Independent Scientific Review Panel
for the Northwest Power and Conservation Council
851 SW 6th Avenue, Suite 1100
Portland, Oregon 97204
isrp@nwcouncil.org

MEMORANDUM May 18, 2004

TO: Mark Fritsch, Fish Production Coordinator, Northwest Power Planning Council

FROM: Rick Williams, ISRP Chair

SUBJECT: ISRP Step Two Review of the Northeast Oregon Hatchery (NEOH) Spring Chinook Master Plan: Monitoring and Evaluation Plan (ISRP 2004-10)

Introduction

Per the Council's request, the ISRP reviewed the Monitoring and Evaluation Plan for Northeast Oregon Hatchery Imnaha and Grande Ronde Subbasin Spring Chinook Salmon (March 1, 2004) as part of the Step 2 Review of the NEOH Spring Chinook Master Plan. This Monitoring and Evaluation Plan is the fourth NEOH core team submittal/response regarding issues that were identified by the ISRP as requiring further explanation or clarification during the ISRP's initial review (July 2000) of the NEOH Master Plan.[1] Earlier responses addressed issues regarding the monitoring and evaluation plan's genetic breeding plan; harvest framework, forecasting, and escapement goals; and other issues. However, those previous NEOH submittals did not constitute a complete monitoring and evaluation plan that provided adequate detail to allow for a technical review.

Recommendation

The ISRP's overall response is that this document is an excellent working draft of a stand-alone M&E Plan for the NEOH hatchery Imnaha and Grande Ronde subbasin spring Chinook salmon program. The NEOH core team was successful in collating a single integrated document, describing general methods and data to be collected, and explaining how these data would be used as an evaluation of the NEOH program. We appreciate the effort that has gone into planning for M&E in the face of numerous ongoing projects, the need to meet 2000 BiOp directives, and the need to evaluate the effectiveness of the supplementation project, harvest, escapement, spawning success, etc.

We further compliment the authors for being among the first to bring the modern EMAP probabilistic sampling procedures into the Columbia Basin. For example, in so far as we are aware, this plan is to be the first use in the Columbia Basin of a rotating panel design to balance

[1] See: ISRP 2003-12: Step Two Review of the Northeast Oregon Hatchery Spring Chinook Master Plan. August 2003. www.nwcouncil.org/library/isrp/isrp2003-12.htm; and follow-up memo: www.nwcouncil.org/library/isrp/isrp2003-12followup.pdf
ISRP 2001-12C: ISRP Step Two Review of the Northeast Oregon Hatchery Spring Chinook Master Plan. December 2001. www.nwcouncil.org/library/isrp/isrp2001-12c.htm

the needs of status (more random sites) and trend (more repeat sites) monitoring. We strongly endorse the authors' development of the EMAP-type probabilistic sampling scheme for redd counts to complement current surveys. The plan appropriately calls for selection of random sites outside the traditional survey areas to be surveyed for redds in each subbasin. In summary, although we raise several yet-to-be-fully-resolved issues and offer other comments for consideration, this Plan provides a good example of a monitoring program that could be used as a model for program development throughout the basin.

We believe that the main remaining issues to be considered and refined in the final Step 3 submittal of the M&E Plan are:

- more thorough prioritization of monitoring and evaluation efforts;
- further scoping of the power and resolution that can be expected for the metrics that are to be measured, given constraints of sampling and inherent variance, and use of this information to inform decisions as to sampling intensity and the priority of evaluation metrics;
- consideration of full use of the suite of descriptors of outcomes (e.g., reports of primary data and thorough statistical description of derived summary metrics), rather than simply hypothesis tests at $p = 0.05$);
- assurance that sample sizes are adequate for the metrics that comprise the core evaluation of the final Plan;
- development of a clear plan for integrating evaluation metrics into adaptive management of the program, including a decision tree or other representation of clearly stated decision triggers and actions that would result in program modification (or even termination, if warranted);
- and development of procedures and protocols for implementing the plan.

Specific Major Issues for which Revision and Response Were Requested

Adequacy of Step 2 draft of the M&E Plan

The NEOH Core Team's earlier description of the M&E plan did not constitute a sufficiently complete plan to allow a level of scientific review expected of the ISRP. The NEOH team was asked to describe clearly the experimental and sampling designs of the overall monitoring and evaluation program and make clear that all the pieces fit together to provide long-term guidance for implementation, monitoring, and evaluation of the NEOH program. They were also asked to make clear what the priority monitoring metrics were, as well as how these would be integrated with specific analyses to result in useful evaluation of outcomes.

The response now constitutes a detailed and complete plan, as requested by the ISRP reviewers, and, as stated above, constitutes an excellent working draft from which to discuss, debate, and hone the final Step 3 Plan that will be implemented. However, the plan is large, so a useful addition may be a flow chart of the related tasks and data needed for groups of the objectives. Tables 5 and 6 provide useful lists of these items but new users may find a visual summary more informative than these long lists of performance measures and projects. This suggestion is merely a potential enhancement, but may provide a useful check on the overall program.

The earlier material that the ISRP reviewed relied heavily on an Appendix A as the basis for an M&E Plan, but did not develop the core, essential elements of the plan. The thoroughness of the current draft clearly demonstrates the adequacy of response to this concern. The only comment in the ISRP's follow-up memo that remains outstanding, concerns "ranking of management questions and objectives by priority." The response addressed this (page 117), but these discussions are ongoing. This remains a potentially important task to complete in case funding is not adequate for the full program; however, we believe that the authors are fully aware of this need and will address it more fully in the Step 3 submittal.

In earlier discussions with the NEOH co-management team, the ISRP suggested nine components that should be included in the M&E Plan: (1) Goals and objectives; (2) Stocks and geography; (3) Experimental design, including both sampling layout and intensity, and specific planned statistical analyses; (4) Critical parameters; (5) Sampling procedures, including consideration of whether sampling is adequate to meet the objectives and the costs of different levels of statistical power; (6) Data management; (7) Program coverage and assessment, i.e., how the ongoing projects that were described combine to meet the needs of the plan, what new elements, if any, are needed, and where there might be redundancy; (8) BiOp and systemwide uncertainties, including the relationship of the NEOH into the big picture of testing supplementation and better describing how hatchery and wild fish interact; and (9) Costs. Eight of the nine items in the list were addressed in an understandable presentation. Program coverage and assessment of current projects was well presented in Table 6 (pages 113-116 of text). The BiOp and systemwide uncertainties were not highlighted, but are included throughout the report. Appendix B presents the Genetic Management Plan for NEOH Spring Chinook. Cost remains to be considered as the Plan is completed, and co-managers will want to weigh this carefully as they choose their priority metrics and finalize details of experimental design and implementation.

Levels of Precision and scale of Data Collections

The NEOH team was asked to set a standard of precision for use in developing the M&E Plan and to state where these standards could and could not be met. The ISRP suggested that the Plan should consider the confidence levels necessary to meet objectives (e.g., 80-90% confidence may suffice), the levels that are achievable given constraints of sampling error, inherent variability of metrics, and cost of getting needed samples. Finally, the 95% confidence intervals and levels commonly used in statistics may not be necessary or achievable for all of the NEOH purposes.

This point continues to merit discussion, but is not likely to change the activities, only the intensity of sampling, associated costs, and some statistical results. The plan continues to use the 95% confidence interval that is commonly used in statistical analyses. However, in biological and field programs, achieving this level of precision can be very expensive, if it is even a realistic goal. Why does the NEOH adopt this high standard; what is the basis of this target and would a slightly lower target not be acceptable? Accepting a lower level of precision may allow funds for other aspects of the program. Two other sampling issues remain:

1. Methods frequently refer to sampling inter-dependent variables (e.g., age-at-maturity, sex ratio, hatchery versus wild, etc.) but then seem to treat them as independent (i.e., estimates based on multiplying them through to estimate a final return number by source, sex, and age). Sample sizes that are stated are frequently quite small which reviewers

suspect reflects treating each of these variables independently when they are actually interdependent. Sampling of multiple categories (multinomial sampling) usually results in quite large sample sizes that are not apparent in this document. The authors should consult a statistician with expertise in environmental sampling.

2. The authors frequently comment that something is a point estimate and/or that there is no precision that can be estimated. However, with the intensity of sampling and data collected, the authors should be able to empirically estimate variances using bootstrap or other repeat sampling methods. This is another issue on which they should consult a statistician with expertise in these methods.

Supplementation and the NEOH Projects

The NEOH team had stated earlier that the NEOH project was not intended to answer whether supplementation projects can support sustaining natural populations. The ISAB/ISRP (and Council) consider the ISS, NEOH, and Yakima programs among the most important studies in the Columbia River Basin specifically funded to test the efficacy of supplementation, so a decision not to pursue answering the bigger question on supplementation would be a significant departure from the ISRP's understanding of one of the primary reasons for funding the suite of NEOH projects at their current level of support. Failure to address the efficacy and limitations of supplementation within these related projects would be an opportunity lost. The current draft addresses these concerns, although we comment below as to a continuing need to explicitly incorporate and prioritize robust core evaluation metrics of supplementation's effects on natural fish.

Genetic Tracking

Tracking demographic and genetic attributes of wild and hatchery components in the suite of NEOH projects will be important in understanding population responses and evaluating the effects of the various projects. The NEOH team was asked to clearly describe ongoing cooperative work with NOAA Fisheries to provide genetic work in Catherine Creek, and the Grande Ronde, Imnaha, and Lostine rivers, as these ongoing activities (and any future genetic studies) that provide information on wild and hatchery interactions should be in the overall M&E plan.

The current draft Plan contains quite extensive and well-presented material on genetic aspects of the program, which appears to have been prepared by or with NOAA staff. The ISRP suggests a few points to consider in finalizing the M&E Plan for the Step 3 submittal. The draft Plan suggests that results from sampling will take quite some time and will not be readily available. The actual time for completion of samples was not defined and should be. Further, the ISRP believes that production and analysis of genetic information should be given some priority, as genetic effects of captive breeding and other hatchery-based conservation and production efforts are critical to study.

The genetic monitoring in this plan is an important scientific issue for the Basin. It would seem important for timely feedback on the monitoring results, particularly within an adaptive management framework. The Plan might consider how to establish a routine for processing and annual analysis and reporting of results. Continued collection of data without ongoing critical

assessment is certainly not desirable anytime, but particularly for consideration of genetic issues that may not be reversible if undesirable effects are revealed.

Hatchery Production Needs
The NEOH co-managers were asked to document in the Plan the level of production needed to meet the objectives of the M&E plan. Matching production needs with production opportunities (i.e., numbers of returning broodstock) will guard against last minute changes that compromise the experimental design of planned projects. Appendix C seems to address this issue.

Potentially Critical Considerations for Future Program Development

The ISRP noted a few issues that we believe require more serious consideration and some modifications of current apparent plans during development of the final Plan for Step 3 submittal. These points follow a logical sequence leading to tests of the hypotheses and have, therefore, been prioritized from highest to lower priority.

1) Sampling of Reference Streams (Highest Priority)
The ISRP previously noted that without references streams the NEOH team would have very limited ability to interpret effects of their programs. Reference streams must be monitored for stock status, as well as inference to hatchery input.

As the ISRP has been concerned before, Pg. 45 Methods 1.d.1 introduces a major issue to address, "Adult abundance estimates in reference streams relies on expansion of extensive area redd counts." However, the expansion of redd counts in the reference stream cannot be (a) independent of the treatment streams, if the fish per redd data used is developed from those streams, or (b) unbiased, if using carcass counts in the reference streams. Carcass counts are very unlikely to provide unbiased estimates of the return and sex composition in the return. This issue would seem to seriously limit the validity of treated and reference stream comparisons and should be addressed. **If this cannot be resolved, it would present a major issue to the Council**. The reference streams must be assessed with the same accuracy and precision that the treated streams are, or the study could be seriously compromised.

The only issue that remained seems to be how complete the sampling and enumeration programs will actually be in the reference streams. More detailed comments are provided below. These reference streams must be enumerated with the same data collection protocols used on the treated streams or the study comparisons are very likely not valid. Intensity of sampling (sample sizes) should also be comparable so that differences of effect sizes are estimated with acceptable precision. If the proponents of this Plan are not serious about sampling these reference streams, then the Plan would have to be modified and re-evaluated. **If this cannot be resolved, it would present a major issue to the Council**.

2) Estimation of Precision of Statistics. All statistics should be presented with estimates of precision (variance). For example, estimates of SARs are subject to variance and measures of precision should be computed (currently, no variance around each estimate of SAR is calculated). The Plan authors should consider bootstrapping of estimates to take sources of

variance due to straying, estimation of proportions in each age class, estimated escapement, etc. For another important example, current redd surveys are considered to be deterministic counts with no variance when in fact there are potential biases and variance involved due to less than 100% detection of redds. There is a strong need for standardization of procedures for redd counts within spawning areas using some type of double sampling with "capture/recapture" analyses to account for redds missed and to yield estimates of the variance of estimated numbers of redds. With the addition of surveys of probabilistic selected sites outside the "known" spawning areas, sampling variance will also be introduced to the procedures.

In a number of locations the authors suggest that variances cannot be calculated and on page 39 they refer to "laws of variances" (referring to expectations?). Sampling programs in Nature may frequently not meet assumptions involved in some of these expansions. We suggest that the authors consider empirical estimates of precision by using multiple re-sampling algorithms (for example bootstrapping) in order to provide an estimate of precision.

3) Sampling Adequacy
There appears to be an important issue related to survival studies in the Imnaha River (Pg. 57). At present the adult enumeration capability in the Imnaha is inadequate to sample returns (downstream of the existing weir at rkm 74). **This seems to be an important project issue for the Council to be aware of and consider.** There did not seem to be any serious proposal in this Plan to correct this limitation.

Another issue concerns recapture sampling in tagging experiments. This could have a significant effect on estimates of population size and sex ratios. The probability of tag recoveries on males is substantially less than for females (excluding Jack males that are acknowledged in the Plan), thus, many programs now estimate the population size of returns by sex and then add the two estimates. This issue may not pertain to spring Chinook in the NEOH environments, but it is a common problem that the authors may want to monitor.

The authors should critically investigate the expected precision of age composition data of adults based on samples of size as small as 20. Roughly, the half width of a 95% confidence interval on the estimated proportion in a given age class is as large as 0.22 = 22% with a sample size of 20.

4) Tests of null hypotheses versus estimation of effects. The NEOH Monitoring and
Each of the three points above is necessary for testing the hypotheses listed in this Plan. The lower priority for testing the hypotheses only reflects the need for the above data, not the importance of this task.

The Evaluation Plan tends to emphasize a hypothesis testing framework, but tests of significant differences from some expectation have a number of limitations. An alternative approach is likely needed for the M&E Plan. The ISRP encourages a shift in philosophy from using statistical methods for "statistical tests of null hypotheses" with consideration of the power to detect important effect sizes to methods for "estimation of effect sizes" with measures of precision and accuracy. Testing of null hypotheses has been the standard in laboratory testing and statistical methods textbooks, but, if used in a very formal manner without estimation of

effect sizes (e.g., confidence intervals), then valuable information is lost. It is acceptable to state null and alternative hypotheses such as:

> Ho: Progeny-per-parent ratio of hatchery-origin fish over time is equal to that of natural-origin fish for each stream.
>
> Ha: Progeny-per-parent ratio of hatchery-origin fish over time is greater than that of natural-origin fish for each stream.

However, emphasis during analysis should be on estimation of the statistics (sizes of ratios and differences, or ratios of ratios in this example), with measures of precision and accuracy, not formal statistical tests of hypotheses with a test of size 5%. Decision-making in response to M&E data will no doubt involve much weighing of interpretation of outcomes, and these will require consideration of biologically meaningful effect sizes, rather than simply statistically significant differences, which may be quite compromised by low power due to high inherent variance or low possible sample size. Robust hypothesis testing at $p = .05$ level is likely often unfeasible, and it is the quantitative measures of effect that are most important. The NEOH co-managers should plan to report primary data and full statistical description of the derived metrics, not simply results of tests of significance. The plan also should explicitly describe the sizes of changes or differences (e.g., the effect sizes between treatment and reference streams) that would be considered biologically meaningful and so trigger increased monitoring scrutiny, program adaptation, or program termination.

Other Issues and Suggestions

We offer the following additional peer review suggestions and questions for consideration by the authors of the excellent first draft of the stand-alone Monitoring and Evaluation Plan as they move to the final Step 3 submittal of the Plan.

Coordination. We are impressed with information collected on M&E activities in and below the Imnaha and Grande Ronde subbasins (e.g., see Management Objective 8 and Table 6). We fully support funding for coordination of these efforts and integration with the NEOH M&E Plan. We encourage continued cooperation and coordination with the NOAA Fisheries RME Plan (NOAA 2003) and BPA project no. 35019 for Pilot Status and Trend Monitoring Program for Salmonids and their Habitat in the Wenatchee, John Day, and Upper Salmon to document progress toward recovery of listed populations. We further support coordination with the relatively recent efforts to coordinate and standardize M&E efforts in the Pacific Northwest by the ad hoc group of fisheries scientists under the banner of the "Pacific Northwest Aquatic Monitoring Partnership." We believe that this point in time is a unique and possibly limited opportunity to accomplish better coordination of M&E activities throughout the Tribal Lands, States, Provinces, the Columbia Basin, and the Pacific Northwest. The benefits of consistent, coordinated monitoring are indisputable. The issues are related to how to best achieve the benefits.

Co-location of survey sites. Primary analysis methods for data collected in M&E Plans will include correlation and regression analyses for evaluating measures of improvement and recovery in fish populations (e.g., survival, abundance, presence/absence, escapement, etc.) and freshwater habitat conditions (e.g., temperature, sediment, cover, water quality, flow, etc.). The utilities of both empirical (statistical) and mechanistic models are greatly enhanced by co-location of data collection on common study sites in the field. We strongly encourage the co-location of data collection efforts while recognizing the need to avoid too much activity on study sites. We have not seen the advantages of co-location of data collection efforts addressed in any of the M&E plans that we have reviewed.

Monitoring of land and water uses. The Plan recognizes that livestock overgrazing, logging activity, mining, channelization, and irrigation water withdrawals limit the quantity and quality of salmon habitat in the Grande Ronde; however, monitoring of these uses of land and water is not mentioned further. The plan should establish sources of data on land and water uses, evaluate the quality of such data, and coordinate activities with responsible agencies.

Implementing the M&E efforts. Given the level of detail in the M&E Plan, the ISRP notes that the NEOH co-managers will likely need to take formal steps to ensure that all the pieces will be collected and fit together. Presumably this is the idea behind Tables 5 and 6, but possibly a checklist for each analytical task (by performance measure, location, and life stage) could also be prepared.

The large amount of planned monitoring effort will provide data, analysis, and information to guide hatchery operations and Chinook management in the subbasins. If carried out as described, the data flow will be large and the learning opportunities substantial. But, experience with planning suggests that a stand-alone document will need to be accompanied by other documents that will aid in establishing operational training, procedures, deadlines, and responsibilities. The Plan in final form may need to evolve to an annotated atlas of other documents, programs, etc., in practice or under development, that provide these important details of implementation process and accountability (conceptually, this would be akin to a website with links that go directly to relevant sections of other documents). For instance, who is going to do which work and by when? Table 5 on pages 106-7 presents a frequency/duration column, but provides no general expectation as to when we might see some answers to the specific hypotheses. Specific tasks are not attributed to co-management entities or milestones. Additionally, it is obvious that detailed written protocols for field crews are needed, with training to include annual testing on the protocols.

Adaptive management, integration of derived information into management decision-making, and need for a key indicator(s) of success. In several locations, this Plan refers to an adaptive management process. However, reviewers did not find an explicit experimental design and set of alternative actions that would result depending on the outcome of periodic, annual, or 5-year reviews. If there is not such a formal framework, it may be worth referring to adaptive responses as opposed to adaptive management; better yet would be provision of a specific decision tree and process for integrating results into management.

It is not entirely clear how interpreted information from the data and analysis will be used to adapt management. In practice, the M&E should in fact influence and shape management and policy decisions. There is commonly a disconnect between management and information and this might be improved by answering questions such as: what are the key decision points to effectively managing for a viable and sustainable resource? What kinds of information or data are critical for making management decisions? Will increased precision in information sufficiently balance expected costs or is reasonable uncertainty tolerable? The M&E in this case attempts to take on a "systems" approach (big, broad, and deep). Therefore, a visual presentation (such as a box and arrow diagram or decision tree) in an appendix on where and when information is needed would assist in guiding the priorities of M&E efforts as well as demonstrate to the non-scientists involved in co-management where and why real and hard data are needed, as well as how these data will help them in the long run. This could aid in prioritization of efforts and key metrics.

A basic success criterion seems missing in this plan to address whether released fish are adding to the natural productivity (as opposed to supplanting natural productivity). For example, following supplemental releases do we see increases in individual and population-level fitness of the populations (e.g., measured by increased number and distribution of redds with wild fish spawning on them, increased effective number of wild breeders, increased proportion of returning adults and jacks spawned by wild parents, and probably other metrics)? This issue is at the core of the controversies surrounding use of supplementation as a strategy. Conceptually, supplementation can be a success only if 1) the overall productivity of wild fish increases throughout the area subject to effects of supplementation; 2) the viability of wild fish is maintained; and 3) a harvestable surplus is maintained (or preferably, grown). The first and third can often be in conflict, making ultimate success challenging. Paradoxically, if supplementation is not creating a larger, more viable wild population (thereby leading to its own obsolescence), it is not really a success.

More Specific Reviewer Comments

Pg. 8. Populations and Subbasins: Is it correct that the TRT identified five (5) independent populations in the Grande Ronde? If so, are these 5 populations actually the subbasins in your plan or do you consider them the Primary Aggregates? This same question developed on page 28 (Spatial Scale) and relates to understanding Table 5 (pages 105-108). There are likely only a few of the Primary Aggregates and Subbasins, so the ISRP would suggest a table of these designations by name to avoid any confusion.

Pg. 14 Management Assumptions: While this is not the conventional way the ISRP considers assumptions being stated, the ISRP finds this to be a good presentation of what the essential steps must be (i.e., the critical uncertainties) for successful use of supplementation to increase these populations. In that sense they are assumptions and then they are structured as the various Monitoring and Evaluation Objectives of the program. This is an understandable presentation of a very complex set of inter-related issues.

Pg. 24 Goal: In one of the reviewers few differences of opinion with the authors, reviewers do not think that this goal is "simple and unambiguous". For example, the term "status" could mean a trend in population abundance or could be the state of a population relative to a management target. The use of "optimal" in hatchery programs always carries the connotation that hatchery returns come first and should be maximized, which is not what the authors are trying to portray. And contrast this with "and minimize adverse ecological impacts". The latter two may not even be compatible. A more useful consideration of the goal may be to identify these trade-offs and describe how the Plan will assess them.

Pg. 33. M&E Objective 1a: This is a primary data collection section and reviewers continue to be concerned for how fish per redds may be used in monitoring (this is not clear to reviewers throughout this text). For example, section 1.a.3 states that fish per redd will be calculated based on the estimated number of adults returning divided by the number of redds counted in the same tributary. It also suggests that this value would be "fairly constant." What evidence is there that this is true? Reviewers think monitoring programs need to be very careful not to assume constants and apply them too freely. There seems to be an immediate example of this in the last sentence of section 1.a.4, but spawner abundance cannot be estimated from redd counts without assuming a value of fish per redd (based on what?). These two sections seem to be circular in logic, but reviewers likely do not understand how the redd expansion may be used. Section 1.a.4 is also an example where the authors state there will be no assessment of variation. If redd counts or fish per redd are based on sampling, then some measure of variation can be determined.

Pg. 37-38, Sections 1.a.5 through 1.a.8: Each of these sections seems to describe independent samples for 4 traits … but they are not truly independent (reviewers are also not sure that 'Adult spawners sex ratio is actually a useful trait of measure). From reviewers' own experience, reviewers would suggest you design the sampling by categories and re-estimate the required sample sizes. Reviewers would suggest that the sampling is logically based on: sex within age classes, age classes with stock source (hatchery versus wild), and stock source. This results in 12 categories (2 stocks, 3 ages, and 2 sexes) and will greatly increase the sample sizes needed, especially at 95% confidence. The best way to proceed may be to consult a statistician with a background in sampling. In a few locations, this Plan states very small sample sizes that reviewers do not believe to be realistic as a sampling target (e.g., page 47 refers to 20 fish as a sample for determining age structure).

Pg. 52 Objective 1.e: The H_o in this objective refers to survival of smolts, but the Descriptive section refers to emigrant survival (as discussed in previous pages of the Plan). The Descriptive section then acknowledges that smolts are used for comparison with the hatchery releases. However, the other emigrant forms may be important alternative life history types in the natural populations and important to the overall survival rate of the wild/natural fish. Does the Plan provide the data necessary to estimate the total survival of all emigrants from the wild/natural fish?

Pg. 58 Objective 2.a: The performance measure refers to size-at-return. Is this correct? Reviewers see little value in that measure without considering age. Reviewers suggest sampling for size-at-age of return, otherwise any change in size may simply reflect the age of fish in a return year.

Pg. 60 and 62: Word Choice: Reviewers have a serious concern for wording on both of these pages. "Monitoring should occur in … reference streams … " but monitoring must occur in reference streams if they are to be useful. What is the reason for saying "should" in this context?

Pg. 69: Genetic Data: It is not clear how genetic data will be used to meet the Management Objective. The authors should more clearly state genetic differences or changes that would trigger adaptive changes in the program. How is "adequate genetic diversity" recognized? Can such diversity be related to individual or population function?

Pg. 72. Sampling for fecundity: In terms of reproductive capacity, three factors may be related to fecundity. Sampling should include numbers of eggs per female (females sampled for size and age), size of the eggs (also sample variability in size of eggs, if an accurate method can be developed), and energy content of the eggs. The latter is seldom considered.

Pg. 79 Objective 5a: The text suggests a bias evident in Figure 6 and that they are examining the reasons for this. However, reviewers do not really see good evidence for a bias (i.e., a consistent directional error) in these figures. They seem more to show only a couple of larger deviations (random error) and overall the regressions are quite balanced regarding any pattern of deviations. What is the basis for concern for bias?

Pg. 81, Dealing with prediction of returns of fish: Although bias is not apparent in these graphs, it does appear that prediction of natural fish returns is relatively poor. Although the overall r^2 for the relationship for natural adult fish is ca. 0.76, the relationship appears to be quite non-linear (so a relatively poor fit) and seems to be strongly influenced by one point, that of highest adult return, which is considerably above all the rest of the returns. Without this point, the relationship of actual to predicted returns would have a much shallower slope.

Pg. 88 and 91, Objectives 6a. and 6b: How is captive brood stock production accounted for in these objectives? Also, it is not clear that objectives 6a – 6c provide adequate metrics to meet the Management Objective?

Pg. 98, Objective 7c: It is not clear how this objective fits into the Plan. What does this objective add to the other detailed surveys already described in the treatment and reference streams? What is gained from this objective is not explained to me.

Reviewers fully agree with an objective directed to data management and reporting (Pg. 100, Objective 8a), but it is not apparent whether there will be one central authority responsible for data integrity, metadata, reporting, and access. Even with distributed databases, it would seem worthwhile to have an explicit responsibility centre for this critical issue.

Pg. 112, Marking text: Why does this page refer only to Fall Chinook? Presumably this may have been imported from another text, but it should be checked and corrected in the final document.

Additional Minor Comments

In the discussion of probabilistic sampling for spatial and temporal distribution of juvenile Chinook salmon, spatial distribution of adult spawning, and physical habitat (Jordan et al. 2002) the reference should be to Tier II evaluations as described by NOAA, CBFWA, and BPA rather than Tier I as written.

There is not a universally accepted technical definition of the word "accuracy." To avoid confusion we would recommend using the common jargon that "accuracy" = "bias" in statistical estimators. In fact, later in the section on Bias, Precision, and Accuracy, the authors equate accuracy with bias. Mean square error (MSE) is of course a useful statistic being a function of both bias and precision, i.e., MSE = ($bias^2$ + variance).

The confidence interval on a parameter is the upper and lower limits of the interval, i.e., the estimate $\pm$ the half width of the interval, not the half width of the interval.

The heading for Monitoring and Evaluation Objective 7a, "Determine status and trends of Chinook salmon habitat in the Imnaha and Grande Ronde subbasins" should include the word "and"; i.e., it should read "Determine status and trends of Chinook salmon and habitat in the Imnaha and Grande Ronde subbasins."

The statistical design of the fish marking should be pulled across from the other related proposals in the subbasin and summarized in the Plan.

Stock assessment and sampling workshops are needed in the Columbia River Basin and such workshops might be developed to inform finalization of this monitoring plan, as well as development of and coordination with other such M&E plans.

The text is in places uneven and in need of editing to correct typographical and cut-and-paste errors. Additional use of summary tables might also improve readability of the final document.

The document might benefit from presentation of a diagram of the life-cycle of the fish, annotated with monitoring metrics that occur at different points during the cycle, and indicating which of these are more important.