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 other 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

[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

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 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, as well as 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 and its appendices as well as on the CSMEP public website.

STATUS AND TRENDS

In FY2006, the CSMEP Status and Trends Subgroup focused on continued development of a simulation model that could be used for evaluating alternative designs (Low, Medium, High) for monitoring status and trends of Snake River spring chinook at the population, MPG and ESU scales, incorporating 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:

- Provides a framework for assessing the variability in the data used to measure Abundance, Productivity, Spatial Structure, and Diversity.
- Misclassification rate of SS/D risk levels is a simple but realistic frame-work for assessing the variability in Spatial Structure/Diversity data which is otherwise difficult to quantify
- 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 given/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 3 simple designs in order to provide direction on which other more complex designs might be worth testing.
- This information can help the manager to determine where it is physically feasible to make improvements and then the model can be used to test how much value would be gained by making those improvements.
- Due to the lack of consistent information regarding the 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 decide what is the best way to achieve that CV depending on the conditions of a specific 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 seven 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 many other areas, for several decades, and they are likely to be continued in order to take advantage of the utility of the long-term databases as well as 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 only on 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 is often infrequent as well as 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 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.

HYDROSYSTEM

In FY2006, the CSMEP Hydro Subgroup focused on two 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? 3. Has hydrosystem complied with performance standards set out in 2000 FCRPS BiOp?
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?

For questions 1 and 2, 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

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

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.
- 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], if the true *TIR* differs from 1.0 by at least 20%. 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 inriver 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 inriver conditions and fish condition. While process variation is largely beyond our control, sampling variation can be reduced by increasing the number of inriver smolts tagged and, perhaps, by increasing the proportion of survivors detected at dams and at the trawl below Bonneville.

---

[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

- None of the preliminary strategies developed by the Hydro Subgroup (low, medium or high, greatly increases the number of inriver migrants. To reduce measurement variation by a factor two, one would need to increase tagged inriver 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 SAR's, upstream survival, and inriver harvest, assuming harvested adults were sampled for PIT tags.

For question 4, we have drawn the following conclusions:

- Wild spring chinook TIR's have generally shown an increasing trend over time – fish transported late in the season have higher SAR's than inriver migrants, while the opposite is true early on. As with inriver 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 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 SAR's. 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.

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 Monitoring and Evaluation approaches.
- *Beyond the CSMEP Harvest Subgroup* …. TAC envisions a need for full ESU/population-based run reconstruction for steelhead.

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 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 engaged regional managers to determine the management objectives for the subbasin; 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 actual statistical analyses that should be undertaken within the proposed CSMEP designs 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 proposed low, medium, high Lemhi designs originally 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

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:

- The 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.
- The 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 late 2006, and that sample “draws” using EMAP will enable cost estimation and statistical power analyses in early 2007.

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).