGR, a subset of total fish tagged. Numbers are divided into hatchery and wild in-river and transported smolts. For this section, where the metric of interest is in-river SAR transport SAR, and TIR, the numbers of most interest are the fish left to migrate in-river following detection or tagging at LGR, and the number of transported smolts.

Evaluation of tradeoffs among the above designs

Obvious $$ vs. precision trade-offs. Increasing tagging/detection effort will substantially increase power, but the details depend on expected TIRs and how those change over a season. Hatchery fish show changes fairly clearly and consistently, but wild fish less so, perhaps due to very small sample sizes. Other trade-offs include spill at dams – higher spill reduces precision and power for both questions 3 and 4, all else being equal, but lower spill reduces the proportion of C0 fish, which may have had significantly higher SARs in years past.

The designs vary in terms of how long it will take to answer the question: more tagging will result in fewer years of study to yield reliable answers, assuming that in future within-season TIRs follow patterns similar to what is seen in existing data.

Additional work required to improve designs

Figure 2.2.19 shows in-season TIRs for hatchery and wild fish for the SQ and High examples. Here, the confidence bounds do not change much, because the number of in-river migrants does not vary much across scenarios (thus the TIR bounds do not vary, since transported migrant increases for wild fish are offset by decreases in in-river numbers).

Thus, even though the number of fish tagged increases substantially, confidence bounds on in-river survival and TIRs do not change very much.

1. Given that the performance measures assessed require fish tagged/detected at LGR, it may make sense to continue or increase tagging efforts at LGR for the high scenario—this is surely cheaper per fish than tagging above LGR, especially for wild fish.
2. While fish tagged at LGR cannot, of course, be used to estimate survival down to LGR, the numbers required for the latter are much smaller than those needed for precise estimates of LGR-BON survival or SARs. A stratified sampling system might be of interest, depending on the questions being asked. That is, one might use wild fish tagged at LGR for TIRs, and wild parr or smolts tagged above LGR for survival down to LGR only.

3. If consistent relationships between wild and hatchery fish can be discerned (CSS has done some work here) substitution of hatchery fish for wild fish may be possible, reducing costs substantially. Perhaps a logistic regression approach may be possible, adding more power to the results.

Table 2.2.7. Descriptions of scenarios for Snake spring/summer Chinook, based on 6/12 table from Dave.

Status Quo	Low	Medium	High
SR Hatchery Chinook:			
200,000 tags @ 5 CSS Hatcheries	10,000 tags @Hat + 10,000 tags @traps to get in-river survival	Distribute tags in proportion to hatchery releases across	Distribute tags in proportionately
10,000 tags @Hat + 10,000 tags @traps to get in-river survival;	20,000 tags @ LGR, all transported	20,000 tags @ LGR, all put in-river	20,000 tags @ LGR, all put in-river
35,000 tags from 3 NPT hatcheries			
TOT tags= 255,000	TOT tags=40,000	TOT tags=275,000	TOT Tags=375,000
SR Wild Chinook:			
40,000 tags @ traps, etc. to get in-river survival	40,000 tags @ traps, etc. to get in-river survival	40,000 tags @ traps, etc. to get in-river survival	146,000 tags @ traps for CSS to get transport & in-river SARs
26,000 tagged at LGR by NOAA	26,000 tagged at LGR by NOAA	26,000 tagged at LGR by NOAA	0 tagged @ LGR (by assumption)
		20,000 tags @ 2 NPT traps to get SARs from sub-basins	40,000 tags @ 2 NPT traps to get SARs from sub-basins
TOT tags=66,000 (29 stream RST's)	TOT tags=66,000 (29 stream RST's)	TOT tags=86,000 (40 stream RST's)	TOT tags=186,000

Table 2.2.8. Fish numbers for scenarios in Table 2.2.7.

	Status Quo	Low	Medium	High
Status Quo are annual averages from 1995-2002				
Fish tagged or detected at LGR				
Hatchery transported:	54,971	20,000	33,966	47,286
Hatchery In-river:	42,904	8,000	54,000	47,333
Wild transported:	8,784	9,000	14,333	49,600
Wild In-river:	27,424	22,000	24,667	24,800
Assumes 0.8 survival, tagging to LGR, and 50% detection prob. @ LGR				
Multipliers vs. SQ (1=same, 2= 2x # available, etc)				
Hatchery transported:	1	0.36	0.62	0.86
Hatchery In-river:	1	0.19	1.26	1.10
Wild transported:	1	1.02	1.63	5.65
Wild In-river:	1	0.80	0.90	0.90

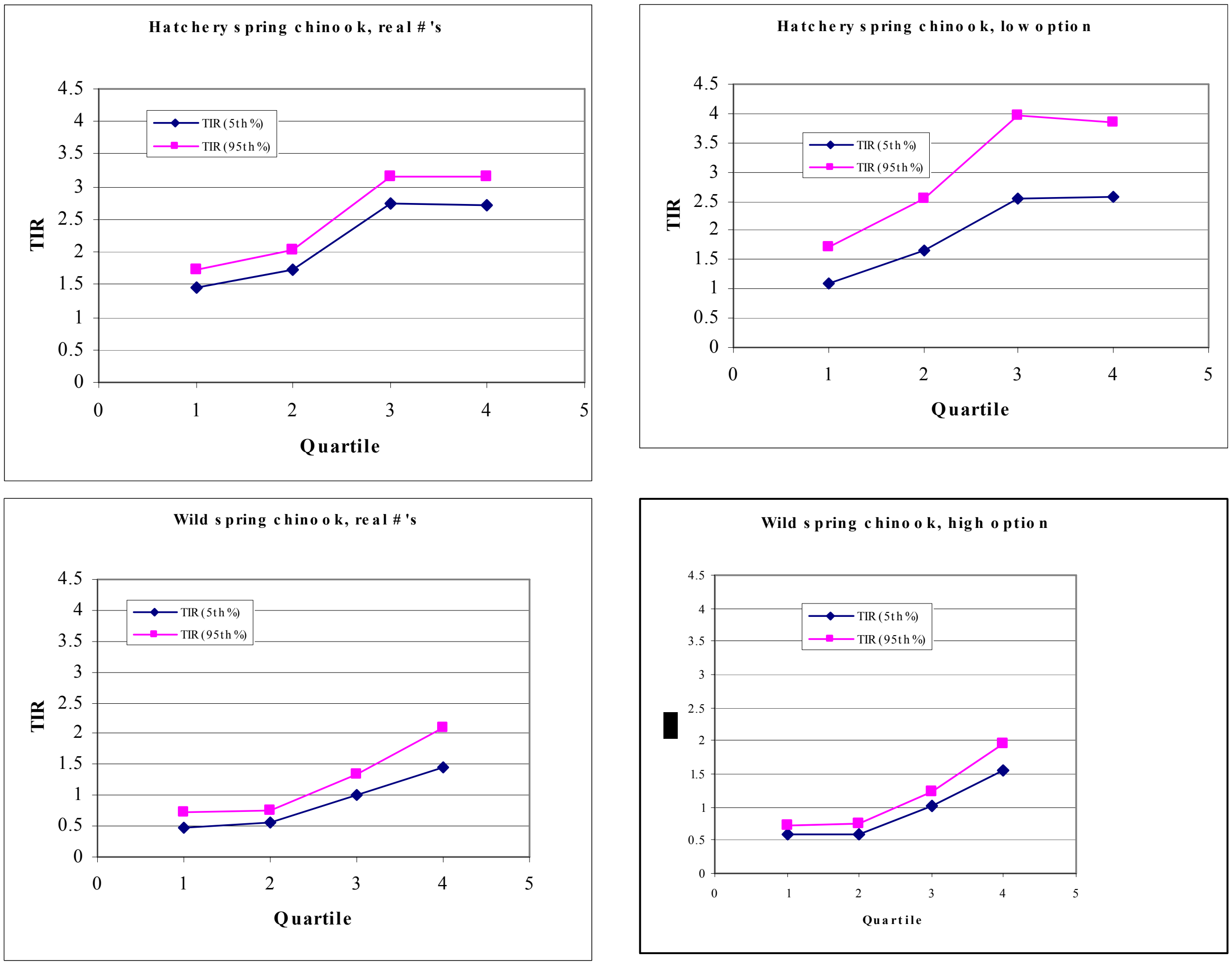


Figure 2.2.19. Confidence bounds on TIR, status quo and high scenarios

2.3 Habitat

Habitat actions are considered a cornerstone of recovery strategies for Columbia River Basin fish stocks but there is a need to more clearly determine the effectiveness of these actions for increasing salmonid survival rates and production. Monitoring designs for evaluating the effectiveness of habitat actions must be able to reliably detect two linked responses:

1. the effect of habitat actions on fish habitat; and
2. the effect of changes in fish habitat on fish populations.

The Habitat Subgroup has recognized that there are serious challenges to the development of a generic template for habitat effectiveness monitoring (i.e., standardized DQO approach can be difficult to apply):

1. Habitat conditions vary greatly across subbasins in terms of their natural biogeoclimatic regimes, the status of their fish populations, the degree of human impact and management, and the number and nature of restoration actions that have been implemented, or are being considered for implementation within them.

2. Habitat effectiveness questions encompass different scales of inquiry, which imply different scales of monitoring.
3. Management objectives for the results of habitat actions are often not clearly articulated and can therefore be difficult to quantitatively evaluate
4. The mechanistic linkages between habitat change and fish response are often poorly understood

CSMEP's Habitat Subgroup has instead been attempting in FY2006 to develop a consistent "process" that can be applied to development of individual monitoring designs dependent on the particular situation. They have been piloting this approach within the Lemhi Subbasin. A full report on the Habitat subgroup's progress in FY2005 in creating a workable design approach for the Lemhi Subbasin (based on a question clarification process to drive development of low, medium and high Habitat designs) is provided on the CSMEP Website. PowerPoint presentations for the Habitat Subgroup's design development (Presentation1, Presentation2) are also available on the site.

The Habitat Subgroup has been comparing their Habitat design development process and the final design recommendations with a parallel design process being undertaken concurrently by ISEMP for the Lemhi HCP. The side-by-side comparison undertaken by CSMEP of these two design efforts (Section 2.3.1) has provided insights into commonalities for design of habitat effectiveness monitoring that will likely occur across subbasins. It has also identified some of the elements that are likely unique to individual subbasins and will not readily lend themselves to standardized design templates. The side-by-comparison also provides an assessment of the pros and cons of "top-down' (i.e., management objectives) vs. "bottom-up" (i.e., scientific questions) design approaches for habitat effectiveness monitoring in the Lemhi Subbasin; an issue that will likely re-emerge in development of an acceptable "process" for improving habitat effectiveness M&E designs within Columbia subbasins. In FY2006 the Habitat Subgroup refined the Lemhi Subbasin design work initially undertaken in FY2005 and began to engage regional managers to determine the management objectives for the subbasin. The intent of this exercise was to "close the loop" i.e., to evaluate how well the CSMEP analysts originally matched up their design questions with the actual management objectives for the Lemhi Subbasin. Part of this "closing the loop" process in FY2006 was a fuller exploration of the actual statistical analyses that should be undertaken within the proposed designs for testing each of the hypotheses formulated by the subgroup within their Lemhi question clarification process (Section 2.3.2). An example of designs for evaluating PIT-tag data that are currently being proposed for the John Day (Section 2.3.3) and that could similarly be applied to habitat action effectiveness studies in the Lemhi subbasin (and other areas) is presented as a subset of the larger design discussion. In FY2006 the CSMEP Habitat Subgroup also began to consider M&E designs for bull trout in the Lemhi Subbasin (Section 2.3.4), and evaluate how these might be integrated with their proposed Low, Medium, High designs originally focused on spring Chinook. Finally, the Habitat Subgroup in FY2006 began to assess what general conditions are necessary for successful design and implementation of habitat effectiveness monitoring (Section 2.3.5) as a precursor to identifying other subbasins that could benefit from similar CSMEP design efforts. Such proactive design efforts for individual subbasins are likely to be of real benefit only when directed to subbasins with major habitat projects planned for the near future and that are supported by a robust, well-funded management program.

2.3.1 Side-by-side comparison of CSMEP and ISEMP Lemhi habitat action effectiveness designs

Introduction

The Collaborative Systemwide Monitoring and Evaluation Project (CSMEP) and Integrated Status and Effectiveness Monitoring Project (ISEMP) have produced separate designs aimed at evaluating the impacts of ongoing and proposed habitat restoration actions in the Lemhi River watershed of Idaho. Both plans specifically address effectiveness of habitat management actions from the perspective of the impact of actions on vital rates (e.g., survival, growth, etc.) of resident and anadromous salmonids. Likewise, each plan contains elements that are intended to be transferable to other locations. Additionally, the plans were developed simultaneously, and included some collaboration among personnel resulting in significant overlap in content. However, the plans maintain a separate history of development, and have some important differences in objectives and scope. It is anticipated that comparing the two designs will indicate what elements of each study are common and so may be transferable to other locations (i.e., which elements can serve as a "template" for other studies), and which elements are different (i.e., are likely to be unique to a location or given habitat action).

Objectives defined under the Lemhi Conservation Plan (LCP) fall into three broad categories:

1. Remove or reduce upstream and downstream migration barriers to fish and provide access to available spawning and rearing habitat by:
    a. providing flow to maintain hydraulic and ecological connectivity between the mainstem and tributaries so that fish have access to historically productive habitat;
    b. providing flow in the lower reach of the Lemhi River so that adults and juveniles can freely migrate in and out of the Lemhi subbasin;
    c. removing physical obstructions (e.g., irrigation berms and push-up dams) that limit localized movements of upstream and downstream migrating fish; and
    d. minimizing entrainment into irrigation ditches that do not provide adequate rearing habitat and do not functionally reconnect with the Lemhi River or tributaries.
2. Maintain or enhance riparian conditions characteristic of good habitat to ensure that adequate vegetation persists to provide shade, increase bank stability and protection, decrease sediment input, and promote the recruitment of large woody debris.
3. Decrease sediment, temperature, provide quality substrate, increase the abundance and quality of off-channel habitat, and increase pool frequency and quality to improve productivity and survival.

Both the Collaborative Systemwide Monitoring and Evaluation Project (CSMEP) Habitat workgroup and NOAA's Integrated Status and Effectiveness Monitoring Project (ISEMP) drafted monitoring plans intended to enable a robust evaluation of the effectiveness of proposed and ongoing habitat actions in the Lemhi. The CSMEP group primarily focused on the effects of tributary reconnections while the ISEMP design focused on a larger set of questions, pertaining to all of the habitat actions proposed for the Lemhi. From their initiation, each design conceptualized information needs in a slightly different manner. The following paragraphs summarize the conceptual approach that each group employed.

CSMEP

The CSMEP design is an attempt to apply a standardized process to the development of a monitoring plan that addresses habitat management action effectiveness. The standardized process employed is the Data Quality Objectives (DQO) process developed originally by the US Environmental Protection Agency (EPA) to match monitoring plan design to management needs. The DQO provides steps to extract from managers the list of decisions that they make, and from these decisions specify the technical questions that must be answered to make those management decisions. The technical questions in turn define a set of data needs, acquisition of which specifies the design requirements for the monitoring plan itself. In the DQO there is an *a priori* connection between indicator values and management decisions. Thus, one important characteristic of the CSMEP project is that management decisions determine the data needs, rather than a specific analytical model or analysis design. The DQO process has been used in diverse monitoring programs outside the Columbia River Basin and is being used by other design groups in CSMEP (i.e., harvest, hatchery, hydro, status and trends), to varying levels of success. It is appropriate to examine if the same process would be helpful for developing action effectiveness monitoring designs in the Lemhi.

The DQO Process (EPA, 1994 & 2006) is a systematic planning process that is the EPA's recommended tool when data are being used to select between two alternative conditions (e.g., compliance or non-compliance with a standard). Monitoring for standard compliance or regulatory needs in general is related to specific levels of indicators or their pattern of change over time. The monitoring designs for compliance monitoring do not rely on the same elements as hypothesis testing or effectiveness monitoring (MacDonald et al., 1991). The potential for a design process for compliance monitoring to generate effectiveness monitoring designs has yet to be established. Indeed, as mentioned above there is no specification of an analytic process; no specification of a statistical test appropriate for the design. This uncertainty was acknowledged in the DQO guidance itself:

> *If a statistical hypothesis is not linked to a clear decision in which the decision maker can identify potential consequences of making a decision error, then some of the activities recommended in this guidance may not apply. (pg.3; EPA, 1994).*

More recent EPA guidance reinforces this caution:

> *The U.S. Environmental Protection Agency (EPA) has developed the Data Quality Objectives Process as the Agency's recommended planning process when environmental data are used to select between two opposing conditions. The Data Quality Objectives Process is used to develop Data Quality Objectives that clarify study objectives, define the appropriate type of data, and specify tolerable levels of potential decision errors that will be used as the basis for establishing the quality and quantity of data needed to support decisions. When this Process is not immediately applicable (i.e., the objective of the program is estimation, research, or any other objective that does not select between two distinct conditions), the Agency requires the use of a systematic planning method for defining performance criteria." (pg. i, EPA 2006)*

Clearly, the cautions specified above are relevant in many current habitat action effectiveness monitoring studies which, and this limits expectations for DQO utility in the Lemhi example.

As a prototype exercise, the CSMEP project focused on the Lemhi basin in Idaho for several reasons. First, the Lemhi had a habitat conservation plan in development that cataloged the management actions planned in the basin. The conservation plan identified groups of projects that could potentially serve as replicates in testing hypotheses about classes of projects, and geographic sub-domains within the Lemhi that could serve as contrasts for tests of aggregated effect of diverse actions. Second, the conservation plan identified stream de-watering as an overarching habitat management issue that was to be addressed with a series of channel reconnection projects phased in over a period of years. This management design provided the prospect to opportunistically overlay an experimental design—accessory to the DQO—to test hypotheses about action effectiveness.

The efforts of the CSMEP Habitat workgroup in the Lemhi subbasin could, therefore, be characterized as an attempt to evaluate whether a standardized organizational process (i.e., the DQO) could generate a technical monitoring design sufficiently robust to test the effectiveness of selected habitat management actions.

ISEMP

The ISEMP project was initiated to meet population and habitat status and trend and habitat action effectiveness information requirements described in the 2000 NOAA Biological Opinion on the Federal Columbia River Power System (FCRPS BiOp). These requirements are intended to determine whether offsite mitigation actions prescribed in that document are realizing their assumed survival benefits. Like CSMEP the ISEMP targets individual subbasins as test beds for alternative monitoring designs in evaluating the effectiveness of the FCRPS BiOp mitigation clearly has a Basinwide scale,. Unlike the CSMEP study in the Lemhi basin, the ISEMP benefited from the existence of predefined information needs associated with the FCRPS BiOp evaluations. Although the ISEMP was initiated in the context of the FCRPS BiOp, it, like CSMEP, is an attempt to address State, Tribal and local monitoring needs simultaneously to those identified by the Federal agencies.

In order to insure maximum design efficiency (i.e., dovetailing existing and proposed monitoring effort) and to insure that federal, tribal, and state information needs were addressed, the ISEMP formed the Research, Monitoring, and Evaluation Technical Oversight Committee (RMETOC). The RMETOC, composed of policy and technical staff of agencies with management jurisdiction in the Lemhi River watershed, is the authority that reviews alternative ISEMP designs for adequacy and insures that implementation of habitat restoration actions and monitoring activities are coordinated. Thus, the ISEMP design benefited from direct collaboration with the agencies that are planning and implementing habitat restoration actions in the Lemhi subbasin. Although the actual implementation of habitat restoration actions in the Lemhi River watershed are subject to the same constraints and limited predictability as any other location (e.g., landowner cooperation, sufficient funding, etc.), coordination between the ISEMP and local managers is expected to prove useful in staging the implementation of habitat restoration actions to maximize contrast in statistical evaluations.

Actions proposed under the Lemhi Conservation Plan (LCP), the guidance document for habitat restoration actions, use a “phased” approach, such that a number of actions will be completed in the first five years of the project (Phase I), and additional actions (Phase II) will be identified after evaluating the effectiveness of Phase I activities. Thus, the ISEMP project conceptualized an approach that would incorporate a predictive component, utilizing information from Phase I, to enable managers to estimate the potential impacts of alternative actions prior to the implementation of Phase II. This requires an ability to identify the factors that limit population growth, such that Phase II actions can be targeted to increase survival at the critical life stage(s), under the implicit assumption that the limiting factor(s) of greatest relevance might differ between phases.

The ISEMP project elected to employ a model (Sharma et al. 2005) that had previously been employed for coastal coho salmon. The model was selected for five primary reasons:

1. the model had undergone peer-review and had proven useful for coastal coho salmon;
2. the model can be populated by diverse classes of data, providing a useful mechanism to test the applicability and cost-effectiveness of different data types (e.g., remote sensing versus on-the-ground sampling);
3. if the model can be manipulated for use with a diversity of species across a large spatial scale, it could potentially be employed as a “template” for habitat action effectiveness efforts;
4. the analytical foundation of the model applies general linear modeling, providing tremendous flexibility in data analysis (e.g., the ability to apply before-after, before-after-control-impact or modifications thereof, such as the staircase design described by Walters et al. 1988); and
5. the model enables simulations of the effectiveness of alternative management actions, thus providing a useful prioritization and planning tool for managers.

Questions

CSMEP

- Have the actions implemented under the LCP expanded the distribution of rearing juvenile salmonids within the basin and increased the density of rearing juvenile salmonids relative to average mainstem densities by X% over 30 years (with Y precision) when the number of spawners, natural disturbances, climate indicators, and habitat conditions not-impacted by the actions have been accounted for?
- Have the actions implemented under the LCP produced at least a Y% increase in the number juvenile spring Chinook salmon leaving the Lemhi River in 30 years (+/- X%) when the number of spawners, natural disturbances, climate indicators, and habitat conditions not-impacted by the actions have been accounted for?
- Have the relative magnitudes of the seasonal migration pulses and size distribution of migrating Chinook juveniles leaving the Lemhi River changed over the life of the LCP?
- Have the actions implemented under the LCP increased the abundance of bull trout in reconnected tributaries relative to unconnected tributaries by X% over 30 years (with Y precision)?
- Have the actions implemented under the LCP increased parr-smolt survival (X% +/-Y precision) of juvenile spring Chinook salmon leaving the Lemhi River in 30 years when the number of spawners, natural disturbances, climate indicators, and habitat conditions not-impacted by the actions have been accounted for?
- Have the returns of adults Chinook salmon to the Lemhi basin increased X% (+/-Y precision, see VSP criteria developed by ICTRT) over the period of the LCP?

Here X and Y are to be identified within an iterative process that uses work to date to inform management expectations for performance. Although the questions refer to "*actions implemented under the LCP"*, it is implicit that this refers initially to channel reconnection projects.

Importantly, none of these questions linked decisions to potential consequences of decision errors and as such lie outside of the strict domain of the DQO. Consequently an *ad hoc* process was recruited to develop and refine habitat effectiveness questions seen as potentially important by CSMEP participants into a format that could then be used to design appropriate monitoring. The *ad hoc* question clarification process was inserted in the CSMEP implementation of the DQO at this point in the Lemhi effectiveness monitoring design.

ISEMP

- Can existing remote sensing and GIS information be used to describe existing habitat conditions, identify habitat attributes that limit salmonid productivity, and prioritize habitat actions?
- Have the life stage specific factors that limit salmonid productivity been correctly identified?
- Can a modeling approach enable managers to prioritize restoration actions to maximize their impact on the population growth rate?
- Have restoration actions changed the distribution of resident and anadromous juveniles and adults?
- Have restoration actions increased juvenile survival from the egg to fry, fry to parr, parr to presmolt, and presmolt to smolt life history stages?

- Have restoration actions improved the condition of juveniles?
- Have restoration actions increased habitat capacity (regardless of utilization)?
- How have the observed changes in juvenile survival, condition, and distribution changed the population growth rate and what type, density, and extent of habitat restoration activities will be required to stimulate a positive population growth rate?

Methods

CSMEP

Data acquisition methods employed in the CSMEP design are standard approaches to fish enumeration and habitat inventories. The CSMEP design identified three levels of intensity of effort, recognizing that resources may be in short supply and that managers may opt to compromise on the total number of questions answered or the power to resolve individual questions. The methods included:

- ground and aerial redd and carcass counts that are repeated within the year;
- screw traps to estimate juvenile densities operated continuously during the season;
- snorkel counts of juvenile densities;
- distributed network of PIT-tag detectors located at major junctions in the stream network and at locations targeted to study designs;
- distributed network of PIT-tag tagging stations facilitated by seining, trapping and snorkeling; and
- habitat inventory surveys performed at distributed sites within the basin on time scales matched to characteristic time scales of change in habitat condition.

The specific locations of these activities are determined based on the degree to which watersheds within the Lemhi basin can act as experimental contrasts given the distribution in space and time of the implementations of habitat management actions. An example distribution of effort is described in Figure 2.3.1.

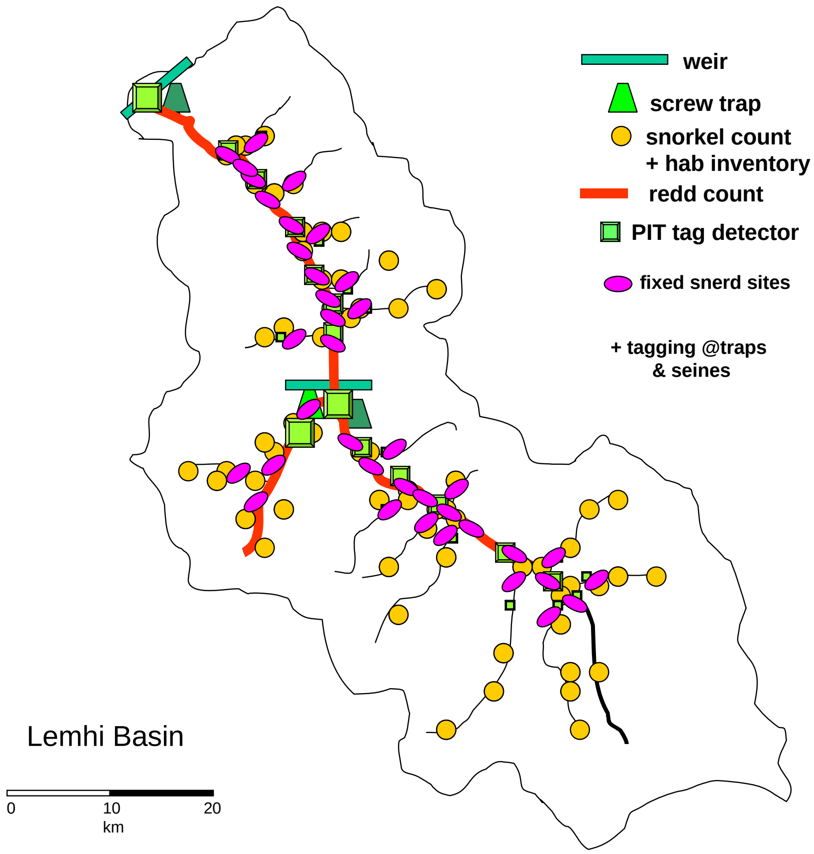


Figure 2.3.1. Example high design for Lemhi HCP. Fixed snerd sites are where snorkellers herd fish into a seine.

There is no single analytical approach that has been specified within the CSMEP project. As pointed out above, the DQO process is based on having an *a priori* expectation for the meaning of various indicator levels—which is not currently available in the context of habitat management actions. Indeed, as mentioned above this expectation of the DQO may not be useful in the context of hypothesis testing. Therefore, various analytical methodologies are being assembled and tested in parallel with this product. At this point we can offer some alternative approaches and comment on their strengths and weaknesses.

Asymmetrical ANOVA/ANCOVA (i.e., BACI type designs; Underwood, 1994) could be used to evaluate the impacts between tributaries that were connected and connected within the upper main-stem Lemhi. The CSMEP design also identified Hayden creek as a larger scale control that may help describe common climatic effects such as precipitation and air temperature and the increase in large scale distribution. The details of how to develop an ANCOVA with a nested design with random and fixed effects will be developed in more detail at a later date. Confidence intervals will be used to assess difference between treated and non-treated areas after accounting for the above components of the ANCOVA. Confidence intervals provide much more information than the simple decision to reject or failure to reject a null hypothesis using p values (e.g., Bradford *et al.,* 2005).

Alternatives to ANCOVA analyses will also be explored. For example, the implementation of habitat reconnections will be staggered in time, a design which lends itself to a Bayesian analysis. Prior probabilities between competing hypotheses would be set equal before the implementation of restoration actions. Posterior probabilities estimated from data collected after reconnections would be used to update the priors before implementation of the next set of reconnections. The output of a Bayesian analysis provides the probability of a hypothesis being correctly expressed in terms much more intuitive than those used for classical frequentist statistics. Other alternatives such as Randomized Intervention Analysis (RIA; Carpenter *et al*., 1989) or dynamic linear modeling to compare competing predictive time series models could also be used to describe results of the HCP restoration actions (Poole *et al*., 1994). Adapting these techniques to the CSMEP Lemhi design will await the time when targeted pilot data can be acquired and the variance structures estimated.

Power analyses will be necessary and can take advantage of some of the prior information collected to provide information about the variability expected from different monitoring protocols. The statistical design being employed will also have to be known to estimate required samples sizes. Specifying these design elements will also await the collection of preliminary data and the estimation of variance structure.

ISEMP

The goals of the Lemhi ISEMP will be addressed using a model-based design that employs a modified version of a watershed model described by Sharma *et al*. (2005). The model views fish abundance and productivity as functions of habitat quantity and quality. In short, the maximum number of fish that can be produced by the watershed at any given life stage is constrained by habitat quantity, while the survival of fish from one life stage to the next is constrained by habitat quality. Habitat quantity and quality take the form of matrices following the fish and habitat relationships described by Bisson *et al*. (1981). Model parameters are initially populated using remote sensing or GIS data in concert with empirical estimates of fish survival and productivity obtained from existing data and the fisheries literature. This coarse scale information will be validated using empirical data collected during habitat surveys and by employing a model-defined sampling program that yields the pertinent life stage specific abundance, survival, and growth data. Once validated, the model will be used to simulate the effects of proposed habitat actions on the quantity and quality of habitat, which can then be used to estimate the potential effects of the actions on fish vital rates. As information on the realized impacts of implemented habitat actions is obtained, the model will be updated using those data and re-parameterized. Depending on the availability of reference sites and the temporal and spatial implementation of restoration actions, data analysis can take many forms ranging from before-after (BA) and before-after-control-impact designs (BACI; Stewart-Oaten *et al.*, 1986), extensive BACI (Hicks *et al.*, 1991), versions thereof that employ general linear modeling such as a staircase design (Walters *et al*., 1988). We view this flexibility as a critical feature of the model owing to the fact that despite best intentions, the timing and location of implementation of restoration activities will be constrained by factors such as the timing of funding opportunities and the ability to obtain and maintain landowner support.

Finally, while habitat restoration actions are anticipated to exert their greatest effect on the freshwater stock recruitment relationship, recovery is typically viewed in the context of the population growth rate calculated over the entire life history of the fish (e.g., number of adults at time t+1 divided by the number of adults at time t that gave rise to them). Within the Lemhi ISEMP, freshwater productivity will be viewed in the context of the full anadromous life cycle using a Beverton-Holt stock recruitment relationship. Survival of emigrating juveniles will be tracked via PIT-tags and external marks during freshwater juvenile life history stages and via PIT-tags through the hydropower system, with tagging rates sufficient to yield smolt to adult return rate information. This information will enable managers to evaluate the degree to which freshwater productivity must be increased to offset hydrosystem mortality and buffer against poor ocean conditions.

Products

CSMEP

CSMEP is a design effort; there is no specific plan for implementation in the Lemhi subbasin. In turn, many of the products that are anticipated from the CSMEP project will be dependant on some pilot implementation and refinement of design. Therefore, a list of products requires some caution and an understanding that currently the products concern the design process rather than technical tools themselves. Also important is that CSMEP is intended to produce generic products that could be applied to other locations, rather than a specific design for a particular location such as the Lemhi basin. To date the principal product is the design process itself. In marked contrast, the main ISEMP product is a determination of whether or not their actions have measurable, beneficial effects on listed stocks, not the plan or the process for formulating it.

Products applicable at broader spatial scales.

1. The CSMEP project will demonstrate the potential utility of the DQO in assessing large-scale hypothesis testing designs.
2. The CSMEP project will demonstrate if a standardized organizational process can produce a technical monitoring design that can test the effectiveness of habitat management actions.
3. The CSMEP project will test whether a single monitoring design will be able to evaluate the question *are a class of restoration actions (e.g., channel reconnection) effective?,* when the study is nested within a basin scale study that includes diverse habitat management actions.
4. The CSMEP project will provide a set of tools to apply the DQO process to the study of project effectiveness that includes methods for specifying the scientific questions addressed and the analytical process for evaluating the data.

Products applicable primarily at the Lemhi River watershed scale.

1. The CSMEP design will evaluate the effects of habitat actions on the distribution of juvenile rearing and adult spawning. In short, we will evaluate whether juveniles and adults access and utilize newly available habitat resulting from tributary reconnections and the construction/ restoration of off-channel habitat (e.g., spring fed irrigation diversions in the lower Lemhi River watershed).
2. Using physical marks (fin clips, Bismark-Brown dye, and PIT-tags) and recapture (via repeated seining surveys and extended length PIT-tag arrays), we will evaluate life stage specific juvenile abundance and growth, and calculate survival from one life stage to the next. General linear modeling utilizing a BA or BACI based design will be used to evaluate how these attributes respond to various proposed habitat actions.

ISEMP

If implemented, the ISEMP will provide products that are transferable, and products that are specific to action effectiveness evaluations in the Lemhi River watershed.

Products applicable at broader spatial scales.

1. The ISEMP will test the transferability of a modified version of a watershed model (Sharma et al. 2005).
2. Remote sensing and GIS-based data on land use/land cover will be contrasted with empirical (on-the-ground) data to determine whether less labor intensive technology can provide sufficient information on habitat complexity and quantity.
3. The ISEMP will test whether an intensive model-based sampling design can provide sufficient information to partition variance (resulting from process error, sampling error, and spatial variation) to enable simulations of the probable effects of proposed suites of habitat actions on freshwater productivity and abundance and the population growth rate.

Products applicable primarily at the Lemhi River watershed scale.

1. The ISEMP design will evaluate the effects of habitat actions on the distribution of juvenile rearing and adult spawning. In short, we will evaluate whether juveniles and adults access and utilize newly available habitat resulting from tributary reconnections and the construction/ restoration of off-channel habitat (e.g., spring fed irrigation diversions in the lower Lemhi River watershed).
2. Using physical marks (fin clips, Bismark-Brown dye, and PIT-tags) and recapture (via repeated seining surveys, rotary screw traps, and extended length PIT-tag arrays), we will evaluate life stage specific juvenile abundance and growth, and calculate survival from one life stage to the next. General linear modeling utilizing a BA or BACI based design will be used to evaluate how these attributes respond to various habitat actions.
3. Finally, the Beverton-Holt relationship within the model will be used to simulate the degree to which freshwater productivity must be improved to achieve a positive population growth rate. Based on the observed outcome of the implemented restoration actions, we will evaluate what (if any) combination of habitat restoration actions is predicted to have a sufficient impact.

Summary

Both ISEMP and CSMEP arrived at monitoring designs for the Lemhi basin. Among the similarities was some overlap in the questions asked. For example, both plans attempt to address the degree to which habitat management actions have affected the distribution and survivorship of anadromous and resident salmonids in the basin. Also both plans look to capitalize on the distribution of habitat management actions planned in the LHP. Similarities also exist in field methods. Specifically, both plans propose PIT-tagging, weirs, habitat inventories and field counts of fish and redds. It is not clear if this similarity is simply due to the generic flexibility of these particular methods, or the narrow range of potential tools currently available. Although the current diversity of tools may be narrow, both plans have demonstrated flexibility in adopting technological progress. For example, both plans have incorporated the tandem extended length PIT-tag arrays designed by NOAA Fisheries, first discussed in the CSMEP group in their NAMPA workshop in April 2005, and deployed for the last year as a component of the ISEMP project in the John Day basin. In any event, these similarities suggest that when planning effectiveness monitoring in other basins one may expect to use similar techniques and protocols.

There were also some significant differences between the project designs. In particular, the CSMEP design is more detailed in terms of process than technical content, while the ISEMP is more detailed in analytical modeling approach and fairly thin on process. As a consequence of this different reliance on process there are some important differences in the questions asked by the projects. ISEMP for example, has more questions aimed at scientific or technical needs—such as the generic utility of remote sensing.

There are also important differences in the specifics of monitoring. For example, the CSMEP relies on comparisons of sub-domains within the Lemhi to contrast the aggregate impact of actions on demographic units; the ISEMP relies on a different analytical approach. This difference in analytical approach results in a different distribution of PIT-tag detectors and tagging stations. Overall while there is overlap in the methods used (i.e., *what* and *how* to make measurements), there are significant differences in *where* and *when* to measure.

Importantly, the limitations of the DQO in addressing habitat management effectiveness monitoring studies were revealed in CSMEP's Lemhi effectiveness monitoring design. Without managers in the room, it is not surprising that our application of the DQO had somewhat different results from the ISEMP, where local stakeholders were directly involved. In addition, a management support tool like the DQO requires a clear functional understanding of the ecological system response to habitat actions. This understanding was not apparent for the Lemhi, and may often be the case for many subbasins in the Columbia and most habitat restoration project types. Within CSMEP's Lemhi basin project this was manifest in the requirement for a new, *ad hoc* process to generate questions that allowed the development of alternative monitoring designs. On the other hand, the ISEMP project started with scientifically competent questions and used these to determine the monitoring design, and inform the management decisions in the basin.

The differences that exist between the CSMEP and ISEMP designs, both in the same Lemhi basin, also suggest that as one moves to other basins where habitat management issues are more diverse there may be potentially larger differences in design. In particular, where and when to deploy monitoring resources will be impossible to predict ahead of consideration of the mature scientifically questions specific to those locations. Consideration of those questions will in turn require a unique rather than template process that is informed by the management history and management plans in those new locations.

The Lemhi monitoring designs generated by the CSMEP and ISEMP have identified a number of pragmatic issues that must be dealt with in any technical "template" for habitat action effectiveness monitoring. First, it is unlikely that opportunities will arise for a truly randomized treatment to be applied to the implementation of habitat restoration actions. In short, the implementation of habitat restoration actions is subject to vagaries that serve to disconnect actions from the monitoring plans that are designed to evaluate them. Therefore, practical action effectiveness monitoring designs must incorporate sufficient analytical flexibility to compensate for less than complete control over action implementation. Second, it is unlikely that disparate existing sampling efforts can provide information at the temporal and spatial scales required for action effectiveness evaluations. Thus, it is likely that the efficient implementation of action effectiveness evaluations will necessitate both new sampling effort and the modification of existing sampling efforts. Finally, it is clear that targeted research that illuminates the mechanistic linkages between habitat actions and fish vital rates is still needed to provide managers with the tools to make the tactical decisions in monitoring required by a management support process like the DQO. It is likely that the implementation of designs such as those developed by CSMEP or ISEMP can identify these mechanistic linkages; and that information may then prove to be a useful foundation for the formulation of transferable monitoring templates.

2.3.2 Experimental designs and analyses for application to the Lemhi subbasin pilot and other habitat action effectiveness studies

Multiple experimental designs exist to assess the impacts of stream restoration efforts. Most of these designs were developed to evaluate the impact of some human perturbation on a resource (Box and Tiao 1975, Green 1979, Steward-Oaten and Bence 2001, Downes et al. 2002). The designs precisely address how the impact is assessed and proper statistical models have been developed to answer these specific questions (Downes et al. 2002). Using the improper statistical model, assumes a different design and question than may have been originally stated. Downes et al. (2002) suggest that it is incorrect to determine the proper statistical model for analysis after the data is collected. The experimental design is driven by the question and the statistical model is driven by the design. The statistical model requires sampling to occur in a certain fashion (e.g., random versus fixed assignments of treatments). The literature discussing these designs is confusing and often conflicting (e.g., Underwood 1994, Steward-Oaten and Bence 2001). Therefore a discussion of the different designs and associated statistical models is warranted because these components drive the sampling design needed to address our question of interest.

The most common designs to evaluate the impacts of restoration actions and the method preferred in this Lemhi design, is to apply a Before and After (BA) treatment comparison. In BA designs, samples are taken at various locations before and after a treatment. This occurs in the same reach or reaches impacted by restoration action, but in some situations are also measured in control areas, referred to as a before-after-treatment-control or BACI design. In most cases, the use of control(s) greatly increases the power of detecting impacts; however, poorly chosen controls sites can decrease the power of detecting an impact (Korman and Higgins 1997, Roni et al. 2003).

The most common approaches to assess the impact of a human action on an ecological process in experimental fashion is related to the family of general linear models such as analysis of variance (ANOVA) models and time-series analyses. The ANOVA approaches are flexible, robust and powerful hypothesis testing procedures (Downes et al. 2002). We will discuss this family of approaches and how they different in their spatial and temporal sampling designs; however, much of this information is discussed in more detail elsewhere (Underwood 1991, 1992, 1994, Steward-Oaten and Bence 2001, Downes et al. 2002, Roni et al. 2005). Intervention analyses (IA) are another family of models that have been widely used to assess environmental impacts (Stewart-Oaten et al. 1986, Carpenter et al. 1989, Steward-Oaten and Bence 2001). These models are based on time series analyses to estimate environmental impacts (Box and Tiao 1975) and do not use controls to measure the expected range of responses that would be expected in the absence of the manipulation as in the ANOVA approaches but used rather as a set of covariates to filter out variability.

Downes et al. (2002) summarized the BACI designs into four types; BACI, BACIP, MBACI, and beyond BACI. In early BACI designs, a measurement was made before an impact at control and impacts locations and then after the impact. It was soon realized that paired measurements through time pre- and post impact was required to prevent spurious results (Green 1979, Stewart-Oaten 1986). Downes et al. (2002) refer to this as BACIP (P for paired measurement through time). Keough and Mapstone (1995) further extended the BACI design to contain multiple controls and if possible multiple impact sites, referred to as MBACI (Downes et al. 2002). In a similar vein, Underwood's (1991, 1992, 1994) beyond-BACI designs employ asymmetrical ANOVA/ANCOVA models to look at spatial and temporal impacts using multiple controls and treatments often evaluated hierarchically over a several time periods. These designs are summarized in Table 2.3.1. Here we add the "staircase" design to this family of models, developed by Walters et al. (1988), used to estimate transient responses to management action.

BACIP designs make comparison between paired measurements in treatment and controls areas. Often there is little flexibility or limited sites where treatments and control areas reside. Therefore, little opportunity exists to randomly choose control and treatments areas. Therefore, treatment and control sites must be treated as fixed affects where inferences should only be attributed to the universe where treatments occurred. In this situation, the question therefore becomes; has the action resulted in a change to the treatment site relative to the control site? Measurements are taken in both the treatment and control sites at multiple times pre- and post- manipulation. In the BACIP designs suggested by Steward-Oaten et al. (1986), samples should be collected at random time periods and far enough apart to reduce temporal autocorrelations. Therefore, a specific sampling strategy for BACIP design emerges where a treatment and control site are measured at random time periods pre- and post- manipulation. The resulting model is:

$$y_{ijpm} = \mu + \mathbf{C}_{i\ldots} + \mathbf{B}_{..p.} + \mathbf{T(B)}_{.jp.} + \mathbf{CB}_{i.p.} + \mathbf{CT(B)}_{ijp.} + \varepsilon_{ijpm} \qquad (1)$$

where:

y_{ijpm}	is the m^{th} observation at Control or Impact site i at Time j in the Before or After period p
μ	is the population grand mean for the measured response variable
$\mathbf{C}_{i\ldots}$	is the time-averaged effect of being in the **C**ontrol or Impact treatment
$\mathbf{B}_{..p.}$	is the spatially averaged effect of the period **B**efore or After the activity starts
$\mathbf{T(B)}_{.jp.}$	is the spatially averaged effect of **T**ime j within **B**efore or After period p
$\mathbf{CB}_{i.p.}$	is the effect of being in either the **C**ontrol or Impact treatment either **B**efore or After the commencement of the activity (i.e., interaction effect)
$\mathbf{CT(B)}_{ijp.}$	is the effect of being in either the **C**ontrol or Impact treatment at Time j within either the **B**efore or After periods
ε_{ijpm}	is the residual value after accounting for all the above effects (and estimable only when multiple subsamples are taken within each location at each time)

This formula is the general equation that explicitly lays out the effects that would be estimated in this design. However, if samples are taken in the Control and Treatment areas at the same time then the equation can be greatly simplified by taken the average difference for the Control area and Treatment area in each period of Before and After the manipulation. We can then estimate the difference between the observation at the Control and Impact treatment areas for each particular time j. The resulting equation is then

$$d_{pj} = \mu + \mathbf{B}_{p.} + \varepsilon_{pj} \qquad (2)$$

where:

d_{jp}	is the j^{th} observation of the difference between the Control and Impact in the period p Before or After the activity
$\mathbf{B}_{p.}$	is the average **difference** between the Control and Impact, Before or After the activity starts
ε_{pj}	is the residual value after accounting for the above effects

MBACI designs have been developed to address questions about the impacts of an action across broader scales. Multiple treatment and control locations are chosen randomly from a group of potential locations; thus the ability to extrapolate to a larger extent. The design by Keough and Mapstone (1995) compares a fixed period of time before the manipulation to (in ideal situations) a similar period of time after the manipulation. All time units (e.g., years) are sampled during these periods. This is modeled using the following equation:

$$y_{inmpj} = \mu + \mathbf{C}_{i....} + l(\mathbf{C})_{in....} + \mathbf{B}_{...p.} + \mathbf{T}(\mathbf{B})_{...pj} + \mathbf{CB}_{i..p.} + \mathbf{CT}(\mathbf{B})_{i..pj} + l(\mathbf{C})\mathbf{B}_{in.p.} + l(\mathbf{C})\mathbf{T}(\mathbf{B})_{in.pj} + \varepsilon_{inmpj} \qquad (3)$$

where:

y_{inmpj}	is the m^{th} observation at location n in Control or Impact treatment i at Time j in the Before or After period p
μ	is the population grand mean for the measured response variable
$\mathbf{C}_{i...}$	is the time and location averaged effect of being in a **C**ontrol or Impact treatment
$\boldsymbol{l}(\mathbf{C})_{in...}$	is the time-averaged effect of being at the n^{th} location within treatment i
$\mathbf{B}_{..p.}$	is the spatially averaged effect of the period **B**efore or After the activity starts
$\mathbf{T}(\mathbf{B})_{...pj}$	is the spatially averaged effect of **T**ime j within **B**efore or After period p
$\mathbf{CB}_{i..p.}$	is the effect of being in either the **C**ontrol or Impact treatment either **B**efore or After the commencement of the activity (i.e., interaction effect)
$\mathbf{CT}(\mathbf{B})_{i..pj}$	is the effect of being in either the **C**ontrol or Impact treatment at Time j within either the **B**efore or After periods
$\boldsymbol{l}(\mathbf{C})\mathbf{B}_{in.p.}$	is the effect of being in the n^{th} location in either the **C**ontrol or Impact treatment at **T**ime j within either the **B**efore or After periods
$\boldsymbol{l}(\mathbf{C})\mathbf{T}(\mathbf{B})_{in.pj}$	is the effect of being at the nth location in either the **C**ontrol or Impact treatment at **T**ime j within either the **B**efore or After period
ε_{ijpm}	is the residual value after accounting for all the above effects (and estimable only when multiple subsamples are taken within each location at each time)

Variants of this design include multiple paired treatment and control locations (e.g., pairs stratified by some environmental variable), and single sample event before and after the manipulation for multiple treatment and control sites.

Several beyond-BACI designs have been presented by Underwood (1991, 1992, 1993, 1994) but in general these designs, like MBACI, are used to address broader questions about the impacts of a type of human event on the landscape. Therefore both treatment and control sites are distributed randomly; however, in contrast to MBACI but similar to BACIP, sampling periods are suggested to be random through time pre- and post-manipulation to make broader statements about action through time. Models evaluating this design are can be written as in the equation 3. More complex models can be used to evaluate hierarchical designs, purposed to evaluate the spatial extent of the perturbation (Underwood 1994). Sampling occurs within a random nested design (Underwood 1994).

Walters et al. (1988) proposed a different approach to determine whether the time trend of responses were a result of natural environmental influences that happened to occur concurrent to the perturbation or to the perturbation itself. Multiple treatment and control sites in this designed are sampled; however, treatments are staggered through time to determine whether responses can consistently be produced by the manipulation regardless of the starting conditions. This staggered approach was termed a "staircase" design Walters et al. (1988). Roni et al. (2005) suggests that tributary reconnection experiments lend themselves well to a staircase designs. In these designs, both time and treatments are considered fixed effects. The equation for a staircase design can generally be written as the BACIP; however, assumptions about time-treatment interactions must explicitly be stated. The most general staircase statistical model can be written as:

$$y_{it} = \mu_i + \tau_t + R_{it} + \varepsilon_{it} \qquad (1)$$

where:

- y_{it} is the t^{th} observation at site i
- μ_i is the mean response for unit i in the absence of treatment
- $\tau_{i...}$ is the time-averaged effect shared by all units independent of whether they are treated
- R_{it} is the effect of treatment on unit i at time t
- ε_{it} is the residual value after accounting for all the above effects (and estimable only when multiple subsamples are taken within each location at each time)

Other models that include time-treatment interactions are also explored in Walters et al. (1988). These designs requires that the first treatment is started after at least one pre-treatment year and that steps are either sequential if treatments per time period is unreplicated or if treatments are replicated for each starting time that sequence be either one step wide or uneven steps.

Intervention analyses have been used to evaluate a time-series of information for the effects of a perturbation on a response. Steward-Oaten and Bence 2001, provide a discussion of these types approaches. They suggest that a BACIP is really a special case of IA. In fact, they refer to the ANOVA approaches discussed above as impact vs. reference sites or IVRS design to distinguish from the Steward Oaten et al. (1986) IA approach were the "BACI" terminology was originated. IVRS approach encompasses the MBACI and the beyond-BACI designs. The IVRS estimates the effects the same as the BACIP/IA approaches but the estimate of variance and degrees of freedom are quite different. The MBACI and beyond-BACI (IVRS) estimates variance over the controls sites whereas IA approaches estimate variance over time. Thus, one is a time-series approach (IA) and other is a spatial approach (IVRS). Stewart-Oaten and Bence (2001) suggests that the requirement of multiple control sites through random selection is generally untenable.

As mentioned earlier IA designs do not require control site(s). If controls are chosen based on similar characteristics as the impact sites, they can be quite useful at reducing the amount of noise observed in a timeseries. For example, in the Lemhi, the geographic proximity of section C to sections A and B, these areas will likely share similar climatic experiences and habitat characteristics (e.g., geology and vegetation). In addition, other similarities are likely to occur such as observations by the same crews using the same protocols. Crews and protocols will likely change over the course of this quasi-experiment, but if they change for all sections at the same time the impact on the overall variability is reduced because we will evaluate the changes in the difference between the control and treatments. The controls act as a group of covariates.

The following example is used to illustrate a Randomized Intervention Analysis (RIA), a non-parametric IA. Figure 2.3.2 is the 1992-2002 time series of survival estimates observed for area B of the Lemhi, with hypothetical survival estimates extended through to 2030. In this simplified example, we assume all tributary reconnects occurred in 2015 with the hypothetical differences due to the treatment added. In this example, it appears that survival estimates fluctuate greatly from year-to-year in both treatment and control areas before and after the treatment. A strong treatment effect is not apparently obvious simply by looking at the time series of information (Figure 2.3.2a). Using an IA approach, we estimate the difference between each control and treatment pair (treatment - control) over the time series (Figure 1b). We then compute the average pre- and postmanipulation differences as D(PRE) and D(POST). These differences are much more apparent on this scale (Figure 2.3.2b). The test statistic is the absolute value of D(PRE)-D(POST) and is compared against a distribution of random permutations of the sequence of control and treatment differences (Carpenter et al. 1989). IA analyses are similar to the RIA but must conform to assumptions of parametric statistics. Cloutman and Jackson (2003) discuss the strength and weaknesses of each approach and suggest using both techniques to analyze these types of manipulations.

Assuming section C could be used as a control site, an important component to this design is the selection of covariates used to help explain some of the variability between control and treatment areas. Figure 2.3.3 provides an example of how the use of covariates can help provide clearer estimates of impact of the restoration action. Figure 2.3.3a shows a hypothetical timeseries of estimates of parr-to-smolt survival rates assuming a treatment (tributary reconnects), temperature, and spawner (density dependent interactions) effects. The two areas appear to have similar responses overtime with the mainstem Lemhi having consistently higher survival rates. The survival rates fluctuate greatly from year-to-year, and a treatment effect is not readily apparent. Again, we estimate the difference in survival rates over the timeseries depicted in Figure 2b. Without consideration of other covariates (e.g., temperature and the number of adult spawners represented by the red line in Figure 2.3.3b) the difference between the two is still not obvious. If we take into account the effects of different temperatures between the control and treatment area one can begin to see a difference between the control and treatment that occurred after the treatment. If we account for differences in both temperature and spawner abundance then the impact of the treatment becomes apparent. Techniques are available to select useful covariates from a list of potential covariates measured through the study period (Milliken and Johnson 2001, Kershner et al. 2004).

An information-theoretic approach to can be used to evaluate model complexity (e.g., BACIP vs. BACIP + covariates; Burnham and Anderson 1998). This approach gives a formal accounting for the relative plausibility of the models estimated, and for trade-offs between the number of parameters and goodness of fit. Thompson and Lee (2002) have applied similar information-theoretic approaches to Snake River spring/summer Chinook spawner-recruit models, while Paulsen and Fisher (2003) apply the approach to models of parr-to-smolt survival.

The information-theoretic approach is described at length in Burnham and Anderson (1998). Briefly, the method consists of the following steps: 1) identify a candidate set of models *a priori*, using information on scientifically plausible relationships between candidate independent variables and the dependent variable of interest; 2) estimate the models using the same dataset (as described above) and the same dependent variable, (e.g., $\mathrm{Ln}(\hat{s}_{i,t})$); 3) for each model, calculate the AICc, which is adjusted or "corrected" for having a small number of observations relative to the number of parameters; 4) among the candidate models, select the model with the lowest AICc. Subtract the lowest AICc from each of the candidate models, yielding a "delta" which will be zero by definition for the model with the lowest AICc; and 5) calculate "AICc weights" for each model, using a simple exponential function of the deltas. The weights are then normalized to sum to one, and their values may be interpreted as the relative plausibility of each model, given the data and the set of candidate models. The models may be non-nested without influencing the comparisons.

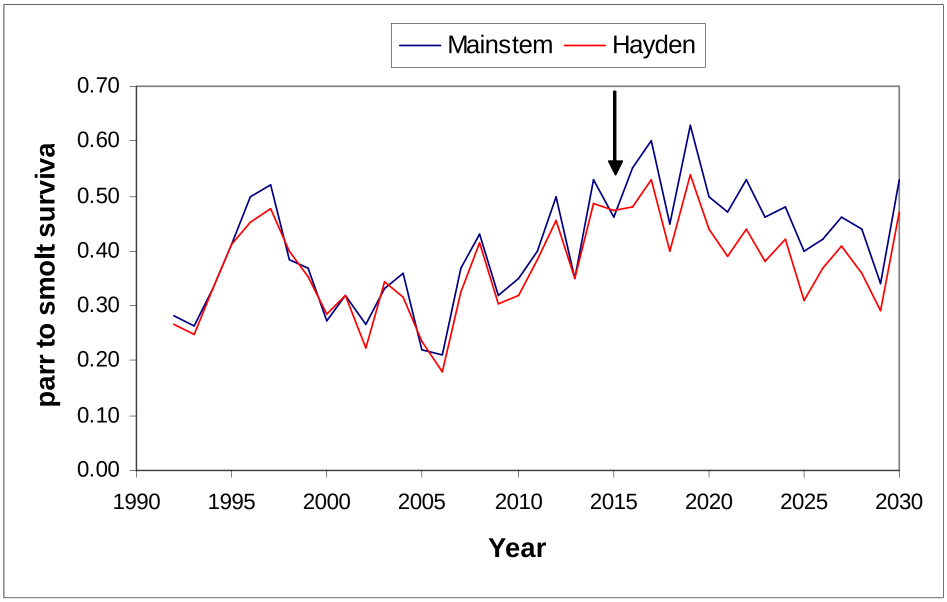
Because the concept may be unfamiliar, we provide a brief overview of the philosophy behind this method of model selection. Consider first how one can interpret the confidence interval on a regression parameter, setting aside the mechanics of how confidence bounds are calculated. Say that from a give set of data one estimates a bivariate regression model, e.g., $\mathrm{y} = \mathrm{a} + \mathrm{bx}_1$ with the point estimate of the intercept, a, is 2.0, with 5% and 95% confidence bounds of 1.8 and 2.2. One way to look at this is that if one had in hand a large number of similar datasets, and estimated the same regression model for each set, the estimated intercept would be between 1.8 and 2.2 for 90% of the models estimated.

The AICc weights may be thought of in a similar way. They are used to rank the models that one has estimated with a given set of data, to select the best model (i.e., the model with the lowest AICc) or the best set of *m* model from the *n* models estimated, $m < n$. To continue the simple example noted above, say that one had estimated a series of three models, using three different independent variables, x_1, x_2, and x_3, one independent variable per model. Assume further that the models' AICc weights were 0.6, 0.3, and 0.1, respectively. Analogously to the above example, the weights can be interpreted as meaning that if one had many similar data sets, and estimated the same regression models for each data set, that the model using x_1 would be the top-ranked model 60% of the time, about twice as often as the second-ranked model, and six times as often as the third-ranked model. In this sense, the model with weight 0.6 is six times more plausible that the model with weigh 0.1.

Our aim with the above discussion is to explore the more common experimental designs in a BA framework to determine the most reasonable approach to elucidate the impact of restoration actions on salmonid populations and their habitat in the Lemhi Basin. These discussions will explicate experimental design assumptions. The appropriate sampling designs to complement the chosen experimental design will be examined further.

Table 2.3.1. Alternative ANOVA analyses of Before-After comparisons (modified from Downes et al. 2002).

	BACI	BACIP	MBACI	Beyond BACI	Staircase design
Question	Did this manipulation result in a change, relative to the control, under this same set of conditions?	Would this manipulation result in a change in the treatment site relative to the control site?	Would this type of manipulation result in a change to sites relative to sites where no manipulation occurred, under these same set of conditions?	Would this type of manipulation result in a change relative to sites where no manipulation occurred? At what scale did the manipulation have an impact?	Would this type of manipulation result in a change if implemented under a range of environmental conditions relative to sites where no manipulation occurred?
Control	1	1	> 1	>1	≥ 1
Impact	1	1	≥ 1	≥ 1	>1
Treatment of space	Fixed Only two locations	Fixed Only two locations	Random Locations are sample of available areas	Random Locations are sample of available areas	Fixed Locations are chosen by design
Treatment of time	Fixed Only two times	Random Times are random selection from periods	Fixed All times within periods are sampled	Random Times are random selection from periods	Fixed Time is evaluated as a shared trend across sites
Value	May be used in meta-analyses, but results are very site specific	Appropriate choice when only one control is possible. Most useful when control vs. treatment location behave similar in absence of treatment.	The best at inferring the impact of a type of action when spatial scale of impact is defined.	Provides much information about expected impacts at possible spatial and temporal scales of from actions.	Provides information on transient response to actions. Help separate action impact from common time trends.



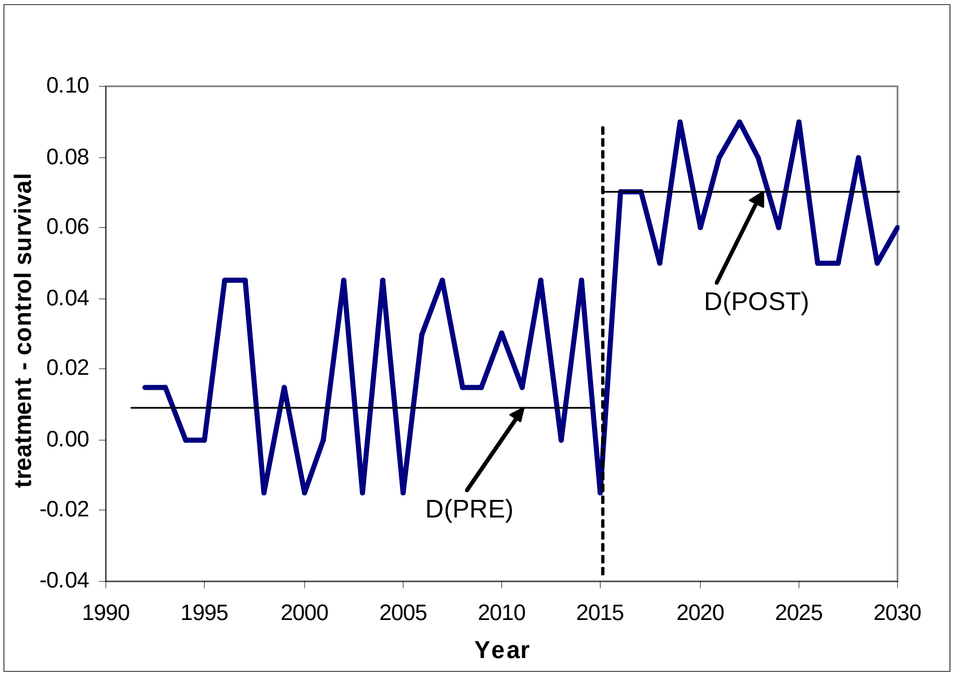
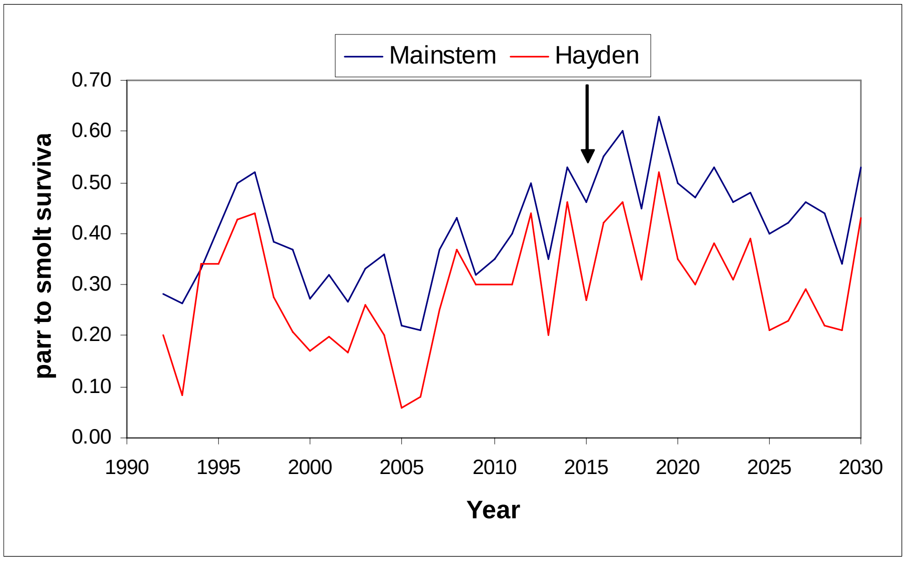


Figure 2.3.2. Value of including a control system for inference. **A)** Hypothetical timeseries of mainstem Lemhi and Hayden Creek parr-to-smolt survival rates, with tributary reconnections assumed to happen simultaneously in 2015 in the mainstem Lemhi. **B)** Difference between mainstem Lemhi and Hayden Creek parr-to-smolts survival rates over the timeseries. D(PRE) and D(POST) represent average survival difference pre- and post-tributary reconnections occurring in 2015 in the mainstem Lemhi.



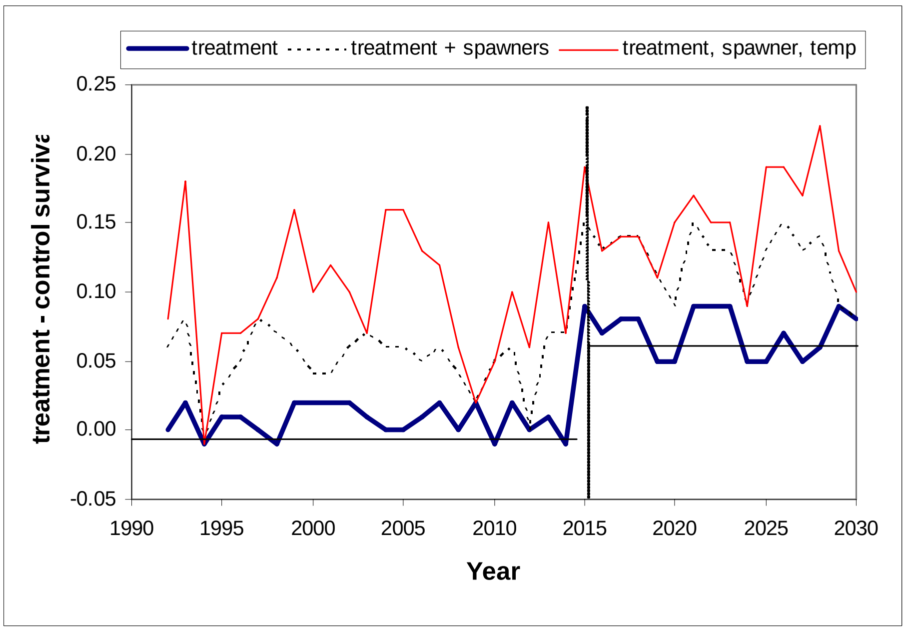


Figure 2.3.3. Value of including covariates for inference. **A)** Hypothetical timeseries of mainstem Lemhi and Hayden Creek parr-to-smolt survival rates, with tributary reconnections assumed to happen simultaneously in 2015 in the mainstem Lemhi. Differences between control and treatment sites include temperature, spawner abundance, and treatment effects. **B)** Difference between mainstem Lemhi and Hayden Creek parr-to-smolts survival rates over the timeseries including temperature, spawner abundance and treatment effects is represented by thin solid red line. Difference including spawner abundance and treatment effects only is represented by dashed line. Differences including only treatment effects is represented by thick solid line. D(PRE) and D(POST) represent average survival difference pre- and post-tributary reconnections occurring in 2015 in the mainstem Lemhi after accounting for spawner and temperature effects.

2.3.3 Alternative mark-recapture models that can be used for the evaluation of PIT-tag information collected from passive instream antennae

Over the past 25 years, biologists have extensively used passive integrated transponder (PIT) tags and mark-recapture methods in fisheries research within the Columbia River system. More recently, there have been substantial increases in the use of PIT-tags in small-river systems, as a result of recent advances and applications in PIT-tag technology. In particular, passive instream antennae (PIA), which are operationally similar to those present at the major hydropower facilities on the Columbia and Snake River systems, have been installed in many tributary systems. PIAs allow for individual recaptures of fish marked with PIT-tags as they migrate through a PIA system within a river channel.

Individual-specific PIT-tags, PIAs, and spatially explicit mark-recapture data can provide estimates of key demographic and vital rate parameters across a range of size-classes and life-history forms. In particular, individual growth information, which can act as a surrogate for fitness and is critical for life-stage population models, can be quantified through marking and recapture events. The combination of marking events and PIA recaptures can provide information regarding the timing and plasticity of movement within and across temporal and spatial scales, and emigration rates, which are critical for estimates of true survival verses apparent survival. Tagging and movement data can also provide insight into the proportion of the population that exhibits resident or migratory life-history expressions. Finally, mark-recapture data can provide estimates of survival, which is a direct indication of fitness, across relevant life-stages and life-history forms and is necessary for understanding the dynamics of populations within population models. Ultimately, this information can provide insights into the dynamics of individual populations and help parameterize site and regional population models that can elucidate effective restoration and recovery strategies (e.g., Al-Chokhachy 2006), and evaluate the effects of different land-use actions (i.e., restoration activities) on population-level factors such as migration, survival, etc. in tributary systems.

Despite the wide spread use of PIT-tags and PIAs in small streams, there have been few efforts to formally evaluate the appropriate design and analytical methods to answer relevant management and biological questions. While mark-recapture data for obligate anadromous species (i.e., Chinook) through the Columbia and Snake systems has been well documented, differences in the behavior and life-history strategies of inland salmonids may predicate the need for additional research considerations. Furthermore, there are multiple models, each with strengths and weaknesses that can be used to estimate survival using mark-recapture data collected with PIAs. The choice of the appropriate model may largely be driven by the data collection methods and the life-history strategies of the populations of interest. For example, many salmonid populations (i.e., rainbow trout/steelhead) exhibit indeterminate life-history strategies (e.g., Thériault and Dodson 2003; Homel 2006) where individuals within a panmictic population may exhibit resident, fluvial, or anadromous life-history patterns. When analyzing mark-recapture data, a difficult step may be identifying the appropriate temporal scale for inference where data collection (recaptures through PIAs) is continuous as a result of the plasticity in life-history expression.

This sub-section describes a mark-recapture project in the South Fork John Day River (SFJD) in Oregon to illustrate the data challenges and opportunities that exist in projects utilizing individual-specific mark-recapture data (i.e., PIT-tags) and PIAs for survival analyses. This sub-section describes the study system and potential research objectives, the mark-recapture data set, the challenges within different mark-recapture models to evaluate survival, and discusses the utility of field data and simulations to design and improve future mark-recapture projects that utilize individual-specific tags and PIAs.

Study and Data Collection

In 2003, researchers from Oregon State University (OSU), Oregon Department of Fish and Wildlife (ODFW), and Utah State University (USU) initiated a steelhead/rainbow trout project using mark-recapture data (PIT-tags and PIAs) in the SFJD river system. The SFJD contains populations of resident, fluvial, and anadromous rainbow trout/steelhead, and delineation between life-history types is often not possible as juveniles. The initial focus of this project was to assess juvenile fish growth and movement across a wide variety of habitat types to evaluate the complex interactions of biotic and abiotic factors on these parameters. The nature of the mark-recapture data in the SFJD allows for the evaluation of other population-level parameters (i.e., survival) over varying temporal and spatial scales. In lieu of this rich data source, there are a number of issues that will directly affect the approach to analyzing this and similar mark-recapture data for estimates of survival where PIAs have been installed.

Within the SFJD study area, there are a variety of methods used across different temporal and spatial scales to mark and recapture fish. Within most years the efforts vary as a result of available resources and the ability to maintain equipment (PIAs and screwtraps) during random environmental events (i.e., high flows). Marking and recapturing occurs continuously throughout each year, but varies with intensity within and across years. Within each year, a variety of techniques are used during intensive marking and recapturing occasions in June and September. During and between these occasions, recaptures continuously occur at screwtraps and PIAs, and additional marking occurs for all unmarked fish collected at screwtraps. Finally, across years, there has been a substantial amount of variability in resource availability, and furthermore recapture probability. For example, in 2005 a river-wide weir was installed in a tributary system, which allowed for high detection probabilities during this period, but in 2006, a PIA with partial river coverage was installed as a result of damage to this system during flood events. Synthesizing mark-recapture data into concise input files can be challenging as a result of the variety of methods and efforts to capture/recapture juvenile steelhead and rainbow trout and the plasticity of movement patterns in this populations.

Mark-Recapture Survival Analyses

Within stream systems, mark-recapture data is generally evaluated with “open” mark-recapture models to estimate population-level parameters such as survival. In general, mark-recapture models are flexible in their ability to incorporate individual and environmental covariates, as well as varying levels of effort across different sampling occasions. For example, it is possible to evaluate how differences in individual parameters, such as growth or condition, and/or changes in the habitat (i.e., water temperatures), affect survival across different age-classes of fish. Thus, researchers can use mark-recapture models to evaluate how survival is linked mechanistically to different environmental and/or habitat characteristics.

Mark-recapture analyses can be evaluated within an information-theoretic approach (ITA; *see* Burnham and Anderson 2002). This approach allows researchers to compare simple null hypothesis tests (i.e., treatment vs. control), as well as more elaborate comparisons of factors affecting individual survival. Within an ITA approach, researchers can compare multiple models, each of which may be plausible, to evaluate the effects of different covariates on survival. For example, managers may be interested in evaluating how fish movement, which may occur across different habitat types, affects survival. Using spatially-explicit mark-recapture data, one can model fish survival, using individual-specific tags and movement information, as a function of different movement distances within a study area; thus, providing information regarding the effects of the habitat characteristics within migratory corridors on fish survival. Within an ITA framework, competing models can be ranked using Akaike’s Information Criterion (AIC), which incorporates maximum-likelihood estimates and the principles of parsimony, to suggest the most plausible model given the data (*see* Burnham and Anderson 2002 for further description).

Our review specifically focuses on two open mark-recapture models, Cormack-Jolly-Seber and Barker models. For each model, we discuss the benefits and challenges of using each approach when considering the analyses of PIT-tag and PIA mark-recapture data.

Cormack-Jolly-Seber Model

The Cormack-Jolly-Seber (CJS) model or variations of the CJS model are commonly used to estimate survival with mark-recapture data. CJS models use maximum likelihood theory to estimate the probability of survival from one occasion to the next occasion. CJS models are sufficient to analyze complex data sets, yet are simple with only two parameters and are relatively insensitive to violations of assumptions (*see* Krebs 1999 for description of assumptions*)*. Under a CJS model, three capture-recapture occasions are required for estimates of survival (Φ) and capture probability (p). Each parameter can vary by time, group, etc. and individual (i.e., condition) and environmental covariates can be incorporated into analyses to examine the effects of different biotic and abiotic factors on survival and capture probability. Generally the precision of survival estimates, and furthermore ability to delineate survival across different groups, is largely driven by recapture probability (Cooch and White 2006).

CJS models require inputs of discrete sampling occasions, which may occur at different temporal resolutions, and generally correspond to the temporal inference of the research (i.e., monthly survival). Analyzing PIT-tag and PIA mark-recapture data presents unique challenges for assimilating data to be analyzed within CJS models. In particular, when PIAs are used within a mark-recapture framework, there can be questions as to the relevant “sampling occasion” as data is collected continuously through time. Splitting sampling occasions into finer temporal periods (i.e., week vs. month) generally results in lower recapture probabilities as fewer individuals may be recaptured within each time period; this generally results in reduced precision in survival estimates and reduced power to detect changes in survival across groups, treatments, etc. On the contrary, grouping continuous recaptures across longer temporal periods into one sampling occasion (i.e., season vs. month) may result in higher capture probabilities (and increased precision) as more individuals are included at this temporal scale. However, this approach presents challenges as to the temporal scale of inference and violates the assumption of sampling occurring over a finite time period. Ideally, there would be some optimum where length of sampling occasion could be varied as a function of the precision and inference.

Barker Model

The Barker model may present a unique alternative to CJS models for evaluating mark-recapture projects that use continuous sampling events through PIAs. The Barker Model is an extension of “known-fate” models that are typically used to estimate survival with radiotelemetry data. Similar to the CJS models, the Barker model also uses maximum likelihood theory, requires three sampling occasions to estimate survival, and the precision of estimates is largely driven by capture probability. However, the Barker model differs significantly from the CJS model in the number of parameters and the data inputs. There are 7 parameters in the Barker model (5 additional parameters to Φ and p), which require larger sample sizes and increased recapture probabilities. Similar to the CJS model, the Barker model also requires distinct sampling occasions but differs from the CJS as information collected between sampling occasions can be included as data inputs. Thus, continuous sampling, which generally occurs through the use of PIAs in stream mark-recapture studies, can be incorporated into survival analyses without binning the data into different sampling occasions. While the Barker model appears to be a more robust approach than CJS analyses for analyzing mark-recapture data where PIAs are used, the increase in the number of parameters may prevent the usage of this model without relatively high capture probabilities and large sample sizes.

Model Evaluations

There is a need to formally analyze PIT-tag and PIA mark-recapture data (the SFJD system represents an ideal test case) to evaluate the use of different mark-recapture models. This will involve three separate steps. First, evaluate the tradeoff between length of sample occasion and precision under a CJS approach. While PIAs provide additional recaptures and can increase recapture probabilities within any study, the structure of continuous data sets present unique challenges with CJS models. Next, compare the precision and efficacy of using CJS models versus the Barker model. In particular, consider precision of estimates, violations of model assumptions, benefits/challenges of using each model and provide recommendations for future research. Finally, perform power analyses across both CJS and Barker models under varying efforts (capture probabilities) and sample sizes (tagging totals) that would be necessary to detect the effects (change in survival) across different management scenarios. The use of power analyses will provide a framework that regional biologists can use in the planning of similar PIT-tag and PIA mark-recapture studies in the Basin.

Across the Pacific Northwest, there has been a dramatic increase in the use of PIT-tag and PIA technology in salmonid research. However, there has been little guidance regarding the effectiveness of different designs and subsequent data analyses to answer key biological and management questions. The large mark-recapture data set collected in the SFJD provides a unique opportunity to evaluate different mark-recapture designs and models to estimate key demographic and vital rate parameters such as survival. The results from this work in the SFJD will allow for more effective and efficient mark-recapture research and provide a set of guidelines and recommendations for similar fisheries monitoring in the Columbia River Basin.

2.3.4 Lemhi Subbasin Bull Trout monitoring design

Introduction

This sub-section provides an outline of monitoring designs, both high and low efforts, for resident salmonids of interest in the Lemhi Basin. While the focus of the HCP is directed towards anadromous species, there needs to be additional consideration of resident salmonids, in particular species listed under the Endangered Species Act (ESA). The life-history strategies of resident salmonids, which spend their entire lives within freshwater habitats, both tributary and mainstem habitats, substantially differ from anadromous species. These differences in life-history strategies predicate the need for consideration of differences in response metrics, study designs, and data collection in order for robust evaluation of the effectiveness of the actions implemented under the Lemhi HCP.

The focus of this review is directed towards bull trout (*Salvelinus confluentus*), which are currently listed as Threatened under the ESA; however the format and applications could be applied to other species of interest in the basin (i.e., westslope cutthroat trout, *Onchorhynchus clarki lewisi*).

Bull Trout

Bull trout are a char species native to the northwestern United States and western Canada. They are primarily an inland species with a distribution from 41° N to 60° N latitude, from the southern limits in the McCloud River in California and the Jarbridge River in Nevada to the headwaters of the Yukon River in Northwest Territories. Across their native range, habitat fragmentation and degradation (Fraley and Shepard 1989; Rieman and McIntyre 1993), barriers to migration (Rieman and McIntyre 1995), and the introduction of non-natives (Leary et al. 1993) have led to isolation and decline of bull trout populations throughout much of the Columbia River Basin. Today, bull trout exist only as sub-populations over a wide range of their former distribution (Rieman et al. 1997), exhibiting potential metapopulation structure (Rieman and McIntyre 1993; Dunham and Rieman 1999).

Across much of their range, bull trout are known to exhibit multiple life-history forms including resident, fluvial, adfluvial, and anadromous; within populations multiple life-history forms can coexist (Rieman and McIntyre 1993; Nelson 2002; Al-Chokhachy 2006). Furthermore, the delineation between resident and migratory life-history expressions may not be discrete (Homel 2006), creating difficult challenges for monitoring these complex populations. Resident bull trout may remain in headwater systems throughout their life-cycle, but can exhibit seasonal movement patterns within a local population. For the migratory life-history strategy, juveniles rear in small, natal tributaries for one to three years before migrating to larger systems for some period of time before returning to their natal streams to spawn (Fraley and Shepard 1989). The migration patterns for juvenile (Homel 2006) and adult (Nelson et al. 2002; Bahr and Shrimpton 2004) bull trout vary considerably across seasons. The majority of downstream movements of juveniles generally occur in spring and late summer, but moderate amounts of downstream movements occur throughout the year (Hemmingsen et al. 2001; Homel 2006). Adult resident and migratory bull trout return to their natal streams for spawning from mid-August, when water temperatures drop below 9° C (Goetz 1989), through the end of October; post-spawning movements are variable throughout the late fall and early winter (Homel 2006).

Bull trout require diverse, yet specific habitat types throughout their lifecycle (Rieman and McIntyre 1993). Both juvenile and adult bull trout are generally associated with deeper, slow-velocity habitats with complex forms of cover (Rieman and McIntyre 1993; Al-Chokhachy 2006). Water temperatures in excess of 16° C are considered to be a limiting factor in the distribution of both juvenile and adult bull trout, with higher temperatures acting as potential thermal barriers to migration (Fraley and Shepard 1989). Increased water temperatures, often as a result of habitat degradation, have been found to significantly reduce bull trout survival (Selong et al. 2001).

Bull trout distribution in the Lemhi Basin

Bull trout are known to be distributed throughout much of the Lemhi Basin (Figure 2.3.4). Six areas within the basin have been identified as potential local populations under the USFWS Draft Recovery Plan, including Hayden Creek, Geerston Creek, Bohannon Creek, Kenny Creek, Pattee Creek, and the Upper Lemhi River (Figure Y; USFWS Draft Recovery Plan 2002); within each local population, small, sub-populations are known to exist in tributary systems. The majority of migratory bull trout in the Lemhi Subbasin exist in Hayden Creek (T. Curet, IDFG, *personal communication*). Although some migratory fish are known to use the mainstem Lemhi, the lack of connectivity with the tributaries and isolation of populations has been identified as limiting factors within the basin (USFWS Draft Recovery Plan 2002).

Existing data

Idaho Fish and Game has collected abundance and distribution data in a number of populations within the Lemhi Basin. Since 2002, redd count data has been collected in Hayden Creek, East Fork Hayden Creek, Big Timber Creek, and Bear Valley Creek. Distribution data have been collected through single and multiple-pass electrofishing censuses from tributary mouth to headwaters in Bohannon, Wimpy, Big Eight-Mile, Big Timber, Kenny, Agency, Little Eight-Mile, Lee, Holly, and Haynes Creeks. Annual snorkel surveys, which are conducted primarily for juvenile anadromous species but may include bull trout data, occur at 11 sites on the mainstem/Big Springs and 7 sites in Hayden/Bear Valley Creeks. Recently (2005), further distribution and abundance data has been collected in the upper Lemhi Basin (Area A). Some additional data exists within the Lemhi basin through work conducted through the Shoshone Bannock tribes and Idaho State University.

Management actions and expected benefits to bull trout populations

The management actions under the Lemhi HCP, including tributary reconnection with the mainstem Lemhi (HCM 1 -01), improvements in fish passage (HCM 2-01, HCM 2 -02, HCM 3 -01, HCM 3 -02, and HCM 6 -01), and improved habitat conditions (HCM 2 -03, HCM 2 -04, HCM 5 -01, and HCM 5 -02; see Katz et al. 2005 for individual conservation measures) may have differential effects for bull trout populations in the Lemhi Basin. Increased connectivity to tributaries and the elimination of passage barriers may significantly increase the distribution and abundance of fluvial bull trout. Furthermore, reconnectivity between tributaries and the mainstem Lemhi may allow for recolonization of unoccupied systems, reduce mortality of individuals migrating during low-flow periods, and increase the variability in movement patterns within and across seasons. The HCP may also increase the amount of suitable habitat for bull trout. Bull trout are generally associated with deeper, slower habitats with cover (Muhlfeld and Marotz 2005; Al-Chokhachy and Budy 2006). Increased flows will likely increase riparian vegetation, large woody debris recruitment, and flood pulses, resulting in the formation of deeper pool habitats, which may increase vital rates (i.e., survival) for bull trout.

Habitat improvements and stream reconnections may also reduce stream temperatures during summer months by increasing flows and decreasing solar inputs through increases in riparian vegetation. This may result in increase in the amount of thermally-suitable habitat across the basin and prevent thermal barriers (e.g., Ebersole et al. 2001) preventing juvenile and adult movements.

Clarified questions and data needs

Below is a list of potential questions related to bull trout populations in the Lemhi Basin as a result of the actions implemented under the HCP. For each of the following questions, there is a list of relevant data needs.

1. Have the actions implemented under the Lemhi HCP increased the density (by X% with some precision) and distribution of juvenile bull trout over 30 years in connected tributary systems when confounding factors, including presence of non-native brook trout (*Salvelinus fontinalis*), natural disturbances, habitat conditions (i.e., complexity), and climate indicators been accounted for?
    i. <u>Sample design (both control and treatment areas)</u>**:**
        1. Density data in fixed sites (systematic sampling) within populations.
        2. Presence/absence data collected (random-design) in occupied and unoccupied streams.
        3. Information regarding climate, habitat conditions, and brook trout distribution and abundance data.
2. Have the actions implemented under the Lemhi HCP increased the abundance of reproductive adults in the reconnected tributaries when confounding factors, including presence of non-native brook trout (*Salvelinus fontinalis*), natural disturbances, habitat conditions (i.e., complexity), and climate indicators been accounted for?
    i. <u>Sample design (both control and treatment areas)</u>**:**
        1. Redd count data.
        2. Abundance data.
        3. Information regarding climate, habitat conditions, and brook trout distribution and abundance data.

3. Have the actions implemented under the Lemhi HCP increased the survival of juvenile and adult bull trout exhibiting fluvial life-history expression when confounding factors, including presence of non-native brook trout (*Salvelinus fontinalis*), natural disturbances, habitat conditions (i.e., complexity), and climate indicators been accounted for?
   i. Sample design (both control and treatment areas)**:**
      1. Survival data and some demographic information (i.e., growth) for all relevant size classes.
      2. Information regarding climate, habitat conditions, and brook trout distribution and abundance data.
4. Have the actions implemented under the Lemhi HCP increased the number of fluvial bull trout migrating from tributaries by X% over 30 years?
   i. Sample design (both control and treatment areas)**:**
      1. Estimates of fluvial bull trout abundance through redd count data (measurements of redd size).
      2. Trend information at weirs and/or passive instream antennae via mark-recapture data: adult fish.
      3. Trend information at screw traps: juveniles.

Study Designs

A critical step in understanding the effects of the actions implemented under the Lemhi HCP will be the decision regarding the appropriate study design (see *Experimental Designs*). The objective of this facet of the study is to utilize a design that will allow for robust assessments of specified response variables for bull trout populations in the Lemhi Basin (*see below*). These set of objectives may differ significantly from those for anadromous salmonids due to differences in life-history strategies, allocation of resources, inference, and recovery criteria. For example, production of anadromous species (i.e., smolt outmigration) can be measured at screw traps; while screw trap data at selected sites may provide some indication of the abundance of migratory bull trout, this metric will not allow for assessments of resident bull trout production. In particular, the current study design, which will be used to measure juvenile production for Chinook salmon and delineates the Lemhi Basin into Section A (migratory corridor or Lower Lemhi), Section B (action area or Upper Lemhi), and Section C (reference area or Upper Lemhi), does not apply to bull trout populations in the basin. A more appropriate design for bull trout incorporates a paired-watershed design between treatment (reconnected tributaries in the lower or upper Lemhi) and control (Hayden Creek area where no reconnections are anticipated under the Lemhi HCP) areas. Furthermore, the distribution of bull trout spawning and rearing habitat may not overlap with Chinook spawning and rearing sites, resulting in the need for additional sites beyond those selected for Chinook.

For robust evaluation of the actions implemented under the Lemhi HCP, both fixed and random sites need to be considered for control and treatment areas. These fixed sites may be entire streams, where census surveys can be conducted (i.e., redd counts) or individual reaches for response variables where censuses are not possible (i.e., juvenile densities).

Response Metrics

Low and high-effort designs are considered in this document. With each of these designs, there will be strong consideration of historical monitoring in the basin, as well as maximizing efficiency of each design by overlapping data collection for bull trout with existing data collection for anadromous species (i.e., Chinook). A detailed breakdown of the comparative costs of the two bull trout designs (low vs. high) for the Lemhi subbasin are provided on the CSMEP website.

Low Design

Under the low design, adult abundance and juvenile bull trout density are considered as response variables (Table 2.3.2). A split panel design (fixed and rotating) will be used to compare treated and untreated watersheds. In the fixed panel, three paired-watersheds (treatment and control tributaries) initially will be randomly selected to be monitored annually over the life of the LCP. Additionally, a rotating panel of 3 new randomly selected paired-watersheds will be additionally sampled each year. The sampling interval for each of these addition tributaries will be four years (approximate generation time for bull trout). The specific number and length of reaches will be based on distribution and spawning data currently being collected by IDFG. While the reconnectivity (treatment) of all tributaries will not occur at the start of the HCP, the rotating panel of anticipated sites will allow for baseline (pre-treatment) data, which will help evaluate the effectiveness of the actions implemented under the Lemhi HCP. In addition, three tributaries currently monitored in Hayden Creek will continued to be monitored. This will allow for a comparison between areas A, B, and C, and may lead insight into metapopulation dynamics at the scale of a subbasin (e.g., Hayden). The hierarchical design, along with random choices between treatment and control watersheds will allow for larger universe of inference and use of ANOVA designs described by Underwood (1994).

Redd counts are widely used across the Pacific Northwest to monitor bull trout populations, as they represent a cost and time effective monitoring tool for managers. Information regarding adult bull trout abundance under the low design will be collected through the use of redd counts. Bull trout redd count data collection may temporally overlap with Chinook redd counts, allowing for increased efficiency in travel to sites, etc. However, the majority of current and proposed Chinook redd counts are in the mainstem Lemhi (Figure 2.3.5), while bull trout spawning occurs primarily in the tributaries. Currently, there is recent redd data for Bear Valley, Hayden, East Fork Hayden, and Big Timber Creeks (Figure 2.3.5).

Snorkeling data will be used to quantify juvenile bull trout densities under the low design. Current general parr monitoring sites occur primarily in the mainstem Lemhi River and Hayden Creek (Figure 2.3.6). Additional sites will need to be added in order to quantify juvenile bull trout densities in the headwaters of selected tributaries.

High Design

Under the *High Design*, estimates of redd count bias, juvenile distribution data, juvenile and adult survival, and abundance of fluvial (migratory) bull trout will be considered as response variables to evaluate the actions implemented under the Lemhi HCP. These *High Design* response variables will be measured in addition to the response variables under the *Low Design* (Table 2.3.2). Under the *High Design*, all paired watersheds will be examined; since the reconnectivity (treatment) of all tributaries will not occur at the start of the LCP, a Staircase design will be the most appropriate strategy for incorporating newly reconnected tributaries (Walters et al. 1988, Roni et al. 2005). In addition, three tributaries currently monitored in Hayden Creek will continued to be monitored as in the low design. The large number of replicates of treatment and controls will provide a powerful test of treatment impacts as well as evaluate time-treatment interactions to address transient responses to the LCP (Walters et al. 1988).

Population estimates will provide a measure of abundance of reproductive individuals, which is critical for meeting recovery goals in the basin (USFWS Draft Recovery Plan). Redd counts provide an index of population size (Williams et al. 2002), but may not effectively represent both large, migratory fish and small, resident adult bull trout (Al-Chokhachy et al. 2005). Additional population estimators (i.e., mark-recapture) will be used to assess redd count bias. During mark-recapture sampling (*see below*) adult bull trout will be marked with visible tags to allow for mark-resight population estimates, conducted through snorkeling density and distribution surveys.

Juvenile distribution, both within and among tributaries and mainstem reaches, will be measured under the *High Design*. Changes in the distribution of juvenile bull trout within tributaries will be quantified using a combination of snorkel and electrofishing surveys (used for mark-recapture; *see below*). Within tributaries, systematic sampling may be required to quantify changes in distribution along the river continuum. Changes in distribution will be quantified via presence/absence data, which will be collected though mark-recapture (*see below)* and density sampling (*see above)*.

Estimates of juvenile and adult survival across control and treatment areas will require the use of mark-recapture data. In control and treatment streams, fish collection, marking (i.e., PIT-tag), and recapture data will be required across the spatial distribution of juvenile and adult bull trout. Systematic sampling, where reaches are evenly distributed throughout a system, may be the most appropriate for accounting for heterogeneity of habitat types (Stevens and Olsen 2004) and distribution of bull trout (e.g., Al-Chokhachy 2006). Annual sampling will allow for estimates of size-specific growth and survival across multiple size classes using open mark-recapture models (i.e., Cormack-Jolly-Seber; Al-Chokhachy 2006).

Changes in abundance of fluvial bull trout will be measured across control and treatment areas. Quantifying these changes will require the collection of information regarding the abundance of both juvenile and adult bull trout at key migration points. In both control and treatment streams, fish collection, marking, and recapture data will be required across the spatial distribution of juvenile and adult bull trout. Fish collection and marking will be accomplished through mark-recapture sampling (*see above*); screw trap data and passive instream antennae, which detect all PIT-tagged fish as they move through the antennae, will provide additional capture/recapture data of tagged (screw trap and antennae) and untagged (screw trap) bull trout, as well as increase capture probabilities for mark-recapture analyses (e.g., Al-Chokhachy 2006). Mark-recapture population estimators (i.e., Petersen Models) will allow for estimates of abundance and trend.

Consideration

Redd Counts

Despite being cost and time effective, caution should be used when using redd counts to monitor bull trout abundance (i.e., as a response variable). In particular, observer variability can reduce the accuracy and precision (Dunham et al. 2001) of redd counts and the ability to detect changes in population trends (Maxell 1999). In addition to observer variability, redd counts may not provide an effective tool for monitoring bull trout populations where both small (i.e., resident) and large (potentially migratory) fish coexist within a single population unit (Al-Chokhachy et al. 2005). Finally, factors such as size differences in redd scour sites between small, resident and large, potentially migratory fish, redd superimposition, and delineation between test digs and redds can reduce the level of certainty in monitoring populations using redd counts (Maxell 1999; Dunham et al. 2001). Therefore, it is necessary to evaluate which portion of the population is represented by redd count data via size-specific abundance estimates (e.g., Al-Chokhachy et al. 2005), and how reliable this information is for gauging the effectiveness of the actions implemented under the Lemhi HCP.

Snorkeling Data

Bull trout are cryptic in their behavior (Thurow 1997), which results in very low detection efficiencies with snorkel surveys (Thurow et al. 2004; Thurow et al. 2006). The low detection efficiencies during snorkel surveys may be especially problematic when response variables such as presence/absence are used as a means to examine changes in distribution; ultimately resulting in high probabilities of Type I and Type II errors (Peterson et al. 2002; Hoffman et al. 2005). Detection efficiencies generally increase with nighttime snorkel surveys (Thurow et al. 2006), as juvenile and adult bull trout are considered primarily nocturnal, and should be implemented in the Lemhi effectiveness mentoring program.

Detecting changes in abundance to measure effectiveness

Effectiveness monitoring designs must consider the tradeoff between sample size and temporal commitment to sampling that will be required to detect changes as a result of the actions implemented under the Lemhi HCP. Using abundance measures to detect change in population size can require high sample sizes and long temporal commitments (Ham and Pearsons 2000). Al-Chokhachy et al. (*In review*) bootstrapped field data for multiple sampling techniques from eastern Oregon and found that detecting moderate changes in bull trout abundance (25%) was not possible over short time intervals (5 years) without extremely high sampling efforts (Figure 2.3.7). Across techniques, snorkel surveys, despite their low sampling efficiencies, were considered the most cost-effective tool (with equal or better power than alternatives; i.e., one-pass removal estimates) to monitor changes in bull trout population abundance (Al-Chokhachy 2006). The need for high sampling efforts was largely driven by the patchy distribution of bull trout. For example, detecting a 25% change over 5 years would require over 40 sample units per system (power = 0.8); however, at 15 year time periods, the required sample size dropped to 15 reaches per year. Detecting 25% change over long time intervals (30 years) was possible with <10 reaches per year. These results suggest that detecting moderate changes in abundance may be possible over the life of the HCP.

Mark-recapture data

Bull trout use of complex habitats with cover (i.e., large woody debris piles) can result in low capture probabilities in mark-recapture studies (Hoffman et al. 2005; Al-Chokhachy 2006). Obtaining desired statistical power to detect changes in survival (e.g., response to actions implemented under Lemhi HCP) may require large sample sizes and high sampling efforts. For example, simulations of mark recapture data for a 10-year study (Figure 2.3.8) illustrate that detecting small (5%) differences in survival between control and treatment areas may not be possible with low capture probabilities (0.2) and relatively low #'s of releases (newly-marked individuals). As expected, the ability to detect changes in survival increased with capture probability and the number of releases per year. In systems where bull trout populations are small, marking >100 fish per year may not be possible (i.e., excessive harassment) without high sampling efforts (spatially and temporally);

Pilot research

Prior to the design and implementation of study designs, robust pilot studies should be considered to better understand the current distribution and abundance of bull trout within prospective control and treatment areas. For example, in tributaries with small populations, few fish may be available (i.e., permitting) for marking, resulting in low statistical power to detect changes in survival as a result of habitat restoration actions. Ultimately the pilot information may allow for early estimates of abundance and subsequent simulation exercises to better understand the tradeoffs between sample size and ability to detect changes in response variable (i.e., survival). When possible, recent historical data (i.e., redd counts) should be considered; however, for many response variables additional data and/or research focused at identifying bias with current data may be required.

Table2.3.2. Data and information pertaining to response variables under Low and High study designs in the Lemhi Basin.

Data	Information	Low	High	Question (s)
Redd counts	Adult Abundance	Yes	Yes	2
Snorkel counts	Juvenile Density	Yes	Yes	1
Redd counts corroborated with additional abundance estimates	Adult Abundance	No	Yes	2, 4
Snorkel counts	Juvenile Distribution	No	Yes	1
Mark-recapture data in tributaries, weirs and screw traps	Juvenile Survival	No	Yes	3
Mark-recapture data in tributaries, weirs and screw traps	Adult Survival	No	Yes	3
Screw trap and weir data	Fluvial abundance	No	Yes	4
Habitat surveys	Covariates for changes in response variables	Yes	Yes	5

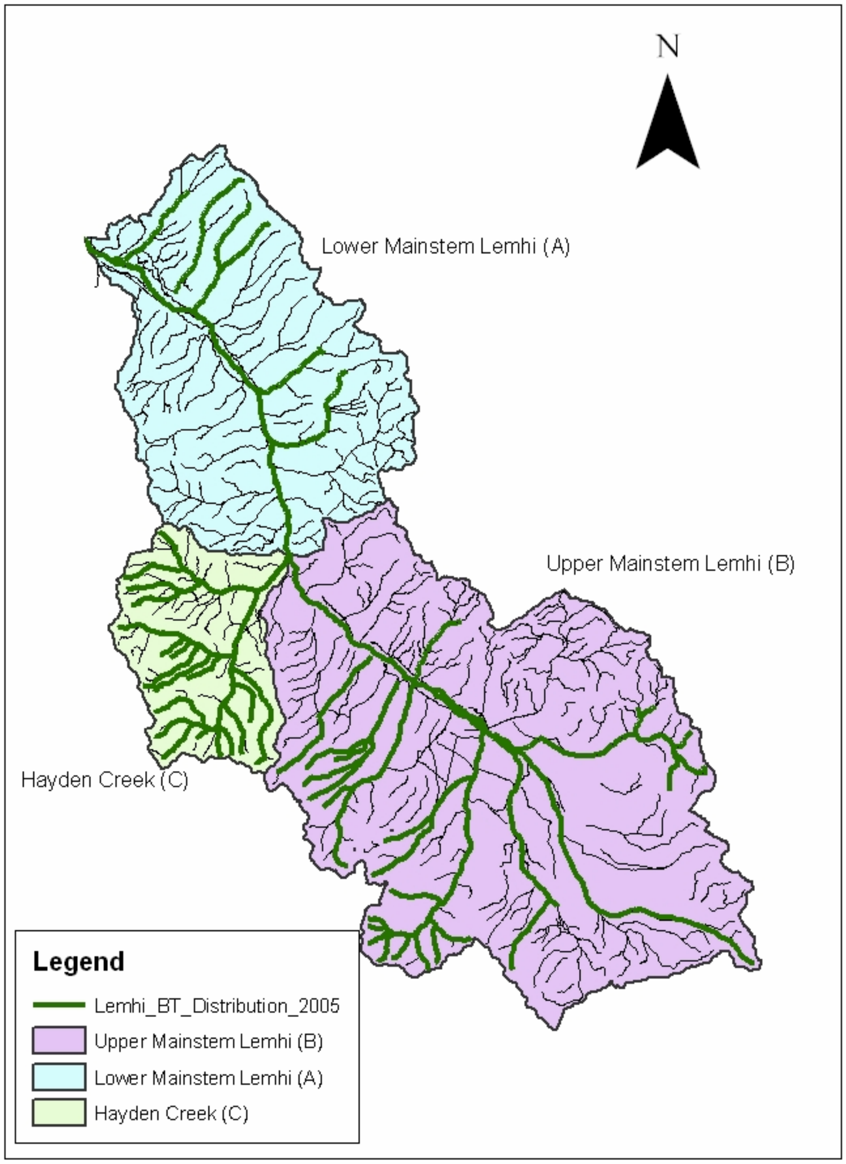


Figure 2.3.4. Distribution of bull trout in the Lemhi Basin.

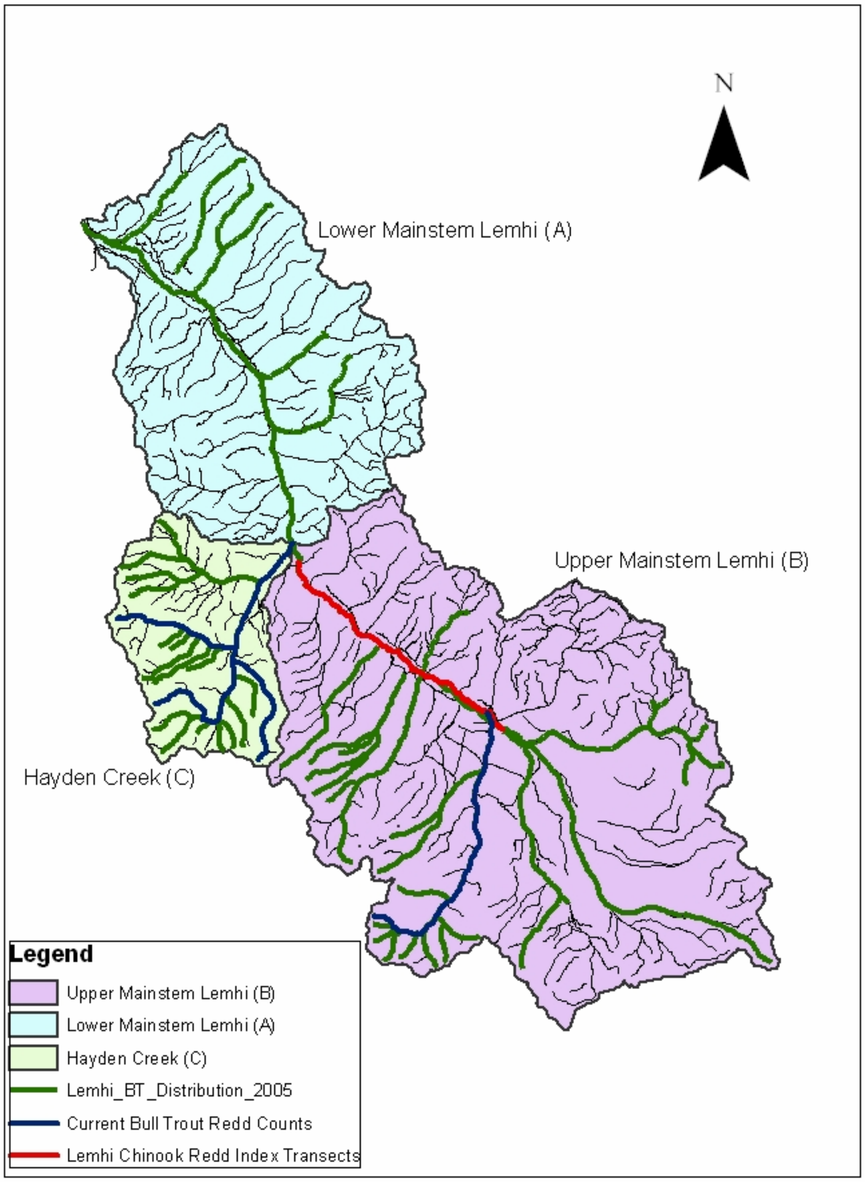


Figure 2.3.5. Map of the distribution of bull trout in tributaries (potential spawning), current bull trout redd counts, and current Chinook redd counts in the Lemhi Basin.

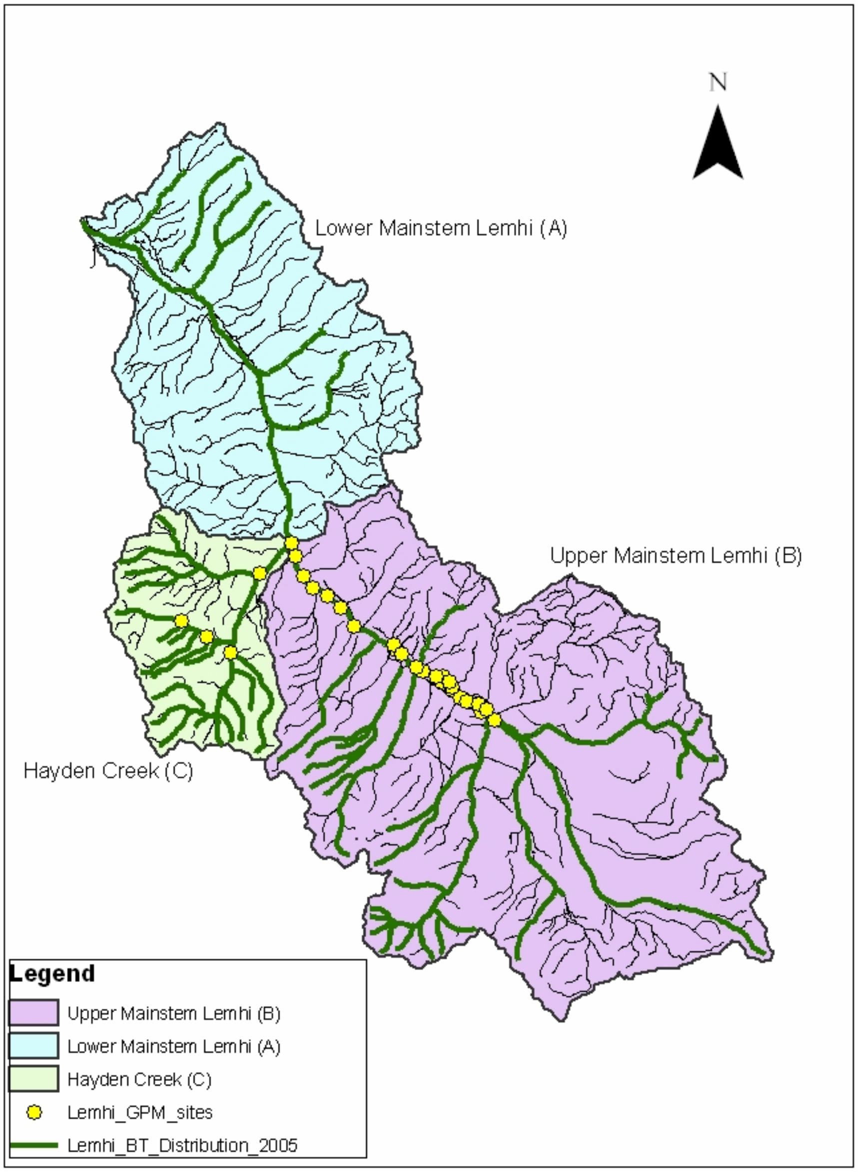


Figure 2.3.6. Spatial distribution of general parr monitoring (GPM) sites and distribution of bull trout in the Lemhi Basin.

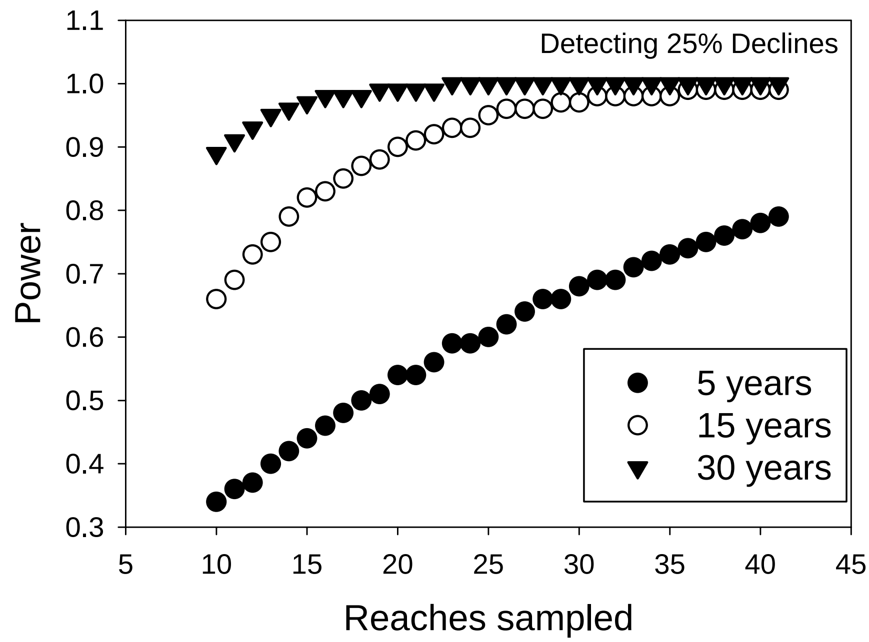


Figure 2.3.7. Power to detect changes in bull trout abundance with snorkel counts over three time intervals (Figure recreated from Al-Chokhachy 2006)

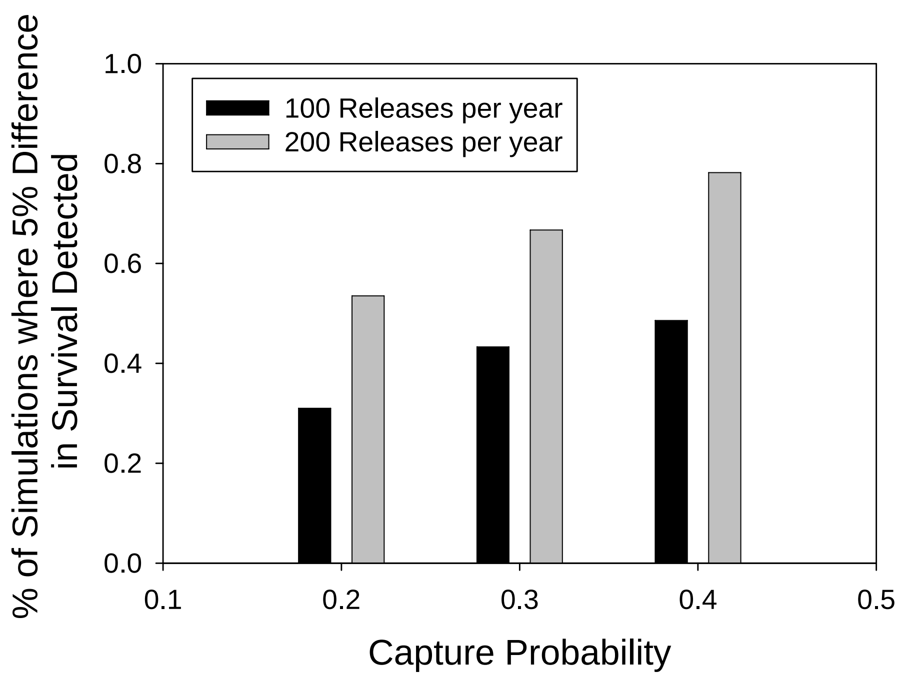


Figure 2.3.8. Simulations results for a 10-year mark-recapture with a 5% difference in survival estimates between control and treatment areas. A detectable effect occurred when models containing different survival between control and treatment differed from the null model under a multi-model approach ($\Delta AIC < 2.00$).

2.3.5 General issues in the effective design and implementation of habitat effectiveness monitoring

Background:

It is important to know how to extend the learning acquired in the CSMEP pilot design project to other locations. One element we need to adapt the Lemhi design to other watersheds is an understanding of the characteristics of the monitoring design process. This is a conceptual summary of those characteristics, rather than a mechanistic focus on details, as the latter are expected to be quite site-specific. The purpose is to stimulate discussion rather than having any pretence to being the final word on the subject. There was no specific watershed in mind when this list was assembled. It would be useful to know how this list performs in identifying such an opportunity.

Technical Design Issues:

In keeping with the Before-After-Control-Impact design we adapted to the Lemhi, three technical criteria should ideally be fulfilled in selecting other potential effectiveness pilots.

- First, we would want to have extensive "before" data on the fish responses of interest—say recruits per spawner and parr-to-smolt survival—for the watershed where habitat actions are scheduled to occur in the future.
- Second, we want comparable information for a "control" watershed where few if any new habitat actions are planned.
- Third, both watersheds would need to be comparable to one another, based on explicit and scientifically defensible criteria.

When comparing multiple watersheds it will also be necessary to show that the time series of fish species of interest are similar. One can only expect the impact of habitat management in one basin to be seen in changes to the correlated fluctuations of fish populations in a paired "reference" watershed. Uncorrelated variation, on the other hand, could result from any number of things, so changes in that variance component can not be attributed to any single difference between the watersheds – such as differences in habitat action distribution. Basins with highly correlated time series, therefore, will provide the strongest inferences about action effectiveness. In all cases, inferences absolutely depend on detailed knowledge of past and future habitat action implementation.

For comparability, there are several additional criteria that one could employ. The first, alluded to above, is that both watersheds would need to have relatively few past actions likely to change the fish metrics of interest. In addition, one should identify physical similarities in the watersheds themselves (e.g., land form, land use, land cover, gradient, hydrograph, etc.). The selection process should include interviews of knowledgeable individuals and organizations, including watershed councils and the like. However, the correlation in fish indices is probably more important that similarity in land use, land form etc. – without the correlations, there is little point in having a control area.

Implementation Issues

Beyond technical monitoring design issues, real implementation would of course need local involvement, from a wide range of stakeholders—resource management agencies, land owners, monitoring organizations, etc. This sort of paired comparison will be very difficult to implement without the active involvement of both funding agencies (in the Columbia River Basin, BPA primarily) and regulators (NOAA, primarily). An important part of maintaining the similarities between pairs of watersheds will be the regulatory authority to prevent "control" or "reference" watersheds from implementing management if opportunities present themselves in the future. It is difficult to foresee that level of authority being exercised for the sake of an effectiveness monitoring experimental design - it is hard to imagine subbasin organizations (model watersheds, etc.) lining up to be "control" sites where no habitat actions are planned. Without clear top-down guidance, where planning, prioritization, implementation and monitoring are all performed on the same large-scale, inferences about habitat action effectiveness will be compromised.

Success in applying top-down designs will likely depend on providing appropriate rewards for participation. Currently, rewards (e.g., delisting) are based on the performance of target fish populations. If monitoring designs are to be integrated into regional planning, then the products of monitoring (i.e., information & learning) need to be rewarded as well. If regional priorities included improving our knowledge of the consequences of management to fish population responses, and reward mechanisms adjusted appropriately, then local agents would be more likely to tolerate large scale habitat management being linked to experimental designs.

2.4 Harvest

Introduction

The CSMEP Harvest Subgroup is assessing the value of harvest monitoring alternatives (bias, precision, and cost) using the *US v Oregon* Technical Advisory Committee (TAC) fishery impact models as a tool to describe how precision and bias of impact estimates may be influenced by changes in harvest monitoring. Snake River Spring/Summer Chinook salmon recovery monitoring has been the focus of our design efforts in FY2006.

Our present goals are to:

- get input from TAC membership on how Columbia Basin harvest monitoring is currently conducted;
- get TAC input on alternative harvest monitoring designs; and
- ultimately we hope that CSMEP's work will assist TAC with an "assessment of data gaps and/or research that would help reduce uncertainty in harvest management and monitoring needed to gauge the success of proposed actions."

Here we report briefly on the Harvest Group's progress during FY2006, describe the analyses and evaluations in which the Group is presently engaged, and outline work the Group would likely continue into FY2007.

During much of FY2006, we continued employing EPA's 7-step DQO process (EPA 2000) to describe the decision making process used to conduct spring Chinook salmon fisheries in the mainstem of the Columbia River and in terminal areas in tributaries of the Snake River (CSMEP 2005). We described the problems encountered in conducting fisheries, namely ensuring that fisheries related mortalities do not exceed prescribed levels for conservation of weak or federal ESA-listed salmon populations or predetermined allocation rates among user groups (*US v Oregon*, Pacific Salmon Treaty, Columbia River Compact). We identified the thresholds (impact rates) at which decisions to close or reshape fisheries occur and the performance measures and metrics needed to monitor and evaluate the magnitude of the impact rates. We described the major fisheries that affect wild Snake River spring/summer Chinook salmon and define the spatial scale of interest. We reviewed the decision rules for each fishery that determine whether and for how long a fishery will be conducted.

Presently, we continue to work on DQO steps 6 and 7 (specify tolerable limits on decision errors and optimize designs for obtaining data), focusing on four major spring Chinook salmon fisheries:

1. A selective commercial drift gillnet fishery in the mainstem Columbia River downstream of Bonneville Dam;
2. A selective recreational fishery in the mainstem of the Columbia River;
3. Nonselective Platform (hook and line or hoop nets) and set gillnet fisheries in the three reservoirs upstream of Bonneville Dam; and
4. Selective tribal and non-tribal fisheries in selected tributaries of the Snake River.

Working with members of TAC, we have begun identifying opportunities and constraints to develop alternative monitoring and evaluation designs relative to current monitoring ("status quo" M&E). We have assembled datasets that represent examples of the statistical properties of the actual harvest monitoring data that are typically collected to estimate harvest impact rates. We have begun to examine the sensitivity of the impact model results to present and alternative inputs. We have begun analyzing the effects of varying rates of sampling effort both temporally and spatially on precision and bias. Through discussion with TAC membership we are preparing to characterize associated costs (e.g., FTEs, number of vehicles, boats, aerial flights, etc.). Examples of some of the metrics and sampling designs we are examining are outlined in Table 2.4.1.

Table 2.4.1. Example fishery monitoring metrics by fishery and purpose. Not all metrics required to assess impacts by stock are presented.

Fishery	Performance Measure	Purpose	Method
LCR Selective Commercial	Released (by stock composition and condition) and Kept Fish	Estimate a Release Rate to expand Landings and determine total fish released	Onboard observations
	Post release mortality	Estimate the number of released fish that die.	Research (Holding studies)
	Landed Fish at Buyers	Determine Total Harvest	Telephone surveys in Oregon; Buyers Report Themselves in Washington
Mainstem Selective Recreational	Catch	Estimate catch and encounter rates	Angler interviews
	Effort	Input for CPUE and estimate of total catch and fish released.	Aerial flights twice weekly for both boat and bank anglers
Treaty Ceremonial and Subsistence	Catch	Estimate catch	Interviews
	Effort	Input for CPUE and estimate of total catch.	Pressure counts at platforms
Warm Springs and Umatilla Commercial Gillnet	Catch	Estimate total commercial catch	Onboard observers

Because these are all mixed-stock fisheries (hatchery and wild origin, listed and unlisted ESUs), determining stock composition becomes a critical element in managing fisheries. Presently, Visual Stock Identification (VSI) in combination with Coded-Wire Tags (CWT) is used to estimate the proportion of lower river (downstream of Bonneville) catch or handle that is destined to return upstream of Bonneville Dam. Once upstream of Bonneville Dam, stock composition of wild fish remains unknown, and whether harvest rates differ among ESUs is also unknown. We have begun examining the potential to use existing PIT tagged fish to understand relative run timing of multiple ESUs and possibly stock composition. We also are examining the implications of varying sampling rates on determining stock origin with Genetic Stock Identification (GSI) techniques.

As we continue these analyses, the opportunity for further integration across the five CSMEP working groups grows. Opportunities may exist for better accounting of survival rates by life stage and cause. Presently, information on adult conversion rates derived from PIT tags is difficult to partition among removals in fisheries and upstream passage losses. For pre-season forecasting by wild stock, knowledge of the hatchery fraction and age at return is required, as are they for status and trends monitoring.

Methods

Here we present the working outline we are following to describe explicitly the inputs required to model each fishery's harvest and impact rates, the current data collection methods, and structure of the models that are presently in use. This is a work in progress and not all work elements are fully addressed here.

1. Review of impact assessments

a. Document the basic field process of fisheries monitoring for impact assessment.
    i. Ocean fisheries
        1) Commercial, Sport, State, National
        2) No work by CSMEP for spring Chinook
    ii. Mainstem Fisheries
        1) Commercial selective

Below Bonneville Dam commercial fishery monitoring currently has 16 observers on four boats (four drivers and 12 observers). ODFW works from Puget Island to Astoria Bridge and WDFW works from Cathlamet to Willamette. Best catches are typically at low tide, 20% of fishers catch 70% of landings. The role of observers is to collect information that informs what the release rate is of non-targeted species (steelhead and adipose-fin-intact Chinook salmon).

Challenges to sampling include working at night, variably low catch rates, and different gears (large mesh gear has higher impact rates to Chinook salmon and lower impact rates to steelhead, whereas smaller mesh tangle nets have lower Chinook impacts, but increase encounter rates for steelhead).

Total catch (pounds) is estimated through telephone surveys of commercial buyers (Oregon). In Washington, commercial buyers are required to voluntarily report the catch they buy with a quick turn around after they receive the catch. Number of fish is estimated through sampling of the landed catch (Pounds of fish divided by average weight per fish is total number). Number of fish handled and released is the number landed times the onboard observers' release per kept ratio.

Example: Impact assessment process/sequence in lower Columbia River selective fisheries. In mark-selective fisheries all wild fish are released, thus mortality of wild fish occurs only among released fish.

- Estimate the upriver run abundance based on retrospective analyses.
- Estimate the stock composition of fish with intact adipose fins.
- Allowable impact rate for given run size is defined by schedule in US v Oregon Columbia River Fisheries Management Plan.
- Use the anticipated impact allowance to inform how to open a test fishery (commercial).

- Onboard harvest monitoring - Onboard observers monitor fishery handle (harvest of marked fish and release of fish that had intact adipose fins). Samplers observe about 10% of fish handled.
- Landed catch sampling at fish buyer. Note only adipose fin-clipped fish are legal to sell.
  - Sample the catch at buyers to estimate average weight. Samplers weigh about 20% of landed catch.
  - Estimate total harvest by species from weights on landing tickets and the average weight of fish at buyer.
  - Use visual stock identification (VSI) to determine the ratio of fish from upper (versus lower) Columbia origin. Corroborate VSI with coded wire tag recoveries.
  - VSI – dark-colored spring Chinook are upriver stocks. Lighter colored fish are stocks from Cowlitz, Kalama, Lewis, Willamette, and Sandy rivers
- Update run size forecast as additional information becomes available Bonneville Dam counts, run timing observations, etc.
- Monitor whether impact rate appears to remain within forecast allowable rate and whether forecast allowable rate appears to track with the CRFMP schedule.
- Continue fishing and monitoring…

2) **Treaty commercial fisheries** in Bonneville, The Dalles, and John Day reservoirs ("Zone 6")
   a) **Platform fisheries -** 3 main platform areas – Bonneville Dam to Bridge of Gods (50% of platforms); Lone Pine (40%); John Day Dam (1/4 mile below JDA: 10%)
      i) Day – 3, 8hr periods, monitor 4 hrs. total during 2 of 3 periods at a site, record number of active gears, direct observation of the catch during monitoring periods
      ii) At start of day - interview fishers for number caught when monitor not there
   b) C & S permit fishery
      i) Umatilla and Warm Springs - Tribal monitors observe 100% of catch
      ii) Yakama and Nez Perce tribes monitored through interview process that includes observed and reported catch.
   c) Treaty Commercial Fishery
      i) Monitored at landing sites, at different areas within pools
      ii) Data on number of nets out and fish per net
      iii) Aerial counts on 2nd or 3rd day of fishery, in AM
         (1) Start counting upstream of Bridge of the Gods – count gear in water up to McNary tailrace
         (2) Have counts in sub-areas to get at distribution (which is patchy)
         (3) Make some correction for diver and drift nets (used at night mostly) – make up small percent of fishery
         (4) Issues: Windy & choppy conditions/glare – difficult to see; Sometimes nets removed to clean – may miss some gear; Need to correlate number of nets to effort throughout period; Hence, counts are indices (index survey).
         (5) Some "ground truthing" during season

3) Treaty Ceremonial and Subsistence and Commercial

The spring fishery develops with different levels of effort, gear types, and purpose. Initially Ceremonial and Subsistence fishing begins (platform and permit gillnet fishing) with the intent to support the community longhouses and churches for the *Washat* service (First Food Feast). As the Chinook run develops, commercial gillnet fisheries begin.

Monitoring of platform fishing is conducted by four CRITFC samplers. Platform fishing effort is distributed approximately 25 sites from Bonneville forebay to Bridge of the Gods, 10 to 20 further upstream to Lone Pine (The Dalles tailrace, Oregon shore), and five in John Day Reservoir – about 50% in Bonneville, 40% in The Dalles, and 10% in John Day reservoirs. The sampling plan was originally developed by Steve Parker and Bill Bosch: 3 8-h periods per day with each 8-h block divided into 2 4-h observation periods in all three pools.

In the C&S permit gillnet fisheries, Warm Springs and Umatilla Tribes sample 100% of their members' catch with onboard observers and most fishing is conducted in Bonneville Reservoir. The Nez Perce Tribe and Yakama Nation monitor harvest through an interview process that includes observed and reported catch.

For the commercial gillnet fishery, effort is monitored through aerial surveys of nets in the water in all three pools proceeding upstream along one shore, then downstream along the other shore. Attempts are made to conduct flights in the middle of the fishing period (peak counts) to ensure fishers had time to deploy nets and that nets have not yet been pulled. Observations could be biased downward due to sun glare or chop on the water. Catch is enumerated through phone surveys of commercial buyers. Note that depending on which tribe's catch is sampled for biological information, samples might not be representative of the upriver run at large, because the tribes typically fish at somewhat different times, whereby the Warm Springs begins first, followed by Umatillas, Yakamas, and Nez Perce tribes. Development of the fish processing facility at Bingen could provide an opportunity for bio-sampling. PIT tags might provide an opportunity to obtain biological information, but only if present in sufficient numbers and if fish remain in the round before they are scanned. In 2006, it appears that approximately 1.6% (unadjusted for fallback or misdetections) of the spring/summer Chinook salmon ascending Bonneville had PIT tags). Detection of PIT tags at Bonneville might also provide an opportunity to refine the knowledge of the overlap in run timing between the spring/summer fish and the summer-run fish.

4) Sport selective fisheries in lower Columbia River

Effort counts (both boat and bank) are conducted by aerial surveys conducted twice weekly. Catch and release rates are determined from angler interviews (complete trips for boats, partial trips for bank). Bio sampling is conducted dockside (creel sampling). Estimates of released fish are based on angler interviews. Potential bias exists if anglers forget how many fish were released or exaggerate actual release rate (plus or minus). Behavioral protocols for modeling the expansion factors for effort rates could be dated (1960 or 70s) and are no longer very accessible in terms of understanding the algorithms (the potential exists to work with agency staff to decipher and document the assumptions of the old mainframe computer program, but only if done before staff retire).

5) Tributary fisheries

a) Primarily sport
b) Harvest relatively small

When hatchery-origin fish are believed to be more abundant than what is needed for broodstock, selective fisheries (retention of adipose fin-clipped fish only) have been opened in the Snake River Basin. Examples include Imnaha River, Clearwater (fish not listed), lower Salmon River, and Little Salmon River. The fisheries are shaped to minimize impacts to wild fish. Sampling methods include creel surveys (Oregon and Idaho) and self check stations (Idaho). In Oregon, the release mortality rate has been assumed to be 10%.

2. Documentation of impact model

a. Review (audit) the mechanics of calculations in impact spreadsheet models
   i. Audit - trace formulas and logic
   ii. Start with selective fishery sport/commercial spring fishery spreadsheet
b. Demonstrate spreadsheet function (and sampling process) through graphic

The concept of impact in Columbia River fisheries management

Harvest action implementation monitoring in Columbia River fisheries is accomplished by monitoring and estimating incidental mortality, or take, of endangered fish stocks associated with each of the annual sport and commercial fisheries throughout the basin. The incidental mortality rate in these fisheries is referred to as an impact rate, or an impact. Impact rates are estimated during each allowed fishery period. For example a two-day commercial gill-net fishery has a specific impact estimate. Commercial, sport, and Tribal fisheries have allowable impact rates that are established in the US v Oregon Columbia River fish management plan (Parties to US v Oregon 2005). Allowable impact rates are scaled to the ESU run size at the mouth of the Columbia River and a fishery can remain open until the allowable impact rate is reached.

Incidental mortality associated with fisheries has three principle components "drop-off mortality," harvest, and post-release mortality (D, H, and R). Drop-off mortality occurs when a fish encounters the gear, escapes (or at least is not handled), and dies from related injury or stress. Drop-off mortality is not addressed in Columbia River fisheries impact assessments, as it is believed to be a negligibly low rate. Harvest mortality, perhaps obviously, is the retained or kept catch. Post-release mortality occurs when a fish is handled and intentionally released. Post-release mortality is a particular concern in mark-selective fisheries; that is fisheries that only allow harvest of fish that have a clipped adipose fin. Mark-selective fisheries allow harvest of hatchery fish and reduce direct mortality of wild or naturally produced stocks. Conceptually, impact rate can be thought of as the product of the probability of being captured and the probability of mortality after release.

3. Sensitivity analysis

a. Develop a tool - an algorithm emulating the spreadsheet calculations and describing associated variance contributions.
b. Focus on performance measures that are perceived to contribute substantial variance.
   i. Released fish ratio
   ii. Post release mortality rate
   iii. Stock composition of catch
c. Items that contribute less variance (perceived)
   i. Preseason abundance estimate
d. Items that contribute little variance
   i. Weight of landed catch. Buyer and seller have vested market interest in bias of these data at time of sale. There is the potential for careless, untimely, or inaccurate reporting of landed weights to state agencies.
   ii. Average weight of landed catch. Sampling target is 20% of the landed catch. Data are available to examine the variance of sampled weights and the actual sampling rates.

4. **Development of status quo, high, medium, and low sampling designs -** List developed among harvest subgroup and query of TAC membership

a. Released fish ratio
   i. Increase sampling rate for onboard observers during selective commercial fisheries
   ii. Address bias concerns in sport angler reporting of released catch. Number of released fish depends on quality of interviews, truthfulness of interview, and memory bias. Magnitude and

direction of biases are unknown. They can only be known by comparing direct observations with interview data.

b. Post release mortality rate
   i. A Double Index Tagging (DIT) program is presently not considered by TAC as a means to describe post release mortality rate using hatchery fish as a surrogate. A DIT program tags a release group of juvenile fish but applies external marks (fin clips) to only a proportion of the release group. By monitoring the mark-no mark ratio of the tag group, the changes in the ratio upon escapement past fisheries can inform differential mortality rates among unmarked and marked fish. While a DIT program could provide information on mortality of unmarked fish in Columbia Basin spring Chinook fisheries, implementing such a program is logistically challenging for reasons including:
      - the need for relatively large tag release groups;
      - coordination of marking and tagging among numerous hatcheries and programs;
      - considerable effort to sample both tagged and untagged fish at sufficient rates in multiple fisheries and hatcheries;
      - potential cross purposes with existing programs (e.g., separating wild and hatchery origin fish in SRB supplementation programs); and
      - social issues with fishers losing opportunity to be able to retain hatchery origin fish in selective fisheries (because some hatchery fish are not fin clipped and cannot be distinguished from wild fish that must be released).
   ii. Field studies - Post release mortality estimates used in selective fisheries are based on just a few field studies and a technical consensus of TAC membership. Small-mesh fishery rate is based on three years of study – but each year the experiment was conducted differently. Sensitivity analysis may help describe the potential magnitude of bias and how much variance might affect estimates of incidental mortality. The TAC, by and large, greets additional field studies of release mortality skeptically, given the inherent expense and difficulty of obtaining definitive results.

c. Stock composition
   i. Mark Rate/ Stock composition – Different hatchery stocks have different mark rates. Need to know –
      1) Number of unmarked hatchery fish released in the fishery
      2) Associated stock composition of unmarked
         a) Upriver wild
         b) Upriver hatchery
         c) Willamette hatchery
         d) Willamette wild
         e) Clackamas wild
         f) Clackamas hatchery
      3) Is there a representative sample spatially and temporally in the fishery?
      4) Does applying juvenile mark rates to pre-season adult forecasts adequately estimate mark rate of returning adults? Could be assisted by real time GSI sampling? Need to know how many unmarked fish monitors could have sampled for GSI with current level of monitoring
   ii. GSI sampling – This methodology holds real promise based on conversations with Pacific States Marine fisheries biologists involved in Klamath River fisheries. Cost of processing samples was about $20 per fish and results (stock origins) were available the following day (personal communication, Carlos Garza, PhD, NOAA Fisheries, Santa Cruz).
   iii. PIT tag sampling – current sampling programs conducted downstream from Bonneville Dam do not carry PIT tag detectors.

1) Commercial fisheries conceivably could outfit buyers with high volume detectors. Concerns were raised that this could substantially slow fish processing creating a cost burden on buyer and perhaps the seller.
2) Sport fish creel samplers can be outfitted with PIT-tag detectors. Concerns were raised that samplers gear is already heavy and current PIT tag detectors add substantial weight. Lighter models are available through various vendors but do not meet the PSMFC standard.

iv. Increased CWT recovery and follow up
1) If called for can we increase the rate of CWT recovery? More creel samplers. More samplers at buyers.

v. Examine random sampling procedures in observer program to address concern about bias in steelhead catch composition.
1) Need to follow up with fishery managers to further understand the reasons this bias is suspected and how to address it

vi. Sport Harvest - Release mortality and stock composition issues are similar to those for commercial harvest

d. Total catch estimation in sport fisheries – precision of estimates using creel samples from a fraction of anglers/location/time
i. Look at main frame program (review/audit)
ii. Is there a sampling protocol/design?
iii. Is sampling conducted according to protocol
iv. Sampling precision?
v. Sampling frame vs. target population?

e. Preseason abundance estimate
i. TAC members generally agree that the preseason forecast is reasonably accurate. Because it is adjusted in season and post season using dam passage and harvest estimates, the effect on final estimates of impact are ameliorated through time.
ii. Consequences of differences between the preseason and final estimates of run size include:
1) Lost opportunities for harvest in fisheries below Bonneville Dam in the case of underestimate.
2) Lost opportunities for harvest in fisheries above Bonneville Dam in the case of overestimate and disproportionate use of allowable impacts by downstream fishers.
3) We have a graphic comparing these estimates over time
4) Saang Yoon Hyun has spent some time examining the process and potential errors in preseason estimates that we encourage him to incorporate.

f. Is there enough sampling effort to accurately estimate Zone 6 Tribal platform fishery?

g. Tributary (Snake River) sampling issues to examine.
i. Creel issues for Clearwater and Salmon no different from above – is creel sampling design adequate (precision/bias)
1) Overall contribution of hatchery and wild stocks (stock composition estimation)?
2) Conversion rates from BON to LGR – need work for some stocks.

ii. Steelhead – fall through winter through spring,
1) Phone survey with spot checking through creel survey
2) Phone survey bias high
3) Accounting for the missing “B-run” fish.

5. **Identification of monitoring program costs-** Through discussion with TAC membership we are preparing to characterize associated costs of fisheries monitoring programs (e.g., FTEs, number of vehicles, boats, aerial flights, etc.).

Results

We have recently developed a preliminary and simplified "impact" model for the lower Columbia River commercial spring fishery. We have also completed a preliminary exercise that evaluates the variability resulting from characteristics of the input data variables and fishing gears (Section 2.4.3.1). We have tracked the rate at which adult PIT-tagged spring and summer Chinook salmon were present in the Columbia River upstream of Bonneville Dam in 2006 by time and ESU of origin. These preliminary results are still in review within the CSMEP Harvest Subgroup and will be refined in FY2007.

Discussion

Further discussion will ensue upon wider review and further development of results. Insights developed to date include but are not limited to:

Collection and use of harvest data

- The mortality rate of federal ESA-listed fish in Columbia River fisheries (incidental in selective fisheries and direct in targeted fisheries) is referred to as an impact rate, or an impact. Conceptually, in selective fisheries, an impact rate can be thought of as the product of the probability of being captured and the probability of mortality after release.
- The variables needed to assess incidental mortality in selective fisheries are: run size, harvest number, stock composition of the catch, marked fish release rate, and post release mortality rate. In assessments of upriver spring Chinook take, TAC views the estimate of harvested fish to be strong and accurate. The preseason forecast is reasonably accurate and is adjusted in season and post season using passage and harvest estimates. Stock composition may be improved by PIT tag monitoring and genetic stock identification (GSI) technology. Onboard observations are used to monitor marked fish release rate. Standard gear-specific values are applied to estimate post release mortality.
- Stock composition in mainstem fisheries is estimated within season by applying juvenile mark rates to pre-season adult forecasts and assumed proportions of wild fish in the juvenile runs. This could be improved or corroborated by real-time Genetic Stock Index sampling or PIT tag sampling.

Challenges of harvest management – Uncertainties in precision of estimates

- In general fisheries managers do not provide precision bounds on estimates of harvest and incidental take.
- Models to assess impacts of Columbia River fisheries on listed fish species are frequently revised (of necessity) and poorly documented (due to lack of staff time). The *US v Oregon* Technical Advisory Committee (TAC) agrees that better documentation is needed and asks if CSMEP staff will be able to take on this task.
- Onboard observation of catch and release numbers in selective commercial fisheries have a potential bias in estimates of steelhead stock composition that might be addressed through stratification or randomizing sampling design.
- Post release mortality estimates used in selective fisheries are based on just a few field studies and a technical consensus of TAC membership. Small-mesh fishery rate is based on 3 yrs of study – but each year the experiment was conducted differently. Sensitivity analysis may help describe the potential magnitude of bias and how much variance might affect estimates of incidental

mortality. The TAC, by and large, greets additional field studies of release mortality skeptically, given the inherent expense and difficulty of obtaining definitive results.

Next steps – Could alternative monitoring approaches improve estimates?

- The CSMEP harvest group is examining existing models to suggest means for incorporating variance estimates.
- The CSMEP harvest subgroup is currently conducting power analyses to describe the effect a range of variation in marked-fish release rate has on estimates of incidental mortality.
- CSMEP auditing of the lower Columbia spring fisheries impact model has identified a probable calculation error that results in small-magnitude errors in estimates of incidental mortality. We will work with TAC membership to confirm and correct the error.
- The results of these analyses will provide a context to develop potential alternative Monitoring and Evaluation approaches.
- *Beyond the CSMEP Harvest Subgroup:* TAC envisions a need for full ESU/population-based run reconstruction for steelhead.

2.4.1 Impact of commercial selective fisheries on wild upriver spring Chinook salmon in the lower Columbia River

Basic definition of Impact

The U.S. v. Oregon Technical Advisory Committee (TAC) defines Impact of a fishery on a wild fish population as total mortality (rate) of the population derived from in-river harvest effects. During return season of a fish population, the TAC updates Impact for harvest decisions. The basic equation of Impact, I at time t in fish run season is:

$$I_t = \frac{\text{number of wild fish that die from fishery effects up to time } t}{\text{number of wild fish return to the Columbia River}} \qquad (1)$$

Impact is dimensionless, and ranges from 0 to 1. Fishery effects in eq. 1 incorporate both direct catch and post-release mortality. Post release mortality may occur when unmarked wild fish are released (as required by regulation in the selective fishery for Chinook salmon) but die due to injury or stress incurred during catch and release.

Quantities that affect Impact of commercial selective fisheries

Commercial harvest of spring run Chinook salmon in the lower Columbia River (zones 1-5) is made by selective fisheries where by-catch of unmarked wild fish must be released. The fishery is complicated by operation of two different fishery gears: 4-1/2” mesh tangle net, and 8 or 9” mesh gill net. Post-release mortality of fish differs by fishing gear (Ashbrook et al. 2004).

To demonstrate a sensitivity analysis of Impact rate we focused on commercial selective fisheries for Upriver spring Chinook salmon that occur in the lower Columbia River. Data on weekly in-river fishery harvest from 2005 are available from the TAC. Two different gears are used in these fisheries and the TAC uses the following equation for calculation of the Impact at time t in the fish run season:

$$I_t = \frac{(H_{t,S} + H_{t,L})}{N_t} \cdot \frac{\left(R_{t,S} M_S + R_{t,L} M_L\right)}{(R_{t,S} + R_{t,L})} \tag{2}$$

where the first fraction of "$(H_{t,S} + H_{t,L})/N_t$" represents the probability of fish caught, and the second fraction of "$\left(R_{t,S} M_S + R_{t,L} M_L\right)/(R_{t,S} + R_{t,L})$" indicates the post-release mortality. See Table 2.4.2 for notations, and also refer to the 2005 dataset (Tom Rien, ODFW) for why we have seven quantities in eq. 2 for the Impact calculation.

Sensitivity analysis

We are interested in identifying which variables affect Impact most. The sensitivity analysis is not a trivial task, because (*a*) seven variables involve the calculation of Impact, and (*b*) data and information on the respective quantities are limited. There would be two approaches for the sensitivity analysis: analytical and numerical methods.

Analytical method

To calculate the mean and variance of Impact in eq. 2, we considered using the Delta method, which is based on a Taylor series expansion (Rao 1973, Benichou and Gail 1989). However, to deal with seven quantities using Delta method seems to be too tedious especially when considering covariances between those quantities. Although some quantities are not correlated (e.g., between M_S and M_L), we cannot ignore covariance between $H_{t,S}$ and N_t, between $H_{t,L}$ and N_t, between $H_{t,S}$ and $R_{t,S}$, and between $H_{t,L}$ and $R_{t,L}$. For example, data used in updating forecast of return size (abundance), N_t include catch data of $H_{t,S}$ and $H_{t,L}$, and thus N_t are correlated with $H_{t,S}$ and $H_{t,L}$, respectively. Also, catch number and by-catch of unmarked wild fish are likely to be correlated; i.e., $H_{t,S}$ (or $H_{t,L}$) is likely correlated with $R_{t,S}$ (or $R_{t,L}$). Note that while we continue to explore using the Delta method, we have initially examined a Monte Carlo method to simulate variability in Impact rate.

Numerical method

As an alternative to the Delta method, we used a Monte Carlo numerical method to calculate the mean and variance of Impact. Letting $\mathbf{X}_t$ be a vector that has N_t, $H_{t,S}$, $H_{t,L}$, $R_{t,S}$, and $R_{t,L}$, we assume $\mathbf{X}_t$ to be multivariate normal. That is,

$$\mathbf{X}_t \sim MVN\left(\boldsymbol{\mu}_t, \Sigma_t\right) \tag{3}$$

where $\boldsymbol{\mu}_t$ = the vector that contains the expected values of elements in $\mathbf{X}_t$; and Σ_t = covariance-variance matrix of elements in $\mathbf{X}_t$. That is, $\boldsymbol{\mu}_t$ is a 5 x 1 column vector,

$$\boldsymbol{\mu}_t = \begin{bmatrix} E(N_t) \\ E(H_{t,S}) \\ E(H_{t,L}) \\ E(R_s) \\ E(R_L) \end{bmatrix}_{5\times 1} \tag{4}$$

and Σ_t is a 5 x 5 matrix,

$$\Sigma_t = \left(s^2_{\text{ij}}\right)_{5\times5} \tag{5}$$

where index *i* refers to N_t, $H_{t,S}$, $H_{t,L}$, $R_{t,S}$, and $R_{t,L}$ in order, and element s^2_{ij} denotes covariance-variance matrix. For example, s^2_{11} is the variance of N_t, and s^2_{24} is the covariance between $H_{t,S}$ and $R_{t,S}$.

We assume that M_S and M_L are independent of each other, and also of the other quantities. Those post-release mortality rates are not correlated with each other, and they are not correlated with forecast of return abundance, catch abundance, and by-catch of unmarked wild fish. Because the domain of post-release mortality is from 0 to 1, we assume it to be a beta random variable.

$$\begin{aligned} M_S &\sim Beta(\alpha_S, \beta_S) \\ M_L &\sim Beta(\alpha_L, \beta_L) \end{aligned} \tag{6}$$

where $\alpha_S > 0$, $\beta_S > 0$, $\alpha_L > 0$, and $\beta_L > 0$.

The Monte Carlo procedure is as follows. Given $\boldsymbol{\mu}_t$ and Σ_t, we can generate many random values of N_t, $H_{t,S}$, $H_{t,L}$, $R_{t,S}$, and $R_{t,L}$ from the multivariate normal distribution in eq. 3. Also, given α_S, β_S, α_L, and β_L, we can generate many random values of M_S and M_L from beta distributions in eq. 6.

- First, we generate hundreds of those random values for the respective seven variable, and store them for each variable: e.g., $N_t^{(1)}$, $N_t^{(2)}$, $\cdots$, $N_t^{(k)}$, $H_{t,S}^{(1)}$, $H_{t,S}^{(2)}$, $\cdots$, $H_{t,S}^{(k)}$, $\cdots$, $M_L^{(1)}$, $M_L^{(2)}$, $\cdots$, $M_L^{(k)}$, where *k* random values per each variable are saved. Note that random values for the first five variables of N_t, $H_{t,S}$, $H_{t,L}$, $R_{t,S}$, and $R_{t,L}$ must be generated and stored as a set per the variables, because they are dependent of each other (eq. 3).
- Second, we pass those random values for each variable to the corresponding variable in eq. 2 to calculate Impact. As a result, we have hundreds of random values of Impact. Finally, we can infer the mean and variance of Impact from these random values of Impact.

The core part for this calculation is data or information on parameters that govern the multivariate normal and beta densities in eqs. 3 and 6. Because data for those parameters are not all available, we use literature information and plausible values as well. The main intent of this document is to describe the methods for this sensitivity analysis for Impact is done. Our intent is not to address bias and precision of Impact.

For demonstration purposes, we assume the calculation of the Impact on 1 April 2005, using updated data from that date, literature information, and plausible values for missing data. Elements of parameters in vector $\boldsymbol{\mu}_t$ and covariance-variance matrix Σ_t are shown in Tables 2.4.3 and 2.4.4. Elements of α_S, β_S, α_L, and β_L in eq. 6 can be calculated with method of moments (MM), using the mean values and variances of M_S and M_L in Table 2.4.3. Finally, feeding those parameter values, we can generate random values of the key seven variables from the corresponding distributions in eqs. 3 and 6, and then calculate random values of Impact with the above Monte Carlo method.

Reference Impact

Our ultimate goal is to examine the relative change in Impact in response to changes in the respective seven variables in eq. 2. Using estimates in Tables 2.4.3 and 2.4.4 for parameters that govern those seven variables, we calculate Impact at 1 April 2005 and use the Impact as its standard or reference. We show the resultant Impact in Figure 2.4.2. Distribution of the reference Impact is seriously skewed, and thus the mean of the reference Impact is different from the mode (Figure 2.4.2). We will examine how the reference Impact change in response to changes in seven variables in eq. 2. Estimates of parameters in Tables 2.4.3 and 2.4.4 are uncertain, and will be updated with managers' opinions or when more information or data are available. The key point in this analysis is not how accurate the reference Impact is, but which variable most influences the Impact.

Table 2.4.2. Notations.

Indices	
t	Time (day)
S	Small mesh size, referring to tangle net
L	Large mesh size, referring to gill net
CV	Coefficient of variation (= $\sqrt{\mathrm{var}}$ /mean)
r	Correlation coefficient
Cov	Covariance
Variables	
I	Impact (rate)
N_t	Abundance of the fish population return (hatchery + wild) to Columbia River mouth. The quantity is used as preseason forecast of the return size before in-season data are collected. As time progresses, in-season forecast of the return size is updated and used for N_t . Thus, subscript time t is added.
$H_{t,S}$	Cumulative catch of the population caught by small-mesh gear (e.g., tangle net gear) to time t.
$H_{t,L}$	Cumulative catch of the population caught by large-mesh gear (e.g., gill net gear) to time *t*
$R_{t,S}$	Cumulative number of wild unmarked fish caught by small-mesh gear and released to time *t*
$R_{t,L}$	Cumulative number of wild unmarked fish caught by large-mesh gear and released to time *t*
M_S	Post-release mortality of fish that are released from small-mesh gear. Post-release mortality does not depend on time and its notation does not have subscript time *t*.
M_L	Post-release mortality of fish that are released from large-mesh gear
$\mathbf{X}_t$	Vector of N_t , $H_{t,S}$, $H_{t,L}$, $R_{t,S}$, and $R_{t,L}$.
Parameters	
$\boldsymbol{\mu}_t$	Vector that has the respective mean values of N_t , $H_{t,S}$, $H_{t,L}$, $R_{t,S}$, and $R_{t,L}$.
Σ_t	Covariance-variance matrix of the vector of N_t , $H_{t,S}$, $H_{t,L}$, $R_{t,S}$, and $R_{t,L}$.

α_S, β_S, α_L, β_L	Parameters that govern beta distributions of M_S and M_L.

Table 2.4.3. Seven variables required for calculation of Impact of selective fisheries of wild Upriver spring Chinook salmon (*Oncorhynchus tshawytscha*) in the lower Columbia River, and the mean, CV, and variance values of those variables at 1 April 2005. Number in parentheses for the first five variables indicates element index in the vector, $\mathbf{X}_t$. Thus, values under "Var" column are diagonal elements (s^2_{ii}) in the covariance-variance matrix in eq. 5, where ii = 11, 22, 33, 44, and 55. Mean values are from data (data source: the U.S. vs. Oregon Technical Advisory Committee (TAC)). CV value of N_t is from empirical experience (Hyun et al. 2006). CV values of $H_{t,S}$, $H_{t,L}$, $R_{t,S}$, and $R_{t,L}$ are plausible values, based on discussion with the TAC. CV values of M_S and M_L are from study of Ashbrook et al. (2004).

Variable	Mean	CV	Var
N_t (1)	106,800	0.50	2,851,560,000
$H_{t,S}$ (2)	2,417	0.06	21,031
$H_{t,L}$ (3)	591	0.06	1,257
$R_{t,S}$ (4)	213	0.02	18
$R_{t,L}$ (5)	778	0.02	242
M_S	0.185	0.72	0.018
M_L	0.400	0.20	0.006

Table 2.4.4. Values of non-diagonal elements in covariance-variance matrix in eq. 5. Diagonal elements are shown in Table 2.4.3. Column *r* denotes correlation coefficients. See Table 2.4.3 for subscript index of element. For example, s^2_{12} = s^2_{21} = covariance between N_t and $H_{t,S}$. These correlations and covariances are plausible values. Total catch of fish by a small-mesh (or a large-mesh gear) is not necessarily correlated with by-catch of unmarked wild fish by a large-mesh gear (or a small-mesh gear). Thus, s^2_{25} = 0; s^2_{34} = 0.

Element	r	Cov
s^2_{12}	0.5	3,872,034
s^2_{13}	0.5	946,782
s^2_{14}	0.3	68,245
s^2_{15}	0.3	249,271
s^2_{23}	0.8	4,114
s^2_{24}	0.8	494
s^2_{25}	0	0
s^2_{34}	0	0
s^2_{35}	0.8	441
s^2_{45}	0.8	53

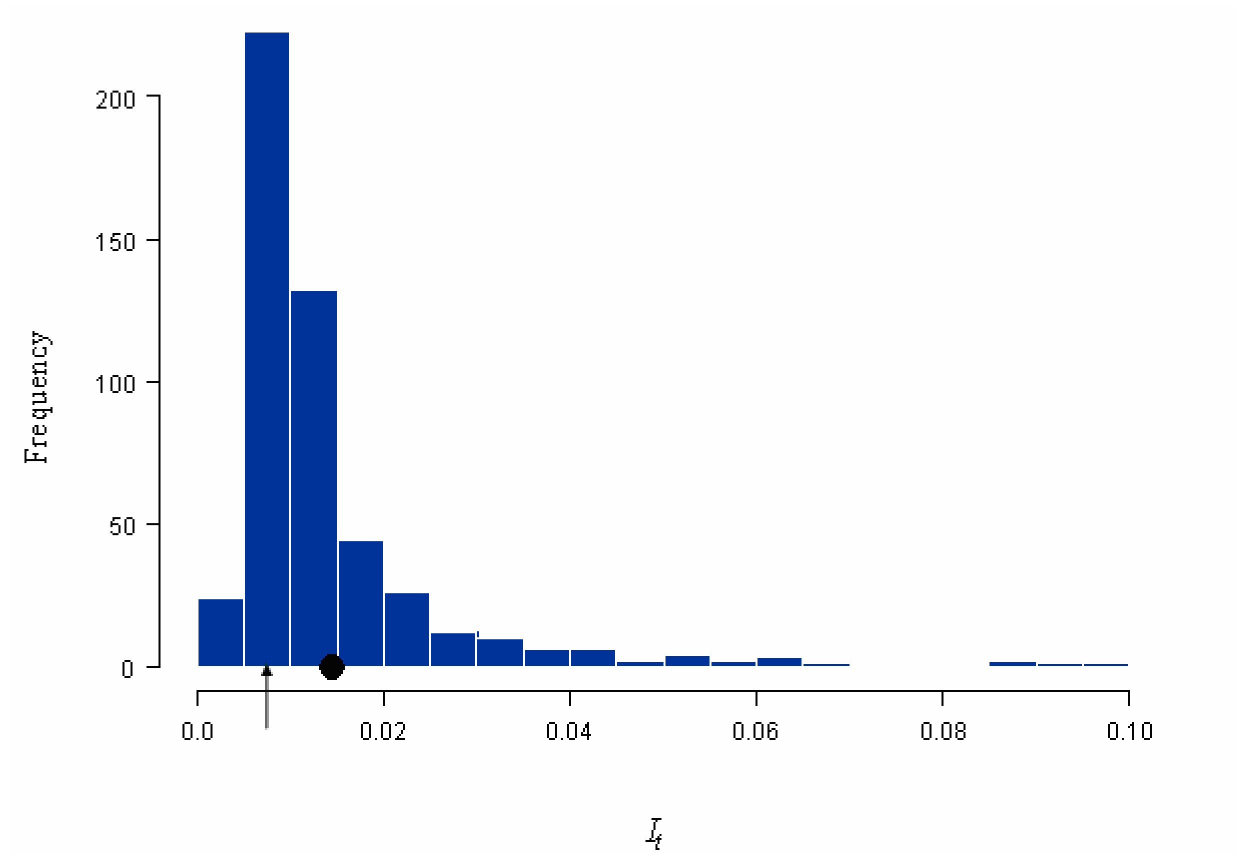


Figure 2.4.2. Illustration of reference Impact at 1 April 2005 for selective fisheries that target wild Upriver spring Chinook salmon (*Oncorhynchus tshawytscha*) in the lower Columbia River. With Monte Carlo methods, we calculate the Impact, using values in Tables 2.4.3 and 2.4.4 for parameters that govern seven variables in eq. 2. Black dot indicates the mean of the Reference Impact ($\approx$0.0145), and arrow indicates the mode ($\approx$ 0.0065). Standard deviation of the Reference Impact is about 0.0186.

2.5 Hatchery

Throughout the FY2004-06 contract period, the Hatchery Subgroup identified a number of questions important to the evaluation of hatchery management, and has reviewed numerous existing and proposed hatchery research, monitoring, and evaluation (RME) plans within the Columbia River Basin. The Hatchery Subgroup's review of regional hatchery M&E suggested that many of the critical CSMEP identified hatchery questions would be sufficiently addressed at the level of individual hatcheries if existing and proposed hatchery RM&E programs—such as that associated with the Northeast Oregon Hatchery project (NEOH)—were funded. Given this, the subgroup in FY2006 began to focus on a subset of key hatchery questions (summarized in Appendix F) that would not or cannot be addressed by individual programs. In short, there are a number of regionally relevant hatchery *effectiveness* questions that simply cannot be answered by existing or proposed RME when programs are viewed as separate entities. We agreed that these regional questions were not only high priority, but they fit within the regional mandate of CSMEP.

For initial development of broader, regional-scale hatchery M&E approaches we prioritized the following hatchery questions for initial evaluation in FY2006:

1. What is the magnitude and distribution of straying of adults from harvest augmentation and supplementation hatcheries?
2. What are the disease agents and pathogens in hatchery fish, to what degree are these agents transmitted to natural fish, and what are the impacts of such transmissions?
3. What is the relative reproductive success of hatchery and natural origin adults under natural conditions?
4. What are the effects of hatchery origin adults on the viability of target and non-target populations?

Question four is considered the most critical for hatchery M&E . Questions one through three are really a subset of the information required to address question four. A number of additional broad hatchery questions were identified (see Appendix xxx), but the four questions above were thought to represent a more “realistic” goal for 2006, and represent some of the most important information needs required to answer question four.

All of the effectiveness questions share a common theme; namely, they all approach uncertainties associated with the operation of hatcheries as a class of actions, and as such, are most applicable when viewed at the scale of the Columbia River Basin as whole. Although somewhat daunting, this approach has some real strengths:

1. Within each question there are strata that can be used to define a sampling design (e.g., we probably want to look for hatchery strays in streams with large hatcheries, small hatcheries, and no hatcheries etc.).
2. Because there are a lot of candidate locations/hatcheries, we can use a statistical design to select which ones should be sampled.
3. Because of points one and two, the results of such a stratified analysis enable us to predict outcomes at unsampled locations/hatcheries.

This approach should allow us to employ an EMAP style approach to probabilistically select locations/hatcheries to implement a study design such that the information produced is applicable to all locations/hatcheries rather than just those where sampling occurs. Question three provides a good illustration of why this approach might be beneficial. Theoretical and empirical literature indicates that the reproductive success of hatchery origin adults can vary from 0-100% relative to their natural origin conspecifics. Much of the effort focused on hatchery reform (at least for supplementation facilities) has been directed at identifying propagation practices that maximize the reproductive success of naturally spawning hatchery origin adults (e.g., use of varying fractions of natural origin adults as broodstock). Currently, a number of reproductive success studies are underway or have been proposed. These studies have and will provide useful information. However, the *status quo* selection of projects has some disadvantages. For example, the non-random selection of such projects could result in an abundance of information for some classes of hatcheries and very little for others (e.g., it could be that these projects are primarily implemented for programs that use only a small fraction of natural origin fish for broodstock, thus yielding limited information on the relative utility of the practice). The alternative approach that we identified towards the end of 2005 would stratify the selection of such projects to answer questions about the influence of various types of hatchery management on the relative reproductive success of hatchery origin adults. The benefit of such an approach is that we could conceivably perform the experiment in fewer locations while simultaneously generating more representative information that could predict what we might expect for programs that remain unsampled. The appeal of this approach lies in the fact that it might be cheaper in the long-run, and that it would give us some information to discuss the potential relative effectiveness of proposed programs, enable a more informed approach to hatchery reform etc.

2.5.1 Progress and challenges in development of hatchery M&E designs

The Data Quality Objectives (DQO; EPA 2006) process employed by CSMEP is an iterative method that attempts to reduce policy level information needs into a tractable study design. A natural byproduct of such a process is the iterative refinement of questions from a more general conceptual state (provided by policy directives) to a more rigorous analytical and technical state (the basis for experimental design). In short, the technical rigor of the questions is expected to improve as analytical solutions are applied to the policy directives.

Over the course of the project, CSMEP's Hatchery Subgroup has struggled to balance the obvious desire to optimize designs for the myriad questions associated with supplementation with the need to identify key uncertainties that are unlikely to be addressed by existing and proposed Research Monitoring and Evaluation (RME) projects. A number of excellent hatchery RME designs have been proposed and implemented over the past several years. Despite the participation of several researchers with exceptional hatchery RME experience, much of the CSMEP's initial hatchery design work was necessarily expended in evaluating the state of hatchery RME science. In short, we had to review existing experimental designs and evaluate successes and failures in the implementation of those designs in order to identify and prioritize where CSMEP design efforts could best be focused. Following this review, the group concluded that many of the uncertainties that accompany supplementation at the scale of individual artificial propagation projects are likely to be sufficiently addressed by proposed and ongoing RME—if those initiatives are adequately funded. However, during this process the subgroup identified a number of large scale uncertainties that will not be addressed by existing or proposed projects, or that could potentially be addressed to greater benefit by optimizing the allocation of effort. Nonetheless, arriving at these conclusions required a great deal of effort, and necessarily limited progress in FY2006 towards the construction of new designs.

Defining supplementation

The regionally accepted definition of supplementation was forwarded by the Regional Assessment of Supplementation Projects (Vogel and Clune 1992), as:

> "*Supplementation is the use of artificial propagation in the attempt to maintain or increase natural production while maintaining the long term fitness of the target population, and keeping the ecological and genetic impacts on non-target populations within specified limits.*"

In fact, the Hatchery Subgroup concluded that sufficient design work has largely been completed to address the impacts of supplementation on target populations; measured as changes in fitness, abundance, and productivity—if ongoing and proposed RME projects are funded. However, the meaningful application of this definition requires that "specified limits" for ecological and genetic impacts on non-target populations can be identified by policy and that RME projects can adequately detect when such limits are exceeded. In this respect, we found that current policy guidance and existing and proposed RME projects are largely insufficient. During the first iteration of the DQO process (see Marmorek et al. 2005 and DQO steps 1-5 presentation) the Hatchery Subgroup found that policy guidance was insufficient at providing "specified limits" for these impacts, and that the region currently lacks the information to describe biologically meaningful thresholds for such impacts.

These conclusions were not immediately apparent; in fact the Hatchery Subgroup quickly became mired in myriad questions required to assess the ecological and genetic impacts of hatchery programs at the scale of individual projects. While necessary and meaningful, this work did little to address the efficacy of artificial propagation as a class of actions aimed at maintaining, restoring, or recovering salmonid populations and harvest. The deficiencies of this approach became apparent only after the group had sufficient time to review available RME efforts. During this review it became apparent that the majority of hatchery projects employed the RASP definition of supplementation, but few identified the "specified limits" for impacts to non-target populations. Those that attempted to define "specified limits" primarily utilized seemingly arbitrary requirements for statistical measures of precision, such as coefficient of variation of the mean, more in an attempt to calculate necessary sampling effort than to identify biologically meaningful thresholds. For example, many programs attempted to identify acceptable levels of straying, often employing language suggesting that no greater than 5% of total hatchery origin adults could return to non-target locations or specifying that strays from a given program could compose no greater than 5% of the escapement in non-target populations. In general, these thresholds were derived from two sources:

1. NOAA Tech Memo NMFS-NWFSC-30 Genetic Effects of Straying of Non-Native Hatchery Fish into Natural Populations (Grant 1997) and
2. NOAA Technical Memo NMFS-NWFSC-42 Viable Salmonid Populations (McElhany et al. 2000)

While we do not dispute the technical value of these documents, the application of these criteria is immensely challenging when viewed from the scale of an individual hatchery RME project. How can an individual hatchery RME project calculate the total number of fish that stray into non-target locations? Likewise, aren't we really interested in ensuring that less that 5% of the total escapement into non-target populations is composed of hatchery strays rather than simply the strays from a single hatchery?

In FY2006, the Hatchery Subgroup finally completed the first iteration of the DQO Steps 6 and 7 process, at which time it was apparent that our focus was at the wrong spatial scale. In short, we recognized that uncertainties exist at the scale of individual artificial propagation projects, but the ongoing and proposed projects that we reviewed were likely to address these uncertainties—if adequately funded. On the other hand, the questions relating to impacts on non-target populations would remain largely unaddressed or at best could be evaluated only via weak inference from the extrapolation of results obtained at the scale of individual programs. We noted that the distribution of monitoring effort was likely insufficient to representatively address many uncertainties, limiting the broad application of the results. Hence, we endeavored to reduce some of the large scale uncertainties to a set of tractable study designs with the goal of evaluating hatcheries as a class of actions (top-down approach) rather than by simply trying to accumulate results from individual projects to address large scale questions (bottom-up approach). A series of reports over the years has called for development of large-scale evaluations of the effects of supplementation as a general restoration strategy (ISAB 2003; ISRP 2005; ISRP and ISAB 2005; NPCC 2004). A related effort has begun to address this specific use of hatchery technology (Galbreath et al. 2006; hereafter referred to as the Supplementation Workshop effort). This effort has only recently been initiated but the intent is that it will be closely coordinated with CSMEP during the FY2007-08 funding cycle.

Identification of large scale uncertainties and design opportunities

At the conceptual level, the impacts of artificial propagation on non-target populations is simply a function of direct and indirect interactions of hatchery and natural origin fish in shared environments and their net impacts on shared resources. There are obvious problems when one moves from this conceptual level into the design of a tractable study that can be implemented. The available data enable limited inference when viewing these issues at the scale of hatcheries as a class of actions, largely as a result of two issues:

1. sampling effort is allocated on the basis of individual projects and may or may not be readily combined to address uncertainties that manifest at larger spatial scales and result from the cumulative impact of all artificial propagation programs; and
2. largely as a function of the first deficiency, we have little information to evaluate the impacts of hatcheries on non-target populations, and thus very limited ability to identify biologically relevant impact thresholds.

Data relevant to the uncertainties of artificial propagation are currently generated primarily by RME projects tailored to evaluate individual programs. As such, information such as that generated by relative reproductive success studies, is gathered opportunistically where infrastructure enables appropriate sampling, individuals are motivated to produce high quality proposals that are subsequently funded, and typically where innovative hatchery practices are being employed. As such, this information, while useful, is not likely to represent the range of hatchery practices, the spatial scale of the Columbia Basin, or the diversity of species influenced by artificial propagation. Likewise, it is extremely rare for this effort to be leveraged towards non-target populations. Similarly, evaluations of stray rates, and the proportion of target and non-target populations composed of strays, are typically limited to locations where sampling infrastructure exists, thus raising obvious questions about whether the generated data are representative.

Given the limitations of current data relative to straying, non-target population composition, and relative reproductive success the hatchery subgroup endeavored to evaluate whether these questions were amenable to a large scale study design. In short, we proposed to evaluate these questions by viewing all hatchery programs as a class of actions with common questions. Viewing hatcheries from the "top-down" perspective enables the application of stratified sampling designs that may be capable of addressing these information needs for all hatcheries simultaneously while sampling activities can be restricted to a subset of the projects. As a simplistic example, one could envision allocating genetically based parentage analysis effort by employing the following strata:

1. proportion of broodstock composed of natural origin adults;
2. proportion of target population escapement composed of hatchery origin adults; and
3. duration of the program.

In addition, a number of non-target populations would be evaluated based on the average composition of hatchery strays (e.g., composed of less than 5% hatchery strays, 5%-15% hatchery strays, and greater than 15% hatchery strays on average). A stratified effort of this type would cover a large range in hatchery programs, thus programs that are not directly evaluated could use the results to bracket their expectations. By stratifying such efforts *a priori* statistical inference can potentially be greatly enhanced relative to the *status quo*.

In short, we believe that a number of the uncertainties relating to artificial propagation can be addressed most efficiently at a larger spatial scale using stratified designs. We have initiated significant design effort on two of these questions: 1) relative reproductive success; and 2) hatchery stray rates and the proportion of non-target populations composed of hatchery strays. However, these designs are far from complete. Nonetheless, we are working with the EPA to utilize EMAP based sampling to provide spatial sampling components for stratification and have begun to develop cost estimates and quantitative methods to evaluate the tradeoffs among alternative designs and differing levels of sampling effort. The Supplementation Workshop effort has similarly concluded that a system-wide stratified monitoring effort of supplementation projects and associated reference populations is likely to answer questions about the impacts of this restoration strategy on naturally reproducing populations. The Supplementation Workshop effort is expected to begin a large-scale study design this year in coordination with the CSMEP Hatchery Subgroup.

2.6 Synthesis and integration

CSMEP's Monitoring Integration Group, with representation from each of the five CSMEP subgroups, has been formed within CSMEP to explore the integration of the individual M&E component parts within a larger monitoring framework (i.e., generate improved efficiencies through integrated designs) for the Snake Basin pilot design. This integration effort across scales and subgroups, illustrated in Figure 2.6.1., is a challenge faced by all subbasins; hence the results will be of general benefit basin wide. The group has begun to develop a comprehensive matrix of shared performance measures and data interdependencies across the different CSMEP subgroups. This evolving Looking Outward Matrix (LOM) is available on the CSMEP Website. The matrix is providing a starting foundation for identifying the priority performance measures for monitoring and the relevant spatial scale(s) of these data for varied subgroup monitoring needs. CSMEP has also begun to explore how to integrate monitoring costs for the shared performance measures to achieve greater efficiencies across monitoring programs. A preliminary draft of an approach to defining and integrating monitoring costs is provided on the CSMEP website.

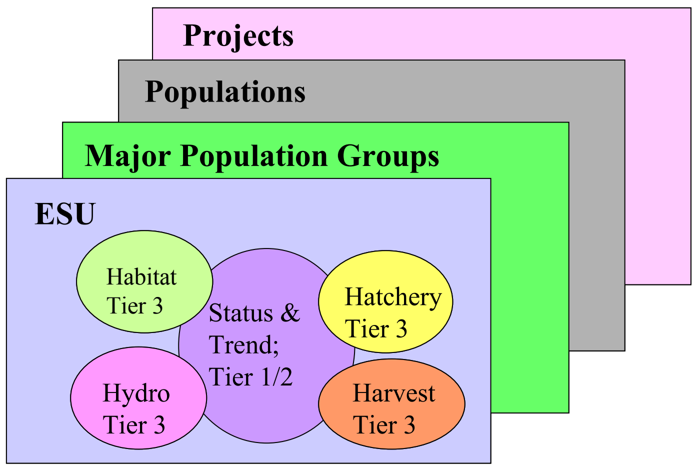


Figure 2.6.1. Integration of M&E across spatial scales and subgroups in the Snake Basin pilot study

CSMEP is also working to ensure that analyses and monitoring designs explored as part of the project are consistent with the overarching objectives of Columbia River Basin monitoring agencies. Table 2.6.1 provides a summary of CSMEP interactions with agency representatives throughout FY2006. CSMEP representatives have participated in a series of PNAMP meetings and workshops and a number of CSMEP participants are also PNAMP members. CSMEP/PNAMP convened a shared workshop in FY2006 on regional M&E with the participation of most of the key monitoring groups in the Basin. The final report (Executive Summary and full report) from this shared workshop are available on both the CSMEP and PNAMP websites.

Table 2.6.1. CSMEP programmatic and technical interactions in FY2006.

Entity	Purpose of Interaction
Pacific Northwest Aquatic Monitoring Partnership (PNAMP)	Explain CSMEP tasks, continue to refine project / program descriptions, harmonize PNAMP and CSMEP workplans. Use PNAMP as conduit to get programmatic support from above for various agencies' staff (e.g., BLM, USFS, DEQ, EPA) to assist StreamNet staff with Task 2. Undertake shared workshops to promote ideas and receive feedback.
AREMP; PIBO; OWEB	Explain CSMEP tasks; more clearly define CSMEP's role in fish habitat monitoring; obtain information on habitat monitoring for integration with our Snake Basin pilot designs
EMAP (ODFW); EPA EMAP (Corvallis)	Explain CSMEP; clarify exactly what they're doing; get inventory and design documents (or URLs) regarding habitat / fish monitoring; collaborate on EMAP designs for Snake Basin pilot areas
NOAAF – Action Agency RME Group	Explain CSMEP; clarify current status (beyond RME plan); get inventory and design documents (or URLs) regarding habitat / fish monitoring; coordinate work plans and priority M&E questions
NOAA – Pilot Projects under 35019; Chris Jordan	Explain CSMEP; clarify exactly what they're doing; get inventory and design documents (or URLs) regarding habitat / fish monitoring pertaining to watersheds of interest:; obtain information on products from RME studies in John Day (OR), Wenatchee, Methow & Okanagan (WA) ,Lemhi and Salmon (ID); contribute to pilot project designs
Technical Recovery Teams (TRTs) for the Interior and Lower Columbia, Willamette	Explain CSMEP; get input on needs of decision-makers clarify exactly what they're doing; get inventory and design documents (or URLs) regarding approaches to monitoring and recovery evaluations; obtain TRT documents and GIS products for Snake Basin design work; get input from TRT to inform S & T designs and simulation models
USFWS Bull Trout Recovery Monitoring and Evaluation Group (RMEG)	Explain CSMEP; clarify exactly what they're doing; get RMEG inventory and design documents regarding approaches to monitoring and recovery evaluations of bull trout; integrate RMEG ideas into CSMEP pilot designs for resident fish M&E
TAC	Explain CSMEP; clarify exactly what they're doing; get TAC input on CSMEP approaches to harvest M&E designs and simulation models
BiOP Remand groups	Explain CSMEP; clarify exactly what they're doing; get Remand groups' input on CSMEP approaches to M&E designs (particularly for Hydrosystem) and simulation models

Ultimately, all M&E decisions involve tradeoffs and a balancing of risks. Insufficient M&E risks repeated implementation of management actions that are actually ineffective, or else not detecting that certain actions actually are effective. Either outcome wastes money and potentially incurs increased risk to fish populations by not expending limited resources most efficiently. For example, at least $14 billion has been spent since 1990 on stream and river restoration projects across the Continental United States, yet only a small fraction of these projects have been monitored for their effects (Bernhardt et al. 2005). Conversely, unnecessary or excessive M&E wastes money that could otherwise be spent on implementing actions that are known to be effective in recovering fish populations. Decision analysis has been shown to be a powerful tool for improving the design of large-scale M&E programs (e.g., Parnell 2002, MacGregor et al. 2002, Walters and Green 1997, Keeley and Walters 1994, Peterman and Antcliffe 1993, Antcliffe 1992, McAllister and Peterman 1992a, b). These studies often show that the optimal design, when the tradeoffs between objectives and across alternatives are considered, is not necessarily the design with the highest statistical power for detecting change or trend in important indicators.

The PrOACT approach (Hammond et al 1999) is being employed by CSMEP as a simplified multi-objective decision analysis that provides a suitable framework for dealing with the large number of objectives associated with the Columbia Basin M&E issues. PrOACT is a process of ***P***roblem definition, determination of ***O***bjectives, development of ***A***lternative actions, calculation or assessment of the ***C***onsequences associated with each alternative across the set of objectives, and the evaluation of ***T***radeoffs across alternatives for particular objective, or between objectives for a particular alternative. PrOACT is an iterative process that involves cycling over the development of M&E alternatives, evaluating them, assessing tradeoffs, revising alternatives and then starting again, starting from a broad set of alternatives that gradually narrows to an acceptable choice or set of choices (Figure 2.6.2).

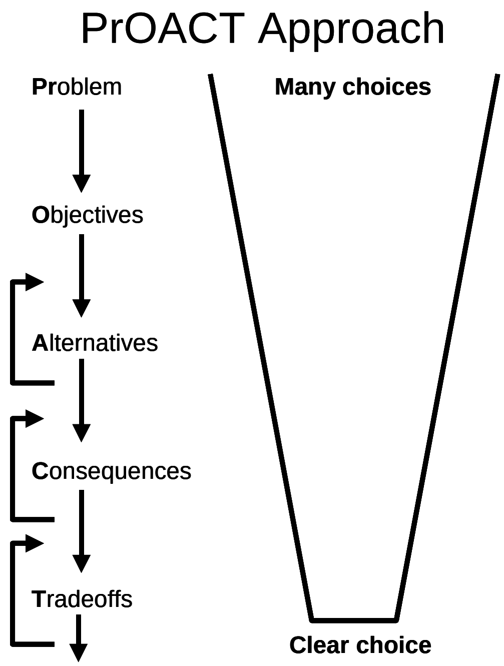


Figure 2.6.2. Flow of PrOACT process for narrowing the range of alternative choices.

CSMEP has been attempting to apply the PrOACT approach for the generation and filtering of their alternative M&E designs across the subgroups based on a suite of criteria which includes: 1) high inferential ability, 2) strong statistical performance, 3) reasonable cost, 4) practical application, and 5) environmental impact (Table 2.6.2).

Table 2.6.2. CSMEP monitoring design objectives and criteria for evaluating alternative designs.

CSMEP Design Objective	Evaluative Criteria for Design Objective
High inferential ability	- Ability to answer key monitoring questions at the appropriate scale
Strong Statistical Performance	- Precision, Statistical power, Coverage, Accuracy
Reasonable Cost	- Cost/year at scale of interest
Practical	- Feasibility of implementation, Widely applicable, Integration with existing programs
Environment:	- Environmental impact of monitoring protocol

Appendix G illustrates the template for an objectives vs. alternatives matrix that CSMEP is attempting to work through for integrated low, medium and high M&E designs in the Snake River Basin employing the criteria outlined in Table 2.6.2. CSMEP is following an approach where the base requirements for status and trends M&E will provide the foundation for low, medium and high design alternatives, while monitoring requirements for the various action effectiveness issues are to be built incrementally onto this foundation (as feasible). CSMEP has been developing (and will continue to expand in FY2007-08) a suite of analytical tools and simulation models that will allow exploration of alternative M&E designs (i.e., statistical power, costs, sampling effort, etc.) within and across the CSMEP subgroups.

3. Subbasin Inventory and Evaluation

3.1 Subbasin inventory work

During FY2006, CSMEP biologists, with the assistance of StreamNet staff, continued with detailed inventories of fish data for pilot subbasins selected in Washington (Okanagan, Methow, Kalama), Oregon (Deschutes, Grande Ronde) and Idaho (Middle Fork Salmon, Upper Fork Salmon). These subbasin inventories describe, in a systematic manner, the kinds of information currently available on the abundance, productivity, spatial distribution and diversity of salmon, steelhead and resident fish species of concern. Efforts in FY2006 have also been focused on completing inventories for remaining subareas and fish species within the Snake Basin Pilot study area not yet captured by CSMEP efforts in FY 2004-05.

3.2 Strengths and weaknesses analyses

Throughout FY2006 CSMEP biologists continued with their evaluations of the strengths and weaknesses of pilot subbasin fish inventory data for addressing the CSMEP Tier 1, 2 and 3 monitoring questions The strengths and weaknesses reviews completed to date (Table 3.1) are identifying areas where fish monitoring is currently being done well, in addition to uncovering inferential weaknesses and data gaps that will be important to address in CSMEP's monitoring design work. A synthesis framework for evaluating strengths and weaknesses across the pilot subbasins has been expanded in FY2006 (i.e., are there strengths and weaknesses in regards to monitoring of particular performance measures for particular species that are common *across* the subbasins?)

Table 3.1. Data strengths and weaknesses analyses completed through FY2006 by subbasin and species (hyperlinked to the Table B2 summaries on the CSMEP website).

State	Subbasin	Species
Idaho	South Fork Salmon River	spring/summer Chinook
	Clearwater, Selway River	Chinook (spring, summer) steelhead (summer) bull trout
Oregon	Imhaha	Chinook (spring) steelhead (summer)
	Lower Columbia	fall Chinook
	Deschutes	steelhead
Washington	Lewis	Chinook (spring, tule and bright fall) steelhead (summer, winter)
	Yakima	coho fall Chinook spring Chinook steelhead (summer)
	Methow	Chinook (spring, summer) steelhead (summer)

3.3 CSMEP public website and web data application

CSMEP spent considerable effort in FY2006 improving (with the assistance of the CBFWA web designer) the user interface of their publicly accessible website. There are now more interesting subgroup topic front-ends and much cleaner, easier and more intuitive access to the multitude of CSMEP products (i.e., analyses, reviews, presentations, reports) being developed within the project.

CSMEP (with StreamNet's assistance) also continued to expand their centralized web-based data application (managed originally by ODFW StreamNet but now being transferred to the regional StreamNet office in Portland) to store and allow access to CSMEP inventory metadata and data assessments. As of the end of FY2006 there were over 1500 metadata records stored on the CSMEP data server.

References

Al-Chokhachy, R., P. Budy and H. Schaller. 2005. Understanding the significance of redd counts: a comparison between two methods for estimating the abundance of and monitoring bull trout populations. North American Journal of Fisheries Management 25:1505-1512.

Al-Chokhachy, R. 2006. An assessment of the dynamics and limiting factors of a fluvial bull trout population. PhD Dissertation. Utah State University, Logan, UT.

Al-Chokhachy, R. and P. Budy. In review. Summer microhabitat use of fluvial bull trout in Eastern Oregon streams. North American Journal of Fisheries Management.

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

Antcliffe, B.L. 1992. Impact assessment and environmental monitoring: the role of statistical power and decisions analysis, Master's thesis, School of Resource and Environmental Management, Simon Fraser University, Burnaby, British Columbia. 166 p.

Ashbrook, C.E., K.W. Yi., J.F. Dixon, A. Hoffmann and G.E. Vander Haegen. 2004. Evaluation of live capture selective harvest methods: 2002. Annual report for BPA contract 2001-007-00. Objective 1.

Bahr, M.A., and J.M. Shrimpton. 2004. Spatial and quantitative patterns of movement in large bull trout (Salvelinus confluentus) from a watershed in north-western British Columbia, Canada re due to habitat selection and not differences in life history. Ecology of Freshwater Fish 13: 294-304.

Benichou, J. and M.H. Gail. 1989. A Delta method for implicitly defined random variables. The American Statistician 43: 41-44.

Berggren, T. , H. Franzoni, L. Basham, P. Wilson, H. Schaller, C. Petrosky, E. Weber and R. Boyce. 2005. Comparative Survival Study (CSS) of PIT-tagged Spring/Summer Chinook and PIT-Tagged Steelhead; Mark/Recapture Activities and Bootstrap Analysis, 2005 Annual Report. Project No. 199602000, 155 pages, Dec. 1, 2005.

Bernhardt, E.S., M.A. Palmer, J.D. Allan, G. Alexander, K. Barnas, S. Brooks, J. Carr, S. Clayton, C. Dahm, J. Follstad-Shah, D. Galat, S. Gloss, P. Goodwin, D. Hart, B. Hassett, R. Jenkinson, S. Katz, G.M. Kondolf, P.S. Lake, R. Lave, J. L.Meyer and T.K. O'Don. 2005. Synthesizing U.S. River Restoration Efforts. Science 308: 636-637.

Bisson, P.A., J.L. Nielsen, R.A. Palmason and L.E. Grove. 1981. A system of naming habitat in small streams, with examples of habitat utilization by salmonids during low streamflow. In N.B. Armantrout ed. Acquisition and Utilization of Aquatic Habitat Inventory Information. Proceedings of a symposium, Oct. 28-30, 1981, Portland, Oregon. Hagen Publishing Co., Billings, Montana. pp. 62-73.

Box, G. E. P., and G. C. Tiao. 1975. Intervention analysis with applications to economic and environmental problems. J. of the American Statistical Association 70:70-79.

Bradford, M.J., J. Korman and P.S. Higgins. 2005. Using confidence intervals to estimate the response of salmon populations (Oncorhynchus spp.) to experimental habitat alterations. Canadian Journal of Fisheries and Aquatic Sciences 62: 2716–2726.

Burnham, K.P., and D.R. Anderson. 1998. Model selection and multimodel inference: a practical information-theroretic approach. Second edition. Spinger-Verlag, New York, NY.

Carmichael, R.W. 2005. Recovery plan for Oregon's Middle Columbia River Steelhead. Progress Report (Draft). Oregon Department of Fish and Wildlife, La Grande OR, December 27, 2005.

Carpenter, S.R., T.M. Frost, D. Heisey and T.K. Kratz. 1989. Randomized intervention analysis and the interpretation of whole-ecosystems. Ecology 70:1142-1152.

Cloutman, D.G., and D.C. Jackson. 2003. A comparison of randomized intervention analysis and two-factor analysis of variance for analysis of angler effort in a floodplain river fishery. Fisheries Management and Ecology, 2003, 10, 173–17.

Catalani, M. 2002. Sampling from a couple of positively correlated beta variates. http://arxiv.org/abs/math.PR/0209090

Cooch, E., and G. White. 2006. Program MARK: "A gentle introduction", 5th Edition.

CSMEP (Collaborative Systemwide Monitoring and Evaluation Project). 2004. Annual Report for FY 2004. Prepared by ESSA Technologies Ltd., Vancouver, B.C. on behalf of the Columbia Basin Fish and Wildlife Authority, Portland, OR. http://www.cbfwa.org/Committees/CSMEP/Documents/Project/CSMEPFY04Annual%20Report013105.doc

CSMEP (Collaborative Systemwide Monitoring and Evaluation Project). 2005. Annual Report for FY 2005. Prepared by ESSA Technologies Ltd., Vancouver, B.C. on behalf of the Columbia Basin Fish and Wildlife Authority, Portland, OR.

CSMEP Hydro subgroup. 2005. Draft document for steps 6 and 7 of the DQO process.

Downes, B.J., L.A. Barmuta, P.G. Fairweather, D.P. Faith, M.J. Keough, P.S. Lake, B.D. Mapstone and G.P. Quinn. 2002. Monitoring ecological impacts: concepts and practice in flowing waters. Cambridge University Press, Cambridge, UK.

Dunham, J.B., B.E. Rieman and K. Davis. 2001. Sources and magnitude of sampling error in redd counts for bull trout. North American Journal of Fisheries Management. 21:343-352.

Dunham, J.B., and B.E. Rieman. 1999. Metapopulation structure of bull trout: influences physical, biotic, and geometrical landscape characteristics. Ecological Applications 9: 642-655.

Ebersole, J.L., W.J. Liss and C.A. Frissell. 2001. Relationship between stream temperature, thermal refugia and rainbow trout Oncorhynchus mykiss abundance in arid-land streams in the northwestern United States. Ecology of Freshwater Fish 10: 1–10.

Fraley, J.J., and B.B. Shepard. 1989. Life history, ecology, and population status of migratory bull trout (Salvelinus confluentus) in the Flathead Lake and River system, Montana. Northwest Science. 63: 133-143.

Galbreath, P.F., P.B. Roger and M.J. Ford. 2006. Report of the Supplementation Monitoring and Evaluation Workshop. April 6-7, 2006 Columbia River Inter-Tribal Fish Commission. Portland OR. 26pp.

Goetz, F. 1989. Biology of the Bull Trout, Salvelinus confluentus, a Literature Review. U.S. Department of Agriculture, Forest Service. Willamette National Forest, Oregon.

Good, T.P., R.S. Waples and P. Adams (editors). 2005. Updated status of federally listed ESUs of West Coast salmon and steelhead. U.S. Dept. Commer, NOAA Tech. Memo. NMFS-NWFSC-66, 598 p.

Gould, W.R., and J.D. Nichols. 1998. The estimation of temporal variability of survival in animal populations. Ecology 79: 2531-2538.

Grant, S.W. (Editor). 1997. Genetic effects of straying of non-native hatchery fish into natural populations: Proceedings of the workshop. U.S. Department of Commerce. NOAA Technical Memo. NMFS-NWFSC-30, 130 p.

Green, B.F. 1979. Sampling design and statistical methods for environmental biologist. Wiley Interscience, New York, New York, USA.

Ham, K.D., and T.N. Pearsons. 2000. Can reduced salmonid population abundance be detected in time to limit management impacts? Canadian Journal of Fisheries and Aquatic Sciences 57:17-24.

Hammond, J.S., R.L. Keeney and H. Raiffa. 1999. Smart Choices: a practical guide to making better decisions. Harvard Business School Press. Boston, Massachusetts.

Hemmingsen, A.R., S.L. Gunckel and P.J. Howell. 2001. Bull trout life history, genetics, habitat needs, and limiting factors in Central and Northeast Oregon. Bonneville Power Administration. Project No. 1994-05400.

Hicks, B.J., J.D. Hall, P.A. Bisson and J.R. Sedell. 1991. Responses of salmonids to habitat changes. Pages 483–518 in W. R. Meehan, editor. Influences of forest and rangeland management on salmonid fishes and their habitats. American Fisheries Society, Special Publication 19, Bethesda, Maryland.

Hilborn, R., and M. Mangel. 1997. The ecological detective: confronting models with data. Mongr. Pop. Biol. No. 28. Princeton University Press, Princeton, N.J.

Hillman, T.W. 2004. Monitoring strategy for the Upper Columbia basin. Draft Report. Prepared by Tracy W. Hillman, BioAnalysts, Inc. Eagle, Idaho. Prepared for Upper Columbia Regional Technical Team, Upper Columbia Salmon Recovery Board, Wenatchee, Washington. February 1, 2004. http://www.cbfwa.org/Committees/RTT/Documents/Reports/020104UCBmonitoringStrategy.doc

Homel, K. 2006. Evaluating genetic population structure and movement patterns of bull trout (Salvelinus confluentus) in Northeast Oregon. Master's Thesis. Utah State University, Logan, UT.

Hyun, S., S. Ellis and P. Roger. 2006. Preseason forecasts of ocean escapements of Columbia River Chinook salmon populations. Pacific Salmon Commission Southern Boundary Project. Final report.

ICTRT (Interior Columbia Basin Technical Recovery Team). 2005. Interior Columbia Basin TRT: viability criteria for application to Interior Columbia Basin Salmonid ESUs. Draft. July, 2005. *URL* http://www.nwfsc.noaa.gov/trt/col_docs/viabilityupdatememo.pdf

Independent Scientific Review Panel (ISRP). 2005. Retrospective Report 1997-2005. Northwest Power and Conservation Council, Portland, Oregon. Report Number 2005-14 dated August 31, 2005.

Independent Scientific Advisory Board (ISAB). 2003. A Review of Salmon and Steelhead Supplementation. ISAB 2003-3. Northwest Power and Conservation Council, Portland, Oregon.

Independent Scientific Advisory Board (ISAB). 2005. Report on Harvest Management of Columbia Basin Salmon and Steelhead. ISAB 2005-4. Northwest Power and Conservation Council, Portland, Oregon.

Independent Scientific Review Panel and Independent Scientific Advisory Board (ISRP/ISAB). 2005a. Review of the Council's draft Columbia River Basin Research Plan (November version). ISRP/ISAB 2005-20. Northwest Power and Conservation Council, Portland, Oregon.

Independent Scientific Review Panel and Independent Scientific Advisory Board (ISRP/ISAB). 2005b. Monitoring and Evaluation of Supplementation Projects. ISRP/ISAB 2005-15. Northwest Power and Conservation Council, Portland, Oregon. http://www.nwcouncil.org/library/isrp/ isrpisab2005-15.pdf

Jacobs, S.E., and T.E. Nickelson. 1999. Use of stratified random sampling to estimate the abundance of Oregon coastal coho salmon. Oregon Department of Fish and Wildlife, Portland, OR.

Jain, S., M. Hoerling and J. Esicheid. 2005. Decreasing reliability and increasing synchronicity of western North American streamflow. Journal of Climate 18: 613-618.

Johnson, N. L., S. Kotz, and N. Balakrishnan. 1994. Continuous univariate distributions, volume 1. 2nd edition. John Wiley & Sons, Inc.

Jones, E.L., T.J. Quinn and B.W. Van Alen. 1998. Observer accuracy and precision in aerial and foot survey counts of Pink Salmon in a southeast Alaska stream. North American Journal of Fisheries Management 18: 832-846.

Jordan C., J. Geiselman, M. Newsome, J. Athearn. (eds.). 2003. Draft Research, Monitoring & Evaluation Plan for the NOAA-Fisheries 2000 Federal Columbia River Power System Biological Opinion. 252 pp.

Keeley, E.R., and C.J. Walters. 1994. The British Columbia watershed restoration program: summary of the experimental design, monitoring, and restoration techniques workshop. 1, British Columbia Ministry of Environment, Lands and Parks, and Ministry of Forests, Vancouver, BC

Kendall, B.E. 1998. Estimating the magnitude of environmental stochasticity in survivorship data. Ecological Applications 8:184–193.

Keough, M.J., and B.D. Mapstone. 1995. Protocols for designing marine ecological monitoring programs associated with BEK mills. Technical Report No. 11. CSIRO, Canberra, ACT.

Korman, J. and P.S. Higgins. 1997. Utility of escapement time series data for monitoring the response of salmon populations to habitat alteration. Can. J. Fish. Aquat. Sci. 54:2058-2097.

Kershner, J.L, B.B. Roper, N. Bouwes, R. Henderson and E. Archer. 2004. An analysis of stream habitat conditions in reference and managed watersheds on some federal lands within the Columbia basin. North American Journal of Fisheries Management 24:1363-1375.

Leary, R.F., F.W. Allendorf and S.H. Forbes. 1993. Conservation genetics of bull trout in the Columbia and Klamath River drainages. Conservation Biology 7: 856-865.

MacDonald, L.H., A. Smart and R.C. Wissmar. 1991. Monitoring guidelines to evaluate effects of forestry activities on streams in the Pacific Northwest and Alaska. EPA/910/9-91-001, U.S. Environmental Protection Agency Region 10, Seattle, WA. 166 pp.

MacGregor, B.W., R.M. Peterman, B.J. Pyper and M.J. Bradford. 2002. A decision analysis framework for comparing experimental designs of projects to enhance Pacific salmon. North Am. J. Fish. Manag. 22: 509-527.

Marmorek, D.R., I.J. Parnell, M. Porter, C. Pinkham, C.A.D. Alexander, C.N. Peters, J. Hubble, C.M. Paulsen and T.R. Fisher. 2004. A Multiple Watershed Approach to Assessing the Effects of Habitat Restoration Actions on Anadromous and Resident Fish Populations. Prepared by ESSA Technologies Ltd., Vancouver, B.C. for Bonneville Power Administration, Portland, OR. 420 pp. http://www.essa.com/watersh-rpt.pdf

Marmorek, D.R., M. Porter , I.J. Parnell and C. Peters, eds. 2004b. Comparative Survival Study Workshop, February 11–13, 2004; Bonneville Hot Springs Resort. Report compiled and edited by ESSA Technologies Ltd., Vancouver, B.C. for Fish Passage Center, Portland, OR and the US Fish and Wildlife Service, Vancouver, WA. 137 pp.

Marmorek, D.R., I. Parnell and M. Porter (eds.). 2005. CSMEP DQO Status and Trends, Habitat, Harvest, Hatchery and Hydrosystem Assessments - Policy Interpretations (Snake River Pilot). Prepared for participants at the CSMEP Monitoring Design Workshop, April 12-14, 2005. Nampa, Idaho.

Maxell, B.A. 1999. A prospective power analysis on the monitoring of bull trout stocks using redd counts. North American Journal of Fisheries Management 19: 860-866.

McAllister, M.K., and R.M. Peterman. 1992a. Decision analysis of a large-scale fishing experiment designed to test for a genetic effect of size-selective fishing on British Columbia pink salmon (Oncorhynchus gorbuscha). Can. J. Fish. Aquat. Sci. 49: 1305-1314.

McAllister, M.K., and R.M. Peterman. 1992b. Experimental design in the management of fisheries: A review. North Am. J. Fish. Manag. 12: 1-18.

McElhany, P., M.H. Ruckelshaus, M.J. Ford, T.C. Wainwright E.P. Bjorkstedt. 2000. Viable salmonid populations and the recovery of evolutionarily significant units. Technical Memorandum NMFS-NWFSC-42. United States Department of Commerce, National Oceanic and Atmospheric Administration. Seattle, WA.

Milliken, G.A., and D.E. Johnson. 2001. Analysis of messy data, Volume 785 III: Analysis of Co-variance. CRC Press. Boca Raton, Florida.

Morris, W.F., and D.F. Doak. 2002. Quantitative Conservation Biology. Sinauer Associates, Sunderland, Massachusetts. 480 pp.

Muhlfeld, C.C., and B. Marotz. 2005. Seasonal movement and habitat use by subadult bull trout in the Upper Flathead River System, Montana. North American Journal of Fisheries Management 25: 797-810.

Nelson, M.L., T.E. McMahon and R.F. Thurow. 2002. Decline of the migratory form in bull charr; Salvelinus confluentus, and implications for conservation. Environmental Biology of Fishes 64: 321-332.

NWPCC (Northwest Power and Conservation Council). 2005. Northwest Power and Conservation Council draft Columbia River Basin Research Plan.

Parnell, I.J. 2002. Use of decision analysis to design a habitat restoration experiment. Master's thesis, School of Resource and Environmental Management, Simon Fraser University, Burnaby, BC. 69 p.

Parties to U. S. v Oregon. 2005. 2005-2007 Interim management agreement for upriver Chinook, steelhead, coho, and white sturgeon. Columbia River Inter-Tribal Fish Commission. Portland, Oregon. 60 pages.

PATH (Plan for Analyzing and Testing Hypotheses). 2000. Preliminary Evaluation of the Learning Opportunities and Biological Consequences of Monitoring and Experimental Management Actions.

Paulsen, C.M., and T.R. Fisher. 2005. Do habitat actions increase juvenile survival? An information-theoretic approach applied to endangered Snake River Chinook salmon. Trans. Am. Fish. Soc 143: 68-85.

Peterman, R.M., and B.L. Antcliffe. 1993. A framework for designing aquatic monitoring programs. Pages 227-251 in A.P. Farrell, editor. The aquatic resources research project: towards environmental risk assessment and management of the Fraser River Basin. Technical Report for the British Columbia Ministry of Environment, Centre for Excellence in Environmental Research, Simon Fraser University, B.C.

Pham-Gia, T. 2000. Distributions of the ratios of independent beta variables and applications. Comm. Statist. Theory Methods 29 (12): 2693–2715.

Poole, A., M. West and J. Harrison. 1994. Applied Bayesian Forecasting and Time Series Analysis, Chapman and Hall, New York, NY.

Porter M. and D.R. Marmorek. 2004. Bull Trout Recovery Monitoring and Evaluation Technical Working Group (RMEG) Workshop 2 (July 12 -13, 2004) - Boise Aquatic Sciences Lab, US Forest Service, Boise, ID - Workshop Report. Report prepared by ESSA Technologies Ltd., Vancouver, BC for the US Fish and Wildlife Service, Vancouver, WA. 75 pp.

Porter M. and D.R. Marmorek. 2005. Bull Trout Recovery Monitoring and Evaluation Technical Working Group (RMEG) Workshop 3 (Dec. 16 - 17, 2004) – USFS Regional Office, 333 SW First Ave., Portland, Oregon - Workshop Report. Report prepared by ESSA Technologies Ltd., Vancouver, BC for the US Fish and Wildlife Service, Vancouver, WA. 113 pp.

Rao, C.R. 1973. Linear statistical inference and its applications (2nd ed.), New York: John Wiley.

Rieman, B.E., and J.D. McIntyre. 1993. Demographic and habitat requirements for conservation of bull trout. USDA General Technical Report INT-302.

Rieman, B.E., and J.D. McIntyre. 1995. Occurrence of bull trout in naturally fragmented habitat patches of varied size. Transactions of the American Fisheries Society 124: 285-296.

Rieman, B.E., D.C. Lee and R.F. Thurow. 1997. Distribution, status, and likely future trends of bull trout within the Columbia River and Klamath basins. North American Journal of Fisheries Management 17:1111-1125.

Rieman, B.E., and G.L. Chandler. 1999. Empirical evaluation of temperature effects on bull trout distribution in the northwest. Final Report to: U.S. EPA, Contract 12957242-01-0, Boise, Idaho.

Roni, P., T.J. Beechie, R.E. Bilby, F.E. Leonetti, M.M. Pollock and G.P. Pess. 2003. A review of stream restoration techniques and a hierarchical strategy for prioritizing restoration in the Pacific Northwest watersheds. North American Journal of Fisheries Management 22:1-20.

Roni, P., M. Liermann, C.E. Jordan and E.A. Steel. 2005. Steps for designing a monitoring and evaluation program for aquatic restoration. Pages 13-34 in Roni, P. (Ed.) Methods for monitoring stream and watershed restoration. American Fisheries Society, Bethesda, Maryland.

Selong, J.H., T.E. McMahon, A.V. Zale and F.T. Barrows. 2001. Effect of temperature on growth and survival of bull trout with application of an improved method for determining thermal tolerance in fishes. Transactions of the American Fisheries Society 130: 1026-1037.

Sharma, R., A.B. Cooper and R. Hilborn. 2005. A quantitative framework for the analysis of habitat and hatchery practices on Pacific salmon. Ecological Modeling. 183: 231-250.

Skalski, J.R., J. Lady, S.G. Smith, 2003. Decision rules for progress and compliance testing for smolt passage survival. In Jordan et al. (2003). www.salmonrecovery.gov/RME_Plan_09-2003.pdf, pp. 189-206

SRFB (Salmon Recovery Funding Board). 2002. The Washington comprehensive monitoring strategy and action plan for watershed health and salmon recovery. Monitoring Oversight Committee.

Stevens, D.L., and A.R. Olsen. 2004. Spatially balanced sampling of natural resources. Journal of the American Statistical Association 99: 262-278.

Stewart-Oaten, A., W.W. Murdoch and K.R. Parker. 1986. Environmental impact assessment: “pseudoreplication” in time? Ecology 67:929–940.

Stewart-Oaten, A., and J.R. Bence. 2001. Temporal and spatial variation in environmental impact assessments. Ecological Monographs 71:305-339

Theriault, V., and J.J. Dodson. 2003. Body size and the adoption of a migratory tactic in brook charr. Journal of Fish Biology 63: 1144-1159.

Thompson, W.L., and D.C. Lee. 2002. A two-stage information-theoretic approach to modeling landscape-level attributes and maximum recruitment of Chinook salmon in the Columbia River Basin. Natural Resource Modeling 15:227-257.

Thurow, R.F. 1997. Habitat utilization and diel behavior of juvenile bull trout (Salvelinus confluentus) at the onset of winter. Ecology of Freshwater Fish 6: 1-7.

Thurow, R.F., J.T. Peterson, C.A. Larson and J.W. Guzevich. 2004. Development of bull trout sampling efficiency models. Final report. USFWS Agreement #134100-2-H001.

Thurow, R.F., J.T. Peterson and J.W. Guzevich. 2006. Utility and validation of day and night snorkel counts for estimating bull trout abundance in first-to-third-order streams. North American Journal of Fisheries Management 26: 217:232

Underwood, A.J. 1991. Beyond BACI: experimental designs for detecting human environmental impacts on temporal variations in natural populations, Australian Journal of Marine and Freshwater Research 42, 569–587.

Underwood, A.J. 1992. Beyond BACI: the detection of environmental impacts on populations in the real, but variable, world, Journal of Experimental Marine Biology and Ecology 161, 145–178.

Underwood, A.J. 1994. On beyond BACI: sampling designs that reliably detect environmental disturbances. Ecological Applications 4:3-15.

US Environmental Protection Agency (EPA). 1994. Guidance for the Data Quality Objectives Process. EPAWA/G-4. http://www.hanford.gov/dqo/project/level4/Epaqag4.pdf

US Environmental Protection Agency (EPA). 2000. Guidance on Systematic Planning Using the Data Quality Objectives Process. EPAQA/G-4. http://www.epa.gov/quality1/qs-docs/g4-final.pdf

United States Fish and Wildlife Service (USFWS). 2002. Bull Trout (Salvelinus confluentus) Draft Recovery Plan. Portland, Oregon.

Vogel, T., and T. J. Clune. 1992. Project Managers, U. S. Department of Energy, Bonneville Power Administration, Division of Fish and Wildlife, Project Number 1985-62, Contract Number DE-AC06-75RL01830, 51 electronic pages (BPA Report DOE/BP-01830-12). http://www.efw.bpa.gov/publications/P01830-12.pdf

Walters, C.J., J.S. Collie and T.Webb. 1988. Experimental designs for estimating transient responses to management disturbances. Canadian Journal of Fisheries and Aquatic Sciences 45:530–538.

Walters, C.J., and R. Green. 1997. Valuation of experimental management options for ecological systems. J. Wildl. Manage. 61: 987-1006.

Appendix A.
Summary of CSMEP Questions[6]

(Used to guide both assessments of the strengths and weaknesses of existing data and the development of robust monitoring designs)

1. Broadscale Fish Distribution and Ecosystem Status

- What is the distribution of adult salmonid fishes across broad regions?
- What is the ecosystem status for Columbia River Basin (CRB) fish populations?

2. Fish Population and Habitat Status and Trends

- What is the size, annualized growth rate, freshwater productivity, age-structure of CRB fish populations?
- How frequently do resident fish spawn, and what life history types make up different populations?
- What is the fraction of potential natural spawners that are of hatchery origin?
- What are the physical habitat condition, biological condition and chemical water quality of CRB fish spawning and rearing habitat?
- Have listed CRB populations recovered sufficiently for delisting and removal of ESA restrictions?

3. Action Effectiveness of Specific Recovery Actions (habitat, hydro, hatchery, or harvest management)

HABITAT

- Have specific habitat projects affected habitat conditions and local fish population survival, abundance or condition?
- Did groups of habitat projects within a subpopulation or sub watershed on aggregate affect fish survival, abundance or condition in a larger demographic unit?
- Are particular classes of habitat projects effective?[7]
- What are the mechanistic connections between habitat actions and fish population responses?
- Have habitat projects achieved the expected improvements in conditions?

HARVEST

- What are the inseason estimates of run size and escapement for each management group and how do they compare to preseason estimates?
- What is the target and nontarget harvest and when is it projected to reach allowable levels?

HATCHERIES

- To what extent can hatcheries be used to assist in meeting harvest management goals while keeping impacts to natural populations within acceptable limits?
- To what extent can hatcheries be used to enhance viability of natural populations while keeping impacts to non-target populations within acceptable limits?

---

[6] The questions span 3 tiers, as defined in Jordan et al. (2002): Tier 1 - broad-scale assessment of fish distributions at a sampling frequency of about 3 to 5 years, and a general assessment of ecosystem status at a sampling frequency of about 5 to 10 years. Tier 2 - statistically based sampling to determine the annual trends in the status of fish populations and their habitat. Tier 3 - research and monitoring to assess, in the form of explicitly posed experiments, the effectiveness of specific recovery actions.

[7] The effects of classes of habitat actions on fish habitat can be evaluated with reach-scale assessments of habitat performance measures. At the scale of a demographic unit however (e.g., a fish population), there are generally several classes of actions being implemented concurrently. Thus in many cases it may not be feasible to isolate the effects of particular classes of habitat actions on fish survival or abundance at the population scale. Even assessing the effects of groups of habitat actions on populations will require a greatly increased degree of regional coordination within and among subbasins in the timing and location of restoration project implementation (Marmorek et al. 2004).

- To what extent can hatcheries be used to conserve the genetic legacy of imperiled fish populations?

HYDROSYSTEM

- Are smolt-to-adult survival rates (SARs) sufficiently high to meet NPCC and recovery goals?
- Has hydrosystem complied with performance standards set out in 2000 FCRPS BiOp?
- What are the patterns in fish survival rates both within the mainstem and subsequent to it, for different species and groups of fish (e.g., transported vs. in-river, hatchery vs. wild, upstream vs. lower river)?
- What's the effect of different within-season transportation management and flow/spill management actions on various measures of fish survival rates?
- To what extent would Removable Spillway Weirs improve fish survival rates, at both the project scale and over the overall life cycle?

For each of the above questions, CSMEP biologists are addressing the following five issues:

1. What are the spatial scales of interest for this question?
2. Has anyone attempted to answer this question before in this sub-basin, or for a larger spatial unit that contains this sub-basin? If Yes, who did this, and how? What methods were used? Provide reference citation. Was accuracy or precision of answer estimated?
3. If answer to #2 was no (or attempt failed), could question be answered with available data? (yes, no, maybe, don't know). Any ideas on how / method? At what level of accuracy AND precision, ideally with quantitative estimates, or if not available qualitative estimates (L, M, H).
4. On what spatial scale <u>could</u> answers be provided with existing information (e.g., tribs, individual pop, pop group, ESU) and over what temporal scale (e.g., last 20 years, last 5 years)?
5. Summarize the overall strengths and weaknesses of existing data for answering this question. What critical improvements are required to overcome weaknesses

Appendix B.
Summary of CSMEP Survey of Regional Entities' Relative Priorities for Addressing Different Management Questions

CSMEP Presentation

Purpose:
To obtain information from policy-level personnel on the relative importance of monitoring questions across various spatial scales for six listed stocks of focus in CSMEP: spring/summer Chinook salmon, fall Chinook salmon, steelhead trout, sockeye salmon, coho salmon, and bull trout. CSMEP scientists will use this information obtained from this survey to guide its work on monitoring designs.

Elements of the Survey:

- Identify priorities in data needs for monitoring the status and trends of listed fishes, as well as the effectiveness of habitat, harvest, hatchery and hydrosystem actions.
- Respondents were asked to rate the importance of 27 monitoring questions.
- For any given question, it was expected that its importance would vary across respondents, species, and spatial scales
- Species of interest included: spring/summer Chinook salmon, fall Chinook salmon, steelhead trout, sockeye salmon, coho salmon, and bull trout
- Spatial scales of interest included: (1) sub-population, 2) population, 3) major population group (MPG), 4) evolutionary significant unit (ESU) or distinct population segment (DPS), and 5) Columbia basin).

Respondents:

- Idaho Department of Fish and Game
- Montana Fish, Wildlife, and Parks
- Oregon Department of Fish and Wildlife
- Washington Department of Fish and Wildlife
- U.S. Fish and Wildlife Service
- Coeur d'Alene Tribe
- Colville Tribe
- Nez Perce Tribe
- Confederated Tribes of the Warm Springs Reservation
- Confederated Tribes and Bands of the Yakama Nation.

Results: (full text of report available on the CSMEP website)

Rating of Tier 1 and 2 monitoring questions for status and trends of fishes across spatial scales.

- For the anadromous species, questions regarding broad scale status and trends were rated as very important, particularly at scales of population, MPG, and ESU.
- Variation in responses for the population and ESU scales were less than for sub-population and MPG scales.
- For anadromous species, questions regarding habitat were rated as less important. Similarly, the question regarding life history types was rated as less important.
- For bull trout, questions regarding broad scale status and trends were rated as very important. Questions about the timing of resident species spawning and life history types were also rated as important, across most spatial scales. Questions regarding habitat were rated as important, particularly at the sub-population and population scales. In general for bull trout, the population scale was rated as most important, but the sub-population, MPG, and ESU/DPS scales also were rated as important. Variability in response was great for bull trout than for the anadromous species.

Rating of Tier 3 questions for action effectiveness monitoring across spatial scales.

- For anadromous species, all questions regarding effectiveness of actions for habitat, harvest, hatcheries, and they hydrosystem were deemed important at one or more scales.
- Hatchery questions were rated as highly important at the population, MPG, and ESU scales.
- Habitat questions were most important at smaller (sub-population and population) scales, except for sockeye salmon.
- For anadromous species, the basin scale became important for harvest questions, while hydrosystem questions were most important at the ESU scale.
- In general, variability among respondents was similar for Tier 3 questions as for Tier 1 and 2 questions, with the exception that hydrosystem questions showed high variability in rated importance.
- For bull trout, respondents rated habitat questions as most important, particularly at sub-population and population scales. Harvest questions were rated as the next most important. Hatchery questions and hydrosystem questions were less important to most, but not all, agencies. There was high variability in ratings for questions regarding the effectiveness of harvest, hatchery, and hydrosystem actions; variability in ratings for habitat action effectiveness was lower.

Some key comments from respondents:

- Several tribes would not rate species with different levels of importance as a matter of tribal policy.
- The survey omitted some anadromous and native fish species that some respondents consider as very important. Respondents named chum, lamprey, sturgeon, coastal cutthroat trout, and, in general, “native species”.

Implications of the survey and emerging priorities

- Respondents placed high importance on questions that address the status and trends of fish. In general, highest priority was placed on answering these questions at the population scale, though the MPG and ESU scales also were rated as important. These results imply that development of a coordinated, consistent, basin-wide approach is appropriate for status and trends monitoring.
- There was substantial variability in the importance ratings for questions pertaining to action effectiveness monitoring. This variability likely reflects the diverse mandates of different agencies and tribes. In addition, variability in response probably reflects regionally varying stressors on aquatic systems. This implies that action effectiveness monitoring will need to vary regionally to reflect the diversity in priorities among agencies and tribes, as well as the diversity in regional stressors and aquatic ecosystem conditions.

Appendix C.
TRT Viability Rules

The population scale viability status is determined by the risk level of: a) the abundance and productivity (AP) and b) the spatial structure and diversity (SSD). The AP risk level is determined by calculating the geometric mean and variance of the abundance over the past 10 years and the geometric mean of the productivity over the past 20 years. The point estimate (abundance, productivity) with error bars is then overlaid on the TRT defined viability curves and the AP risk level is determined by the location of the estimate and error bars on the viability curve plot (Figure C1).

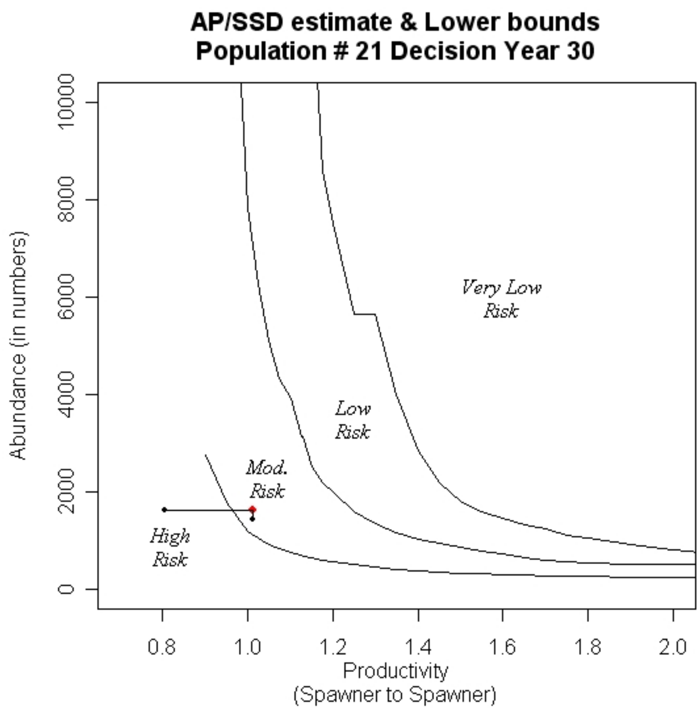


Figure C1. Population viability curves and abundance/productivity risk levels.

The spatial structure and diversity for a population is summarized by 12 metrics defined by the TRT. Each metric considers a different aspect of the SSD information from distribution of fish to genetic composition (Table 12 of TRT). For each population each of the 12 metrics is assessed a risk level (VL, L, M, H) and these 12 risk levels are then combined into a single integrated population score for each population. The viability status of the population is defined as either: not viable (NV), maintained (M), viable (V), or highly viable (HV) based on the AP and SSD risk levels (Table 13 of TRT). The major population group (MPG) is then defined as viable (V) or not viable (NV) based on the viability status of the combination of populations in the MPG. The Snake River Spring/Summer Chinook Salmon ESU is considered viable only if all of the MPGs are viable (Pete Hassemar-TRT, TRT viability criteria).

Appendix D.
Towards Integrated Status Quo, Low, Medium, High Designs for CSMEP Status & Trends and Hydrosystem Questions

Status & Trend	Spawner-Recruit Information				PIT-tag information (links to subgroup needs[8]) and other data required to answer question			
	Status Quo	Low	Medium	High	Status Quo	Low	Medium	High
Abundance of Fish	Weir w M-R in 10 popns, no M-R in 2, census in 1.	- none	Weir M-R in one population for each of 7 MPGs.	Weir M-R in one population for each of 30 populations; census in 1 population			PIT-tagging for MR of selected population in each MPG	PIT-tagging for MR of selected populations
Abundance / Spatial Distn of Redds	Fixed multi redd counts in 22, single redd counts in 7, census in 1, none in 2 (extirpated).	Fixed single redd counts for all 31 populations.	Statistically representative multiple redd counts for all 31 populations	Census using multiple redd counts for all 31 populations				
Age Structure of Spawners	Length @ age in 17 popns; hard parts in 4 popns.	Hard parts in one population for each of 7 MPGs; length at age in all 31 populations.	Hard parts in all 31 populations	Tags + Hard parts in all 31 populations				Mark fish at hatchery w PIT-tag; detect at tributary weirs for each of 31 populations; detect carcasses
Origin of Spawners	Examine hatchery marks on carcasses in 18 popns.	Examine hatchery marks on carcasses in all 31 popns.	Hatchery marks; handle fish at weirs in one population for each of 7 MPGs; examine carcasses in all 31 populations	Hatchery marks; handle fish at weirs in all 31 popns.		Mark fish at hatchery w PIT-tag; detect from carcass survey??	Mark fish at hatchery w PIT-tag; detect at tributary weirs for each of 7 MPGs; detect carcasses	Mark fish at hatchery w PIT-tag; detect at tributary weirs for each of 31 populations; detect carcasses
Sex Ratio of Spawners	Carcass survey in 22 popns; female per redd expansion in 3.	Female per redd expansion in all 31 populations	Carcass survey in all 31 popns.	Carcass survey in all 31 popns.				
Abundance / Spatial Distn of Juveniles	Juv. trap in 15 populations	Snorkel survey in all 31 popns.	Juv. trap in one population for each of 7 MPGs; snorkel survey in all 31 popns.	Juv. trap and snorkel survey in all 31 popns.				
Egg->Smolt	See Claire's table for existing data						More precise estimates of smolt output for MPG populations[9]	More precise estimates of smolt output for selected populations
Survival of Juveniles	M-R in 15 populations.	M-R in all 31 popns.	M-R in all 31 popns.	M-R in all 31 popns.		Mark parr w PIT-tags, recapture at smolt traps	Mark parr w PIT-tags, recapture at smolt traps; get SARs for some (??) of 40 streams where sample size large enough	Mark parr w PIT-tags, recapture at smolt traps; get SARs for 6 MPGs (including Clearwater), but not by population (see Hydro designs on pg. 2)

[8] As described in Charlie Paulsen's integrative Table I1 on page I-9 in Appendix I of CSMEP Annual Report for FY05 (WORD file) or page 111 of the pdf file.

[9] Using PIT-tags can get estimates of both fall emigrating pre-smolts and spring emigrating smolts to obtain more precise estimate of total smolt output.

Status & Trend	Spawner-Recruit Information				PIT-tag information (links to subgroup needs[8]) and other data required to answer question			
	Status Quo	Low	Medium	High	Status Quo	Low	Medium	High
SARs, T/I and D (Questions 1, 3 and 5 in Hydro subgroup) + in-river survival estimates (Question 2) + upstream-downstream comparisons (Question 4)	as described for Status & Trend in above rows	Run reconstruction SARs: partition wild and hatchery smolts and adults at upper dam (e.g., Petrosky et al. 2001; Raymond 1988) + as described in above rows for SR stocks + redd counts from 3 John Day stocks	not required for questions 1, 2, 3, 5 if PIT studies implemented + As above, weirs at one population / MPG in SR to get more representative wild stock composition. + One population / regional stock group in Lower & Mid Columbia [10]: John Day, Deschutes (Warm Springs), Yakima, Wind, Klickitat	not required for questions 1, 2, 3, 5 if PIT studies implemented + Use weirs for SR populations and add weirs for John Day, Wind and Klickitat if required for Status & Trend or other reasons. But weirs are not essential for addressing Hydro Question 4 if High level PIT-tagging is implemented (more precise).	*SR Hatchery Chinook:* • 200,000 tags @ 5 CSS Hatcheries • 10,000 tags @Hat + 10,000 tags @traps to get in-river survival; • 35,000 tags from 3 NPT hatcheries • TOT tags= 255,000 *SR Wild Chinook:* • 40,000 tags @ traps, etc. to get in-river survival • 26,000 tagged at LGR by NOAA • TOT tags=66,000 (29 stream RSTs) *Lower and Mid-Col R Hatchery Chinook:* • 15,000 Carson Hatchery (Wind R) • 15,000 Leavenworth Hatchery (SMP – Wenatchee R) • 40,000 tags CleElum Hatchery (Yakima hatchery evaluation) • TOT tags=70,000 *Lower and Mid-Col R Wild Chinook:* • 6,000 PIT-tags @ John Day River	Background level of PIT-tagging. *SR Hatchery Chinook:* • 10,000 tags @Hat + 10,000 tags @traps to get in-river survival • 20,000 tags @ LGR, all put in-river • TOT tags=40,000 *SR Wild Chinook:* • Same as Status Quo • TOT tags=66,000 (29 stream RSTs) *Lower and Mid-Col R Hatchery Chinook:* • Same as Status Quo but drop Carson • TOT tags=55,000	*SR Hatchery Chinook:* • Distribute tags in proportion to hatchery releases across all SR hatcheries; distribute fish (i.e.,% transported) according to run at large • 20,000 tags @ LGR, all put in-river • TOT tags=275,000 *SR Wild Chinook:* • 66,000 tags @ traps for CSS to get transport & in-river SARs + • 20,000 tags @ 2 NPT traps to get SARs from sub-basins • TOT tags=86,000 (40 stream RSTs) *Lower and Mid-Col R Hatchery Chinook:* • Same as Status Quo • TOT tags=70,000 *Lower and Mid-Col R Wild Chinook:* • 6,000 PIT-tags @ John Day River	*SR Hatchery Chinook:* • Distribute tags in proportionately as for Medium; increase # • 20,000 tags @ LGR, all put in-river • TOT TAGS=375,000 *SR Wild Chinook:* • 146,000 tags @ traps for CSS to get transport & in-river SARs + • 40,000 tags @ 2 NPT traps to get SARs from sub-basins • TOT tags=186,000 (29 stream RSTs + 8 large traps to cover 6 MPG strata, incl. Clearwater; not by population) *Lower and Mid-Col R Hatchery Chinook:* • Status Quo + • 15,000 tags@Warm Spg H for Deschutes • 15,000 tags@Irrigon H for Umatilla R • TOT tags=100,000 *Lower and Mid-Col R Wild Chinook:* • 6,000 PIT-tags @ each of John Day, Warm Springs, Kilickitat, Yakima, Umatilla [11]

[10] No defined MPGs for spring/summer Chinook in Lower and Mid-Columbia as stocks not listed. John Day has multiple redd counts; Warm Springs has a weir; Yakima has both dam counts of returning spawners (Rosa, Prosser) and multiple pass redd counts

[11] PIT-tagging would be new to Warm Springs and Klickitat, and an adjustment to existing PIT-tagging in Yakima and Umatilla. Yakima and Umatilla are both supplemented populations.

Appendix E. Costing for PIT-tagging studies based on SMP and CSS

							cost per tag $2.10				
Year of cost	**Organization**	**Location of tagging**	**Species [Notes]**	**Total labor, equip., etc. cost other than tag**	**Estimated Incremental Labor Cost for PIT-tagging**	**# of tags**	**$ of tags**	**Total Actual Cost**	**Computed Total Cost**	**Inferred Labor Cost/ PIT Tag**	
	Hatchery Fish Trapping										
2006	IDFG	RAPH, MCCA	Chinook Yrlgs	$96,440		104,000	$213,200	$309,640	$314,840	$0.93	
2005	IDFG	PAHP	Steelhead	$55,773		40,000	$84,000	$139,773	$139,773	$1.39	
2006	ODFW	LOOH (Imnahw&Cathep)	Chinook Yrlgs	$63,446		42,000	$86,100	$149,546	$151,646	$1.51	
2006	USFWS	DWOR, CARS	Chinook Yrlgs	$68,797		67,000	$137,350	$206,147	$209,497	$1.03	
2005	USFWS	DWOR	Steelhead [1]	$36,715		40,000	$84,000	$120,715	$120,715	$0.92	
									Average	**$1.16**	
	Wild Fish Trapping										
2006	IDFG	saltrp, snktrp, clwtrp	Chinook+steelhead [2]	$359,074	$50,000	29,050	$61,005	$420,079	$420,079	**$1.72**	
2006	IDFG	8+ traps funded under other programs [3]				14,500	$30,450	$30,450	$30,450	$0.00	
2006	ODFW	grntrp	Chinook+steelhead [4]	$261,203		6,550	$13,755	$274,958	$274,958	$0.00	need data
								$725,487			

Notes

[1]. Labor costs relatively low because Dworshak hatchery is already tagging chinook

[2]. Funded by Smolt Monitoring Program. Traps in Salmon, Clearwater and Grande Ronde (all Incline Plane), and Snake (Dipper/Scoop w Rotary Drum) all PIT-tag smolts in spring

CSS charged $21K for about 40% of the PIT tags at this trap, so cost for all tags about $50K

[3]. Funded separately through supplementation or hatchery programs. These ID traps include PIT-tagging of both pre-smolts in summer and fall, as well as smolts in spring.

Juvenile wild spring/summer chinook tagged at tributary screw traps in Idaho, 2006 outmigration

IDFG data extraction (P. Bunn) - (Petrosky planning summary, not error checked)

Species	subcode	Subbasin	Stream	Method	7/1/05-8/31/05 Summer	9/1/05-12/31/05 Fall	1/06-6/30/06 Spring	Total			Total / Subbasin
11W	CW	S Fk Clearwater	American River	SCREWT	4	599		603			
11W	CW	Clearwater	Clear Creek	SCREWT		4		4			4
11W	CW	Lochsa	Colt Kill Creek	SCREWT	15	351	138	504			
11W	CW	Lochsa	Crooked Fork Creek	SCREWT			27	27			
11W	CW	Lochsa	Crooked Fork Trap	SCREWT	309	2739	233	3281			
11W	CW	Lochsa	Fish Creek Trap	SCREWT		568	4	572			4384
11W	CW	S Fk Clearwater	Crooked River	SCREWT	32	118	15	165			
11W	CW	S Fk Clearwater	Newsome Creek	SCREWT		3203	143	3346			
11W	CW	S Fk Clearwater	Red River	SCREWT	230	1802	12	2044	n	10	6158
11W	CW	Selway	Meadow Creek	SCREWT		3085	2566	5651	sum	16197	5651
11W	GRI	Grande Ronde	Catherine Creek	SCREWT		500	349	849			5189
11W	GRI	Grande Ronde	Wallowa River	SCREWT		521	723	1244			
11W	GRI	Grande Ronde	Lookingglass Creek	SCREWT	494	610	99	1203			
11W	GRI	Grande Ronde	Lostine River	SCREWT		495	517	1012			
11W	GRI	Grande Ronde	Minam River	SCREWT		499	382	881	n	6	
12W	GRI	Imnaha	Imnaha Trap	SCREWT	18	1808	994	2820	sum	8009	2820
11W	S	Lemhi	Lemhi River	SCREWT	2	1168	62	1232			
11W	S	Lemhi	Lemhi River weir	SCREWT	20	1367	401	1788			3020
11W	S	Middle Fk Salmon	Marsh Creek Trap	SCREWT	1179	935	182	2296			2296
12W	S	Pahsimeroi	Pahsimeroi River	SCREWT	248	1736	1588	3572			3572
12W	S	S Fk Salmon	Johnson Creek Trap	SCREWT	971	4374	2540	7885			
12W	S	S Fk Salmon	Lake Creek	SCREWT	737	782	176	1695			
12W	S	S Fk Salmon	Lick Creek	SCREWT	1			1			
12W	S	S Fk Salmon	Secesh River Screw Trap	SCREWT	1410	2704		4114			
12W	S	S Fk Salmon	South Fk Salmon River Trap	SCREWT	1386	3066	1453	5905			
12W	S	S Fk Salmon	Secesh River	SCREWT	1400	886	130	2416			22016
11W	S	Upper Salmon	East Fork Salmon River	SCREWT	179	1033	34	1246			
11W	S	Upper Salmon	Sawtooth Trap	SCREWT	1741	2628	1604	5973	n	13	
11W	S	Upper Salmon	W Fk Yankee Fork	SCREWT	5	23		28	sum	38151	7247
				sum	10381	37604	14372	62357			
				20%ile	13	499	51	545			
				median	239	935	208	1246		avg # fish	2150
				80%ile	1220	2689	1178	3436		avg cost	30.23
							summer fish	28408			
								46%			

Cost Estimates for CSMEP Alternatives

Status Quo	29 streams with screw traps tagging 62K wild chinook			
	assumed all traps PIT tagging most of catch		increasing # fish would require more traps	
	assumed cost to operate 1 trap	65,000	operating for 9 months/year; 2 staff/crew	
	tag costs @ 2.10	130,200		
	cost per 29 traps	1,885,000	$ 30.23	labour cost / tag
	Total cost	2,015,200	PIT tagging currently for a number of natural production and hatchery supplementation M&E objectives	
Medium design: 86K tags				
	# traps needed (assume median catch/trap)	40	1.39	
	tag costs @ 2.10	180,600		
	cost per 40 traps	2,614,677	30.40	labour cost / tag
	Total cost	2,795,277	costs may be much less if selective for large production areas	
High design: 186K tags - focussed on additional tributary traps				
	# traps needed (average catch/trap)	87	3.00	
	tag costs @ 2.10	390,600		
	cost per 87 traps	5,655,000	30.40	labour cost / tag
	Total cost	6,045,600	costs may be much less if selective for large production areas	

High design: 186K tags - focussed on mainstem screw traps to sample MPGs

assume could get close to catch numbers to represent SARs, T/C and D estimates by MPG

new traps at:

1. upper Salmon below EFSR (sample 5 populations)
2. upper Salmon below North Fork (sample 3 additional populations)
3. Salmon River below MFSR (MFSR MPG + Panther Cr) (SFSR MPG covered by pilot study)
4. Lochsa near mouth
5. Selway near mouth
6. S. Fk Clearwater near mouth
7. Grande Ronde above Wallawa ??
8. Grande Ronde low in system ?? (Imnaha covered?)

# traps needed (average catch/trap)	37	feasibility not really assessed	
tag costs @ 2.10	390,600		
cost per 37 (29+8) traps	2,405,000	12.93	labour cost / tag
Total cost	2,795,600		

Appendix F. Subset of regional-scale hatchery effectiveness questions for CSMEP assessment in FY2006 and beyond

Background

In 2005, the hatchery subgroup identified a number of questions critical to the evaluation of hatchery management, and reviewed numerous existing and proposed hatchery Research, Monitoring, and Evaluation (RME) plans within the Columbia River Basin to determine whether they have the potential to generate the information necessary to address those questions. Following this review, the subgroup concluded that existing and proposed hatchery RME plans, if fully funded and implemented as designed, are likely to address the majority of the management questions identified by the subgroup. However, the subgroup also concluded that a number of questions regarding the *effectiveness* of hatcheries as a class of actions are unlikely to be adequately addressed by existing and proposed hatchery RME. These effectiveness questions (listed below) will likely be efficiently and comprehensively addressed only through the implementation of a stratified and representative study design that spans the entire Columbia River Basin. As such, the study designs to address these questions are best developed within a collaborative process that can rely on the expertise of the multiple tribal, state, and federal agencies with operational jurisdiction and familiarity with implementation of artificial production and the operation of artificial production facilities. This expertise exists within CSMEP and has been useful in assimilating the high level of diversity represented by individual programs to identify pertinent questions that are not currently addressed (representatively) by existing hatchery RME projects. With appropriate stratification, this diversity can be leveraged to identify the mechanistic linkages of individual programs to broader monitoring questions that evaluate the effectiveness of hatchery strategies, as a class of management actions, at the regional scale. These broader-scale hatchery program effectiveness questions (as opposed to individual hatchery operation questions) will become the focus of CSMEP designs intended to address larger scale multi-hatchery questions (listed below) that can be stratified across the region. Please note that the order of the questions within the tables does not reflect a prioritization, all questions presented below were deemed to be high priority.

Questions relevant to harvest augmentation hatcheries follow a logical progression; assuming that these programs are intended to augment harvest, presumably lost as a result of habitat modification (e.g., hydropower development), they should provide a demonstrable contribution to harvest (questions 1 and 2). The contribution to harvest must be large enough to offset the potentially deleterious effects of the operation of such facilities. This requires an assessment of the effects of harvest augmentation hatcheries on the viability of natural populations. The degree to which harvest augmentation hatcheries are expected to effect natural populations is assessed at a coarse scale by the distribution (question 3) and magnitude (question 4) of hatchery strays. Given an understanding of stray rates, the impacts of hatchery strays on the viability of natural populations (question 7) is a function of their reproductive success (question 5, which then dictates the magnitude of expected ecological interactions between juveniles with hatchery ancestry and natural origin juveniles as well as the genetic impacts of introgression) and the potential for disease transfer (question 6).

Harvest Augmentation Hatcheries: *To what extent can hatcheries be used to assist in meeting harvest management goals while keeping impacts to natural populations within acceptable limits?*

	Regional Question	Priority
1	What are annual harvest contributions and catch distribution of hatchery produced fish?	H
2	To what degree do hatchery programs meet harvest objectives?	H
3	What is the distribution of hatchery strays into natural populations?	H
4	What are the proportions of stray hatchery fish in non-target natural populations?	H
5	What is the relative reproductive success of hatchery origin adults relative to natural origin adults?	H
6	What are the disease agents and pathogens in hatchery fish, to what degree are these agents transmitted to natural fish, and what are the impacts of such transmissions?	H
7	What are the impacts of hatchery strays on non-target populations?	H

Supplementation Hatcheries: *To what extent can hatcheries be used to enhance viability of natural populations while keeping impacts to non-target populations within acceptable limits?*

	Regional Question	Priority
1	What are the status and trends of habitat targeted by supplementation projects and what is/are the life-stage specific factors that limit productivity?	H
2	What is the reproductive success of naturally spawning hatchery fish relative to natural origin fish in target populations?	H
3	What are the disease agents and pathogens in hatchery fish, to what degree are these agents transmitted to natural fish, and what are the impacts of such transmissions?	H
4	What are the relative effective population sizes and genetic diversity of hatchery supplemented vs. un-supplemented populations before, during, and after supplementation?	H
5	What proportion of hatchery origin juveniles return as adults to target versus non-target populations?	H
6	Do hatchery origin juveniles from supplementation programs stray at a greater rate than their natural origin conspecifics?	H
7	What are the proportions of natural spawning stray hatchery fish in non-target natural populations and their impact on the viability of natural populations?	H
8	What is the reproductive success of naturally spawning hatchery fish relative to natural origin fish in non-target populations?	H
9	What are the effects of hatchery supplementation on the productivity, abundance, and viability of non-target natural and hatchery-influenced populations?	H

Supplementation hatcheries act as refuge to offset mortality in early life-history stages. The ability of hatcheries to decrease early life-history mortality, though not ubiquitous (Miller 1990), is well supported (Hard et al. 1992), and a routine metric considered in many of the monitoring and evaluation programs that accompany hatcheries. Juveniles from supplementation programs are typically released into habitats to which they are expected to return and spawn, thereby potentially increasing natural production. Thus a common metric of the supplementation hatcheries is a comparison of the progeny per parent ratios of the hatchery relative to natural production. Because this has been a key metric of numerous monitoring and evaluation projects, the ability of hatcheries to achieve a higher adult to adult return rate, relative to natural production, although again not ubiquitous, is well established (Waples et al. 2001). Given that supplementation programs can successfully increase escapement relative to natural spawning, it follows that targeted habitat must be capable of supporting increased escapement. Monitoring and evaluation activities that accompany numerous supplementation projects have illustrated that targeted streams are underseeded, suggesting that "excess capacity" is available for production (e.g., Arnsberg et al. 1992). Nonetheless, it has also been shown that spawning and early rearing (i.e., egg to emigrant) habitat is not the limiting factor for many populations that are supplemented (Petrosky et al. 2001), and the status and trends of habitat at the life-history stages that limit survival may or may not be known (e.g., mainstem, estuary, and marine). Thus, question 1 seeks to determine which habitats limit productivity, the life history stages that are expressed in those habitats, and the status and trends of habitat(s) that limit productivity.

Assuming that supplementation programs increase survival from the juvenile to adult life-history stages, achieving the goal of increasing natural production requires that hatchery origin adults successfully reproduce and that their progeny are viable and survive at rates similar to their conspecifics that do not have hatchery ancestry. Given that their natural origin conspecifics might be expected to exhibit optimal reproductive success, it is reasonable to compare the reproductive success of hatchery origin adults to their natural origin conspecifics (question 2). Likewise, the survival of juveniles with hatchery ancestry can be meaningfully compared to their conspecifics that lack hatchery ancestry.

Assuming that supplementation provides a demographic benefit from the perspective of productivity (as measured by question 2), hatchery origin juveniles have the potential to serve as disease vectors, potentially offsetting otherwise positive demographic benefits (question 3). Broodstock collection, mortality within the hatchery, and post-release mortality can potentially decrease genetic diversity of targeted populations (Hard et al. 1992); likewise, the implementation of specific breeding protocols, decreased genetic drift owing to reduced random mortality, and increased abundance potentially resulting from supplementation can maintain or increase genetic diversity (Hedrick and Hedgecock 1994). Question 4 evaluates the variance among the effective population sizes of hatchery and conspecific natural populations to evaluate whether supplementation, as a class of recovery actions, is most likely to have a positive or negative effect on the maintenance of genetic variation.

Despite the fact that supplementation programs strive to produce juveniles that are genetically, behaviorally, and functionally identical to their natural origin conspecifics, the fact remains that straying of hatchery origin adults can potentially have deleterious consequences for natural origin populations (e.g., Grant 1997). Therefore, the distribution and magnitude of straying of hatchery origin adults originating from supplementation programs is of interest (question 5). Because supplementation is a key component of multiple recovery plans it is also meaningful to determine whether the stray rates of adults originating from supplementation programs is greater than their natural origin conspecifics (question 6); particularly given that changes in the life-history stage of released juveniles, release timing, and method of release can potentially decrease stray rates (Quinn1993; Unwin and Quinn 1993; Hard and Heard 1999). At a coarse scale, the impacts of hatchery strays is a function of the magnitude of straying (question 7), the reproductive success of strays (question 8), and the resulting effects on the viability of non-target populations (question 9).

Work Elements

Study design alternatives to address the questions identified in the previous sections will be addressed initially at the scale of the Snake River Basin. The intent is that design templates formulated at the scale of the Snake River Basin will then be expanded by the inclusion of at least one priority subbasin in both Washington and Oregon. Finally, during the FY2007-08 contract period revised designs and strata will be progressively expanded to representatively cover the Columbia River Basin. In addition to expanding the spatial scale of hatchery study designs, the hatchery subgroup likewise will increase collaboration with the ISRP, ISAB, and the Federal Caucus RME workgroup. Closer collaboration with these entities is anticipated to increase the rigor of work products, and ensure that designs are focused on key questions of regional significance. Work performed in the FY2007 to 2008 contract period will address specific information needs, including:

1. A review of existing effort and the compilation of alternative designs to provide a coordinated marking and mark recovery strategy, at the scale of the Columbia River Basin, for hatchery and natural origin juveniles that can provide mark recoveries with known efficiency and expansion rates based on known mark effort:
    a. relevant to harvest and
    b. relevant to the calculation of stray rates for index hatchery and natural populations.
2. A compilation of the distribution and frequency with which communicable diseases occur at all hatcheries operating in the Columbia River Basin and a probabilistic (status) and fixed frame (trend) survey of disease prevalence and presence in natural populations, with a specific focus on how the transmission of pathogens (vertical versus horizontal) can or cannot be addressed with common study design alternatives.
3. A stratified[12] sampling effort to evaluate the relative reproductive success of hatchery origin adults, under natural conditions, via genetic assay.
4. A stratified[13] sampling effort to evaluate the effects of hatchery origin adults on the metrics that relate to viability (e.g., VSP criteria; McElhany et al. 2000).
    a. abundance
    b. productivity
    c. spatial structure
    d. diversity.

Methods

For each work element, data will be collected from existing RME projects and reduced to describe the current state of information. Stratification will be evaluated using simulations based on existing data, and gaps in existing information will be identified relevant to the following tasks:

1. Evaluate the ability of existing RME to **representatively** populate strata relevant to marking effort and mark recovery efforts within the Columbia River Basin.
2. Compile hatchery disease records, accumulate data from diseases surveys of natural populations, evaluate gaps in existing information, and identify strata appropriate to fill monitoring gaps across species, hatchery programs, and the spatial extent of the Columbia River Basin.
3. Identify "sentinel" locations that are representatively stratified throughout the Columbia River Basin (possibly a subset of locations identified in question 1) to evaluate the proportion of total escapement composed of strays, evaluate relative adult reproductive success, monitor trends in productivity and life-history diversity, and evaluate trends in habitat.

[12] Strata to be evaluated include proportion of broodstock composed of natural origin adults, duration of the hatchery project, effective population size of the aggregate (hatchery and natural) population.

[13] Strata should include populations with varying degrees of hatchery influence – both target and non-target populations for supplementation programs, and locations where inadvertent escapement of augmentation hatchery origin adults might be expected to have a range of demographic impacts.

Products

For each task, a summary report will evaluate:

1. whether existing data can provide an unbiased evaluation of the key metrics;
2. a comparison of the statistical power and assumptions associated with alternative study designs; and
3. a suite of recommended alternatives ranked as incremental gains in precision versus cost.

Arnsberg, B. D., W. P. Connor, and E. Connor. 1992. Mainstem Clearwater River study: Assessment for salmonid spawning, incubation, and rearing. Final Report by the Nez Perce Tribe, Contract DEAI79- 87-BP37474 to Bonneville Power Administration, Portland, Oregon.

Grant, S.W. (Editor). 1997. Genetic effects of straying of non-native hatchery fish into natural populations: Proceedings of the workshop. U.S. Department of Commerce. NOAA Technical Memo. NMFS-NWFSC-30, 130 p.

Hard, J.J. and W.R. Heard. 1999. Analysis of straying variation in Alaskan hatchery Chinook salmon (*Oncorhynchus tshawytscha*) following transplantation. Canadian Journal of Fisheries and Aquatic Sciences. 65: 578-589.

Hedrick, P.W., and D.P. Hedgecock. 1994. Effective population size in winter-run Chinook salmon. Conservation Biology. 8(3): 890-892.

Appendix G. Draft CSMEP Objectives x Alternatives Matrix for Snake Basin M & E Designs

*For Numerical or quantitative evaluations (**N**), units should be shown. Need to assess which quantitative PMs are feasible, given tools. For Qualitative evaluations (**Q**): **5** = excellent; **4** = very good; **3**= good; **2**= fair; **1**=poor; **?**= Unknown; **n.a.** not applicable*

Decision Domain	Design Objectives	Performance Measures	Design Alternatives[14]			
			Status Quo	Low	Medium	High
Overall PMs	Cost	costs / year ($)[15] (*N*)				
		coverage / cost (%/$)(*N*)				
		ability to leverage other funding				
		opportunities for collaboration				
	Practicality (*Q*ualitative)	feasibility of implementation (*Q*)				
		safety (*Q*)				
		flexibility[16] (*Q*)				
		regional applicability[17] (*Q*)				
		integration w existing pgms (*Q*)				
	Environmental impact (*Q*)	impact of monitoring (*Q*)				
Status & Trend	Inferential ability (*Q*)	ability to answer questions & make decisions at right scales (*Q*)				
		ability to estimate long term trends, continue time series (*Q*)				
		ability to aggregate data to multiple scales (*Q*)				
	Statistical reliability (*N*)	Pr(making right listing decision given TRT rules)				
		index of precision[18]				

[14] To begin with, there may be more than one alternative per category (e.g., L1, L2, L3 for the low alternatives), but ultimately we should converge to a small number of alternatives.

[15] Costs / year should include one time capital costs amortized over the length of the monitoring program (assume 20 years??)

[16] Flexibility to accommodate different types of indicators (e.g., physical habitat, juvenile fish), or different monitoring protocols.

[17] Applicability to different regions or environmental conditions.

[18] Index of precision could be presented as (% error acceptable for decisions / % error of method); a higher value is better. For example, if ±50% is acceptable level of error, and a given method provides ±50% error, then index of precision = 1.0. Error rates of ±25% would yield an index of 2.0; error rates of ±75% would yield an index of 0.67. However,

Decision Domain	Design Objectives	Performance Measures	Design Alternatives[14]			
			Status Quo	Low	Medium	High
		accuracy or bias (± %)[19]				
		coverage[20]				
		statistical power[21]				
Hydro Action Effectiveness	Inferential ability (*Q*)	ability to answer questions & make decisions at right scales (*Q*)				
		ability to estimate long term trends, continue time series (*Q*)				
		ability to aggregate data to multiple scales (*Q*)				
	Statistical reliability (*N*)	index of precision				
		accuracy or bias (± %)				
		coverage				
		statistical power: # years to test hypothesis				
		Pr(making correct conclusion of whether TIR>0)				
Hatchery Action Effectiveness	Inferential ability (*Q*)	ability to answer questions & make decisions at right scales (*Q*)				
		ability to estimate long term trends, continue time series (*Q*)				
		ability to aggregate data to multiple scales (*Q*)				
	Statistical reliability (*N*)	index of precision				
		accuracy or bias (± %)				
		coverage (true straying vs. estimating straying)				
		statistical power (# years required to test null H of no difference in reproductive success w given stat power)				
Harvest Action	Inferential ability (*Q*)	ability to answer questions & make decisions at				

the percent error that is acceptable for decisions may not be known by decision makers or their advisors. In this case, one could simply use % error as the performance measure for precision (lower is better).

[19] Accuracy or bias = % of true mean, as estimated by best possible measurement.

[20] Coverage (% of the time or frequency that true value falls within 95% CI of measured value) depends on both accuracy and precision of method.

[21] Expression of statistical power will vary with decision and hypothesis. Each domain needs to state hypothesis, associated indicator, effect size or range of interest, and statistical power (e.g., test H:T/I > 1.0 with $\alpha=\beta=0.1$). Alternative PMs for statistical power: power to detect an effect size of interest within a 10 year period; # years required to detect an effect size of interest with 0.8 statistical power; or # years required to test stated hypothesis at stated level of power.

Decision Domain	Design Objectives	Performance Measures	Design Alternatives[14]			
			Status Quo	Low	Medium	High
Effectiveness		right scales (*Q*)				
		ability to estimate long term trends, continue time series (*Q*)				
		ability to aggregate data to multiple scales (*Q*)				
	Statistical reliability (*N*)	index of precision				
		accuracy or bias (± %)				
		coverage (true run size vs. estimated run size; true vs. estimated non-target harvest)				
		frequency & magnitude of over-harvest & under-harvest decisions				
		statistical power				

Endnotes:

[i] To begin with, there may be more than one alternative per category (e.g., L1, L2, L3 for the low alternatives), but ultimately we should converge to a small number of alternatives.

[ii] Costs / year should include one time capital costs amortized over the length of the monitoring program (assume 20 years??)

[iii] Flexibility to accommodate different types of indicators (e.g., physical habitat, juvenile fish), or different monitoring protocols.

[iv] Applicability to different regions or environmental conditions.

[v] Index of precision could be presented as (% error acceptable for decisions / % error of method); a higher value is better. For example, if ±50% is acceptable level of error, and a given method provides ±50% error, then index of precision = 1.0. Error rates of ±25% would yield an index of 2.0; error rates of ±75% would yield an index of 0.67. However, the percent error that is acceptable for decisions may not be known by decision makers or their advisors. In this case, one could simply use % error as the performance measure for precision (lower is better).

[vi] Accuracy or bias = % of true mean, as estimated by best possible measurement.

[vii] Coverage (% of the time or frequency that true value falls within 95% CI of measured value) depends on both accuracy and precision of method.

[viii] Expression of statistical power will vary with decision and hypothesis. Each domain needs to state hypothesis, associated indicator, effect size or range of interest, and statistical power (e.g., test H:T/I > 1.0 with $\alpha=\beta=0.1$). Alternative PMs for statistical power: power to detect an effect size of interest within a 10 year period; # years required to detect an effect size of interest with 0.8 statistical power; or # years required to test stated hypothesis at stated level of power.