MONITORING STRATEGY FOR THE UPPER COLUMBIA BASIN

Second Draft Report

August 1, 2006

Prepared by:
Tracy W. Hillman
BioAnalysts, Inc.
Boise, Idaho

Prepared for:
Upper Columbia Salmon Recovery Board,
Bonneville Power Administration, and
National Marine Fisheries Service

TABLE OF CONTENTS

Preface

The first draft of this strategy was prepared without the benefit of rigorous field testing. After two years of testing the strategy in the Wenatchee subbasin and drawing upon information gathered from the Okanogan Basin Monitoring and Evaluation Program (OBMEP), Aquatic and Riparian Effectiveness Monitoring Plan (AREMP), the Pacific Northwest Aquatic Monitoring Partnership (PNAMP), Washington Salmon Recovery Funding Board (SRFB), the Collaborative Systemwide Monitoring and Evaluation Project (CSMEP), and the John Day Protocol Test, several important changes have been made to the strategy. Chief among these were redefining the sampling frame, revising habitat measuring protocols, and updating biological protocols. The statistical and sampling designs were modified slightly.

This strategy is a working draft that will continue to evolve as new information becomes available. The choice of sampling designs, indicators, and protocols identified and described in this strategy will change with advancing scientific understanding and shifts in societal concerns. It is important to note that the methods described in this strategy may not identify causal mechanisms—rigorous experiments may be needed to demonstrate causality. The methods described in this strategy are about shoring up correlational investigations of potential actions. Experiments designed to identify causal mechanisms apply treatments in a random fashion to prevent systematic bias in the estimation of experimental effects. In most restoration cases the treatments to be monitored will not be applied randomly and therefore one cannot be sure if the treatment or some other explanation caused the response.

The development of this strategy benefited from the helpful discussions and comments from biologists throughout the Northwest. A relatively small group of biologists were instrumental in testing and helping develop this strategy. They included John Arterburn, Roy Beaty, Bob Bugert, Brian Cates, Bruce Crawford, Jim Geiselman, Jackie Haskins, Chris Jordan, Steve Katz, Joe Lange, Steve Lanigan, Phil Larsen, Steve Leider, Ken MacDonald, Glenn Merritt, Mark Miller, Dylan Monahan, John Monahan, Charlie Paulsen, Carolyn Pearson, Chuck Peven, Rob Plotnikoff, Kate Terrell, Paul Wagner, Mike Ward, Keith Wolf, Richard Woodsmith, Frank Young, and several ambitious field technicians. Inasmuch as I would like to hold these folks responsible for any errors, omissions, or shortcomings in the strategy, I accept full responsibility for the report. This work was funded by the Bonneville Power Administration, NOAA Fisheries, and the Upper Columbia Salmon Recovery Board.

SECTION 1: INTRODUCTION

Managers often implement actions within tributary streams to improve the status of fish populations and their habitats. Until recently, there was little incentive to monitor such actions to see if they met their desired effects. In cases where actions were monitored, investigators often used inappropriate experimental designs, resulting in failures to assess effects of habitat improvements on fish (Bayley 2002; Currens 2002). Now, however, many programs require that funded actions include valid monitoring efforts, coordinated indicators and measurements to reduce duplication, and a process for standardized reporting and strategic planning. Within the Upper Columbia Basin[1], Washington, several different organizations, including federal, state, tribal, local, and private entities currently implement tributary actions and conduct monitoring studies. Because of different goals and objectives, different entities use different monitoring approaches and protocols. In some cases, different entities are measuring the same (or similar) things in the same streams with little coordination or awareness of each others efforts. The Upper Columbia Regional Technical Team (RTT) is aware of this problem and desires a monitoring strategy or plan that reduces redundancy, increases efficiency, and meets the goals and objectives of the various entities.

There are several different groups within the region that have drafted integrated monitoring strategies. These include the Independent Scientific Advisory Board (ISAB); Action Agencies (Bonneville Power Administration, Army Corps of Engineers, and the Bureau of Reclamation) and NOAA Fisheries Regional Monitoring Program; Aquatic and Riparian Effectiveness Monitoring Plan (AREMP); Pacfish/Infish Biological Opinion (PIBO); the Pacific Northwest Aquatic Monitoring Partnership (PNAMP); Collaborative Systemwide Monitoring and Evaluation Project (CSMEP); Washington Salmon Recovery Funding Board (SRFB); Oregon Plan Monitoring Plan Strategy for Salmon and Watersheds; and the Okanogan Basin Monitoring and Evaluation Program (OBMEP) to name a few. Although the programs have different objectives and describe monitoring in slightly different terms, they all address a similar goal. That is, they all intend to assess the effectiveness of restoration projects and management actions on tributary habitat and fish populations. Consequently, the overall approaches among the programs are similar. Indeed, the Action Agencies/NOAA Fisheries Program calls for monitoring all tributary actions with intensive, standardized protocols and data collection methods. For each tributary action, a list of specific indicators, ranging from water quality to watershed condition, are to be measured.

As noted above, various entities, including the Washington Salmon Recovery Fund Board, are funding and implementing various restoration projects and actions within the Upper Columbia Basin. These projects will be monitored to assess their effectiveness. Other groups, such as the U.S. Forest Service, U.S. Fish and Wildlife Service, Washington Department of Ecology, Washington Department of Fish and Wildlife, The Colville Tribes, The Yakama Nation, Chelan County, and Chelan and Douglas County Public Utility Districts, will continue their ongoing monitoring of fish and habitat in the basin. In addition, NOAA Fisheries, with funding from the Bonneville Power Administration, will continue the implementation of the Integrated Status and Effectiveness Monitoring Program (ISEMP) in the Wenatchee and Entiat basins. Because of all the activities

[1] As described in Section 2, the Upper Columbia Basin includes all tributaries and the Columbia River between the Yakima River and Chief Joseph Dam.

occurring within the Upper Columbia Basin, it is important that a monitoring plan capture the needs of all entities, avoids duplication of sampling efforts, increases monitoring efficiency, and reduces overall monitoring costs.

The monitoring strategy described in this document is not another regional monitoring plan. Rather, this plan draws from existing strategies and outlines an approach specific to the Upper Columbia Basin. The plan described here addresses the following basic questions:

1. What are the current habitat conditions and abundance, distribution, life-stage survival, and age-composition of fish[2] in the Upper Columbia Basin (status monitoring)?
2. How do these factors change over time (trend monitoring)?
3. What effects do tributary habitat actions have on fish populations and habitat conditions (effectiveness monitoring)?

The plan is designed to address these questions and at the same time eliminate duplication of work, reduce costs, and increase monitoring efficiency. Importantly, the strategy will also provide a way to assess the recovery of ESA-listed fish species. The implementation of valid statistical designs, probabilistic sampling designs, standardized data collection protocols, consistent data reporting methods, and selection of sensitive indicators will increase monitoring efficiency (Currens et al. 2000; Bayley 2002).[3] For this plan to be successful, all organizations involved must be willing to cooperate and freely share information. Cooperation includes sharing monitoring responsibilities, adjusting or changing sampling methods to comport with standardized protocols, and adhering to statistical design criteria. In those cases where the standardized method for measuring an indicator is different from what was used in the past, it may be necessary to measure the indicator with both methods for a few years so that a relationship can be developed between the two methods. Scores generated with a former method could then be adjusted to correct for any bias.

For convenience, this report is divided into eight major parts. The first part (Section 2) describes the area in which this plan will be implemented. Section 3 identifies valid statistical designs for status/trend and effectiveness monitoring. Section 4 discusses issues associated with sampling design, emphasizing how one selects a sample and how to minimize measurement error. Section 5 examines how sampling should occur at different spatial scales. Section 6 describes the importance of classification and identifies a suite of classification variables. Section 7 identifies and describes biological and physical/environmental indicators, while Section 8 identifies methods for measuring each indicator variable. These sections provide the foundation for implementing an efficient monitoring plan in the Upper Columbia Basin. The last section deals with how the program will be implemented. It provides a checklist of questions that need to be addressed in order to implement a valid plan. The appendices attached to this document describe how the plan will be implemented within each of the major subbasins or monitoring zones (see Section 2) within the Upper Columbia Basin.

---

[2] Although this plan targets ESA-listed fish species (i.e., spring Chinook, summer steelhead, and bull trout), it is also applicable to other non-listed species.

[3] An efficient monitoring plan reduces "error" to the maximum extent possible. One can think of error as unexplained variability (see Section 4.3), which can reduce monitoring efficiency through the use of invalid statistical designs, biased sampling designs, poorly selected indicators, biased measurement protocols, and non-standardized reporting methods.

As much as possible discussions have been kept fairly general. Because this report discusses some issues that are quite involved, footnotes are used to define technical terms, offer further explanation, offer alternative explanations, or to describe a given topic or thought in more detail. It is hoped the reader will not be too distracted by the extensive use of footnotes. In some instances, it was necessary to provide considerable detail within the text (e.g., discussion on choosing sample sizes).

As a final note, this document does not include a detailed Quality Assurance/Quality Control (QA/QC) Plan or data management plan. Although the monitoring strategy includes a description of recommended sampling and experimental designs, indicators, and a general description of sampling protocols, it does not address in detail an evaluation of data, quality control,[4] or qualifications and training of personnel. These are important components of a valid monitoring program that will be developed after the monitoring strategy is finalized.

[4] Quality control refers to specific actions required to provide information for the quality assurance program. Included are standardizations, calibration, replicates, and control and check samples suitable for statistical estimates of confidence of the data.

SECTION 2: PROJECT AREA

This monitoring plan will be implemented within the Upper Columbia River Basin, which includes all tributaries and the Columbia River between the Yakima River and Chief Joseph Dam (Figure 1). This area forms part of the larger Columbia Basin Ecoregion (Omernik 1987). The Wenatchee and Entiat rivers are in the Northern Cascades Physiographic Province, and the Okanogan and Methow rivers are in the Okanogan Highlands Physiographic Province. The geology of these provinces is somewhat similar and very complex, developed from marine invasions, volcanic deposits, and glaciation. The river valleys in this region are deeply dissected and maintain low gradients except in extreme headwaters. The climate includes extremes in temperatures and precipitation, with most precipitation falling in the mountains as snow. Melting snowpack, groundwater, and runoff maintain stream flows in the area. Mullan et al. (1992) described this area as a harsh environment for fish and stated that "it should not be confused with more studied, benign, coastal streams of the Pacific Northwest."

The Upper Columbia River Basin consists of five major "subbasins" (Wenatchee, Entiat, Chelan, Methow, and Okanogan basins) and several smaller watersheds (Figure 1). This area captures the distribution of the Upper Columbia River Basin Summer Steelhead (listed as endangered in 1996 and reclassified as threatened in 2006). It also captures the Upper Columbia River Spring Chinook Salmon Evolutionarily Significant Unit (ESU) (listed as endangered in 1999) and the Upper Columbia Recovery Unit for the Columbia River Bull Trout Distinct Population (listed as threatened in 1998). Recently, the Interior Columbia Basin Technical Recovery Team identified independent populations of summer steelhead and spring Chinook within the Upper Columbia ESUs (ICBTRT 2003). They identified three independent populations of spring Chinook within the Upper Columbia ESU; Wenatchee, Entiat, and Methow populations. For summer steelhead, they identified five independent populations within the ESU; Wenatchee, Entiat, Methow, Okanogan, and Crab Creek populations. Although they identified five geographic areas for the independent populations of steelhead within the ESU, steelhead may also exist within smaller tributaries to the Columbia River, such as Squilchuck, Stemilt, Colockum, Tarpiscan, Tekison, Quilomene/Brushy, Palisade, Douglas, Foster, and Swakane creeks, and the Chelan River and tailrace[5]. These tributaries to the Upper Columbia River will be included in the monitoring plan.

For the purpose of monitoring status and trend of habitat conditions and fish populations, this plan divides the Upper Columbia Basin into five, "status/trend monitoring zones." Four of these zones include the Wenatchee, Entiat, Methow, and Okanogan subbasins. These zones comport with the geographic locations of independent populations of summer steelhead and spring Chinook. Except for the Okanogan, they also correspond with bull trout core areas (USFWS 2002). The fifth zone captures the Chelan River/tailrace and all the smaller watersheds that drain into the Columbia River.

[5] A few steelhead spawn in the Chelan tailrace. At the present time, steelhead do not spawn in the Chelan River (bypass) because there is no year-round flow there. However, Chelan County Public Utility District proposes to provide year-round flows to the Chelan River so that steelhead and Chinook can spawn in the river downstream from the natural barriers. In addition, they intend to provide suitable spawning and early rearing habitat for steelhead and summer/fall Chinook in the river downstream from the natural barriers in the river.

Each zone will be monitored independently. That is, a fixed number of sites will be selected probabilistically within each of the five monitoring zones. Sites selected in one zone will not affect the number or locations of sites selected in another zone. This approach will provide information necessary to assess recovery of independent ESA-listed populations within the Upper Columbia Basin.

One can think of the mainstem Columbia River as a sixth status/trend monitoring zone. Although this plan does not describe in detail monitoring within the mainstem Columbia River, annual work conducted by the mid-Columbia Public Utility Districts assesses numbers of upstream-migrating adults and index numbers and conditions of juvenile migrants. Juvenile survival will be assessed for at least three years in the mainstem Columbia River. These studies result from Relicensing Agreements and Hydro Power Habitat Conservation Plans.

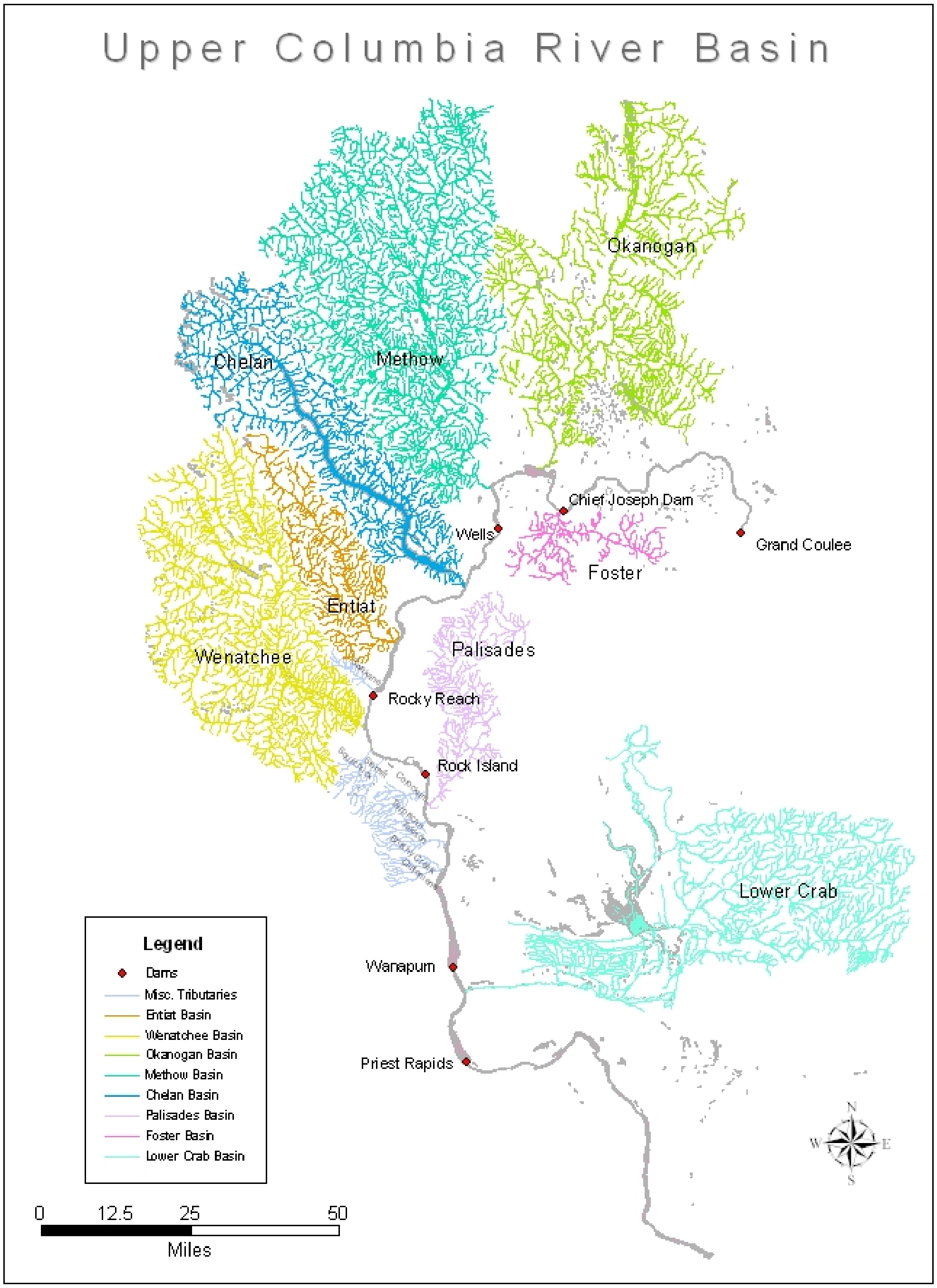

Figure 1. Major tributaries within the Upper Columbia River Basin.

SECTION 3: STATISTICAL DESIGN

This document defines "statistical design" as the logical structure of a monitoring study. It does not necessarily mean that all studies require rigorous statistical analysis. Rather, it implies that all studies, regardless of the objectives, must be designed with a logical structure that reduces bias and the likelihood that rival hypotheses are correct.[6] This section takes a classical, frequentist approach that relies on confidence intervals and *P* (probability) values from statistical tests of hypotheses. The Bayesian approach for interpreting probabilities and making decisions is not considered in this report. The logic of designing a valid monitoring program, however, is relatively independent of the debate between frequentists and Bayesians and my preference is for the classical methods.

The purpose of this section is two-fold. First, it identifies the minimum requirements of valid statistical designs and second it identifies the appropriate designs for status/trend and effectiveness monitoring. The following discussions draw heavily on the work of Hairston (1989), Thompson et al. (1998), Hicks et al. (1999), Krebs (1999), Manly (1992, 2001), Downes et al. (2002), Hillman and Giorgi (2002), and Roni (2005).

The "validity" of monitoring designs is mentioned throughout this report. The validity of a monitoring design is influenced by the degree to which the investigator can exercise experimental control; that is, the extent to which rival variables or hypotheses can be controlled or dismissed. Experimental control is associated with randomization, manipulation of independent variables, sensitivity of dependent (indicator) variables to management activities (treatments), and sensitivity of instruments or observations to measure changes in indicator variables. There are two criteria for evaluating the validity of any effectiveness monitoring design: (1) does the study infer a cause-and-effect relationship (*internal validity*) and (2) to what extent can the results of the study be generalized to other populations or settings (*external validity*)? Ideally, when assessing cause-and-effect, the investigator should select a design strong in both internal and external validity. With some thought, one can see that it becomes difficult to design a study with both high internal and external validity.[7] Because the intent of effectiveness monitoring is to demonstrate a treatment effect, the study should err on the side of internal validity. Without internal validity the data are difficult to interpret because of the confounding effects of uncontrolled variables. Listed below are some common threats to validity.

- Sampling units that change naturally over time, but independently of the treatment, can reduce validity. For example, fine sediments within spawning gravels may decrease naturally over time independent of the treatment. Alternatively, changes in land-use activities upstream from the study area and unknown to the investigator may cause levels of fine sediments to change independent of the treatment.
- The use of unreliable or inconsistent sampling methods or measuring instruments can reduce validity. That is, an apparent change in an indicator variable may actually be

[6] Rival hypotheses are alternative explanations for the outcome of an experimental study. In effect, rival hypotheses state that observed changes are due to something other than the management action under investigation.

[7] Studies with high internal validity (laboratory studies) tend to have low external validity. In the same way, studies with high external validity (field studies) tend to have lower internal validity.

nothing more than using an instrument that was not properly calibrated. Changes in indicator variables may also occur if the measuring instrument changes or disturbs the sampling site (e.g., core sampling).

- Measuring instruments that change the sampling unit before the treatment is applied can reduce validity. That is, if the collection of baseline data alters the site in such a way that the measured treatment effect is not what it would be in the population, the results of the study cannot be generalized to the population.
- Differential selection of sampling units can reduce validity, especially if treatment and control sites are substantially different before the study begins. This initial difference may at least partially explain differences after treatment.
- Biased selection of treatment sites can reduce validity. The error here is that the investigator selects sites to be treated in such a way that the treatment effects are likely to be higher or lower than for other units in the population. This issue is complicated by the fact that treatment areas are often selected precisely because they are thought to be problematic.
- Loss of sampling units during the study can reduce validity. This is most likely to occur when the investigator drops sites that shared characteristics such that their absence has a significant effect on the results.
- Multiple treatment effects can reduce validity. This occurs when sampling units get more than one treatment, or the effects of an earlier treatment are present when a later treatment is applied. Multiple treatment effects make it very difficult to identify the treatment primarily responsible for causing a response in the indicator variables.
- The threats above could interact or work in concert to reduce validity.

In most cases, there are simple design elements or requirements that reduce threats to internal and external validity. What follows is a brief description of those elements.

3.1 Minimum Requirements

What are the required elements of a "valid" monitoring study? In general, the more complex the study, the more complex the requirements, but the minimum requirements include *randomization*, *replication*, *independence*, and *controls/references*.

Randomization—Randomization should be used whenever there is an arbitrary choice to be made of which units will be measured in the sampling frame, or of the units to which treatments will be assigned. The intent is that randomization will remove or reduce systematic errors (bias) of which the investigator has no knowledge. If randomization is not used, then there is the possibility of some unseen bias in selection or allocation. In some situations, complete randomization (both random selection of sampling units and random assignment of treatments) is not possible. Indeed, there will be instances where the investigator cannot randomly assign management activities to survey areas (e.g., removal of mine contaminants from a stream). In this case replication in time and space is needed to generalize inferences of cause-effect relationships.[8] Here, confidence in the inference comes from replication outside the given study area. The rule of thumb is simple: randomize whenever possible.

Replication—Replication is needed to estimate "experimental error," which is the basic unit of measurement for assessing statistical significance or for determining confidence limits. Replication is the means by which natural variability is accounted for in interpreting results. The only way to assess variability is to have more than one replicate for each treatment, including the controls (see Section 4). In the absence of replication, there is no way, without appealing to non-statistical arguments, to assess the importance of observed differences among experimental units. Depending on the objectives of the study, spatial and/or temporal replication may be necessary.

Independence—It is important that the investigator select replicates that are spatially and temporally independent. A lack of independence can confound the study and lead to "pseudoreplication" (Hurlbert 1984). The basic statistical problem of pseudoreplication is that replicates are not independent, and the first assumption of statistical inference is violated. The simplest and most common type of pseudoreplication occurs when the investigator only selects one replicate per treatment. It can be argued that case studies, where a single stream or watershed has been monitored for several years, suffer from pseudoreplication. Therefore, one might conclude that no inference is possible. However, the motive behind a single-replicate case study is different from that behind statistical inference. The primary purpose of a case study is to reveal information about biological or physical processes in the system. This information can then be used to formulate and test hypotheses using real statistical replicates. Indeed, case studies provide the background information necessary to identify appropriate management actions and to monitor their effectiveness.

Investigators need to be aware of spatial pseudoreplication and how to prevent it or deal with it. Spatial pseudoreplication can occur when sampling units are spaced close together. Sampling units close together are likely to be more similar than those spaced farther apart.[9] Spatially dependent sites are "subsamples" rather than replicates and should not be treated as independent replicates. Confounding also occurs when control sites are not independent of treatment sites. This is most likely to occur when control sites are placed downstream from treatments sites (although the reverse can also occur; see Underwood 1994). Understandably, there can be no detection of a management action if the treatment affects both the test and control sites similarly.

Similar, although less often recognized problems occur with temporal replication. In many monitoring studies it is common for sampling to be done once at each of several years or

[8] This does not mean that one cannot infer a cause-effect relationship in the study area. The point here is that without random assignment of management activities, it is questionable if results can be generalized to other sites outside the study area.

[9] A common concern of selecting sampling units randomly is that there is a chance that some sampling units will be placed next to each other and therefore will lack independence. Although this is true, if the investigator has designed the study so that it accounts for the obvious sources of variation, then randomization is always worthwhile as a safeguard against the effects of unknown factors.

seasons. Any differences among samples may then be attributed to differences among years or seasons. This could be an incorrect inference because a single sample collected each year or season does not account for within year or season variability. Take for example the monitoring of fine sediments in spawning gravels in, say, the Chiwawa River. An investigator measures fine sediments at five random locations (spatial replication) during six consecutive years during the second week of July. A simple statistical analysis of the data could indicate that mean percentages of fine sediments decreased significantly during the latter three years. The investigator may then conclude that fines differed among years.

The conclusion may be incorrect because the study lacked adequate temporal replication. Had the investigator taken samples several times during each year (thereby accounting for within year variability), the investigator may have found no difference among years. A possible reason for the low values during the last three years is because the investigator collected samples before the stream had reached baseflow (i.e., there was a delay in the time that the stream reached baseflow during the last three years compared to the first three years). The higher flows during the second week of July in the last three years prevented the deposition of fines in spawning gravels. An alternative to collecting several samples within years or seasons is to collect the annual sample during a period when possible confounding factors are the same among years. In this case, the investigator could have collected the sample each year during baseflow. The results, however, would apply only to baseflow conditions.

The use of some instruments to monitor physical/environmental indicators may actually lead to pseudoreplication in monitoring designs. This can occur when a "destructive" sampling method is used to sample the same site repeatedly. To demonstrate this point one can look at fine-sediment samples collected repeatedly within the same year. In this example, the investigator designs a study to sample five, randomly-selected locations once every month from June through November (high flows or icing preclude sampling during other months). The investigator randomly selects the week in June to begin sampling, and then samples every fourth week thereafter (systematic sampling). To avoid systematic bias, the same well-trained worker using the same equipment (McNeil core sampler) collects all samples. After compiling and analyzing the data, the investigator may find that there is no significant difference in percent fines among replicates within the year. This conclusion is tenuous because the sampling method (core sampler) disturbed the five sampling locations, possibly reducing fines that would have been measured in following surveys. A more appropriate method would have been to randomly select five new sites (without replacement) during each survey period.

Although replication is an important component of monitoring and should be included whenever possible, it is also important to understand that using a single observation per treatment, or replicates that are not independent, is not necessarily wrong. Indeed, it may be unavoidable in some field studies. What is wrong is to ignore this in the analysis of the data. There are several analyses that can be used to analyze data that are spatially or temporally dependent (see Manly 2001). Because it is often difficult to distinguish between true statistical replicates and subsamples, even with clearly defined objectives, investigators should consult with a professional statistician during the development of monitoring studies.

Controls/References—Controls are a necessary component of effectiveness research because they provide observations under normal conditions without the effects of the management action or treatment. Thus, controls provide the standard by which the results are compared.[10] The exact nature of the controls will depend on the hypothesis being tested. For example, if an investigator wishes to implement a rest-rotation grazing strategy along a stream with heavy grazing impacts, the investigator would monitor the appropriate physical/environmental indicators in both treatment (modified grazing strategy) and control (unmodified intensive grazing) sites. Because stream systems are quite variable, the study should use "contemporaneous controls." That is, both control and treatment sites should be measured at the same time.

Temporal controls can be used to increase the "power" of the statistical design. In this case the treatment sites would be measured before and after the treatment is applied. Thus, the treatment sites serve as their own controls. However, unless there are also contemporaneous controls, all before-after comparisons must assume homogeneity over time, a dubious assumption that is invalid in most ecological studies (Green 1979). Examples where this assumption *is* valid include activities that improve fish passage at irrigation diversions or screen intake structures. These activities do not require contemporaneous controls. However, a temporal control is needed to describe the initial conditions. Therefore, a before-after comparison is appropriate. The important point is that if a control is not present, it is impossible to conclude anything definite about the effectiveness of the treatment.

There is another aspect to the use of controls. "Degraded" controls that do not receive restoration efforts are insufficient, by themselves, to test the full restoration model. That is, controls tell us that our boat has left the shore, but we are uncertain if it is heading in the right direction. Also needed are comparative data from a target or "reference" condition that the restored areas should head toward. Thus it is important to evaluate the direction of change as well as the degree of change. For this task, three distinct types of "conditions" are needed: (1) those degraded areas to be restored (called treatments); (2) those starting out in a similar degraded state but that will not be restored (called controls); and (3) those representing the target state (called references). Reference conditions can come from actual stream areas representing desirable conditions or targets established through modeling, professional judgment, or regulatory standards.

---

[10] Lee (1993, pg 205) offers a quote from Tufte that adequately describes the importance of controls in study designs. Lee writes, "One day when I was a junior medical student, a very important Boston surgeon visited the school and delivered a great treatise on a large number of patients who had undergone successful operations for vascular reconstruction. At the end of the lecture, a young student at the back of the room timidly asked, 'Do you have any controls?' Well, the great surgeon drew himself up to his full height, hit the desk, and said, 'Do you mean did I not operate on half of the patients?' The hall grew very quiet then. The voice at the back of the room very hesitantly replied, 'Yes, that's what I had in mind.' Then the visitor's fist really came down as he thundered, 'Of course not. That would have doomed half of them to their death.' God, it was quiet then, and one could scarcely hear the small voice ask, 'Which half?' (Tufte 1974, p.4--attributed to Dr. E. Peacock, Jr., chairman of surgery, University of Arizona College of Medicine, in Medical World News, Sept. 1, 1974, p. 45.)"

It should be clear that the minimum requirements of valid monitoring include randomization, replication, independence, and controls/references. In some instances monitoring studies may lack one or more of these ingredients. Such studies are sometimes called "quasi-experiments." Although these studies are often used in environmental science, they have inherent problems that need to be considered during data analysis. There is no space here to discuss these problems; however, many of them are fairly obvious. The reader should consult Cook and Campbell (1979) for a detailed discussion of quasi-experimental studies.

3.2 Recommended Statistical Designs

A perfect study design would take into account all sources of variability associated with fluctuations in indicator variables. In the absence of perfection, the best approach is to use a design that accounts for all known sources of variation not directly associated with treatment (management action) differences. A reasonable rule is to use the simplest design that provides adequate control of variability. The design should also provide the desired level of precision[11] with the smallest expenditure of time and effort. A more complex design has little merit if it does not improve the performance of statistical tests or provide more precise parameter estimates. Furthermore, an efficient design usually leads to simpler data analysis and cleaner inferences. Described below are valid designs for both effectiveness and status/trend monitoring.

> **Effectiveness Monitoring**—Because effectiveness monitoring attempts to explain cause-and-effect relationships (e.g., effect of a tributary project on fish abundance), it is important to include as many elements of valid statistical design as possible. An appropriate design recommended by the Action Agencies/NOAA Fisheries (2003), ISAB (2003), and WSRFB (2003) is the Before-After-Control-Impact or BACI design (Stewart-Oaten et al. 1986, 1992; Smith et al. 1993). This type of design is also known as a Control-Treatment Paired or CTP design (Skalski and Robson 1992), or Comparative Interrupted Time Series design (Manly 1992). Although names differ, the designs are essentially the same. That is, they require data collected simultaneously at treatment and control/reference sites before and after treatment. These data are paired in the sense that the treatment and control sites are as similar as possible and sampled simultaneously. Replication comes from collecting such paired samples at a number of times (dates) both before and after treatment. Spatial replication is possible if the investigator selects more than one treatment and control/reference site.[12] The pretreatment sampling serves to evaluate success of the pairings and establishes the

[11] It is important to understand the distinction between precision and accuracy in measurements. Accuracy refers to how close the measurement is to the true value. Accurate measurements are unbiased, meaning they are neither consistently above nor below the true value. Precision refers to the agreement among a series of measurements and the degree to which these measurements can be discriminated. Accuracy is more important than precision. It is better to use an accurate instrument that is precise to only 1 decimal place than an inaccurate balance that is precise to 5 decimal places. Extra decimal places do not bring you any closer to the true value if your instrument is flawed or biased.

[12] The use of several test and control/reference sites is recommended because it reduces spatial confounding. In some instances it may not be possible to replicate treatments, but the investigator should attempt to replicate control/reference sites. These "Beyond BACI" designs and their analyses are described in more detail in Underwood (1996).

relationship between treatment and control sites before treatment. This relationship is later compared to that observed after treatment.

The success of the design depends on indicator variables at treatment and control sites "tracking" each other; that is, maintaining a constant proportionality. The design does not require exact pairing; indicators simply need to "track" each other. Such synchrony is likely to occur if similar climatic and environmental conditions equally influence sampling units (NRC 1992). Precision of the design can be improved further if treatment and control stream reaches are paired according to a hierarchical classification approach (see Section 6). Thus, indicator variables in stream reaches with similar climate, geology, geomorphology, and channel types should track each other more closely than those in reaches with only similar climates.

It is important that treatment and control/reference sites be independent; treatment at one site cannot affect indicators in another site. The NRC (1992) recommends that control data come from another stream or from an independent reach in the same stream. After the pretreatment period, sites to be treated should be selected randomly.[13] Randomization eliminates site location as a confounding factor and removes the need to make model-dependent inferences (Skalski and Robson 1992). Hence, conclusions carry the authority of a "true" experiment and will generally be more reliable and less controversial. Post-treatment observations should be made simultaneously in both treatment and control sites.

Several different statistical procedures can be used to analyze BACI designs. Manly (1992) identified three methods: (1) a graphical analysis that attempts to allow subjectively for any dependence among successive observations, (2) regression analysis, which assumes that the dependence among successive observations in the regression residuals is small enough to ignore, and (3) an analysis based on a time series model that accounts for dependence among observations. Cook and Campbell (1979) recommend using autoregressive integrated moving average models and the associated techniques developed by Box and Jenkins (1976). Skalski and Robson (1992) introduced the odd's-ratio test, which looks for a significant change in dependent variable proportions in control-treatment sites between pretreatment and post-treatment phases. A common approach, recommended by WSRFB (2003), includes analysis of difference scores. Differences are calculated between paired control and treatment sites. These differences are then analyzed for a before-after treatment effect with a two-sample t-test, Welch modification of the t-test, or with nonparametric tests like the randomization test, Wilcoxon rank sum test, or the Mann-Whitney test (Stewart-Oaten et al. 1992; Smith et al. 1993). Downes et al. (2002) provide an excellent discussion on analysis of BACI and BACI-related designs (see their Chapter 7). Choice of test depends on the type of data collected and whether those data meet the assumptions of the tests.

In some cases, the investigator will not be able to randomly assign treatments to sampling locations. Despite a lack of randomization of treatment conditions, if the treatment conditions are replicated spatially or temporally, a sound inference to effects may be

[13] In most cases treatments will not be randomly assigned to sites. Thus, the studies will be "causal-comparative," rather than "true" experimental studies.

possible. Although valid statistical inferences can be drawn to the sites or units, the authority of a randomized design is not there to "prove" cause-effect relationships. Skalski and Robson (1992) describe in detail how to handle BACI designs that lack randomization.

Status/Trend Monitoring—Because the intent of status/trend monitoring is simply to describe existing conditions and document changes in conditions over time, it does not require all the elements of valid statistical design found in effectiveness monitoring studies. For example, controls are not required in status/trend monitoring. Controls would be important if one desires to assess cause-and-effect relationships (goal of effectiveness monitoring), which is not the purpose of status/trend monitoring. However, status/trend monitoring does require temporal and spatial replication and probabilistic sampling.

Monitoring the status and trends of recovery units, Evolutionarily Significant Units (ESUs), populations, subpopulations, and habitat characteristics is an important component of the Action Agencies/NOAA Fisheries RME Plan, which will be implemented within the Upper Columbia Basin. The RME Plan calls for the implementation of the U.S. Environmental Protection Agency's Environmental Monitoring and Assessment Program (EMAP) design, which is a spatially-balanced, site-selection process developed for aquatic systems. The state of Oregon has successfully implemented an EMAP-based program for coastal coho salmon (Moore 2002). The monitoring program has also been implemented successfully in subbasins in the upper Columbia (e.g., Okanogan Basin). The monitoring program is spatially explicit, unbiased, and has reasonably high power for detecting trends. The design is sufficiently flexible to use on the scale of multiple large river basins and can be used to estimate the numbers of adult salmon returning each year, the distribution and rearing density of juvenile salmon, productivity and relative condition of stream biota, and freshwater habitat conditions. In addition, the EMAP site-selection approach supports sampling at varying spatial extents.

Specifically, EMAP is a survey design that describes current status and detects trends in a suite of indicators. These two objectives have conflicting design criteria; status is ordinarily best assessed by including as many sample units as possible, while trend is best detected by repeatedly observing the same units over time (Overton et al. 1990; Roper et al. 2003). EMAP addresses this conflict by using rotating panels (Stevens 2002). Each panel consists of a collection of sites that will have the same revisit schedule over time. For example, sites in one panel could be visited every year, sites in another revisited every five years, and sites in still another revisited every ten years. This plan recommends that each of the five status/trend monitoring zones (see Section 2) include six panels, with one panel defining sites visited every year and five panels defining sites visited on a five-year cycle (Table 1). If each panel consists of 25 independent sites, one would need a total of 150 sites within each monitoring zone. The process by which sites are selected for each panel and the statistical methods used to analyze data are described in Section 4.

Table 1. Rotating panel design for status/trend monitoring within a given status/trend monitoring zone (e.g., Wenatchee Basin). Shading indicates the years in which sites within each panel are sampled. For

example, sites in panel 1 are visited every year, while sites in panel 2 are visited only in years 1, 6, 11, and 16, assuming a 20-year sampling frame.

Panel	**Year**																			
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1																				
2																				
3																				
4																				
5																				
6																				

SECTION 4: SAMPLING DESIGN

Once the investigator has selected a valid statistical design, the next step is to select "sampling" sites. *Sampling* is a process of selecting a number of units for a study in such a way that the units represent the larger group from which they were selected. The units selected comprise a *sample* and the larger group is referred to as a *population*.[14] All the possible sampling units available within the area (population) constitute the *sampling frame*.[15] The purpose of sampling is to gain information about a population. If the sample is well selected, results based on the sample can be generalized to the population.[16] Statistical theory assists in the process of drawing conclusions about the population using information from a sample of units.

Defining the population and the sample units may not always be straightforward, because the extent of the population may be unknown, and natural sample units may not exist. For example, a researcher may exclude livestock grazing from sensitive riparian areas in a watershed where grazing impacts are widespread. In this case the management action may affect aquatic habitat conditions well downstream from the area of grazing. Thus, the extent of the area (population) that might be affected by the management action may be unclear, and it may not be obvious which sections of streams to use as sampling units.

When the population and/or sample units cannot be defined unambiguously, the investigator must subjectively choose the potentially affected area and impose some type of sampling structure. For example, sampling units could be stream habitat types (e.g., pools, riffles, or glides), fixed lengths of stream (e.g., 150-m long stream reaches), or reach lengths that vary according to stream widths (e.g., see Simonson et al. 1994). Before selecting a sampling method, the investigator must define the population, size and number of sample units, and the sampling frame.

4.1 <u>Methods of Selecting a Sample</u>

Selection of a sample is a crucial step in monitoring fish populations and physical/environmental conditions in streams. The "goodness" of the sample determines the general applicability of the results. Because monitoring studies usually require a large amount of time and money, non-

---

[14] This definition makes it clear that a "*population*" is not limited to a group of organisms. In statistics, it is the total set of elements or units that are the target of our curiosity. For example, habitat parameters will be monitored at sites selected from the *population* of all possible stream sites in the watershed.

[15] The *sampling frame* is a "list" of all the available units or elements from which the sample can be selected. The sampling frame should have the property that every unit or element in the list has some chance of being selected in the sample. A sampling frame does not have to list all units or elements in the population.

[16] The error of extrapolating from a poor sampling design is nicely summarized by Mark Twain: "In the space of one hundred and seventy six years, the Lower Mississippi has shortened itself by two hundred and forty two miles. That is an average of a trifle over one mile and a third every year. Therefore, any calm person, who is not blind or idiotic, can see that in the Old Oolitic Silurian Period, just a million years ago next November, the Lower Mississippi was upwards of one million three hundred thousand miles long, and stuck out over the Gulf of Mexico like a fishing rod. There is something fascinating about science. One gets such wholesale returns of conjecture out of such a trifling investment of fact."

representative results are wasteful. Therefore, it is important to select a method or combination of methods that increases the degree to which the selected sample represents the population. Described below are the five most commonly used sampling designs for monitoring fish populations and physical/environmental conditions: random sampling, stratified sampling, systematic sampling, cluster sampling, and multi-stage sampling. See Scheaffer et al. (1990) for a more detailed discussion of sampling methods.

> **Random sampling**—A simple random sample is one that is obtained in such a way that all units in the defined sampling frame have an equal and independent chance of being selected. Stated differently, every unit has the same probability of being selected and the selection of one unit in no way affects the selection of another unit. Random sampling is the best single way to obtain a representative sample.[17] Random sampling should lead to small and unsystematic differences between the sample and the population because differences are a function of chance and not the result of any conscious or unconscious bias on the part of the investigator. Random sampling is also required by inferential statistics. This is important because statistics permit the researcher to make inferences about populations based on the behavior of samples. If samples are not randomly selected, then one of the major assumptions of inferential statistics is violated, and inferences are correspondingly tenuous.
>
> The process of selecting a random sample involves defining the sampling frame, identifying each unit within the frame, and selecting units for the sample on a completely chance basis. If the sampling frame contains units numbered from 1 to **N**, then a simple random sample of size **n** is obtained without replacement by drawing **n** numbers one by one in such a way that each choice is equally likely.
>
> **Stratified sampling**—Stratified sampling is the process of selecting a sample in such a way that identified strata in the sampling frame are represented in the sample.[18] This sampling method addresses the criticism that simple random sampling leaves too much to chance, so that the number of sampling units in different parts of the population may not match the distribution in the population.
>
> Stratified sampling involves dividing the units in the sampling frame into non-overlapping strata, and selecting an independent random sample from each of the strata. An example would be to stratify a stream based on habitat types (i.e., pools, riffles, glides, etc.) and then randomly selecting **n** units within each habitat type. This would ensure that each habitat type is represented in the sample. There are a couple of advantages of stratified sampling: (1) if the sampling units within the strata are more similar than units in general, the estimate of the overall population mean will have a smaller standard error than a mean calculated with simple random sampling; and (2) there may be value in having separate estimates of population parameters for the different strata. Stratification requires the investigator to

---

[17] No sampling technique guarantees a representative sample, but the probability is higher for random sampling than for other methods.

[18] The number of units selected from each strata could be equal (i.e., n is the same for all strata), or the number could be proportional to the size of the strata. Equal-sized samples would be desired if one wanted to compare the performance of different strata.

consider spatial location, areas within which the population is expected to be uniform, and the size of sampling units. Generally, the choice of how to stratify is just a question of common sense.

In some situations there may be value in analyzing a simple random sample as if it were obtained by stratified random sampling. That is, one takes a simple random sample and then places the units into strata, possibly based on information gathered at the time of sampling. The investigator then analyzes the sample as if it were a stratified random sample. This procedure is known as *post-stratification*. Because a simple random sample should place sample units in different strata according to the size of those strata, post-stratification should be similar to stratified sampling with proportional allocation, provided the total sample size is reasonably large. This may be valuable particularly when the data may be used for a variety of purposes, some of which are unknown at the time of sampling.

Systematic sampling—Systematic sampling is sampling in which units are selected from a list by taking every k^{th} unit. If k = 4, one would sample every 4^{th} unit; if k = 10, one would sample every 10^{th} unit. The value of **k** depends on the size of the sampling frame (i.e., the total number of units) and the desired sample size. The major difference between systematic sampling and the methods discussed above is that all units of the population do not have an independent chance of being selected. Once the first unit is selected, all remaining units to be included in the sample are automatically determined. Nevertheless, systematic sampling is often used as an alternative to simple random sampling or stratified sampling for two reasons. First, the process of selecting sample units is simpler for systematic sampling. Second, under certain circumstances, estimates for systematic sampling may be more precise because the population is covered more evenly. Systematic sampling is not recommended if the population being sampled has some cyclic variation (e.g., regular occurrence of pools and riffles along the course of a stream). Simple random sampling and stratified sampling are not affected by patterns in the population.

Cluster sampling—Cluster sampling is sampling in which groups, not individual units, are randomly selected. Thus, cluster sampling involves sampling clusters of units rather than single units. All units of selected groups have similar characteristics. For example, instead of randomly selecting pools throughout a watershed, one could randomly select channel bed-form types (e.g., plane-bed, step-pool, etc.) within the watershed and use all the pools within those randomly-selected channel types. Cluster sampling is more convenient when the population is very large or spread out over a wide geographic area. This advantage is offset to some extent by the tendency of sample units that are close together to have similar measurements. Therefore, in general, a cluster sample of **n** units will give estimates that are less precise than a simple random sample of **n** units. Cluster sampling can be combined with stratified sampling (see Scheaffer et al. 1990 for more details).

Multi-Stage Sampling—Multi-stage sampling is sampling in which clusters or stages (and clusters within clusters) are randomly selected and then sample units are randomly selected from each sampled cluster. With this type of sampling, one regards sample units as falling within a hierarchical structure. The investigator randomly samples at each of the various levels within the structure. For example, suppose that an investigator is interested in describing changes in fine sediments in stream riffles after livestock grazing is removed from sensitive riparian areas in a large watershed. The investigator may be able to divide the watershed into different geological/geomorphic units (primary sampling units) and then divide each geological/geomorphic unit into channel types (secondary sampling unit). Finally, the investigator may divide each channel type into habitat types (e.g., pools, riffles, glides, etc.). The investigator would obtain a "three-stage" sample of riffle habitats by first randomly selecting several primary sampling units (geological/geomorphic units), next randomly selecting one or more channel types (second-stage units) within each sampled primary unit, and finally randomly selecting one or more riffles (third-stage units) from each sampled channel type. This type of sampling is useful when a hierarchic structure exists, or when it is simply convenient to sample at two or more levels.

It is important to note that some monitoring programs include a combination of sampling designs. As you will see later in this section, the EMAP approach is a combination of random and systematic sampling. Juvenile fish monitoring in the Chiwawa Basin included a combination of stratified random sampling and two-stage sampling (Hillman and Miller 2002). These complex sampling designs require an understanding of the more basic designs.

4.2 Choosing Sample Size

It is now necessary to address the question, "to have a high probability of detecting a management or treatment effect (effectiveness monitoring) or a change in current conditions (status/trend monitoring), what sample size should the investigator use?" This is one of the most important questions of a monitoring plan.[19] If the sample is too small, the results of the study may not be generalized to the population. In addition, the wrong decision may be made concerning the validity of the hypothesis. Therefore, it is important that the investigator select a sample size that will increase the validity of the hypothesis. Fortunately, there are a number of equations and tables that can assist in selecting sample sizes. Before these are considered, it is appropriate to discuss the factors that one needs to consider when selecting a total sample size.

In general, the total sample size for status/trend monitoring depends upon the population size (total number of units in the sampling frame), population variance or standard deviation, and the level of error that the investigator considers acceptable. Quite often the population standard deviation is unknown. In this situation, the investigator can replace the population standard deviation with the sample standard deviation, which may be available from previous studies (an informal "meta-analysis"). Scheaffer et al. (1990) and Browne (2001) describe methods for guessing the population standard deviation when little prior information is available.[20] The level of error is selected by the investigator and should be based on the objectives of the study. Many studies set the error at 0.05.

[19] Although the paradigm, "for a difference X, I need sample size Y" is important, perhaps an equally important paradigm is, "for a sample size Y, I get information Z." In some cases, an investigator is only able to study a small sample. In this sense, sample size is viewed not as a unique "right" number, but rather as a factor needed to assess the utility of a study (e.g., see Parker and Berman 2003).

[20] For simple random sampling, the guess is one-fourth the range of possible values. The idea being that for many distributions, the effective range is the mean plus and minus about two standard deviations. This type of approximation is often sufficient because it is only necessary to get the sample size roughly right.

Scheaffer et al. (1990) provide equations for estimating sample sizes for simple random, stratified, systematic, and cluster sampling. There are also a number of computer packages that can be used to estimate sample sizes, such as PASS 2000 (Power Analysis and Sample Size), which is produced by NCSS Statistical Software (2000), SYSTAT, and Methodologist's Toolchest, which is produced by Idea Works (1997).[21]

Effectiveness monitoring, on the other hand, almost always requires the testing of statistical hypotheses, which means that additional factors must be considered when selecting a total sample size. Indeed, statistical significance is usually the desired outcome of effectiveness monitoring (i.e., statistical significance indicates that the management action did what it was suppose to do).[22] Therefore, when selecting a total sample size for effectiveness monitoring, the investigator must carefully evaluate all the factors that influence the validity of statistical hypotheses. These factors include significance level, effect size, variability, and statistical power.[23] What follows is a brief description of each of these factors. First, however, I briefly describe the errors of inference.

There are four possible outcomes of a statistical hypothesis test. If the hypothesis of no difference (null hypothesis) is really true, then two outcomes are possible: not rejecting the null hypothesis is a correct inference, while rejecting it constitutes a Type I error. That is, a Type I error occurs when the investigator concludes that a difference between or among treatments is real when in fact it is not. Similarly, if the null hypothesis is really false, the correct inference is to reject it, and failing to do so constitutes a Type II error. To recap, a Type I error occurs when the investigator concludes that a difference is real when in fact it is not. A Type II error occurs when the investigator concludes that there is no difference when in fact a difference exists. In statistical terms, the probability of committing a Type I error is α, while the probability of a Type II error is β. The power of the test (**1-** β) is the probability of correctly rejecting the null hypothesis when it is really false.

Both types of errors can be costly in monitoring studies where management actions involve the effects of commercial activities, such as timber harvesting or road building, on stream ecosystems. For example, a Type I error may lead to unnecessary limitations on commercial activities, while a Type II error may result in the continuation of activities damaging to the stream ecosystem. While it is impossible to calculate the probability that a hypothesis is true using classical statistical tests, the probability of incurring either a Type I or a Type II error can be controlled to acceptable levels. For example, Type I error is typically limited by the conventional significance level of statistical tests to a frequency of less than five errors per 100 tests performed ("**critical** α"<0.05). In other words, a critical α of 0.05 means that if the null hypothesis was really true and the experiment was repeated many times, the null hypothesis would be rejected incorrectly in at most 5% of the replicate experiments. In contrast, "statistical power analysis" is used to estimate and limit Type II error. With this understanding, I now present a brief description of the factors that affect sample size.

---

21 The use of trade or firm names in this paper is for reader information only and does not imply endorsement by an agency of any product or service.

22 As pointed out earlier, not all effectiveness research requires the testing of statistical hypotheses. For example, improving fish passage at a culvert or irrigation diversion does not require one to test a statistical hypothesis. It does require that the results of the action comply with the desired outcome.

23 Total sample size is also affected by the choice of experimental design and statistical analysis. Because these two factors are used to explain or partition variability, they are included in the discussion on variability.

Significance Level—The significance level is a critical value of α, which is the maximum probability of a Type I error that the researcher is willing to accept. When a P-value is less than 0.05 (the usual critical value of α), the researcher rejects the null hypothesis with the guarantee that the chance is less than 1 in 20 that a true null hypothesis has been rejected. Of course, this guarantee about the probability of making a Type I error is valid only if the assumptions of the test are met. The probability of a Type I error (significance level) is completely under the control of the investigator and is inversely related to total sample size. However, increasing critical α-level is not the most effective way to reduce total sample size or to gain statistical power (Lipsey 1990). Generally one increases the significance level when the cost of Type II errors is much larger than the cost of Type I errors.

Effect size—The effect size is the size of change in the parameter of interest that can be detected by an experiment. In statistical jargon, effect size is the difference between the equality components of the null and alternative hypotheses, usually chosen to represent a biologically or practically significant difference.[24] For example, a practical significant effect size of interest might be the difference between the maximum acceptable percentage of fine sediments in spawning gravels and the current percentage of fines in spawning gravels. The investigator must select an effect size to calculate total sample size.

Selection of significant effect size can be straightforward for some designs. In the example above, the practical significant effect size was the difference between a population mean and a known constant (e.g., maximum acceptable percentage of fines in spawning gravels). Similarly, when comparing two population means or two correlation coefficients, the estimate of effect size is simply the difference between the two values. However, formulas for effect size become more complex in designs that involve many relationships among statistical parameters, such as analysis of variance or multiple regression.

In other cases the selection of an appropriate effect size is difficult because it is very subjective. Ideally the effect size to be detected should be practically significant, but quite often this value cannot be expressed quantitatively because of a lack of information. In the absence of information, Cohen (1988) proposes small, medium, and large standardized effect sizes. Standardized effect sizes include measures of variance as well as summaries of the magnitude of treatment effects. For example, the standardized effect size for the difference between two means is expressed as the effect size (μ_1-μ_2), divided by the common standard deviation (σ). According to Cohen (1988), small effects sizes [$(\mu_1-\mu_2)/\sigma = 0.2$] are subtle, medium effect sizes (0.5) are large enough to be perceived in the course of normal experience, and large effect sizes (0.8) are easily perceived at a glance. One should use caution when selecting standardized effect sizes based on Cohen. His standardized effect sizes are derived from behavioral studies, which may not represent ecological studies. In

[24] Often, statistical significance and biological significance differ. For example, a temperature difference of 0.2ºC may be significant statistically, but not biologically. On the other hand, a 1.0ºC may be biologically significant, but because of a small sample size, the difference is not significant statistically. It is important that the investigator design the study to assess biological or practical significance.

general, sample size is inversely related to effect size. In other words, a larger sample size is needed to detect a small significant effect size.

Variability—Variability is a measure of how much scores (e.g., water temperatures) differ (vary) from one another. A measure of variability simply indicates the degree of dispersion among the set of scores. If the scores are similar, there is little dispersion and little variability. If the scores are dissimilar, there is a high degree of dispersion (variability). In short, a measure of variability does nothing more than indicate the spread of scores. The variance and the standard deviation are often used to describe the variability among a group of scores. An estimate of the population variability is generally needed to calculate sample size. As indicated earlier, if the population standard deviation is not available, one can use the sample standard deviation (from other studies or pilot studies) as an estimate of the population standard deviation, or one can guess the variability using methods described in Scheaffer et al. (1990).[25] In general, the greater the variability the larger the sample size needed to detect a significant difference.

Statistical Power—Statistical power is the probability that a statistical test will result in statistical significance (Cohen 1988). More technically, statistical power (**$1-\beta$**) is the probability of detecting a specified treatment effect (management action) when it is present. Its complement, β, is the probability of a Type II error. Statistical power is directly proportional to sample size. That is, greater statistical power requires a larger sample size. Cohen (1988) suggested that experiments should be designed to have a power of 0.80 (β = 0.20). This comports with Peterman (1990) and Green (1994), who suggest that fisheries researchers should prefer β at least <0.2, or power ≥0.8. If the investigator desires to be as conservative about making Type II as Type I errors, β should equal α, or desired power = 0.95 if $\alpha = 0.05$ (Lipsey 1990).

In summary, significance level, effect size, variability, and statistical power affect the total sample size needed for most effectiveness monitoring studies. Because of the time and cost of sampling fish and physical/environmental conditions in tributary habitats, it should be the desire of the investigator to sample the minimum possible number of units. There are several ways that one can reduce sample size. One can reduce statistical power, increase effect size, decrease the variance of the observed variables, or increase the probability of making a Type I error. Although any one of these can be used to reduce the total sample size, it is not necessarily wise (or even possible) to manipulate all of them.

Alpha is completely under the control of the researcher and there may be good reasons to choose critical α-levels other than 0.05. However, changing the critical α-level is not the most effective way

[25] If there are no estimates of variability, one can use the "signal-to-noise ratio" to estimate sample size (see Green 1994). The signal-to-noise ration is the ratio of the effect size to standard deviation. This approach may be appealing because an estimate of population variability seems to disappear, as does the need to estimate it. However, this plan does not recommend using this ratio to calculate sample size because it really does matter what the standard deviation is. The standard deviation is partly natural variation, but it also contains sampling and analysis error. The latter sources of error will affect the estimate of total sample size. Furthermore, to some degree the investigator can control the size of the standard deviation (by using valid designs and selecting sensitive indicators and reliable measurements). Therefore it is best to have some estimate of population standard deviation.

to reduce sample size (Lipsey 1990). In addition, it is unwise to reduce statistical power (1-β), unless there is good reason to do so. The objective of the study should guide the value of α and β. Data snooping or exploratory research, for example, will often be more cost-effective if α is set relatively high and β relatively low, because the objective is to detect previously unknown relationships. In addition, one should consider the prior probability that each hypothesis is true. A hypothesis that seems likely to be true, based on previous work, should be treated more cautiously with respect to erroneous rejection than a hypothesis that seems less credible (Lipsey 1990). Mapstone (1995) offers a method of selecting α and β based on the relative weighting of the perceived consequences of Type I and Type II errors. It is recommended that investigators review the methods proposed in Mapstone (1995).

Increasing effect size and/or decreasing variability may be the most effective ways to reduce sample size. However, the investigator has little flexibility in selecting significant effect sizes. Effect size is based on "practical significance" or the difference between some desirable condition and current conditions. It is inappropriate to "stretch" the effect size beyond what is considered practically significant. Consequently, the investigator is left primarily with reducing variability as a means of reducing sample size. Because physical/environmental variables often exhibit large variances, strategies for reducing variability are especially important for reducing sample size (and achieving high statistical power). Variability is generally reduced by improving measurement precision, selecting dependent (indicator) variables that are sensitive to the management action, and by various techniques of experimental design (e.g., blocking,[26] stratification, or covariate analysis). Later sections of this report identify sensitive indicator variables (Section 7) and reliable methods for measuring those variables (Section 8).

There are a number of aids that the investigator can use to estimate total sample size. Cohen (1988) provides tables and equations for calculating sample sizes. Various computer packages also estimate sample sizes, such as PASS 2000, SYSTAT, and Methodologist's Toolchest. It is recommended that the investigator use the method that meets their particular needs.

4.3 Measurement Error

Measurements and estimates are never perfect. Indeed, most fish population and habitat variables are difficult to measure, and the errors in these measurements are often large. It is tempting to ignore these errors and proceed as though the estimates reflect the true state of the resource. One should resist this temptation because it could lead to missing a treatment effect, resulting in a waste of money and effort. Investigators need to be aware of the types of errors and how they can be identified and minimized. This is important because total sample size and statistical power are related to variability. By reducing measurement error and bias, one effectively reduces variability, resulting in greater statistical power. This section identifies and describes the various types of errors and describes ways to minimize these errors.

[26] Although un-replicated random block designs are useful methods of reducing variability, they are not recommended for monitoring tributary conditions because they fail to deal with interactions between treatments (management actions) and blocks. The assumption of no interaction is unrealistic in environmental studies (Underwood 1994).

In general, "error" indicates the difference between an estimated value (from a sample) and its "true" or "expected" value. The two common types of error are *random error* and *systematic error*. Random error (a.k.a. chance error) refers to variation in a score or result that displays no systematic *bias*[27] when taking repeated samples. In other words, random error is the difference between the estimate of a population parameter that is determined from a random sample and the true population value, absent any systematic bias. One can easily detect the presence of random errors by simply repeating the measurement process several times under similar conditions. Different results, with no apparent pattern to the variation (no bias) indicate random error. Although random errors are not predictable, their properties are understood by statistical theory (i.e., they are subject to the laws of probability and can be estimated statistically). The standard deviation of repeated measurements of the same phenomenon gauges the average size of random errors.[28]

Random errors can occur during the collection and compilation of sample data. These errors may occur because of carelessness in recording field data or because of missing data. Recording errors can occur during the process of transferring information from the equipment to field data sheets. This often results from misplacing decimal points, transposing numbers, mixing up variables, or misinterpreting hand-written records. Although not always the fault of the investigator, missing data are an important source of error.

Systematic errors or bias, on the other hand, are not subject to the laws of probability and cannot be estimated or handled statistically without an independent estimate of the bias. Systematic errors are present when estimates consistently over or underestimate the true population value. An example would be a poorly calibrated thermometer that consistently underestimates the true water temperature. These errors are often introduced as a result of poorly calibrated data-recording instruments, miscoding, misfiling of forms, or some other error-generating process. They may also be introduced via interactions among different variables (e.g., turbidity is usually highest at high flows). Systematic error can be reduced or eliminated through quality control procedures implemented at the time data are collected or through careful checking of data before analysis. For convenience, systematic errors are divided into two general classes: those that occur because of inadequate procedures and those that occur during data processing. Each is considered in turn.

Biased Procedures—A biased procedure involves problems with the selection of the sample, the estimation of population parameters, the variables being measured, or the general operation of the survey. For example, selecting sample units based on access can increase systematic error because the habitat conditions near access points may not represent the overall conditions of the population. Changing sampling times and sites during the course of a study can introduce systematic error. Systematic errors can grow imperceptibly as equipment ages or observers change their perspectives (especially true of "visual" measurements). Failure to calibrate equipment introduces error, as does demanding more

27 *Bias* is a measure of the divergence of an estimate (statistic) from the population parameter in a particular direction. The greater the divergence the greater the bias. Nonrandom sampling often produces such bias.

28 It is important not to confuse standard deviation with standard error. The *standard error of a sample average* gauges the average size of the fluctuation of means from sample to sample. The *sample standard deviation* gauges the average size of the fluctuations of the values within a sample. These two quantities provide different information.

accuracy than can be expected of the instrument or taking measurements outside the range of values for which the instrument was designed.

Processing Errors—Systematic errors can occur during compiling and processing data. Errors can occur during the transfer of field records to computer spreadsheets. Investigators can also introduce large systematic errors by using faulty formulas (e.g., formulas for converting variables). Processing errors are the easiest to control.

The investigator must consider all these sources of error and develop a plan (quality control plan) that minimizes measurement bias. Certainly some errors are inevitable, but a substantial reduction in systematic errors will benefit a monitoring study considerably. The following guidelines will help to reduce systematic errors.

(1) Measures based on counts (e.g., Redds, LWD, Pools)

- Make sure that new personnel are trained adequately by experienced workers.
- Reduce errors by taking counts during favorable conditions and by implementing a rigorous protocol.
- If an over or underestimate is assumed, attempt to assess its extent by taking counts of populations of known size.

(2) Measures based on visual estimates (e.g., snorkel surveys, bank stability)

- Make sure that all visual estimates are conducted according to rigorous protocols by experienced observers.
- Attempt to assess observer bias by using trained personnel to check observations of new workers.

(3) Measures based on instruments (e.g., dissolved oxygen, temperature)

- Calibrate instruments before first use and periodically thereafter.
- Personnel must be trained in the use of all measuring devices.
- Experienced workers should periodically check measurements taken by new personnel.
- Use the most reliable instruments.

(4) Re-measurement of indicators

- Use modern GPS technology, photographs, permanent station markers (e.g., orange plastic survey stakes or rebar[29]), and carefully marked maps and diagrams to relocate previous sampling units.
- Guard against the transfer of errors from previous measurements.
- Make sure that bias is not propagated through the use of previous measurements as guides to subsequent ones.

[29] Metal detectors can be used to relocate rebar.

(5) Handling of data

- Record data directly into electronic form where possible.
- Back-up all data frequently
- Design manual data-recording forms and electronic data-entry interfaces to minimize data-entry errors.
- Use electronic data-screening programs to search for aberrant measurements.
- Frequently double-check the transfer of data from field data forms to computer spreadsheets.

Before leaving this discussion, it is important to describe briefly how one should handle outliers. Outliers are measurements that look aberrant (i.e., they appear to lie outside the range of the rest of the values). Outliers increase the variance in the data. Inflated variances decrease the power of the test and increase the chance of a Type II error (failure to reject a false null hypothesis). Because they stand apart from the others, it appears as if the investigator made some gross measurement error. It is tempting to discard them not only because they appear unreasonable, but because they also draw attention to possible deficiencies in the measurement process.[30] Before discarding an apparent outlier, the investigator should look thoroughly at how they were generated. Quite often apparent outliers result from simple errors in data recording, such as a misplaced decimal point. On the other hand, they may be part of the natural variability of the system and therefore should not be ignored or discarded.[31] If one routinely throws out aberrant values, the resulting data set will give false impressions of the structure of the system. Therefore, as a general rule, investigators should not discard outliers unless it is known for certain that measurement errors attend the estimates.

4.4 Recommended Sampling Designs

Using the basic tools described above, valid sampling designs can be identified for status/trend and effectiveness monitoring in the Upper Columbia Basin. The recommended sampling designs, if implemented correctly, should reduce bias and error.

Effectiveness Monitoring—This plan recommends that sampling units for effectiveness monitoring be selected according to a stratified random sampling design. The plan requires that streams or stream segments to be treated with some action(s) will be classified according to a hierarchical classification system (see Section 6). Once classification identifies non-overlapping strata, sampling sites are then selected randomly within each stratum. The same process occurs within control or reference areas, which are similar to treatment areas based on classification. The number of sites selected will depend on effect size, variability, power, and significance levels. The number of sites within each stratum should be proportional to the size of the stratum. That is, a larger stratum would receive more sites than a smaller

30 Deleting observations simply because they are "messy" is laundering or doctoring the results and legitimately could be considered scientific fraud.

31 Outliers can reflect real biological processes, and careful thought about their meaning can lead to new ideas and hypotheses. Interesting ecology and evolution often happens in the statistical tails of a distribution.

stratum. The investigator will record the location of the upstream and downstream end of each sampling site with GPS, photographs, permanent station markers, and site diagrams.

<u>Status/Trend Monitoring</u>—Because the plan follows EMAP, which requires spatially balanced samples, sites will be selected according to the generalized random tessellation stratified design (GRTS) (Stevens 1997; Stevens and Olsen 1999; Stevens and Urquhart 2000; Stevens 2002). Briefly, the GRTS design achieves a random, nearly regular sample point pattern via a random function that maps two-dimensional space onto a one-dimensional line (linear space). A systematic sample is selected in the linear space, and the sample points are mapped back into two-dimensional space. The GRTS design is used to select samples for all panels.

As a starting point, this plan recommends a sample size of 25 sites per panel. This means that GRTS will select a total of 150 sites (6 panels x 25 sites per panel = 150 sites) for each of the five status/trend monitoring zones, regardless of the size of the zones. Two panels of sites will be monitored each year (see Section 3.2), resulting in a total of 50 sites sampled annually within each zone. Some of the sites may fall in areas that are physically inaccessible or cannot be accessed because of landowner denial. Therefore, GRTS will select an additional 150 sites (100% oversample) for each zone, any one of which can replace an inaccessible site.

Within each monitoring zone, the sampling frame for the 150 sites (and the 150 oversample sites) will consist of portions of first through fifth-order[32] streams (based on 1:25,000 scale USGS topographic maps) with reach gradients less than 12%[33]. These stream segments were selected because most spawning and rearing of salmonids occur in these areas (Roni et al. 1999). However, spawning and rearing of resident and anadromous salmonids are not evenly distributed among stream orders. Resident populations may occur above natural barriers that preclude anadromous salmonids. Therefore, this plan recommends that sites be selected from stream orders supporting both anadromous and resident populations. As a starting point, 72% of the sites (18 sites per panel) will fall within anadromous fish streams; 28% (7 sites per panel) within resident fish streams. Table 2 shows the distribution of sites among stream orders for both anadromous and resident fish.

In order to estimate precision, 10% of the sites within each of the five monitoring zones will be sampled by two independent crews each year for five years. This means that each year, five randomly selected sites within each zone will be surveyed by two different crews. Sampling by the two independent crews will be no more than two-days apart. This will minimize the effects of site changes on estimates of precision. These sites will also be used to compare protocols (e.g, comparison of the protocols in this plan with the Hankin-Reeves methods currently used by the Forest Service).

---

32 Stream order is based on Strahler (1952). This method of ordering streams is described in Gordon et al. (1992).

33 Here, a reach is defined as a 300-m long stretch of stream. Therefore, all 300-m long reaches with a sustained gradient of >12% will be excluded from the sampling frame.

Data collected within the EMAP design will be analyzed according to the statistical protocols outlined in Stevens (2002). The Horvitz-Thompson or π-estimator is recommended for estimation of population status. Multi-phase regression analyses are recommended for estimating the distribution of trend statistics. These approaches are explained fully in Diaz-Ramos et al. (1996) and Stevens (2002).

Table 2. Distribution of sample sites among stream orders for resident and anadromous populations.

Stream order	Number of sites per panel	Percent
Anadromous 1	2	8
Anadromous 2	4	16
Anadromous 3	5	20
Anadromous 4	5	20
Anadromous 5	2	8
Resident 1	4	16
Resident 2	2	8
Resident 3	1	4

SECTION 5: SAMPLING AT DIFFERENT SPATIAL SCALES

Because monitoring will occur at a range of spatial scales, there may be some confusion between the roles of status/trend monitoring and effectiveness monitoring. Generally, one thinks of status/trend monitoring as monitoring that occurs at coarser scales and effectiveness monitoring at finer scales. In reality, both occur across different spatial scales, and the integration of both is needed to develop a valid monitoring program (ISAB 2003; AA/NOAA Fisheries 2003; WSRFB 2003).

The scale at which status/trend and effectiveness monitoring occurs depends on the objectives of the study, the size or distribution of the target population, and the indicators that will be measured. In status/trend monitoring, for example, the objective may be to measure egg-parr survival of spring Chinook salmon in the Wenatchee Basin. Because the Wenatchee Basin consists of one population of spring Chinook (ICBTRT 2003), the entire basin is the spatial scale at which egg-parr survival is monitored. In contrast, if the objective is to assess egg-parr survival of spring Chinook in the Chiwawa Basin (a sub-group of the Wenatchee population), the spatial scale at which monitoring occurs includes only the Chiwawa Basin, a much smaller area than the entire Wenatchee Basin. Thus, status/trend monitoring can occur at various scales depending on the distribution of the population of interest.

In the same way, effectiveness monitoring can occur at different spatial scales. That is, one can assess the effect of a tributary action on a specific Recovery Unit or ESU (which may encompass several populations), a specific population (may include several sub-populations), at the sub-group level (may encompass a watershed within a basin), or at the reach scale. Clearly, the objectives and hence the indicators measured dictate the spatial scale at which effectiveness monitoring is conducted. For example, if the objective is to assess the effects of nutrient enhancement on egg-smolt survival of spring Chinook in the Chiwawa Basin (a sub-group of the Wenatchee spring Chinook population), then the spatial scale covered by the study must include the entire area inhabited by the eggs, fry, parr, and smolts. If, on the other hand, the objective is to assess the effects of a sediment reduction project on egg-fry survival of a local group of spring Chinook (i.e., Chinook within a specific reach of stream), then the study area would only encompass the reach of stream used by spawners of that local group.

In theory there might be no limit to the scale at which effectiveness monitoring can be applied, but in practice there is a limit. This is because as the spatial scale increases, the tendency for multiple treatments (several habitat actions) affecting the same population increases (Table 3). That is, at the spatial scale representing a Recovery Unit, ESU, or population, there may be many habitat actions within that area. Multiple treatment effects make it very difficult to assess the effects of specific actions on an ESU (see Section 3). Even though it may be impossible to assess specific treatment effects at larger spatial scales, it does not preclude one from conducting effectiveness monitoring at this scale. Indeed, one can assess the combined or cumulative effects of tributary actions on the Recovery Unit, ESU, or population. However, additional effectiveness monitoring may be needed at finer scales to assess the effects of individual actions on the ESU or population.

Table 3. Relationship between biological indicators, spatial scales, and our ability to assess effects of specific management actions. Examples of each scale are shown in parentheses. Table is modified from Action Agencies/NOAA Fisheries RME Plan (2003).

Biological Indicators	Example of spatial scales	Ability to assess effects of specific tributary actions
ESU (Upper Columbia Spring Chinook) ↓	**Basins** (Upper Columbia) ↓	**Low**
Population (Wenatchee Spring Chinook) ↓	**Basin** (Wenatchee) ↓	↓
Sub-Population (Chiwawa River Spring Chinook) ↓	**Watershed** (Chiwawa River) ↓	
Local Group	**Reach** (5 km of the Chiwawa River)	**High**

Given the potential problems of multiple treatment effects, there are two general strategies for conducting effectiveness monitoring at different spatial scales. One strategy is a "project-based" approach, which addresses the effects of individual tributary projects at smaller spatial scales (e.g., stream or stream reach). This approach is identified in the Action Agencies/NOAA Fisheries Plan as the "Bottom-Up" approach. It is designed to assess the effects of specific projects in isolation of other tributary actions. That is, results from this type of effectiveness monitoring would not be confounded by actions occurring elsewhere in the basin. This approach requires that the investigator maintain control of all actions that occur within the assessment area (stream, watershed, or basin).

The second strategy is an "intensive" approach that addresses the cumulative effects of tributary actions at larger spatial scales (e.g., watershed or basin). This approach is identified in the Action Agencies/NOAA Fisheries Plan as the "Top-Down" Approach. The WSRFB (2003) refers to it as "Intensive (Validation) Monitoring." This approach requires intensive and extensive sampling of several indicator variables within the watershed or basin. Although the effects of individual projects on fish populations may not be assessed unequivocally, their cumulative effects can be measured.

Both approaches (project-based and intensive) require valid statistical and sampling designs. That is, both approaches require controls (reference conditions), replication, and probabilistic sampling. This plan recommends the use of BACI designs (see Section 3) with stratified random sampling (see Section 4) for both approaches. Both approaches will likely be implemented within subbasins in the Upper Columbia Basin.

SECTION 6: CLASSIFICATION

Both status/trend and effectiveness monitoring require landscape classification. The purpose of classification is to describe the "setting" in which monitoring occurs. This is necessary because biological and physical/environmental indicators may respond differently to tributary actions depending on landscape characteristics. An hierarchical classification system that captures a range of landscape characteristics should adequately describe the setting in which monitoring occurs. The idea advanced by hierarchical theory is that ecosystem processes and functions operating at different scales form a nested, interdependent system where one level influences other levels. Thus, an understanding of one level in a system is greatly informed by those levels above and below it.

A defensible classification system should include both ultimate and proximate control factors (Naiman et al. 1992). Ultimate controls include factors such as climate, geology, and vegetation that operate over large areas, are stable over long time periods, and act to shape the overall character and attainable conditions within a watershed or basin. Proximate controls are a function of ultimate factors and refer to local conditions of geology, landform, and biotic processes that operate over smaller areas and over shorter time periods. These factors include processes such as discharge, temperature, sediment input, and channel migration. Ultimate and proximate control characteristics help define flow (water and sediment) characteristics, which in turn help shape channel characteristics within broadly predictable ranges (Rosgen 1996).

This plan proposes a classification system that incorporates the entire spectrum of processes influencing stream features and recognizes the tiered/nested nature of landscape and aquatic features. This system captures physical/environmental differences spanning from the largest scale (regional setting) down to the channel segment (Table 4). The Action Agencies/NOAA Fisheries RME plan proposes a similar classification system. By recording these descriptive characteristics, the investigator will be able to assess differential responses of indicator variables to proposed actions within different classes of streams and watersheds. Importantly, the classification work described here fits well with Level 1 monitoring under the ISAB (2003) monitoring and evaluation plan. Classification variables and recommend methods for measuring each variable are defined below.

Table 4. List of classification (stratification) variables, their corresponding measurement protocols, and temporal sampling frequency. The variables are nested according to spatial scale and their general characteristics. Table is modified from Action Agencies/NOAA Fisheries RME Plan (2003).

Spatial scale	General characteristics	Classification variable	Recommended protocols	Sampling frequency (years)
Regional setting	Ecoregion	Bailey classification	Bain and Stevenson (1999)	20
		Omernik classification	Bain and Stevenson (1999)	20
	Physiography	Province	Bain and Stevenson (1999)	20
	Geology	Geologic districts	Overton et al. (1997)	20
Drainage basin	Geomorphic features	Basin area	Bain and Stevenson (1999)	20
		Basin relief	Bain and Stevenson (1999)	20
		Drainage density	Bain and Stevenson (1999)	20
		Stream order	Gordon et al. (1992)	20
Valley segment	Valley characteristics	Valley bottom type	Cupp (1989); Naiman et al. (1992)	20
		Valley bottom width	Naiman et al. (1992)	20
		Valley bottom gradient	Naiman et al. (1992)	20
		Valley containment	Bisson and Montgomery (1996)	20
Channel segment	Channel characteristics	Elevation	Overton et al. (1997)	10
		Channel type (Rosgen)	Rosgen (1996)	10
		Bed-form type	Bisson and Montgomery (1996)	10
		Channel gradient	Overton et al. (1997)	10
	Riparian veg.	Primary vegetation type	Platts et al. (1983)	5

As noted above, all watersheds that will be monitored will be classified according to their landscape characteristics. Table 4 lists the "core" set of classification variables. Investigators may elect to describe additional classification variables depending on the objectives of the study. Only a general description of each classification variable is provided here. Because time and space do not allow for a detailed description of methods, this plan only identifies recommended methods and instruments. The reader should refer to the cited documents for detailed descriptions of methods and measuring instruments.

The classification work described here relies heavily on remote-sensed data and GIS. The majority of this work can be conducted in an office with GIS. It is important, however, to spend some time in the field verifying spatial data. This plan recommends that at least 10% of the channel segments identified in a subbasin be verified in the field. These segments can be selected randomly. Additional verification may be needed for those segments that cannot be accurately delineated from remote-sensed data. Variables such as primary riparian vegetation type, channel type, and bed-form type will be verified during field surveys (described in Section 8).

6.1 Regional Setting

Ecoregions

Ecoregions are relatively uniform areas defined by generally coinciding boundaries of several key geographic variables. Ecoregions have been defined holistically using a set of physical and biotic factors (e.g., geology, climate, landform, soil, vegetation, and water). Of the systems available, this plan includes the two most commonly used ecoregion systems, Bailey (1978) and Omernik (1987). Bailey's approach uses macroclimate and prevailing plant formations to classify the continent into various levels of detail. Bailey's coarsest hierarchical classifications include domains, divisions, provinces, and sections. These regional classes are based on broad ecological climate zones and thermal and moisture limits for plant growth (Bailey 1998). Specifically, domains are groups of related climates, divisions are types of climate based on seasonality of precipitation or degree of dryness or cold, and provinces are based on macro features of vegetation. Provinces include characterizations of land-surface form, climate, vegetation, soils, and fauna. Sections are based on geomorphology, stratigraphy and lithology, soil taxa, potential natural vegetation, elevation, precipitation, temperature, growing season, surface water characteristics, and disturbance. Information from domains, divisions, and provinces can be used for modeling, sampling, strategic planning, and assessment. Information from sections can be used for strategic, multi-forest, statewide, and multi-agency analysis and assessment.

The system developed by Omernik (1987) is used to distinguish regional patterns of water quality in ecosystems as a result of land use. Omernik's system is suited for classifying aquatic ecoregions and monitoring water quality because of its ecological foundation, its level of resolution, and its use of physical, chemical, and biological information. Like Bailey's system, this system is hierarchical, dividing an area into finer regions in a series of levels. These levels are based on characterizations of land-surface form, potential natural vegetation, land use, and soils. Omernik's system has been extensively tested and found to correspond well to spatial patterns of water chemistry and fish distribution (Whittier et al. 1988).

Until there is a better understanding of the relationships between fish abundance/distribution and the two classes of ecoregions, investigators should use both classifications. Chapter 3 in Bain and Stevenson (1999) outlines protocols for describing ecoregions. Published maps of ecoregions are available to assist with classification work.[34] This work will be updated once every 20 years.

Physiographic Province

Physiographic province is the simplest division of a land area into hierarchical natural regions. In general, delineation of physiographic provinces is based on topography (mountains, plains, plateaus, and uplands) and, to a lesser extent, climate, which governs the processes that shape the landscape (weathering, erosion, and sedimentation). Specifically, provinces include descriptions of climate,

---

34 Bailey's digital-compressed ARC/INFO ecoregion maps are available at http://www.fs.fed.us/institute/ecolink.html. Omernik's digital level III ecoregion maps of the conterminous U.S. are available at http://www.epa.gov/OST/BASINS/gisdata.html (download BASINS core data) with documentation at http://www.epa.gov/envirofw/html/nsdi/nsditxt/useco.txt.

vegetation, surficial deposits and soils, water supply or resources, mineral resources, and additional information on features particular to a given area (Hunt 1967). Physiographic provinces and drainage basins have traditionally been used in aquatic research to identify fish distributions (Hughes et al. 1987; Whittier et al. 1988).

Chapter 3 in Bain and Stevenson (1999) outlines methods for describing physiographic provinces. Physiographic maps are available to aid classification work.[35] Investigators will update physiographic provinces once every 20 years.

Geology

Geologic districts are areas of similar rock types or parent materials that are associated with distinctive structural features, plant assemblages, and similar hydrographic character. Geologic districts serve as ultimate controls that shape the overall character and attainable conditions within a watershed or basin. They are corollary to subsections identified in the U.S. Forest Service Land Systems Inventory (Wertz and Arnold 1972). Watershed and stream morphology are strongly influenced by geologic structure and composition (Frissell et al. 1986; Nawa et al. 1988). Structural features are the templates on which streams etch drainage patterns. The hydrologic character of landscapes is also influenced by the degree to which parent material has been weathered, the water-handling characteristics of the parent rock, and its weathering products. Like ecoregions, geologic districts do not change to other types in response to land uses.

Geologic districts can be identified following the methods described in Overton et al. (1997).Published geology maps aid in the classification of rock types. This work will be updated once every 20 years.

6.2 Drainage Basin

Geomorphic Features

This plan includes four important geomorphic features of drainage basins: basin area, basin relief, drainage density, and stream order. Basin area (a.k.a. drainage area or catchment area) is the total land area (km^2), measured in a horizontal plane, enclosed by a drainage divide, from which direct surface runoff from precipitation normally drains by gravity into a wetland, lake, or river. Basin relief (m) is the difference in elevation between the highest and lowest points in the basin. It controls the stream gradient and therefore affects flood patterns and the amount of sediment that can be transported. Hadley and Schumm (1961) demonstrated that sediment load increases exponentially with basin relief. Drainage density (km) is an index of the length of stream per unit area of basin and is calculated as the drainage area (km^2) divided by the total stream length (km). This ratio represents the amount of stream necessary to drain the basin. High drainage density may indicate high water yield and sediment transport, high flood peaks, steep hills, and low suitability for certain land uses (e.g., agriculture). The last geomorphic feature, stream order, is based on the premise that the order

---

[35] Detailed information about physiographic provinces of the U.S. can be found at http://www.salem.mass.edu/~lhanson/. Digital maps can be found at http://water.usgs.gov/GIS/.

number is related to the size of the contributing area, to channel dimensions, and to stream discharge. Stream ordering follows the Strahler ordering system. In that system, all small, exterior streams are designated as first order. A second-order stream is formed by the junction of any two first-order streams; third-order by the junction of any two second-order streams. In this system only one stream segment has the highest order number.

Chapter 4 in Bain and Stevenson (1999) outlines standard methods for estimating basin area, basin relief, and drainage density. Gordon et al. (1992) describes the Strahler stream-ordering method. Investigators will use USGS topographic maps (1:100,000 scale) and GIS to estimate these parameters. This work will be updated once every 20 years.

6.3 Valley Segment

Valley Characteristics

The plan incorporates four important features of the valley segment: valley bottom type, valley bottom width, valley bottom gradient, and valley confinement. Valley bottom types are distinguished by average channel gradient, valley form, and the geomorphic processes that shaped the valley (Cupp 1989a,b; Naiman et al. 1992). They correspond with distinctive hydrologic characteristics, especially the relationship between stream and alluvial ground water (Table 5). Valley bottom width is the ratio of the valley bottom[36] width (m) to active channel width (m). Valley gradient is the slope or the change in vertical elevation (m) per unit of horizontal valley distance (m). Valley gradient is typically measured in lengths of about 300 m (1,000 ft) or more. Valley confinement refers to the degree that the valley walls confine the lateral migration of the stream channel. The degree of confinement can be classified as strongly confined (valley floor width < 2 channel widths), moderately confined (valley floor width = 2-4 channel widths), or unconfined (valley floor width > 4 channel widths).

The latter three variables, valley bottom width, valley gradient, and confinement, are nested within valley bottom types. Therefore, these three variables will be described for each valley bottom type identified within the drainage basin (i.e., the valley bottom type defines the scale at which these variables are described).

Investigators should follow the methods of Cupp (1989a,b) and Naiman et al. (1992) to describe valley bottom types. Naiman et al. (1992) also describe methods for measuring valley bottom width and valley bottom gradient. Bisson and Montgomery (1996) outline methods for measuring valley confinement. GIS will aid in estimating these parameters. These variables will be updated once every 20 years.

[36] Valley bottom is defined as the essentially flat area adjacent to the stream channel.

Table 5. Examples of valley bottom types and valley geomorphic characteristics in forested lands of Washington. Table is from Naiman et al. (1992).

Valley bottom type[a]	Valley bottom gradient[b]	Sideslope gradient[c]	Valley bottom width[d]	Channel patterns	Strahler stream order	Landform and geomorphic features
F1 Estuarine delta	≤0.5%	<5%	>5X	Unconstrained; highly sinuous; often braided	Any	Occur at mouth of streams on estuarine flats in and just above zone of tidal influence
F2 Alluviated lowlands	≤1%	>5%	>5X	Unconstrained; highly sinuous	Any	Wide floodplains typically formed by present or historic large rivers within flat to gently rolling lowland landforms; sloughs, oxbows, and abandoned channels commonly associated with mainstream rivers
F3 Wide mainstream valley	≤2%	<5%	>5X	Unconstrained; moderate to high sinuosity; braids common	Any	Wide valley floors bounded by mountain slopes; generally associated with mainstream rivers and the tributary streams flowing through the valley floor; sloughs and abandoned channels common.
F4 Wide mainstream valley	≤1-3%	≤10%	>3X	Variable; generally unconstrained	1-4	Generally occur where tributary streams enter low-gradient valley floors; ancient or active alluvial/colluvial fan deposition overlying floodplains of larger, low-gradient stream segments; stream may actively downcut through deep alluvial fan deposition.
F5 Gently sloping plateaux and terraces	≤2%	<10%	1-2X	Moderately constrained; low to moderate sinuosity	1-3	Drainage ways shallowly incised into flat to gently sloping landscape; narrow active floodplains; typically associated with small streams in lowlands, cryic uplands or volcanic flanks.
M1 Moderate sloping plateaux and terraces	2-5%	<10-30%	<2X	Constrained; infrequent meanders	1-4	Constrained, narrow floodplains bounded by moderate gradient sideslopes; typically found in lowlands and foothills, but may occur on broken mountain slopes and volcano flanks.
M2 Alluviated, moderate slope bound	≤2%	<5%, gradually increase to 30%	2-4X	Unconstrained; moderate to high sinuosity	1-4	Active floodplains and alluvial terraces bounded by moderate gradient hillslopes; typically found in lowlands and foothills, but may occur on broken mountain slopes and volcano flanks.
V1 V-shaped moderate-gradient bottom	2-6%	30-70%	<2X	Constrained	≥2	Deeply incised drainage ways with steep competent sideslopes; very common in uplifted mountainous topography; less commonly associated with marine or glacial outwash terraces in lowlands and foothills.
V2 V-shaped high-gradient bottom	6-11%	30-70%	<2X	Constrained	≥2	Same as above, but valley bottom longitudinal profile steep with pronounced stair-step characteristics.

Table 5. (continued)

Valley bottom type[a]	Valley bottom gradient[b]	Sideslope gradient[c]	Valley bottom width[d]	Channel patterns	Strahler stream order	Landform and geomorphic features
V3 V-shaped, bedrock canyon	3-11%	70%+	<2X	Highly constrained	≥2	Canyon-like stream corridors with frequent bedrock outcrops; frequently stair-stepped profile; generally associated with folded, faulted or volcanic landforms.
V4 Alluviated mountain valley	1-4%	Channel adjacent slopes <10%; increase to 30%+	2-4X	Unconstrained; high sinuosity with braids and side-channels common	2-5	Deeply incised drainage ways with relatively wide floodplains; distinguished as "alluvial flats" in otherwise steeply dissected mountainous terrain.
U1 U-shaped trough	<3%	<5%; gradually increases to 30%+	>4X	Unconstrained; moderate to high sinuosity; side channels and braids common	1-4	Drainage ways in mid to upper watersheds with history of glaciation, resulting in U-shaped profile; valley bottom typically composed of glacial drift deposits overlain with more recent alluvial material adjacent to channel.
U2 Incised U-shaped valley, moderate-gradient bottom	2-5%	Steep channel adjacent slopes, decreases to <30%, then increases to >30%	<2X	Moderately contrained by unconsolidated material; infrequent short flats with braids and meanders	2-5	Channel downcuts through deep valley bottom glacial till, colluvium, or coarse glacio-fluvial deposits; cross-sectional profile variable, but generally weakly U-shaped with active channel vertically incised into valley fill deposits; immediate side-slopes composed of unconsolidated and often unsorted coarse-grained deposits.
U3 Incised U-shaped valley, high-gradient bottom	6-11%	Steep channel adjacent slopes, decreases to <30%, then increases to >30%	<2X	Moderately constrained by unconsolidated material; infrequent short flats with braids and meanders	2-5	Channel downcuts through deep valley bottom glacial till, colluvium, or coarse glacio-fluvial deposits; cross-sectional profile variable, but generally weakly U-shaped with active channel vertically incised into valley fill deposits; immediate side-slopes composed of unconsolidated and often unsorted coarse-grained deposits.
U4 Active glacial out-wash valley	1-7%	Initially <5%, increasing to >60%	<4X	Unconstrained; highly sinuous and braided	1-3	Stream corridors directly below active alpine glaciers; channel braiding and shifting common; active channel nearly as wide as valley bottom.
H1 Moderate-gradient valley wall/head-water	3-6%	>30%	<2X	Constrained	1-2	Small drainage ways with channels slightly to moderately entrenched into mountain toe-slopes or head-water basins.
H2 High-gradient valley wall/head-water	6-11%	>30%	<2X	Constrained; stair-stepped	1-2	Small drainage ways with channels moderately entrenched into high gradient mountain slopes or headwater basins; bedrock exposures and outcrops common; localized alluvial/colluvial terrace deposition.

Table 5. (concluded)

Valley bottom type[a]	Valley bottom gradient[b]	Sideslope gradient[c]	Valley bottom width[d]	Channel patterns	Strahler stream order	Landform and geomorphic features
H3 Very high-gradient valley wall/head-water	11%+	>60%	<2X	Constrained; stair-stepped	1-2	Small drainage ways with channels moderately entrenched into high gradient mountain slopes or headwater basins; bedrock exposures and out-crops common; localized alluvial/colluvial terrace deposition.

[a]Valley bottom type names include alphanumeric mapping codes in italic (from Cupp 1989a, b).
[b]Valley bottom gradient is measured in length of about 300 m (1,000 ft).
[c]Sideslope gradient characterizes the hillslopes within 1,000 horizontal and about 100 m (300 ft) vertical distance from the active channel.
[d]Valley bottom width is a ratio of the valley bottom width to active channel width.

6.4 Channel Segment

Channel Characteristics

The plan includes four important characteristics of the channel segment: elevation, channel gradient, channel type, and bed-form type. These characteristics are nested within valley bottom types and therefore should be described for each valley bottom type identified within the drainage basin. Elevation (m) is the height of the stream channel above or below sea level. Channel gradient is the slope or the change in the vertical elevation of the channel per unit of horizontal distance. Channel gradient can be presented graphically as a stream profile.

Channel type follows the classification technique of Rosgen (1996) and is based on quantitative channel morphology indices.[37] These indices result in objective and consistent identification of stream types. The Rosgen technique consists of four different levels of classification. Level I describes the geomorphic characteristics that result from the integration of basin relief, landform, and valley morphology. Level II provides a more detailed morphological description of stream types. Level III describes the existing condition or "state" of the stream as it relates to its stability, response potential, and function. Level IV is the level at which measurements are taken to verify process relationships inferred from preceding analyses. All monitoring in subbasins in the Upper Columbia Basin will include at least Level I (geomorphic characterization) classification (Table 6).

37 Indices include entrenchment, gradient, width/depth ratio, sinuosity, and dominant channel material.

Table 6. General stream type descriptions and delineative criteria for Level I channel classification. Table is from Rosgen (1996).

Stream type	General description	Entrenchment ratio	W/D ratio	Sinuosity	Slope %	Landform/ soils/features
Aa+	Very steep, deeply entrenched, debris transport, torrent streams.	<1.4	<12	1.0-1.1	>10	Very high relief. Erosional, bedrock or depositional features; debris flow potential. Deeply entrenched streams. Vertical steps with deep scour pools; waterfalls.
A	Steep, entrenched, cascading, step/pool streams. High energy/debris transport associated with depositional soils. Very stable if bedrock or boulder dominated channel.	<1.4	<12	1.0-1.2	4-10	High relief. Erosional or depositional and bedrock forms. Entrenched and confined streams with cascading reaches. Frequently spaced, deep pools in associated step/pool bed morphology.
B	Moderately entrenched, moderate gradient, riffle-dominated channel, with infrequently spaced pools. Very stable plan and profile. Stable banks.	1.4-2.2	>12	>1.2	2-4	Moderate relief, colluvial deposition, and/or structural. Moderate entrenchment and W/D ratio. Narrow, gently sloping valleys. Rapids predominate with scour pools.
C	Low gradient, meandering, point-bar, riffle/pool, alluvial channels with broad, well defined floodplains.	>2.2	>12	>1.4	<2	Broad valleys with terraces, in association with floodplains, alluvial soils. Slightly entrenched with well-defined meandering channels. Riffle/pool bed morphology.
D	Braided channel with longitudinal and transverse bars. Very wide channel with eroding banks.	n/a	>40	n/a	<4	Broad valleys with alluvium, steeper fans. Glacial debris and depositional features. Active lateral adjustment, with abundance of sediment supply. Covergence/divergence bed features, aggradational processes, high bedload and bank erosion.

Table 6. (concluded)

Stream type	General description	Entrenchment ratio	W/D ratio	Sinuosity	Slope %	Landform/ soils/features
DA	Anastomosing (multiple channels) narrow and deep with extensive, well-vegetated floodplains and associated wetlands. Very gentle relief with highly variable sinuosities and width/depth rations. Very stable streambanks.	>2.2	Highly variable	Highly variable	<0.5	Broad, low-gradient valleys with fine alluvium and/or lacustrine soils. Anastomosed (multiple channel) geologic control creating fine deposition with well-vegetated bars that are laterally stable with broad wetland floodplains. Very low bedload, high wash load sediment.
E	Low gradient, meandering riffle/pool stream with low sidth/depth ratio and little deposition. Very efficient and stable. High meander width ratio.	>2.2	<12	>1.5	<2	Broad valley/meadows. Alluvial materials with floodplains. Highly sinuous with stable, well-vegetated banks. Riffle/pool morphology with very low width/depth ratios.
F	Entrenched meandering riffle/pool channel on low gradients with high width/depth ratio.	<1.4	>12	>1.4	<2	Entrenched in highly weathered material. Gentle gradients, with a high width/depth ratio. Meandering, laterally unstable with high bank erosion rates. Riffle/pool morphology.
G	Entrenched "gully" step/pool and low width/depth ratio on moderate gradients.	<1.4	<12	>1.2	2-4	Gullies, step/pool morphology with moderate slopes and low width/depth ratio. Narrow valleys, or deeply incised in alluvial or colluvial materials, i.e., fans or deltas. Unstable, with grade control problems and high bank erosion rates.

Bed-form type follows the classification proposed by Montgomery and Buffington (1993). This technique is comprehensive and is based on hierarchies of topographic and fluvial characteristics. This system provides a geomorphic, process-oriented method of identifying valley segments and stream reaches. It employs descriptors that are measurable and ecologically relevant. Montgomery and Buffington (1993) identified three valley segment types: colluvial, alluvial, and bedrock. They subdivided the valley types into one or more stream-reach types (bed-form types) depending on whether substrates are limited by the supply of sediment or by the fluvial transport of sediment (Table 7). For example, depending on sediment supply and transport, Montgomery and Buffington (1993) recognized six alluvial bed-form types: braided, regime, pool/riffle, plane-bed, step-pool or cascade. Both colluvial and bedrock valley types consist of only one bed-form type. Only colluvial bed-forms occur in colluvial valleys and only bedrock bed-forms occur in bedrock valleys.

Table 7. Characteristics of different bed-form types. Table is modified from Montgomery and Buffington (1993).

Valley types	Bed-form types	Predominant bed material	Dominant roughness elements	Typical slope (%)	Typical confinement	Pool spacing (channel widths)
Colluvial	Colluvial	Variable	Boulders, large woody debris	>20	Strongly confined	Variable
Bedrock	Bedrock	Bedrock	Streambed, banks	Variable	Strongly confined	Variable
Alluvial	Cascade	Boulder	Boulders, banks	8-30	Strongly confined	<1
	Step-pool	Cobble/boulder	Bedforms (steps, pools) boulders, large woody debris, banks	4-8	Moderately confined	1-4
	Plane-bed	Gravel/cobble	Boulders and cobbles, banks	1-4	Variable	None
	Pool-riffle	Gravel	Bedforms (bars, pools) boulders and cobbles, large woody debris, sinuosity, banks	0.1-2	Unconfined	5-7
	Regime	Sand	Sinuosity, bed-forms (dunes, ripples, bars), banks	<0.1	Unconfined	5-7
	Braided	Variable	Bedforms (bars, pools)	<3	Unconfined	Variable

Methods for measuring elevation and channel gradient can be found in Overton et al. (1997). Bisson and Montgomery (1996) describe in detail the method for identifying channel bed-form types, while Rosgen (1996) describes methods for classifying channel types. All classification work will include at least Level I (geomorphic characterization) channel type classification. Depending on the objectives of the monitoring program, additional levels of classification may be necessary. These variables will be updated once every 10 years.

Riparian Vegetation

Because riparian vegetation has an important influence on stream morphology and aquatic biota, the plan incorporates primary vegetation type as a characteristic of riparian vegetation. Primary vegetation type refers to the dominant vegetative cover along the stream. At a minimum, vegetation should be described as barren, grasses or forbs, shrubs, and trees. If remote sensing allows, it would be better to further classify the types of shrubs and trees. For example, trees could be described as cottonwoods, fir, cedar, hemlock, pine, etc. Primary vegetation type should be described for a riparian width of at least 30 m along both sides of the stream. More desirably, primary vegetation type should be described for the entire floodplain.

Remote sensing will be used to describe the primary vegetation type along streams within valley bottom types. Remote sensing may include aerial photos, LANDSAT ETM+, or both.

SECTION 7: SELECTION OF INDICATORS

This section identifies the "core" set of biological and physical/environmental indicator variables that will be measured within all watersheds and streams that receive status/trend and effectiveness monitoring. The "core" list of variables represents the minimum, required variables that will be measured. Investigators can measure additional variables depending on their objectives and past activities. For example, reclamation of mining-impact areas may require the monitoring of pollutants, toxicants, or metals. Some management actions may require the measurement of thalweg profile, placement of artificial instream structures, or livestock presence. Adding these indicators will supplement the core list.

Indicator variables identified in this plan are consistent with those identified in the Action Agencies/NOAA Fisheries RME Plan and with most of the indicators identified in the WSRFB (2003) monitoring strategy. Indicators were selected based on a review of the literature (e.g., Bjornn and Reiser 1991; Spence et al. 1996; Gregory and Bisson 1997; Bauer and Ralph 1999; Downes et al. 2002; and Roni 2005) and several regional monitoring programs (e.g., PNAMP, CSMEP; OBMEP; PIBO, AREMP, EMAP, WSRFB, and the Oregon Plan). To be selected, indicators had to meet various purposes including assessment of fish production and survival, identifying limiting factors, assessing effects of various land uses, and evaluating habitat actions. Criteria for selecting indicators were based on the following characteristics:

- They should be sensitive to land-use activities or stresses.
- They should be relevant to the questions asked.
- They should be consistent with other regional monitoring programs.
- They should be ecologically and/or socially significant.
- They should be efficient to measure.
- They should be associated strongly with the restoration action.

Indicators selected were consistent with most of the variables identified by the NMFS (1996) and USFWS (1998) as important attributes of "properly functioning condition." Indeed, NMFS and USFWS use many of these indicators to evaluate the effects of land-management activities for conferencing, consultations, and permits under the ESA.

The indicators selected were also consistent with "key" parameters used in the Ecosystem Diagnosis and Treatment model. Analyses by Mobrand Biometrics indicated that certain physical/environmental parameters have a relatively important influence on modeled salmon production. These parameters included channel configuration, gradient, pool/riffle frequency, migration barriers, flow characteristics, water temperature, riparian function, fine sediment, backwater areas, and large woody debris (LWD) (K. Malone, Mobrand Biometrics, personal communication).

Identified and described below are the "core" set of biological and physical/environmental indicators that will be monitored in the Upper Columbia Basin.

7.1 Biological Variables

The biological variables that will be measured in the Upper Columbia River Basin can be grouped into five general categories: adults, redds, parr, smolts, and macroinvertebrates. Each of these general categories consists of one or more indicator variables (Table 8). These biological indicators in concert will describe the characteristics of populations or sub-groups of fish in the monitoring zones (see Section 2) and will provide information necessary for assessing recovery of listed stocks.

Table 8. Biological indicator variables to be monitored in the Upper Columbia River Basin.

General characteristics	Specific indicators
Adults	Escapement/Number
	Age structure
	Size
	Sex ratio
	Origin (hatchery or wild)
	Genetics
	Fecundity
Redds	Number
	Distribution
Parr/Juveniles	Abundance
	Distribution
	Size
Smolts	Number
	Size
	Genetics
Macroinvertebrates	Transport
	Composition

Adults

Escapement:

The plan includes escapement of mature adults as an important biological indicator of population health. “Total” escapement is the total number of mature adults that enter or occur within a stream or watershed. “Spawning” escapement is the number of adults that spawn in a stream or watershed.[38] Numbers of mature adults within a stream or watershed is a function of all the factors that affect the life history of the population.

Spawners:

The plan includes six indicators associated with the characteristics of the spawning populations: age structure, size, sex ratio, origin, genetics, and fecundity. Age structure describes the ages of adult fish within the spawning population. For anadromous species, age structure includes the number of years the fish spent in freshwater and number of years in salt water.[39] Size describes the lengths and weights of adult fish within the spawning population. Sex ratio is the ratio of males to females within the spawning population. Origin identifies the parentage (hatchery or wild) of individuals within the spawning populations, while genetics defines not only the parentage but also within and between population variability. Fecundity is the number of eggs produced by a female.[40]

Redds

Abundance/Distribution:

Abundance describes the number of redds (nests) of fish species within a subbasin. Total numbers (based on a complete census) will be estimated for fall-spawning anadromous species, while numbers of redds of other species (e.g., steelhead and bull trout) will be estimated within index areas and sites selected randomly (following EMAP) (see Thompson et al. 1998). Distribution indicates the spatial arrangement (e.g., random, even, or clumped) and geographic extent of redds within the basin.

Parr

Abundance/Distribution:

Abundance describes the number of juvenile fish within specified stream reaches. Distribution is the spatial arrangement of juvenile fish within populations. It also captures the geographic range of individuals within the watershed or basin.

---

[38] Pre-spawn loss is the difference between total escapement and spawning escapement.

[39] Hatchery programs can affect the age structure of anadromous fish. For example, naturally produced summer Chinook (ocean-type Chinook) typical enter saltwater as age-0 fish. Summer Chinook raised in hatcheries may not be released until the fish are age-1. Therefore, hatchery produced summer Chinook may have a different age structure than naturally produced summer Chinook.

[40] By definition, *fecundity* refers to the number of eggs readied for spawning by a female (Royce 1996). *Relative fecundity* is the number of eggs per unit of weight, while *total fecundity* is the number of eggs laid during the lifetime of the female. This plan refers to fecundity as the number of eggs per size (length and weight) of female.

Condition:

The condition (or well-being) of fish can be assessed by measuring the length (fork length for salmonids, FL mm; total length for all other species, TL mm) and weight (0.1 g) of juvenile fish. The plan includes Fulton-type condition as the metric for well-being of juvenile fish (Anderson and Neumann 1996). The Fulton-type condition factor is of the form:

$$K = (W/L^3) \times 100{,}000,$$

where K = Fulton-type condition, W = weight in grams, and L = length in millimeters. The constant 100,000 is a scaling constant used to convert small decimals to mixed numbers so that the numbers can be more easily comprehended.

Smolts

Abundance:

Abundance of smolts is an estimate of the total number of smolts produced within a watershed or basin. The estimate should be for an entire population or sub-group.

Condition:

The Fulton-type condition factor describes the well-being of smolts within a population or sub-group.

Genetics:

Genetic characterization (via DNA microsatellites) describes within- and between-population genetic variability of smolts.

Macroinvertebrates

Transport:

The plan includes export of invertebrates (aquatic and terrestrial) and coarse organic detritus from headwaters to habitats downstream as an important attribute of productivity. The movement of prey items and organic detritus among habitats has a strong influence on fish populations, food webs, community dynamics, and ecosystem processes (Wipfli and Gregovich 2002).

Composition:

The plan includes benthic macroinvertebrate composition as an important indicator of aquatic invertebrates in streams. Benthic macroinvertebrate assemblages in streams reflect overall biological integrity of the benthic community. Because benthic communities respond to a wide array of stressors in different ways, it is often possible to determine the type of stress that affects a macroinvertebrate community.

7.2 Physical/Environmental Variables

The physical/environmental variables that will be measured in the Upper Columbia Basin can be grouped into seven general categories: water quality, habitat access, habitat quality, channel condition, riparian condition, flow/hydrology, and watershed condition. Each of these categories consists of one or more indicator variables (Table 9). In sum, these categories and their associated indicators address watershed processes and "input" variables (e.g., artificial physical barriers, road density, and other anthropogenic disturbances) as well as "outcome" variables (e.g., temperature, sediment, woody debris, pools, riparian habitat, etc.), as outlined in Hillman and Giorgi (2002) and the Action Agencies/NOAA Fisheries RME Plan.

What follows is a brief description of each physical/environmental indicator variable. Section 8 identifies recommended methods for measuring each indicator. Unless indicated otherwise, most of the information presented below has been summarized in Meehan (1991), MacDonald et al. (1991), Armantrout (1998), Bain and Stevenson (1999), OPSW (1999), Hillman and Giorgi (2002), and the Action Agencies/NOAA Fisheries RME Plan (2003).

Water Quality

Water Temperature:

The plan includes two temperature metrics that will serve as specific indicators of water temperature: maximum daily maximum temperature (MDMT) and maximum weekly maximum temperature (MWMT). MDMT is the single warmest daily maximum water temperature recorded during a given year or survey period. MWMT is the mean of daily maximum water temperatures measured over the warmest consecutive seven-day period. MDMT is measured to establish compliance with the short-term exposure to extreme temperature criteria, while MWMT is measured to establish compliance with mean temperature criteria.

Turbidity:

The plan includes turbidity as the one sediment-related specific indicator under water quality. Turbidity refers to the amount of light that is scattered or absorbed by a fluid. Suspended particles of fine sediments often increase turbidity of streams. However, other materials such as finely divided organic matter, colored organic compounds, plankton, and microorganisms can also increase turbidity of streams.

Table 9. A "core" list of physical/environmental indicator variables to be monitored within subbasins in the Upper Columbia Basin. Table is modified from Action Agencies/NOAA Fisheries RME Plan (2003).

General characteristics	Specific indicators[1]
Water Quality	MWMT and MDMT
	Turbidity
	Conductivity
	pH
	Dissolved oxygen
	Nitrogen
	Phosphorus
Habitat Access	Road crossings
	Diversion dams
	Fishways
Habitat Quality	Dominant substrate
	Embeddedness
	Depth fines
	LWD (pieces/km)
	Pools (pools/km)
	Residual pool depth
	Fish cover
	Side channels and backwaters
Channel condition	Stream gradient
	Wetted width
	Bankfull width
	Width/depth ratio
	Bank stability
Riparian Condition	Riparian structure
	Riparian disturbance
	Canopy cover
Flows and Hydrology	Streamflow
Watershed Condition	Watershed road density
	Riparian-road index
	Land ownership
	Land use

[1] Other indicators can be measured depending on the objectives of the study. For example, various metals and pollutants, herbicides and pesticides, thalweg profile, presence of livestock, artificial instream structures, bedload, etc. can be measured at each sampling site.

Contaminants and Nutrients:

The plan includes five specific indicators associated with contaminants and nutrients: conductivity, pH, dissolved oxygen (DO), nitrogen, and phosphorus. Most of these indicators are commonly measured because of their sensitivity to land-use activities, municipal and industrial pollution, and their importance in aquatic ecosystems.

The plan included conductivity, pH, and DO because these parameters are often incorporated into water quality monitoring programs (e.g., OPSW 1999; Bilhimer et al. 2003). Conductivity (or specific conductance) refers to the ability of water to conduct an electric current. The conductivity of water is a function of water temperature and the concentration of dissolved ions. It is measured as micromhos/centimeter (μmhos/cm).[41]

pH is defined as the concentration of hydrogen ions in water (moles per liter). It is a measure of how acidic or basic water is—it is not a measure of acidity or alkalinity (acidity and alkalinity are measures of the capacity of water to neutralize bases and acids, respectively). The logarithmic pH scale ranges from 0 to 14. Pure water has a pH of 7, which is the neutral point. Water is acidic if the pH value is less than 7 and basic if the value is greater than 7.

DO concentration refers to the amount of oxygen dissolved in water. Its concentration is usually measured in mg per liter (mg/L). The capacity of water to hold oxygen in solution is inversely proportional to the water temperature. Increased water temperature lowers the concentration of DO at saturation. Respiration (both plants and animals) and biochemical oxygen demand (BOD) are the primary factors that reduce DO in water. Photosynthesis and dissolution of atmospheric oxygen in water are the major oxygen sources.

The plan includes nitrogen and phosphorus as indicators of nutrient loading in streams. Nitrogen in aquatic ecosystems can be partitioned into dissolved and particulate nitrogen. Most water quality monitoring programs focus on dissolved nitrogen, because it is more readily available for both biological uptake and chemical transformations. Both dissolved and particulate nitrogen can be separated into inorganic and organic components. The primary inorganic forms are ammonia (NH_4^+), nitrate (NO_3^-), and nitrite (NO_2^-). Nitrate is the predominant form in unpolluted waters. This plan calls for the measurement of ammonia, nitrate/nitrites, and total nitrogen.

Phosphorus can also be separated into two fractions, dissolved and particulate. Dissolved phosphorus is found almost exclusively in the form of phosphate ions (PO_4^{-3}), which bind readily with other chemicals. There are three main classes of phosphate compounds: orthophosphates, condensed phosphates, and organically-bound phosphates. Each can occur as dissolved phosphorus or can be bound to particulate matter. In general, biota use only orthophosphates. Total phosphorus and orthophosphates will be measured under this plan.

41 Conductivity may also be reported in millisiemens/meter, where 1 millisiemen/m equals 0.1 μmhos/cm.

Habitat Access

Artificial Physical Barriers:

The plan includes three specific indicators associated with artificial physical barriers: road crossings (culverts), dams, and fishways. Roads and highways are common in the Upper Columbia River Basin and where they intersect streams they may block fish passage. Culverts can block passage of fish particularly in an upstream direction (WDFW 2000). In several cases, surveys have shown a difference in fish populations upstream and downstream from existing culverts, leading to the conclusion that free passage is not possible (Clay 1995). Dams and diversions that lack fish passage facilities can also block fish passage. Unscreened diversions may divert migrating fish into ditches and canals. Entrained fish can end in irrigated fields and orchards. Fishways are man-made structures that facilitate passage of fish through or over a barrier. Although these structures are intended to facilitate passage, they may actually impede fish passage (Clay 1995; WDFW 2000).

Habitat Quality

Substrate:

The Plan includes three specific indicators of substrate: dominant substrate, embeddedness, and depth fines. Dominant substrate refers to the most common particle size that makes up the composition of material along the streambed. This indicator describes the dominant material in spawning and rearing areas. Embeddedness is a measure of the degree to which fine sediments surround or bury larger particles. This measure is an indicator of the quality of over-wintering habitat for juvenile salmonids. Depth fines refer to the amount of fine sediment (<0.85 mm) within the streambed. Depth fines will be estimated at a depth of 15-30 cm (6-12 inches) within spawning gravels.

Large Woody Debris:

The plan includes the number of pieces of large woody debris (LWD) per stream kilometer as the one specific indicator of LWD in streams. LWD consists of large pieces of relatively stable woody material located within or spanning the bankfull channel and appearing to influence bankfull flows. LWD is also referred to as large organic debris (LOD) and coarse woody debris (CWD). LWD can occur as a single piece (log), an aggregate (two or more clumped pieces, each of which qualifies as a single piece), or as a rootwad.

The definition of LWD differs greatly among institutions. For example, NMFS (1996) defined LWD east of the Cascade Mountains as any log with a diameter greater than 30 cm (1 ft) and a length greater than 10.6 m (35 ft). Armantrout (1998) and BURPTAC (1999) defined LWD as any piece with a diameter >10 cm and a length > 1 m. Schuett-Hames et al. (1994) defined it as any piece with a diameter >10 cm and a length >2 m, while Overton et al. (1997) defined LWD as any piece with a diameter >10 cm and a length >3 m or two-thirds of the wetted stream width. Crews on the Wenatchee National Forest currently define LWD as any piece with a diameter >15 cm and a length >6 m.

This plan defines large woody debris as a piece of wood with a diameter of at least 0.1 m (4 in) and a length of at least 1.5 m (5 ft). Each piece of large woody debris will be categorized according to its diameter (measured at the large end) and length. There are four diameter classes: 0.1 m < 0.2 m (4 in < 12 in); 0.2 m < 0.6 m (12 in < 24 in); 0.6 m < 0.8 m (24 in < 32 in); and > 0.8 m (> 32 in) and three length classes: 1.5 m < 5.0 m (5 ft < 17 ft); 5.0 m < 15 m (17 ft < 50 ft); and > 15 m (> 50 ft).

Pool Habitat:

The plan includes two specific indicators associated with pool habitat: number of pools per kilometer and residual pool depth. A pool is slow-water habitat with a gradient less than 1% that is normally deeper and wider than aquatic habitats upstream and downstream from it (Armantrout 1998). To be counted, a pool must span more than half the wetted width, include the thalweg, be longer than it is wide, and the maximum depth must be at least 1.5 times the crest depth. Plunge pools are included in this definition even though they may not be as long as they are wide. Residual pool depth refers to the maximum depth of a pool if there is little or no flow in the channel. It is calculated as the difference between the maximum pool depth and the maximum crest depth (Overton et al. 1997).

Fish Cover:

Fish cover consists of such things as algae, macrophytes, moss, large woody debris, brush/small woody debris, in-channel live trees or roots, overhanging vegetation, undercut banks, large substrate, and artificial structures that offer concealment cover for fish and macroinvertebrates. This information is used to assess habitat complexity, fish cover, and channel disturbance.

Off-Channel Habitat:

Off-channel habitat consists of side-channels, off-channel pools, off-channel ponds, and oxbows. A side channel is a secondary channel that contains a portion of the streamflow from the main or primary channel. Off-channel pools occur in riparian areas adjacent to the stream channels and remain connected to the channel. Off-channel ponds are not part of the active channel but are supplied with water from over bank flooding or through a connection with the main channel. These ponds are usually located on flood terraces and are called wall-based channel ponds when they occur near the base of valley walls. Finally, oxbows are bends or meanders in a stream that become detached from the stream channel either from natural fluvial processes or anthropogenic disturbances.

Channel Condition

Stream gradient:

Stream gradient is the slope (change in vertical elevation per unit of horizontal distance) of the water surface within a site or reach. Although gradient is not usually affected by land-use activities, it is a major classification variable that indicates potential water velocities and stream power, which in turn control aquatic habitat and sediment transport within the reach. It is also an index of habitat complexity, as reflected in the diversity of water velocities and sediment sizes within the stream reach.

Wetted Width:

Wetted width is the width of the water surface measured perpendicular to the direction of flow. Wetted width is used to estimate water surface area, which is then used to calculate the density of fish (i.e., number of fish divided by the water surface area sampled)[42] within the site or reach.

Bankfull Width:

Bankfull width is the width of the channel (water surface) at the bankfull stage, where bankfull stage corresponds to the channel forming discharge that generally occurs within a return interval from 1.4 to 1.6 years and may be observed as the incipient elevation on the bank where flooding begins (Wolman and Leopold 1957; Leopold et al. 1964; Williams 1978). There are several indicators that one can use to identify bankfull stage. The active floodplain surface (topographic break between the channel bank and floodplain) is the best indicator of bankfull stage. It is the flat, depositional surface adjacent to many stream channels. These are most prominent along low-gradient, meandering reaches, but are often absent along steeper mountain streams. Where floodplains are absent or poorly defined, other useful indicators may serve as surrogates to identify bankfull stage (Harrelson et al. 1994). Those include:

- The height of depositional features (especially the top of the pointbar, which defines the lowest possible level for bankfull stage);
- A change in vegetation (especially the lower limit of perennial species);
- Slope or topographic breaks along the bank;
- A change in the particle size of bank material, such as the boundary between coarse cobble or gravel with fine-grained sand or silt;
- Scour line of roots and banks;
- Undercuts in the bank, which usually reach an interior elevation slightly below bankfull stage; and
- Stain lines or the lower extent of lichens on boulders.

[42] By definition, the measure of the number of fish per unit area is called "crude density" (Smith and Smith 2001). However, not all of the water surface area provides suitable habitat for fish. Density measured in terms of the amount of area suitable as living space is "ecological density."

One should use all of these indicators to help identify the bankfull elevation.

Width/Depth Ratio:

The width/depth ratio is an index of the cross-section shape of a stream channel at bankfull level. The ratio is a measure of the response of a channel to changes in bank conditions. Increases in width/depth ratios, for example, indicate increased bank erosion, channel widening, and infilling of pools. Because streams almost always are several times wider than they are deep, a small change in depth can greatly affect the width/depth ratio.

Streambank Condition:

The plan includes streambank stability as the one specific indicator of streambank condition. Streambank stability is an index of firmness or resistance to disintegration of a bank based on the percentage of the bank showing active erosion (alteration) and the presence of protective vegetation, woody material, or rock. A stable bank shows no evidence of breakdown, slumping, tension cracking or fracture, or erosion (Overton et al. 1997). Undercut banks are considered stable unless tension fractures show on the ground surface at the bank of the undercut.

Riparian Condition

Riparian structure:

Riparian structure describes the type and amount of various types of vegetation within the riparian zone (within 10 m). Information on riparian structure can be used to evaluate the health and level of disturbance of the stream corridor. In addition, it provides an indication of the present and future potential for various types of organic inputs and shading.

Riparian disturbance:

Riparian disturbance refers to the presence and proximity of various types of human land-use activities within the riparian area (within 30 m from the edge of the channel). Activities include such things as walls, dikes, riprap, dams, buildings, pavement, paved roads and trails, unpaved roads and trails, railroads, pipes, trash, parks, lawns, mining, agriculture, pastures, and logging. All these activities have an effect on the riparian vegetation, which in turn affects the quantity and quality of aquatic habitat for listed fish species.

Canopy cover:

Riparian canopy cover over a stream is important not only in its role in moderating stream temperatures through shading, but it also serves to control bank stability and provides inputs of coarse and fine particulate organic materials. Organics from riparian vegetation become food for stream organisms and structure to create and maintain complex channel habitat.

Flows and Hydrology

Streamflows:

The plan includes three specific indicators of streamflows: change in peak flow, change in base flow, and change in timing of flow. Peak flow is the highest or maximum streamflow recorded within a specified period of time. Base flow is the streamflow sustained in a stream channel and is not a result of direct runoff. Base flow is derived from natural storage (i.e., outflow from groundwater, large lakes, or swamps), or sources other than rainfall. Timing of flow refers to the time when peak and base flows occur and the rate of rises and falls in the hydrograph. These indicators are based on “annual” flow patterns.

Watershed Conditions

Road Density:

A road is any open way for the passage of vehicles or trains. The plan includes both road density and the riparian-road index (RRI) as indicators of roads within watersheds. Road density is an index of the total miles of roads within a watershed. It is calculated as the total length of all roads (km) within a watershed divided by the area of the watershed (km^2). The RRI is expressed as the total mileage of roads (km) within riparian areas divided by the total number of stream kilometers within the watershed (WFC 1998). For this index, riparian areas are defined as those falling within the federal buffers zones; that is, all areas within 300 ft of either side of a fish-bearing stream, within 150 ft of a permanent nonfish-bearing stream, or within the 100-year floodplain.

Watershed Disturbance:

The plan includes land ownership and land use as the two indicators of watershed disturbance. Land ownership describes the surface status of the basin. That is, it delineates the portions of the basin owned by federal, state, county, tribal, and private entities. Land use, on the other hand, delineates the portions of the basin that are subject to specific land uses, such as urban, agriculture, range, forest, wetlands, etc.

7.3 Recommended Indicators

As noted earlier, the biological and physical/environmental indicators identified in this section represent a “core” list of variables that will be measured in subbasins in the Upper Columbia Basin. This plan does not preclude the investigator from measuring other indicator variables. Which variables will be measured depends on the questions asked by the program, type of monitoring (status/trend vs. effectiveness), the target fish species and life histories, and the type of tributary action implemented. Identified below are the appropriate indicators for each type of monitoring.

Effectiveness Monitoring—This plan does not recommend that all indicators listed in Tables 8 and 9 be measured for each tributary action. Different biological indicators will be measured depending on the questions asked and the fish species of interest (Table 10). Any of the biological indicators identified in Table 8 could be measured for actions that affect anadromous species (spring Chinook, summer/fall Chinook, steelhead, and sockeye salmon). For resident species (bull trout and cutthroat trout), however, indicators related to smolts and fecundity would not be measured.

The plan recommends that only those physical/environmental indicators that are linked directly to the proposed action be measured. In other words, the most useful indicators are likely to be those that represent the first link in the chain of cause-and-effect. Because different projects have different objectives and desired effects, the investigator only needs to measure those indicators directly influenced on the chain of causality between the habitat action and the effect (Table 11). This approach differs from the Action Agencies/NOAA Fisheries Plan, which requires all indicators be measured, regardless of the type of habitat action implemented.

Status/Trend Monitoring—All physical/environmental indicators identified in Table 9 will be measured as part of status/trend monitoring within the five monitoring zones (see Section 2) in the Upper Columbia Basin. In contrast, different biological indicators will be measured depending on the target fish species (Table 10). As with effectiveness monitoring, all biological indicators identified in Table 8 will be measured for anadromous species. Indicators related to smolts and fecundity will not be measured for resident species.

Table 10. Biological indicator variables that will be measured (marked with an "X") for anadromous (spring Chinook, summer/fall Chinook, steelhead, and sockeye salmon) and resident (bull trout and cutthroat trout) fish species during status/trend and effectiveness monitoring in subbasins in the Upper Columbia Basin.

General characteristics	Specific indicators	Anadromous species	Resident species
Adults	Escapement/Number	X	X
	Age structure	X	X
	Size	X	X
	Sex ratio	X	X
	Origin (hatchery or wild)	X	X
	Genetics	X	X
	Fecundity	X	
Redds	Number	X	X
	Distribution	X	X
Parr/Juveniles	Abundance	X	X
	Distribution	X	X
	Size	X	X
Smolts	Number	X	
	Size	X	
	Genetics	X	
Macroinvertebrates	Transport	X	X
	Composition	X	X

Table 11. Rankings of the usefulness of physical/environmental indicators to monitoring effects of different tributary habitat actions. Rankings vary from 1 = highly likely to be useful; 2 = moderately likely to be useful; and 3 = unlikely to be useful or little relationship, although the indicator may be useful under certain conditions or may help interpret data from a primary indicator. Table is modified from Hillman and Giorgi (2002). The different classes of habitat actions are from the Action Agencies/NOAA Fisheries RME Plan.

General characteristics	Specific indicators	Classes of habitat actions							
		Diversion screens	Barrier removal	Sediment reduction	Water quality improvement	Nutrient enhancement	Instream flows	Riparian habitat	Instream structure
Water quality	MWMT/MDMT	3	2	3	1	2	1-2	1	3
	Turbidity	3	1-2	1	1	1	1-2	2	3
	Conductivity	3	2	2	1	1	2	2	3
	pH	3	3	3	1	1	3	2-3	3
	Dissolved Oxygen	3	2-3	2-3	1	1	1-2	2-3	3
	Nitrogen	3	3	3	1	1	3	2	3
	Phosphorus	3	3	3	1	1	3	2	3
Habitat access	Road crossings	3	1	3	3	3	3	3	3
	Diversion dams	1-2	1	3	3	3	2	3	3
	Fishways	2-3	1	3	3	3	3	3	3
Habitat quality	Dominant substrate	3	2	1	3	3	1-2	2	1-2
	Embeddedness	3	1-2	1	1-2	3	1-2	2	1-2
	Depth fines	3	1-2	1	1-2	2	2	2	1-2
	LWD	3	3	3	3	3	2	1	1
	Pools	3	1-2	1-2	3	3	1-2	1-2	1
	Residual pool depth	3	1-2	1	3	3	1	1-2	1
	Fish cover	3	2	1	1-2	1-2	1	1-2	1
	Off-channel habitat	3	2	2	3	3	1	1-2	1
Channel condition	Stream gradient	2	2	2	2	2	2	2	2
	Width/depth	3	1-2	1-2	3	3	1-2	1-2	1
	Wetted width	3	1-2	1-2	3	3	1-2	1-2	1
	Bankful width	3	1-2	1-2	3	3	1-2	1-2	1
	Bank stability	3	2	1-2	3	3	2	1	1
Riparian condition	Riparian structure	3	3	2	2-3	3	2	1	1-2
	Riparian disturbance	3	3	2	2-3	3	2	1	1-2
	Canopy cover	3	3	2	2-3	3	2	1	1-2
Flows/hydrology	Streamflows	3	1-2	3	3	3	1	2	1-2
Watershed condition	Road density	3	3	1-2	2	3	2-3	2-3	2
	Riparian-road index	3	3	1-2	2	3	2-3	1	2
	Land ownership	2	2	1	1	2-3	1	1	2
	Land use	1-2	1-2	1	1	2-3	1	1	2

SECTION 8: MEASURING PROTOCOLS

An important component of all regional monitoring strategies (ISAB, Action Agencies/NOAA Fisheries, OBMEP, PNAMP, CSMEP, and SRFB) is that the same measurement method be used to measure a given indicator. The reason for this is to allow comparisons of biological and physical/environmental conditions within and among watersheds and basins.[43] This section identifies methods to be used to measure biological and physical/environmental indicators. The methods identified in this plan are for the most part consistent with those described in the Action Agencies/NOAA Fisheries RME Plan, EMAP, and SRFB protocols.

The Action Agencies/NOAA Fisheries monitoring group reviewed several publications, including the work of Johnson et al. (2001) that describe methods for measuring indicators. Not surprisingly, there can be several different methods for measuring the same variable. For example, channel substrate can be described using surface visual analysis, pebble counts, or substrate core samples (either McNeil core samples or freeze-core samples). These techniques range from the easiest and fastest to the most involved and informative. As a result, one can define two levels of sampling methods. Level 1 (extensive methods) involves fast and easy methods that can be completed at multiple sites, while Level 2 (intensive methods) includes methods that increase accuracy and precision but require more sampling time. The Action Agencies/NOAA Fisheries monitoring group selected primarily Level 2 methods, which minimize sampling error.

Before identifying measuring protocols, it is important to define a few terms. These terms are consistent with the Action Agencies/NOAA Fisheries RME Plan.

Reach (effectiveness monitoring) – for effectiveness monitoring, a stream reach is defined as a relatively homogeneous stretch of a stream having <u>similar</u> regional, drainage basin, valley segment, and channel segment characteristics and a repetitious sequence of habitat types. Reaches are identified by using a list of classification (stratification) variables (from Table 4). Reaches may contain one or more sites. The starting point and ending point of reaches will be measured with Global Positioning System (GPS) and recorded as Latitude and Longitude.

Reach (status/trend monitoring) – for status/trend monitoring, a reach is a length of stream (20 times the mean bankfull width, but not less than 150-m long or longer than 500 m)[44] selected with a systematic randomized process (GRTS design). GRTS selects a point on the "blue-line" stream network represented on a 1:100,000 scale USGS map. This point is referred to as the "X-site." The X-site identifies the midpoint of the reach. That is, the sampling reach extends a distance of 10 times the average bankfull width upstream and downstream from the X-site, measured along the

43 Bonar and Hubert (2002) and Hayes et al. (2003) review the benefits, challenges, and the need for standardized sampling.

44 This reach length differs from Simonson et al. (1994) and Reynolds et al. (2003), which use 40x the wetted width. The use of 20x the bankfull width is consistent with AREMP and PIBO protocols. This protocol also allows one to assess channel conditions even if the channel is dry. There are naturally dry channels within the project area.

thalweg[45]. Biological and physical/environmental indicators are measured within the reach. The X-site and the upstream and downstream ends of the reach will be measured with GPS and recorded as Latitude and Longitude. For purposes of re-measurements, these points will also be photographed, marked with permanent markers (e.g., orange plastic survey stakes or rebar[46]), and carefully identified on maps and site diagrams. Reach lengths and boundaries will be "fixed" the first time they are surveyed and they will not change over time even if future conditions change.

Site (effectiveness monitoring) – a site is an area of the effectiveness monitoring stream reach that forms the smallest sampling unit with a defined boundary. Site length depends on the width of the stream channel. Sites will be 20 times the average bankfull width with a minimum length of 150 m and a maximum length of 500 m. Site lengths are measured along the thalweg. The upstream and downstream boundaries of the site will be measured with GPS and recorded as Latitude and Longitude. For purposes of re-measurements, these points will also be photographed, marked with permanent markers (e.g., orange plastic survey stakes or rebar), and carefully identified on maps and site diagrams. Site lengths and boundaries will be "fixed" the first time they are surveyed and they will not change over time even if future conditions change.

Transect – a transect is a straight line across a stream channel, perpendicular to the flow, along which habitat features such as width, depth, and substrate are measured at pre-determined intervals. Effectiveness monitoring sites and status/trend monitoring reaches will be divided into 11 evenly-spaced transects by dividing the site into 10 equidistant intervals with "transect 1" at the downstream end of the site or reach and "transect 11" at the upstream end of the site or reach.

Habitat Type – Habitat types, or channel geomorphic units, are discrete, relatively homogenous areas of a channel that differ in depth, velocity, and substrate characteristics from adjoining areas. This plan recommends that the investigator identify and map habitat types throughout the reach or site following the Level II classification system in Hawkins et al. (1993). That is, habitat will be classified as turbulent fast water, non-turbulent fast water, scour pool, or dammed pool (see definitions in Hawkins et al. 1993). By definition, for a habitat unit to be classified, it should be longer than it is wide. Plunge pools, a type of scour pool, are the exception, because they can be shorter than they are wide.

45 "Thalweg" is defined as the path of a stream that follows the deepest part of the channel (Armantrout 1998).

46 Metal detectors can be used to relocate rebar.

8.1 Biological Indicators

This section identifies the methods and instruments that should be used to measure biological indicators. Table 12 identifies indicator variables, example protocols, and sampling frequency. The reader is referred to the cited documents for a more detailed description of each method.

Table 12. Recommended protocols and sampling frequency for biological indicator variables.

General characteristics	Specific indicators	Recommended protocol	Sampling frequency
Adults	Escapement/Number	Dolloff et al. (1996); Reynolds (1996); Van Deventer and Platts (1989)	Annual
	Age structure	Borgerson (1992)	Annual
	Size	Anderson and Neumann (1996)	Annual
	Sex ratio	Strange (1996)	Annual
	Origin (hatchery or wild)	Borgerson (1992)	Annual
	Genetics	WDFW Genetics Lab	Annual
	Fecundity	Cailliet et al. (1986)	Annual
Redds	Number	Mosey and Murphy (2002)	Annual
	Distribution	Mosey and Murphy (2002)	Annual
Parr/Juveniles	Abundance/Distribution	Thurow (1994); Reynolds (1996); Van Deventer and Platts (1989)	Annual
	Size	Anderson and Neumann (1996)	Annual
Smolts	Number	Murdoch et al. (2000)	Annual
	Size	Anderson and Neumann (1996)	Annual
	Genetics	WDFW Genetics Lab	Annual
Macroinvertebrates	Transport	Wipfli and Gregovich (2002)	Annual/Monthly
	Composition	Peck et al. (2001)[1]	Annual

[1]The Peck et al. (2001) report is a draft document, which states that it should not be cited or quoted. However, it provides an appropriate method for estimating benthic macroinvertebrate composition.

Adults

Escapement:

The plan includes escapement/number of mature adults as an important biological indicator of population status. "Total escapement" of anadromous fish into subbasins within the Upper Columbia Basin can be estimated roughly as the number of fish counted at mainstem Columbia River dams. For example, escapement of spring Chinook into the Wenatchee Basin can be estimated as the difference between fish counts at Rock Island Dam and Rocky Reach Dam (with some correction for fallback). Counts at dams should be made with video that operates continuously during the upstream migration of anadromous salmonids. Counts of adults at weirs are more accurate and should be used whenever possible. This method is recommended if accurate estimates of escapements into specific watersheds are necessary.

"Spawning escapement" can be estimated as the number of redds times a "fish-per-redd" estimate. WSRFB (2003) uses 2.2 Chinook per redd, assuming one redd per female. For steelhead, they assume 1.23 redds per female. A more accurate method currently used by WDFW in the Upper Columbia Basin is based on the sex ratio of broodstock (not recovered carcasses) collected randomly over the run (A. Murdoch, personal communication, WDFW). For example, if the sex ratio of a random sample of the run is 1.5:1.0, the expansion factor for the run would be 2.5 fish/redd. This method is used for all supplemented stocks within the Upper Columbia Basin. Another method, which can be used if the sex ratio is unknown, is the "Modified Meekin Method" (A. Murdoch, WDFW, personal communication). This method takes the 2.2 adults/redd (from Meekin 1967) and increases it by the proportion of jacks in the run. For example, if jacks make up 10% of the run, the modified adults/redd would be 2.42 (2.2 x 1.1 = 2.42 adults/redd). This plan recommends that spawning escapement be estimated based on sex ratios. Both total and spawning escapement will be reported as "whole" numbers.

Numbers of resident adult fish should be estimated within status/trend monitoring reaches and effectiveness monitoring sites using underwater observations (snorkeling) or electrofishing surveys. Snorkeling, which is a quick, nondestructive method that is not restricted by deep water and low conductivities,[47] is the "primary" sampling method in this plan. Snorkel surveys will follow the protocols identified in Thurow (1994). Accurate estimates of adult bull trout may require nighttime snorkeling. However, Hillman and Chapman (1996) counted more adult bull trout during the day than at night in the Blackfoot River, Montana, because adult bull trout were unable to conceal themselves, making them readily visible to snorkelers. Both daytime and nighttime surveys should be conducted for at least two years to see which survey time (daytime or nighttime) provides the best estimate of resident adult fish. For each fish observed during day or night surveys, snorkelers will estimate fish size to the nearest 2 cm and report numbers as fish/ha.

Electrofishing is the "secondary" method and will be used within a sub-sample of snorkel sites. The plan recommends that at least five randomly-selected sites (10% of the status/trend sites sampled annually) within each monitoring zone be sampled with both snorkeling and electrofishing.[48] The purpose for this is to establish a relationship between the methods and to collect fish for assessment of condition (length and weight), age, gender, and genetics.[49]

---

[47] Hillman and Miller (2002) reported that snorkel estimates were more accurate than electrofishing estimates in the Chiwawa River, a Wenatchee River tributary, because low conductivity (35 μmhos) in the river reduced the efficiency of electrofishing. They noted that electrofishing estimates were at best 68% of snorkel estimates.

[48] Sampling within a site should occur within the same day and sites should be blocked to prevent movement into and out of the site during and between sampling.

[49] Because ESA-listed species occur within monitoring areas in the Upper Columbia Basin, federal agencies may limit the amount of electrofishing that can be conducted within the basin. If electrofishing cannot be used, this plan recommends that snorkelers collect fish at night with hand-held dip nets. Using an appropriate sampling design, fish can be selected so as to not bias the sample.

electrofishing will follow the protocols outlined in Reynolds (1996) and NMFS (2000). For salmonids, fork length (anterior tip to the median caudal fin rays) will be measured to the nearest 1 mm and weighed to the nearest 0.1 g. For all other fish, total length (anterior tip to the longest "compressed" caudal fin rays) will be measured to the nearest 1 mm and weighed to the nearest 0.1 g. This plan recommends the removal-depletion method of electrofishing, with at least three complete passes.[50] Population numbers and 95% confidence intervals can be estimated with the maximum-likelihood formula (Van Deventer and Platts 1989).[51] Numbers of fish will be reported as fish/ha.

Spawners:

The plan includes six indicators associated with the characteristics of the spawning populations: age structure, size, sex ratio, origin, genetics, and fecundity. For anadromous fish, most of these characteristics will be collected from live fish trapped at weirs or from carcasses sampled during spawning surveys. Scales will be pulled from live fish and carcasses. Scales will be read to determine age structure and origin (wild or hatchery). Presence or absence of an adipose fin will also determine origin. Age analysis will be completed following methods described by Borgerson (1992). Size will be measured to the nearest 1 mm and reported as both fork length (anterior tip to the median caudal fin rays) and hypural length (mid-eye to hypural plate) (Anderson and Neumann 1996). The latter is necessary because some carcasses will have decomposed to a point where fork length cannot be measured accurately. The gender of each fish sampled will be recorded (Strange 1996). Fecundity (total number of eggs produced by a given size female) will be estimated for fish collected for hatchery broodstock and from dead pre-spawn females collected during spawning surveys (Cailliet et al. 1986). Finally, genetic samples will be collected and analyzed according to the protocols being refined at the WDFW Genetics Lab. To avoid resampling the same carcass, the head of all sampled carcasses will be removed.

Many of the characteristics identified above for anadromous fish will be collected from resident fish during electrofishing surveys (or nighttime dip netting), collection at weirs, and during spawning surveys. Characteristics such as age structure (from scales or length frequencies), size (fork length or total length; mm), weight (0.1 g), origin, and genetic samples can be collected from adults trapped at weirs and during electrofishing (or nighttime dip netting) surveys. Origin can be assessed by examining fins, with hatchery fish tending to have deformed or eroded fins. Gender can be recorded for those fish found dead during spawning surveys. The protocols identified above can be used to measure characteristics of resident fish.

Redds

[50] The removal-depletion method has been criticized because of heterogeneity in sampling efficiencies (probability of capture) (see Peterson et al. 2004; Thurow et al. 2004). Unbiased abundance estimates can be generated by dividing the number of fish collected with three-pass electrofishing by the predicted capture efficiency. Predicted capture efficiencies are based on species-specific capture efficiency models and site-specific habitat data. Thurow et al. (2006) provide an example of adjusting three-pass electrofishing data.

[51] Van Deventer recently updated the maximum-likelihood model. The MicroFish model can be found at: http://www.microfish.org/

Abundance/Distribution:

This plan includes abundance and distribution of salmonid redds as indicators of population status. The plan calls for a complete census of redds of fall-spawning anadromous fish (e.g., Chinook and sockeye salmon). For other species (e.g., steelhead and bull trout), numbers of redds will be counted annually within already-established index areas and in reaches that will be selected using probabilistic sampling (GRTS; see Section 4.4). At least 25 reaches, each 1.6-km long (1.0 mile), will be surveyed throughout the spawning period within all monitoring zones, except in the Entiat Basin. Because the spawning distribution of fish in the Entiat Basin is relatively small, each survey reach there will be 1.0-km long. As in other monitoring zones, 25 reaches will be surveyed in the Entiat Basin each year.

To assess changes in spawning distribution, a five-year rotating panel design with 25 reaches per year will be implemented (see Section 3.2). Thus, a different set of 25 reaches will be sampled each year during the first five years. Throughout the spawning period, investigators will conduct weekly redd surveys following the example of Mosey and Murphy (2002). Each week new redds will be counted, mapped, and marked.[52] Marking is needed to avoid recounting redds during subsequent surveys. Abundance of redds will be reported as the number of redds within a population or sub-group. Abundance will also be reported as the number of redds per km within each population or sub-group.

Parr

Abundance/Distribution:

The plan includes the abundance and distribution of juvenile fish as an indicator of population status. Juvenile numbers will be estimated with snorkeling and electrofishing within status/trend monitoring reaches and effectiveness monitoring sites. Snorkeling is the "primary" sampling method in this plan and will follow the protocols identified in Thurow (1994). Accurate estimates of juvenile salmonids may likely require nighttime snorkeling. Therefore, both daytime and nighttime surveys should be conducted for at least two years to see which survey time (daytime or nighttime) provides the best estimate of juvenile fish. For each fish observed during day or night surveys, snorkelers will estimate fish size to the nearest 2 cm and report numbers as fish/ha.

Electrofishing is the "secondary" method and will be used within a sub-sample of snorkel sites (same five sites within each monitoring zone used to sample adult fish). Electrofishing will follow the protocols outlined in Reynolds (1996) and NMFS (2000). This plan recommends the removal-depletion method of electrofishing, with a minimum of three complete passes (see footnotes 50 and 51). Population numbers and 95% confidence intervals can be estimated with the maximum-likelihood formula (Van Deventer and Platts 1989). Numbers will be reported as fish/ha.

---

[52] Because of inclement weather and high streamflows, surveys for steelhead redds may not be made on regularly timed intervals. Adjusting surveys to fit environmental conditions may be necessary.

Juvenile fish collected during electrofishing (or nighttime dip netting) will be measured (see below) and at least 5,000 juvenile Chinook and 5,000 steelhead within each population will be implanted with PIT tags. The sample size of 5,000 for anadromous populations in the Upper Columbia Basin was estimated by the Action Agencies/NOAA Fisheries Monitoring Group. This is a very rough estimate of the minimum number needed to estimate life-stage survival rates and life-history characteristics.

Condition:

The plan includes Fulton-type condition as the metric for well-being of juvenile fish. Juvenile fish collected during electrofishing and with rotary traps (or other appropriate traps) will be measured (fork length for salmonids and total length for all other fish; mm) and weighed (0.1 g). Fulton-type condition will be estimated with methods described in Anderson and Neumann (1996).

Smolts

Abundance:

Abundance of smolts is an estimate of the total number of smolts produced within a watershed or basin. Investigators will use floating screw traps (or other appropriate traps depending on stream conditions) to collect downstream migrating smolts. Traps will operate for at least the entire period of the smolt migration. Trapping efficiency, based on mark/recapture will be estimated throughout the trapping period. Methods for operating the trap, estimating efficiency, and the frequency at which efficiency tests are conducted are described in Murdoch et al. (2000). Numbers of smolts will be reported for populations or sub-groups.

Condition:

The Fulton-type condition factor describes the well-being of smolts within a population or sub-group. Smolts collected with traps will be measured (fork length; mm) and weighed (0.1 g). Fulton-type condition will be estimated with methods described in Anderson and Neumann (1996).

Genetics:

Genetic characterization (via DNA microsatellites) describes within- and between-population genetic variability of smolts. DNA samples from a systematic sample of smolts[53] will be collected and analyzed according to the protocols being refined at the WDFW Genetics Lab.

Macroinvertebrates

Transport:

The plan includes export of invertebrates (aquatic and terrestrial) and coarse organic detritus from headwaters to downstream habitats as an attribute of freshwater productivity. Investigators will follow the methods described in Wipfli and Gregovich (2002) to assess energy sources for downstream food webs. The method requires the placement of sampling stations near tributary junctions of fishless and fish-bearing streams. Specially-modified drift nets (Wipfli and Gregovich 2002) will capture invertebrates and particulate organic matter. Invertebrate transport will be reported as numbers per day and dry mass (mg) per day. Debris transport will be reported as dry mass (g) per m^3 water and dry mass (g) per day. This work will be conducted monthly.

Composition:

The plan also includes benthic macroinvertebrate composition as an attribute of freshwater productivity. Investigators will follow the "targeted-riffle-sample" method described in Peck et al. (2001). This method requires at least eight independent kick-net[54] samples from riffles within sites or reaches. The eight samples are combined, sieved to remove debris and sediments, and then processed in a lab. Samples will be analyzed according to the River InVertebrate Prediction and Classification System (RIVPACS) (Hawkins et al. 2001).

8.2 <u>Physical/Environmental Indicators</u>

This section identifies the methods and instruments needed to measure physical/environmental indicators. Table 13 identifies indicator variables, example protocols for measuring indicators, and sampling frequency. There is no space here to describe each method in detail; therefore, the reader is referred to the cited documents for detailed descriptions of methods and measuring instruments. Suggested deviations, additions, or changes to the cited protocols are described under each indicator. Importantly, and for obvious reasons, all habitat sampling would follow fish sampling (snorkeling and electrofishing) within status/trend monitoring reaches and effectiveness monitoring sites.

---

53 The total number of smolts needed to characterize within and between-population genetic variability is presently unknown. Therefore, "**k**" (i.e., the k^{th} smolt sampled) remains undefined.

54 The kick net is a D-frame sampler with a 30.5-cm wide base, a muslin bottom panel, a net with a mesh size of 500 µm, and a detachable bucket with a 500-µm mesh end (see Figure 11-1 in Peck et al. 2001).

Table 13. Recommended protocols and sampling frequency of physical/environmental indicator variables. Some of the protocols recommended in this table have been modified based on protocol comparison tests. Modified protocols are noted with an asterisks (*) and modifications are described in the text.

General characteristics	Specific indicators	Recommended protocols	Sampling frequency
Water Quality	MWMT/MDMT	Zaroban (2000)	Hourly
	Turbidity	OPSW (1999)	Hourly
	Conductivity	OPSW (1999)	Daily
	pH	OPSW (1999)	Daily
	Dissolved oxygen	OPSW (1999)*	Daily
	Nitrogen	OPSW (1999)	Monthly
	Phosphorus	OPSW (1999)	Monthly
Habitat Access	Road crossings	Parker (2000); WDFW (2000)	Annually
	Diversion dams	WDFW (2000)	Annually
	Fishways	WDFW (2000)	Annually
Habitat Quality	Dominant substrate	Peck et al. (2001)*	Annually
	Embeddedness	Peck et al. (2001)*	Annually
	Depth fines	Schuett-Hames (1999)	Annually
	LWD (pieces/km)	BURPTAC (1999)*	Annually
	Pools per kilometer	Hawkins et al. (1993); Overton et al. (1997)	Annually
	Residual pool depth	Overton et al. (1997)	Annually
	Fish cover	Peck et al. (2001)	Annually
	Off-channels habitats	WFPB (1995)*	Annually
Channel condition	Stream gradient	Peck et al. (2001)*	Annually
	Wetted width	Peck et al. (2001)	Annually
	Bankfull width	Peck et al. (2001)	Annually
	Width/depth ratio	Peck et al. (2001)*	Annually
	Bank stability	Moore et al. (2002)	Annually
Riparian Condition	Structure	Peck et al. (2001)	Annually
	Disturbance	Peck et al. (2001)*	Annually
	Canopy cover	Peck et al. (2001)	Annually
Flows and Hydrology	Streamflow	Peck et al. (2001)	Continuous
Watershed Condition	Watershed road density	WFC (1998); Reeves et al. (2001)	5 years
	Riparian-road index	WFC (1998)	5 years
	Land ownership	n/a	5 years
	Land use	Parmenter et al. (2003)	5 years

Water Quality

Water Temperature:

The plan includes two temperature metrics that will serve as specific indicators of water temperature: maximum daily maximum temperature (MDMT) and maximum weekly maximum temperature (MWMT).[55] Data loggers can be used to measure MWMT and MDMT. Zaroban (2000) describes pre-placement procedures (e.g., selecting loggers and calibration of loggers), placement procedures (e.g., launching loggers, site selection, logger placement, and locality documentation), and retrieval procedures. This manual also provides standard methods for conducting temperature-monitoring studies associated with land-management activities and for characterizing temperature regimes throughout a watershed.

The number of loggers used will depend on the number of reaches and treatment and control sites. For monitoring the effectiveness of actions that affect water temperatures, at a minimum, at least one logger will measure water temperatures at the downstream end and one at the upstream end of each reach that contains treatment or control sites. Additional measurements may be needed within reaches (at treatment sites) depending on the objectives and scope of the study. For status/trend monitoring, one logger will be placed at the downstream end of the distribution of each population or sub-group (e.g., near smolt traps).

Data loggers will record temperatures hourly to the nearest 0.1˚C throughout the year. Investigators will also measure water temperatures with a calibrated thermometer at each site or reach sampled for fish. These snap-shot measurements will be used to assess the reliability of fish sampling techniques.[56]

Turbidity:

The plan includes turbidity as the one sediment-related specific indicator under water quality. Investigators will measure turbidity with monitoring instruments calibrated on the nephelometric turbidity method (NTUs). Daily means (and SD) and maximum turbidities should be calculated. Chapter 11 in OPSW (1999) provides a standardized method for measuring turbidity, data quality guidelines, equipment, field measurement procedures, and methods to store and analyze turbidity data.

For monitoring the effectiveness of actions that affect turbidity, at a minimum, turbidity will be measured at the downstream end and at the upstream end of each reach that contains treatment or control sites. Additional instruments may be needed to measure turbidity at treatment sites within reaches depending on objectives and scope of the study. For

[55] Additional metrics such as number of days that daily maximum temperatures exceed certain temperature criteria should also be calculated. For example, the number of days that daily maximum temperatures exceed 15°C (for bull trout) and 20°C (for Chinook) are useful metrics.

[56] Both electrofishing and snorkeling are affected by water temperature. Hillman et al. (1992) demonstrated that snorkel counts are less reliable at cold water temperatures.

status/trend monitoring, one instrument will be placed at the downstream end of the distribution of each population or sub-group (e.g., near smolt traps).

Monitoring instruments will measure turbidity hourly to the nearest 1 NTU throughout the year. Investigators will also measure turbidity with a portable turbidimeter within each site or reach sampled for fish. Because both electrofishing and snorkeling are affected by turbidity, these snap-shot measurements will be used to assess the reliability of the fish sampling techniques.

Contaminants and Nutrients:

The plan includes five specific indicators associated with contaminants and nutrients: conductivity, pH, dissolved oxygen (DO), nitrogen, and phosphorus. OPSW (1999) identifies standard methods for measuring conductivity (Chapter 9), pH (Chapter 8), DO (Chapter 7)[57], and nitrate/nitrites, ammonium, total nitrogen, total phosphorous, and orthophosphates (Chapter 10). OPSW (1999) also includes criteria for data quality guidelines, equipment, field-measurement procedures, and methods to store and analyze water quality data.

For monitoring the effectiveness of actions that affect these parameters, at a minimum, conductivity, pH, and DO will be measured hourly at the downstream end and upstream end of each reach that contains treatment or controls sites. At a minimum, nitrogen and phosphorus will be measured monthly (12 times per year) at both ends of a reach. Additional measurements (both in time and space) may be needed at treatment and control sites depending on objectives and scope of the study. For status/trend monitoring, conductivity and pH will be collected once per daily (at 1200 hrs), DO twice per day (1200 and 2400 hrs), and nitrogen and phosphorus monthly at the downstream end of the distribution of each population or subpopulation.

Water quality instruments can be used to monitor conductivity, pH, and DO. Conductivity will be measured to the nearest 0.1 μmhos/cm, pH to the nearest 0.1 unit, and DO to the nearest 0.1 mg/L. Indicators associated with nitrogen and phosphorus will be collected as grab samples. Ammonia is recorded in mg NH_3-N/L, nitrite in mg NO_2^- -N/L, nitrate in mg NO_3^- -N/L, and total nitrogen in mg N/L. Both total phosphorus and orthophosphates are recorded in mg P/L. Because conductivity affects electrofishing success, a portable conductivity meter will be used to measure conductivity within each site or reach sampled for fish.

[57] Although OPSW (1999) indicates that the Winkler Titration Method is the most accurate method for measuring DO concentration, this plan recommends the use of an electronic recording device with an accuracy of at least ±0.2 mg/L.

Habitat Access:

Artificial Physical Barriers:

The plan includes three specific indicators associated with artificial physical barriers: road crossings (culverts), dams, and fishways. Remote sensing (aerial photos, LANDSAT ETM+, or both) will be used as a first cut to identify possible barriers. Investigators will then conduct field surveys using the WDFW (2000) protocols to evaluate possible barriers. The WDFW (2000) manual provides guidance and methods on how to identify, inventory, and evaluate culverts, dams, and fishways that impede fish passage. WDFW (2000) also provides methods for estimating the potential habitat gained upstream from barriers, allowing prioritization of restoration projects. The manual by Parker (2000) focuses on culverts and assesses connectivity of fish habitats on a watershed scale. These manuals can be used to identify all fish passage barriers within monitoring reaches. Assessment of fish passage barriers will occur once annually during base-flow conditions.

Habitat Quality

Substrate:

The plan includes three specific indicators of substrate: dominant substrate, depth fines, and embeddedness. Peck et al. (2001) provides a method for describing substrate composition within each site or reach. Substrate composition will be assessed within the bankfull width (not wetted width) along the "channel bottom" in the site or reach, regardless if the channel is wet or dry. Investigators will measure substrate at eleven equidistant points along each of the 11 "regular" transects, plus along an additional 10 transects placed mid-way between each of the 11 transects. The eleven points along a transect correspond to 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 100% of the bankfull width. The investigator will visually estimate the size of a particle at each of the points along the 21 transects (total sample size of 231 particles). Classification of bed material by particle size will follow Table 14. For each sampling site or reach, the investigators will report the dominant substrate size. Additionally, investigators can calculate reach-level means, standard deviations, and percentiles for substrate size classes (see methods in Kaufmann et al. 1999). Substrate will be characterized annually during base-flow conditions.

Investigators will measure depth fines with McNeil core samplers.[58] Methods for conducting core sampling can be found in Schuett-Hames et al. (1999). For effectiveness monitoring, four randomly-selected samples (subsamples) will be taken from each spawning area (pool tailout or riffle) within each site (samples will not be taken from sites that lack spawning areas). For status/trend monitoring, four subsamples from one randomly-selected spawning area within "accessible" (see footnote 58) reaches will be collected. Where possible, core

---

[58] Core sampling can be conducted in streams with shallow spawning areas. Spawning areas that are too small or too deep cannot be sampled effectively with core samplers (see Schuett-Hames et al. 1999). In addition, because of the extensive equipment needed to conduct substrate core sampling, core sampling within sites located long distances from access points (>0.75-1.0 km) may be skipped. Every effort, however, should be made to collect the data.

samples will also be collected in spawning areas at or near the long-term, water-quality monitoring sites. These long-term monitoring sites will be located at the downstream end of the distribution of populations or sub-groups (near smolt traps). The volumetric method will be used for processing samples sorted via a standard set of sieves. The volumetric method measures the milliliters (to the nearest 1 ml) of water displaced by particles of different size classes. At a minimum, the following sieves will be used to sort particles: 64.0 mm, 16.0 mm, 6.4 mm, 4.0 mm, 1.0 mm, 0.85 mm, 0.50 mm, 0.25 mm, and 0.125 mm. Fines will be measured once annually during base-flow conditions.

Peck et al. (2001) also provides methods for measuring embeddedness. As with substrate composition, embeddedness will be assessed within the bankfull width (not wetted width) along the "channel bottom," regardless if the channel is dry or wet. Embeddedness will be estimated at eleven equidistant points along the 11 "regular" transects (total sample size of 121). At each sampling point along a transect, all particles larger than sand within a 10-cm diameter circle will be examined for embeddedness. Embeddedness is the fraction of particle surface that is surrounded by (embedded in) sand or finer sediments (< 2 mm). By definition, sand and fines are embedded 100%, while bedrock is embedded 0%. Investigators will record the average percent (%) embeddedness of particles in the 10-cm circle. Embeddedness will be measured once annually during base-flow stream conditions.

Table 14. Classification of stream substrate channel materials by particle size. Table is from Peck et al. (2001).

Class name	Size range (mm)	Description
Bedrock (smooth)	>4,000	Smooth surface rock larger than a car
Bedrock (rough)	>4,000	Rough surface rock larger than a car
Hardpan		Firm, consolidated fine substrate including sand solidified in moss
Boulders	>250-4,000	Basketball to car size
Cobbles	>64-250	Tennis ball to basketball size
Gravel (coarse)	>16-64	Marble to tennis ball size
Gravel (fine)	>2-16	Ladybug to marble size
Sand	>0.06-2	Smaller than ladybug size, but visible as particles
Fines	<0.06	Silt, clay, muck (not gritty between fingers)

Large Woody Debris:

Large woody debris (LWD) consists of large pieces of relatively stable woody material located within the bankfull channel or spanning the channel. Investigators will simply count the number of LWD pieces within sites or reaches (wet or dry) in forested streams (e.g., see BURPTAC 1999). Pieces are counted throughout the entire reach or site, not just along transects. Each piece of large woody debris will be categorized according to its diameter (measured at the large end) and length. There are four diameter classes:

(1) 0.1 m < 0.2 m (4 in < 12 in);
(2) 0.2 m < 0.6 m (12 in < 24 in);
(3) 0.6 m < 0.8 m (24 in < 32 in); and
(4) > 0.8 m (> 32 in)

and three length classes:

(1) 1.5 m < 5.0 m (5 ft < 17 ft);
(2) 5.0 m < 15 m (17 ft < 50 ft); and
(3) > 15 m (> 50 ft).

Investigators will record the count of LWD pieces within each size category. This indicator will be measured once annually during base-flow conditions.

Pool Habitat:

The plan includes two indicators associated with pool habitat: number of pools per km and residual pool depth. Investigators will count the number of pools throughout a monitoring reach or site. To be counted, a pool must span more than half the wetted width, include the thalweg, be longer than it is wide, and the maximum depth must be at least 1.5 times the crest depth. Plunge pools are included in this definition even though they may not be as long as they are wide. Hawkins et al. (1993) and Overton et al. (1997) provide good descriptions of the various types of pools and how to identify them. Pools are counted throughout the entire reach or site, not just along transects.

Overton et al. (1997) describe methods for measuring residual pool depth. Residual pool depth is simply the difference between the maximum pool depth and the crest depth. Measurements differ, however, depending on the type of pool. For dammed pools, residual depth is the difference between maximum pool depth and maximum crest depth at the head of the pool. For scour pools, on the other hand, residual pool depth is the difference between maximum pool depth and maximum crest depth at the tail of the pool. Depths are measured to the nearest 0.01 m. For effectiveness monitoring, residual pool depth will be measured in all pools within treatment and control sites. For status/trend monitoring, residual pool depth will be measured in all pools within a reach. Both pools per km and residual pool depth will be measured once annually during base-flow conditions.

Fish Cover:

Fish cover is measured within the wetted width of a site or reach. Fish cover is not measured in dry channels. It is visually estimated at 5 m upstream and 5 m downstream (10-m total length) at each of the 11 transects following procedures described in Peck et al. (2001). Cover types consist of filamentous algae, aquatic macrophytes (including wetland grasses), moss, large woody debris, brush and small woody debris, in-channel live trees or roots, overhanging vegetation (within 1 m of the water surface but not in the water), undercut banks, boulders, and artificial structures (e.g., concrete, cars, tires, rip-rap, etc.). For each cover type, the investigator will record areal cover as: 0 (zero cover), 1 (<10% cover), 2 (10-40% cover), 3 (40-75% cover), and 4 (>75% cover). Fish cover will be estimated annually during base-flow conditions.

Off-Channel Habitat:

Off-channel habitat consists of side-channels, off-channel pools, off-channel ponds, and oxbows. Following the definitions for each off-channel habitat type (see Section 7), the investigator will enumerate the number of each type of off-channel habitat within a monitoring reach or site. Off-channel habitats will be enumerated throughout the entire site or reach, not just along transects. In addition, investigators will measure the lengths of side channels in the site or reach. Investigators will record the number of off-channel habitat types and the lengths of side channels (measured to the nearest 0.5 m) within the site or reach. Sampling will occur once annually during base-flow conditions.

Channel Condition

Stream Gradient:

The water surface gradient or slope is an indication of potential water velocities and stream power. Water surface slope will be reported as a percentage[59] and will be measured according to the protocol described in Peck et al. (2001) with some modifications. Rather than measure percent slope directly with a clinometer or Abney level, as recommended in Peck et al. (2001), this plan calls for the measurement of water surface elevations with a hand level. That is, water surface elevation will be measured between each of the 21 transects (includes both "regular" and "additional" transects) using a hand level (5X magnifying level recommended) and a telescoping leveling rod (graduated in cm). Beginning at the downstream-end of the reach or site, water surface elevation is measured by "backsighting" downstream between transects (results in at least 20 measurements per reach or site). If a meander bend, vegetation, or some other object blocks your line-of-sight along the channel[60], it will be necessary to establish a "supplemental" point between transects. The investigator records the elevation (measured to the nearest cm) and horizontal distance

[59] Although this plan recommends reporting slope as a percentage, one can easily convert between percentage, decimal, and degrees with the following formulas: (1) Percent slope = slope (in decimal form) x 100; (2) Slope (in decimal form) = tan (slope in degrees); and (3) Slope (in degrees) = $\tan^{-1}$ (slope in decimal form).

[60] If you have to sight across land to measure surface elevation between two transects, then you need to make supplementary measurements (do not "short-circuit" meander bends).

between transects or supplementary points (measured to the nearest cm). Percent water surface slope is then calculated as the fall per unit distance (rise over run), times 100. Sampling will occur once annually during base-flow conditions.

Wetted Width:

Wetted width is the width of the water surface measured perpendicular to the direction of flow. Peck et al. (2001) describes the recommend method for measuring this indicator. Wetted width will be measured to the nearest 0.1 m at the 21 transects (11 "regular" and 10 "additional" transects) in each reach or treatment and control sites. Sampling will occur once annually during base-flow conditions.

Bankfull Width:

Bankfull width is the width of the channel (water surface) at bankfull stage. Peck et al. (2001) describe methods for measuring bankfull width. Bankfull width will be measured to the nearest 0.1 m at the 21 transects in each reach (for status/trend monitoring) or treatment and control sites (for effectiveness monitoring), regardless if the channel is wet or dry. Sampling will occur once annually during base-flow conditions.

Width/Depth Ratio:

The width/depth ratio is an index of the cross-section shape of a stream channel at bankfull level. The ratio is expressed as bankfull width (geomorphic term) divided by the mean cross-section bankfull depth. Peck et al. (2001) offer the recommended protocol for measuring bankfull widths and depths. This indicator will be measured at the 21 transects (includes the 11 "regular" and 10 "additional" transects) within each reach (for status/trend monitoring) or treatment and control site (for effectiveness monitoring), regardless if the channel is wet or dry. Eleven bankfull depths will be recorded along each transect. Width and depth will be recorded to the nearest 0.1 m. Sampling will occur once annually during base-flow conditions.

Streambank Condition:

The plan includes streambank stability as the one specific indicator of streambank condition. Moore et al. (2002) describe the recommended method for assessing stream bank stability. The method estimates the percent (%) of the lineal distance that is actively eroding at the active channel height on both sides of the transect regardless if the channel is wet or dry. Active erosion is defined as recently eroding or collapsing banks and may have the following characteristics: exposed soils and inorganic material, evidence of tension cracks, active sloughing, or superficial vegetation that does not contribute to bank stability. Stability is assessed at or below the bankfull level. Bank stability will be measured once annually during base-flow conditions at the 11 "regular" transects within each reach (for status/trend monitoring) or treatment and control site (for effectiveness monitoring).

Riparian Condition

Structure:

Riparian structure identifies the type and amount of various kinds of riparian vegetation. Peck et al. (2001) offer methods for describing riparian structure. Riparian structure will be assessed within a 10 m x 10 m plot on both ends of each of the 11 "regular" transects, regardless if the channel is wet or dry. Within each riparian plot, the investigator will divide the vegetation into three layers: canopy layer (>5-m high), understory layer (0.5-5-m high), and the ground-cover layer (<0.5-m high). Areal cover will be estimated within each of the three vegetation layers. Aerial cover is recorded as "0" if no cover; "1" if <10% cover; "2" if 10-40%; "3" if 40-75%; or "4" if >75% cover. The type of vegetation will be described in both the canopy and understory layers. Vegetation types include deciduous, coniferous, broadleaf evergreen, mixed, and none. Kaufmann et al. (1999) describes methods for analyzing riparian structure data. This indicator will be measured once annually during base-flow conditions.

Disturbance:

Riparian disturbance will be measured as the presence and proximity of various types of human land-use activities in the riparian area. Peck et al. (2001) provide the recommended method for assessing this indicator. The presence/absence and proximity of 12 categories of human influences will be described within 5 m upstream and 5 m downstream from each of the 11 "regular" transects, regardless if the channel is wet or dry. Human influences include:

(1) Walls, dikes, revetments, riprap, and dams;
(2) Buildings;
(3) Pavement/cleared lot;
(4) Paved roads and trails or railroads;
(5) Unpaved roads and trails;
(6) Inlet or outlet pipes;
(7) Landfills or trash;
(8) Parks or maintained lawns;
(9) Row crops;
(10) Pastures, rangeland, hay fields, or evidence of livestock,
(11) Logging; and
(12) Mining.

Proximity classes include:

(1) Present at or along the stream bank;
(2) Present between the bank and 10 m from the bank;
(3) Present between 10 m and 30 m from the bank; and
(4) Not present.

Kaufmann et al. (1999) describes methods for analyzing riparian disturbance data. Riparian disturbance will be measured once annually during base-flow conditions.

Canopy Cover:

Peck et al. (2001) describe the recommended method for measuring canopy cover. Canopy cover will be measured at each of the 11 “regular” transects in wet or dry channels using a Convex Spherical Densiometer (model B). Six measurements are collected at each “regular” transect (four measurements in four directions at mid-channel and one at each bank). The mid-channel measurements estimate canopy cover over the channel, while the two bank measurements estimate cover within the riparian zone. The two bank measurements are particularly important in wide streams, where riparian canopy may not be detected at mid-channel. The investigator records the number of grid intersection points (0-17) that are covered by vegetation at the six points along each transect. Mean densiometer readings and standard deviations are calculated according to methods described in Kaufmann et al. (1999). Canopy cover will be measured once annually during base-flow conditions.

Flows and Hydrology

Streamflows:

Changes in streamflows will be assessed by collecting flow data at the downstream end of monitoring reaches and/or at the downstream end of the distribution of each population or sub-group. Investigators will use USGS or WDOE flow data where available to assess changes in peak, base, and timing of flows. For those streams with no USGS or State stream-gauge data, investigators will use the velocity-area method described in Peck et al. (2001) to estimate stream flows. Water velocities will be measured to the nearest 0.01 m/s with a calibrated water-velocity meter rather than the float method. Wetted width and depth will be measured to the nearest 0.1 m. Flows will be reported as m^3/s.[61]

Watershed Conditions

Road Density:

The plan includes road density and the riparian-road index (RRI) as indicators of roads within watersheds. Using remote sensing, investigators will measure the road density and riparian-road index within each watershed in which monitoring activities occur. Road density will be calculated with GIS as the total length (km) of roads within a watershed divided by the area (km^2) of the watershed. The riparian-road index will be calculated with GIS as the total kilometers of roads within riparian areas divided by the total number of stream kilometers within the watershed. WFC (1998) provides an example of calculating the riparian-road index in the Umpqua Basin. Both road density and the riparian-road index will be updated once every five years.

Watershed Disturbance:

---

[61] The following formula can be used to convert cfs (cubic feet per second) to cms (cubic meters per second): cms = cfs x 0.02832.

The plan includes both land ownership and land use as the two indicators of watershed disturbance. Using remote sensing techniques and available GIS layers, the investigator will map the spatial extent of land ownership and land uses within each watershed that includes monitoring reaches or sites. These indicators will be updated once every five years.

8.3 Recommendations

This plan requires that the protocols identified above be used to measure biological and physical/environmental indicators. It is understood that some of these methods will differ from those currently used by entities that will be implementing this plan. Indeed, some of the entities that will implement this plan may have collected data for several years with protocols different from those identified in this plan. It is not the intent of this plan to have those entities immediately switch protocols. Rather, this plan encourages entities to use both methods for a few years.[62] This will allow them to compare the performance of different methods and to develop relationships between different protocols. As noted in Section 4.4, at least 10% of the sites within each subbasin (Wenatchee, Entiat, Methow, and Okanogan) will be used to compare sampling protocols. This means that five sites in each of the subbasins will be sampled with more than one protocol.

The precision (repeatability) of measurements will be assessed by repeatedly sampling the same sites with independent crews. As noted in Section 4.4, 10% of the sites within each subbasin (same sites used to compare protocols) will be sampled by two independent crews each year for five years. Sampling by the two independent crews will be no more than two-days apart. This will minimize the effects of site changes on estimates of precision.

As noted earlier, water quality and quantity indicators will be measured at permanent, long-term monitoring locations and some (temperature, turbidity, and conductivity) will be measured during each visit to a monitoring site or reach for the purpose of evaluating fish sampling methods. For status/trend monitoring, at a minimum, permanent locations occur at the downstream end of the distribution of target populations or sub-groups. For effectiveness monitoring, if the intent of the management action (treatment) is to affect water quality, then water quality sampling occurs at the upstream and downstream ends of reaches that contain treatment or control sites. As outlined in Section 8.2, this plan recommends that water quality indicators be measured with instruments that record water quality conditions. Hydrolab® has a water quality instrument (DataSonde 4a)[63] that measures most of the water quality indicators identified in this plan (Table 15). Other instruments that record water quality indicators could be used provided they achieve the same (or better) accuracy standards as the DataSonde.

Table 15. Water quality indicators, range, accuracy, and resolution of the DataSonde 4a developed by Hydrolab.

---

[62] The number of years needed to compare performance and to develop relationships between methods will be determined as data are collected. At a minimum, entities implementing this plan should expect to use both methods for at least five years (n = 5).

[63] Information on Hydrolab and the DataSonde 4a can be found at http://www.hydrolab.com. As noted earlier, the use of trade names in this paper is for reader information only and does not imply endorsement by an agency of any product.

Indicator	Range	Accuracy	Resolution
Temperature	-5˚ to 50˚C	±0.10˚C	0.01˚C
Turbidity	0 to 1000 NTU	±5% of range	0.1 to 1 NTU
Conductivity	0 to 100 mS/cm	±0.001 mS/cm	4 digits
pH	0 to 14 units	±0.2 units	0.01 units
Dissolved oxygen	0 to 50 mg/L	±0.2 mg/L	0.01 mg/L

As a final note, the protocols identified in this section for measuring physical/environmental indicators were designed for sampling wadeable streams. Because the sampling frame for subbasins within the Upper Columbia Basin include first through fifth-order streams, sites in large, non-wadeable rivers such as the Wenatchee and Methow rivers may be included in the sample. Most of the protocols identified in Table 13 can be used in non-wadeable streams with some creative modifications. For example, dominant substrate and embeddedness can be assessed using snorkel or SCUBA gear.[64] Measurements of widths and depths can be collected from boats or rafts. Depth fines is the only indicator that cannot be collected in most non-wadeable streams.

[64] Water clarity during base-flow conditions within the Wenatchee, Methow, and Entiat rivers is suitable for underwater observation work.

SECTION 9: IMPLEMENTATION

The preceding sections serve notice that considerable care must be put into the appropriate methods and logic structure of a status/trend and effectiveness monitoring plan. The intent of this section is to distill the information presented in this document into a concise outline that an investigator can follow to develop a statistically-valid monitoring plan. For convenience, this summary is offered as a checklist of steps that will aid the investigator in developing a valid monitoring program. Although these steps are generic, the investigator should address each one in order to demonstrate complete understanding of status/trend and effectiveness monitoring.

This section consists of three parts. The first part outlines the steps needed to setup and implement the monitoring plan. The second and third parts outline the steps needed to design status/trend monitoring studies and effectiveness monitoring studies, respectively.

9.1 <u>Program Setup</u>

In order to setup the monitoring program, it is important to follow a logical sequence of steps. By proceeding through each step, the investigator will better understand the goals of monitoring and its strengths and limitations. These steps should aid the investigator in implementing a valid monitoring program that reduces duplication of sampling efforts, and thus overall costs, but still meets the needs of the different entities. The plan assumes that all entities involved with implementing the plan will cooperate and freely share information.

Setup Steps:

1. Identify the populations and/or subpopulations of interest (e.g., spring Chinook, steelhead, bull trout).
2. Identify the geographic boundaries (areas) of the populations or sub-groups of interest.
3. Describe the purpose for selecting these populations or sub-groups (what are the concerns?).
4. Identify the objectives for monitoring.
5. Select the appropriate monitoring approach (status/trend or effectiveness monitoring or both) for addressing the objectives.
6. Identify and review existing monitoring and research programs in the area of interest.
7. Determine if those programs satisfy the objectives of the proposed program.
8. If data gaps exist, implement the appropriate monitoring approach by following the criteria outlined in Sections 9.2 and 9.3.
9. Classify the landscape and streams in the area of interest (see Section 6).
10. Describe how data collection efforts will be shared among the different entities.
11. Identify a common database for storing biological and physical/environmental data.
12. Estimate costs of implementing the program.
13. Identify cost-sharing opportunities.

9.2 Status/Trend Monitoring

If the objective of the monitoring program is to assess the current status of populations and/or environmental conditions, or to assess long-term trends in these parameters, then the following steps will help the investigator design a valid status/trend monitoring program.

Problem Statement and Overarching Issues:

1. Identify and describe the problem to be addressed.
2. Identify boundaries of the study area.
3. Describe the goal or purpose of the study.
4. List hypotheses to be tested.

Statistical Design (see Section 3):

1. Describe the statistical design to be used (e.g., EMAP design).
2. List and describe potential threats to external validity and how these threats will be addressed.
3. If this is a pilot test, explain why it is needed.
4. Describe descriptive and inferential statistics to be used and how precision of statistical estimates will be calculated.

Sampling Design (see Sections 4 & 5):

1. Describe the statistical population(s) to be sampled.
2. Define and describe sampling units.
3. Identify the number of sampling units that make up the sampling frame.
4. Describe how sampling units will be selected (e.g., random, stratified, systematic, etc.).
5. Describe variability or estimated variability of the statistical population(s).
6. Define Type I and II errors to be used in statistical tests (the plan recommends no less than 0.80 power).

Measurements (see Sections 7 & 8):

1. Identify indicator variables to be measured.
2. Describe methods and instruments to be used to measure indicators.
3. Describe precision of measuring instruments.
4. Describe possible effects of measuring instruments on sampling units (e.g., core sampling for sediment may affect local sediment conditions). If such effects are expected, describe how the study will deal with this.
5. Describe steps to be taken to minimize systematic errors.
6. Describe QA/QC plan, if any.
7. Describe sampling frequency for field measurements.

Results:

1. Explain how the results of this study will yield information relevant to management decisions.

9.3 Effectiveness Monitoring

If the objective of the monitoring program is to assess the effects of tributary habitat actions (e.g., improve stream complexity), then the following steps will help the investigator design a valid effectiveness monitoring program (these steps are modified from Paulsen et al. 2002). Because effectiveness monitoring encompasses the essence of cause-and-effect research (i.e., attempts to control for sources of invalidity), the steps below are more extensive and intensive than those in the status/trend monitoring program.

Problem Statement and Overarching Issues:

1. Identify and describe the problem to be improved or corrected by the action being monitored.
2. Describe current environmental conditions at the project site.
3. Describe factors contributing to current conditions (e.g., road crossing causing increased siltation).
4. Identify and describe the habitat action(s) (treatments) to be undertaken to improve existing conditions.
5. Describe the goal or purpose of the habitat action(s).
6. Identify the hypotheses to be tested.
7. Identify the independent variables in the study.

Statistical Design (see Section 3):

1. Describe the statistical design to be used (e.g., BACI design).
2. Describe how treatments (habitat actions) and controls will be assigned to sampling units (e.g., random assignment).
3. Show whether or not the study will include “true” replicates or subsamples.
4. Describe how temporal and spatial controls will be used and how many of each type will be sampled.
5. Describe the independence of treatment and control sites (are control sites completely unaffected by habitat actions?).
6. Identify covariates and their importance to the study.
7. Describe potential threats to internal and external validity and how these threats will be addressed.
8. If this is a pilot test, explain why it is needed.
9. Describe descriptive and inferential statistics to be used and how precision of statistical estimates will be calculated.

Sampling Design (see Sections 4 & 5):

1. Describe the statistical population(s) to be sampled.
2. Define and describe sampling units.
3. Describe the number of sampling units (both treatment and control sites) that make up the sampling frame.
4. Describe how sampling units will be selected (e.g., random, stratified, systematic, etc.).
5. Define "practical significance" (e.g., environmental or biological effects of the action) for this study.
6. Describe how effect size(s) will be detected.
7. Describe the variability or estimated variability of the statistical population(s).
8. Define Type I and II errors to be used in statistical tests (the plan recommends no less than 0.80 power).

Measurements (see Sections 7 & 8):

1. Identify and describe the indicator (dependent) variables to be measured.
2. Describe methods and instruments to be used to measure indicators.
3. Describe the precision of measuring instrument(s).
4. Describe possible effects of measuring instruments on sampling units (e.g., core sampling for sediment may affect local sediment conditions). If such effects are expected, describe how the study will deal with this.
5. Describe steps to be taken to minimize systematic errors.
6. Describe QA/QC plan, if any.
7. Describe sampling frequency for field measurements.

Results:

1. Explain how the results of this study will yield information relevant to management decisions.

These steps should be carefully considered when designing a monitoring plan to assess the effectiveness of any habitat action, regardless of how simple the proposed action may be. Even monitoring the effectiveness of irrigation screens requires careful consideration of all steps in the checklist. In some cases, the investigator may not be able to address all steps with a high degree of certainty, because adequate information does not exist. For example, the investigator may lack information on population variability, effect size, "practical significance," or instrument precision, which makes it difficult to design studies and estimate sample sizes. In this case the investigator can address the statements with the best available information, even if it is based on professional opinion, or design a pilot study to answer the questions.

SECTION 10: REFERENCES

Action Agencies (Bonneville Power Administration, U.S. Bureau of Reclamation, and U.S. Army Corps of Engineers) and NOAA Fisheries. 2003. Draft research, monitoring and evaluation plan for the NOAA-Fisheries 2000 Federal Columbia River Power System Biological Opinion. Bonneville Power Administration, U.S. Bureau of Reclamation, U.S. Army Corps of Engineers, and NOAA Fisheries, Portland, OR. Web link: http://www.nwr.noaa.gov/1hydrop/hydroweb/fedrec.htm

Anderson, R. O. and R. M. Neumann. 1996. Length, weight, and associated structural indices. Pages 447-482 *in*: B. R. Murphy and D. W. Willis, editors. Fisheries techniques, 2nd edition. American Fisheries Society, Bethesda, MD.

Armantrout, N. B., compiler. 1998. Glossary of aquatic habitat inventory terminology. American Fisheries Society, Bethesda, MD.

Bain, M. B. and N. J. Stevenson, editors. 1999. Aquatic habitat assessment: common methods. American Fisheries Society, Bethesda, MD.

Bailey, R. G. 1978. Description of ecoregions of the United States. U.S. Forest Service, Intermountain Region, Ogden, UT.

Bailey, R. G. 1998. Ecoregions map of North America: explanatory note. U.S. Forest Service, Miscellaneous Publication 1548, Washington, D.C.

Bayley, P. B. 2002. A review of studies on responses of salmon and trout to habitat change, with potential for application in the Pacific Northwest. Report to the Washington State Independent Science Panel, Olympia, WA.

Bevenger, G. S. and R. M. King. 1995. A pebble count procedure for assessing watershed cumulative effects. Research Paper RM-RP-319, USDA Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO.

Bilhimer, D., J. Carroll, S. O'Neal, and G. Pelletier. 2003. Draft quality assurance project plan: Wenatchee River temperature, dissolved oxygen, pH, and fecal coliform total maximum daily load year 2 technical study. Washington State Department of Ecology, Olympia, WA.

Bisson, P. A. and D. R. Montgomery. 1996. Valley segments, stream reaches, and channel units. Pages 23-52 *in*: R. R. Hauer and G. A. Lamberti, editors. Methods in stream ecology. Academic Press, New York, NY.

Bjornn T. C. and D. W. Reiser. 1991. Habitat requirements of salmonids in streams. American Fisheries Society Special Publication 19:83-138.

Bonar, S. A. and W. A. Hubert. 2002. Standard sampling of inland fish: benefits, challenges, and a call for action. Fisheries 27:10-16.

Borgerson, L. A. 1992. Scale analysis. Oregon Department of Fish and Wildlife, Fish Research Project F-144-R-4, Annual Progress Report, Portland, OR.

Box, G. E. P. and G. M. Jenkins. 1976. Time-series analysis: forecasting and control. Holden-Day, San Francisco, CA.

Browne, R. H. 2001. Using the sample range as a basis for calculating sample size in power calculations. The American Statistician 55:293-298.

BURPTAC (Beneficial Use Reconnaissance Project Technical Advisory Committee). 1999. 1999 beneficial use reconnaissance project workplan for wadable streams. Idaho Division of Environmental Quality, Boise, ID. Web link: http://www.deq.state.id.us/water/surface_water/99_burp_workplan.pdf

Cailliet, G. M., M. S. Love, and A. W. Ebeling. 1986. Fishes, a field and laboratory manual on their structure, identification, and natural history. Wadsworth Publishing Company, Belmont, CA.

Clay, C. H. 1995. Design of fishways and other fish facilities. Second edition. Lewis Publishers, Boca Raton, Florida.

Cohen, J. 1988. Statistical power analysis for the behavioral sciences. Lawrence-Erlbaum, Hillsdale, NJ.

Cook, T. D. and D. T. Campbell. 1979. Quasi-experimentation: design and analysis issues for field settings. Houghton Mifflin Company, Boston, MA.

Cupp, C. E. 1989a. Identifying spatial variability of stream characteristics through classification. Master's thesis. University of Washington, Seattle, WA.

Cupp, C. E. 1989b. Valley segment type classification for forested lands of Washington. Washington State Timber/Fish/Wildlife Agreement, TFW-AM-89-001, Department of Natural Resources, Olympia, WA.

Currens, K. P., H. W. Li, J. D. McIntyre, D. R. Montgomery, and D. W. Reiser. 2000. Recommendations for monitoring salmonid recovery in Washington State. Independent Science Panel Report 2000-2, Olympia, WA.

Currens, K. P., H. W. Li, J. D. McIntyre, D. R. Montgomery, and D. W. Reiser. 2002. Responses of salmon and trout to habitat changes. Independent Science Panel Technical Memorandum 2002-2, Olympia, WA.

Diaz-Ramos, S., D. L. Stevens, and A. R. Olsen. 1996. EMAP statistical methods manual. U.S. Environmental Protection Agency, EPA/620/R-96/XXX, Corvallis, OR.

Dolloff, A., J. Kershner, and R. Thurow. 1996. Underwater observation. Pages 533-554 *in*: B. R. Murphy and D. W. Willis, editors. Fisheries techniques, 2nd edition. American Fisheries Society, Bethesda, MD.

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. Cambridge University Press, Cambridge, UK.

Frissell, C. A., W. J. Liss, C. E. Warren, and M. D. Hurley. 1986. A hierarchical framework for stream habitat classification; viewing streams in a watershed context. Environmental Management 10:199-214.

Gordon, N. D., T. A. McMahon, and B. L. Finlayson. 1992. Stream hydrology an introduction for ecologists. John Wiley and Sons, New York, NY.

Green, R. H. 1979. Sampling design and statistical methods for environmental biologists. John Wiley and Sons, Inc., New York, NY.

Green, R. H. 1994. Aspects of power analysis in environmental monitoring. Pages 173-182 *in*: D. J. Fletcher and B. F. J. Manly, editors. Statistics in ecology and environmental monitoring. University of Otago Press, Dunedin.

Gregory, S. V. and P. A Bisson. 1997. Degradation and loss of anadromous salmonid habitat in the Pacific Northwest. Pages 277-314 in: D. J. Stouder, P. A. Bisson, and R. J. Naiman, editors. Pacific salmon and their ecosystems, status and future options. Chapman and Hall, New York, NY.

Hadley, R. F. and S. A. Schumm. 1961. Sediment sources and drainage basin characteristics in upper Cheyenne River basin. U.S. Geological Survey, Water-Supply Paper 1531-B, Reston, Virginia.

Hairston, N. G. 1989. Ecological experiments: purpose, design, and execution. Cambridge University Press, New York, NY.

Harrelson, C. C., C. L. Rawlins, and J. P. Potyondy. 1994. Stream channel reference sites: an illustrated guide to field technique. USDA Forest Service, Rocky Mountain Forest and Range Experiment Station, General Technical Report RM-245, Fort Collins, CO.

Hawkins, C. P., and ten others. 1993. Hierarchical approach to classifying stream habitat features. Fisheries 18:3-12.

Hawkins, C., J. Ostermiller, M. Vinson, and J. Stevenson. 2001. Steam algae, invertebrate, and environmental ssampling associated with biological water quality assessments: field protocols. Utah State University, Logan, UT.
Web link: http://www.usu.edu/buglab/monitor/USUproto.pdf

Hayes, D. and 14 others. 2003. Developing a standardized sampling program: the Michigan experience. Fisheries 28:18-25.

Hicks, L. L., J. Light, G. Watson, B. Sugden, T. W. Hillman, and D. Berg. 1999. Adaptive management: concepts and applications to Plum Creek's Native Fish Habitat Conservation Plan. Native Fish Habitat Conservation Plan Technical Report No. 13, Plum Creek Timber Company, Seattle, WA.

Hillman, T. W., J. W. Mullan, and J. S. Griffith. 1992. Accuracy of underwater counts of juvenile Chinook salmon, coho salmon, and steelhead. North American Journal of Fisheries Management 12:598-603.

Hillman, T. W. and D. W. Chapman. 1996. Comparison of underwater methods and electrofishing for estimating fish populations in the Upper Blackfoot River Basin. BioAnalysts, Inc. Report to the Seven-Up Pete Joint Venture, Lincoln, MT.

Hillman, T. W. and A. E. Giorgi. 2002. Monitoring protocols: effectiveness monitoring of physical/environmental indicators in tributary habitats. BioAnalysts, Inc. Report to Bonneville Power Administration, Portland, Or. Web link:
http://www.efw.bpa.gov/cgi-bin/FW/welcome.cgi?ViewMode=External

Hillman, T. W. and M. D. Miller. 2002. Abundance and total numbers of Chinook salmon and trout in the Chiwawa River Basin, Washington. BioAnalysts, Inc. Report to Chelan County Public Utility District, Wenatchee, WA.

Hughes, R. M., E. Rexstad, and C. E. Bond. 1987. The relationship of aquatic ecoregions, river basins and physiographic provinces to the ichthyogeographic regions of Oregon. Copeia 2:423-432.

Hunt, C. B. 1967. Physiography of the United States. W. H. Freeman, San Francisco, CA.

Hurlbert, S. J. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54:187-211.

ICBTRT (Interior Columbia Basin Technical Recovery Team). 2003. Independent populations of Chinook, steelhead, and sockeye for listed evolutionarily significant units within the interior Columbia River domain. Working draft. NOAA Fisheries Northwest Fisheries Science Center, Seattle, WA.

Idea Works, Inc. 1997. Methodologists toolchest. Version 1.2. Distributed by Scolari, Sage Publications Software, Beverly Hills, CA.

ISAB (Independent Scientific Advisory Board). 2003. A review of strategies for recovering tributary habitat. Independent Scientific Advisory Board for the Northwest Power Planning Council, ISAB 2003-2, Portland, OR. Web link: http://www.nwppc.org/library/isab/Default.htm

Johnson, D. H., and nine others. 2001. Inventory and monitoring of salmon habitat in the Pacific Northwest—directory and synthesis of protocols for management/research and volunteers in Washington, Oregon, Montana, Idaho, and British Columbia. Review draft. Washington Department of Fish and Wildlife, Olympia, WA. Web link: http://www.wa.gov/wdfw/hab/sshiap/dataptcl.htm

Kaufmann, P. R., P. Levine, E. G. Robinson, C. Seeliger, and D. V. Peck. 1999. Quantifying physical habitat in wadeable streams. EPA/620/R-99/003, U.S. Environmental Protection Agency, Washington, D.C. Web link: http://www.epa.gov/emap/html/pubs/docs/groupdocs/surfwatr/field/phyhab.html

Krebs, C. J. 1999. Ecological methodology. Second edition. Benjamin/Cummings, Menlo Park, CA.

Lazorchak, J. M., D. J. Klemm, and D. V. Peck (editors). 1998. Environmental monitoring and assessment program—surface waters: field operations and methods for measuring the ecological condition of wadeable streams. EPA/620/R-94/004F, U.S. Environmental Protection Agency, Washington, D.C. Web link: http://www.epa.gov/emap/html/pubs/docs/groupdocs/surfwatr/field/ws_abs.html

Lee, K. N. 1993. Compass and gyroscope: integrating science and politics for the environment. Island Press, Washington, D.C.

Leopold, L. B., M. G. Wolman, and J. P. Miller. 1964 Fluvial processes in geomorphology. W. H. Freeman, San Francisco, CA.

Lipsey, M. W. 1990. Design sensitivity: statistical power for experimental research. Sage Publications, Beverly Hills, CA.

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

Manly, B. F. J. 1992. The design and analysis of research studies. Cambridge University Press, New York, NY.

Manly, B. F. J. 2001. Statistics for environmental science and management. Chapman and Hall, New York, NY.

Mapstone, B. D. 1995. Scalable decision rules for environmental impact studies: effect size, Type I, and Type II errors. Ecological Applications 5:401-410.

Meehan, W. R., editor. 1991. Influences of forest and range management on salmonid fishes and their habitats. American Fisheries Society Special Publication 19, Bethesda, MD.

Meekin, T. K. 1967. Report on the 1966 Wells Dam Chinook tagging study. Washington Department of Fisheries report to Douglas County Public Utility District, Contract Number 001-01-022-4201.

Montgomery, D. R. and J. M. Buffington. 1993. Channel classification, prediction of channel response, and assessment of channel condition. Washington State Timber/Fish/Wildlife Agreement, TFW-SH10-93-002, Department of Natural Resources, Olympia, WA. Web site: http://www.nwifc.wa.gov/cmerdoc/TFW_SH10_93_002.pdf

Montgomery, D. R. and J. M. Buffington. 1997. Channel-reach morphology in mountain drainage basins. Geological Society of American Bulletin 109:596-611.

Moore, K., K. Jones, and J. Dambacher. 2002. Methods for stream habitat surveys; aquatic inventory project. Oregon Department of Fish and Wildlife, Aquatic Inventories Project, Natural Production Program, Corvallis, OR. Web link: http://oregonstate.edu/Dept/ODFW/freshwater/inventory/methods.html

Moore, K. 2002. Draft Oregon plan for salmon and watersheds monitoring strategy. Oregon Plan, Salem, OR.

Mosey, T. R. and L. J. Murphy. 2002. Spring and summer Chinook spawning ground surveys on the Wenatchee River Basin, 2001. Chelan County Public Utility District, Wenatchee, WA.

Mullan, J. W., K. R. Williams, G. Rhodus, T. W. Hillman, and J. D. McIntyre. 1992. Production and habitat of salmonids in Mid-Columbia River tributary streams. U.S. Fish and Wildlife Service Monograph I, Leavenworth, WA.

Murdoch, A., K. Petersen, T. Miller, M. Tonseth, and T. Randolph. 2000. Freshwater production and emigration of juvenile spring Chinook salmon from the Chiwawa River in 1998. Report No. SS99-05, Washington Department of Fish and Wildlife, Olympia, WA.

Naiman, R. J., D. G. Lonzarich, T. J. Beechie, and S. C. Ralph. 1992. General principles of classification and the assessment of conservation potential in rivers. Pages 93-123 *in*: P. J. Boon, P. Calow, and G. E. Petts, editors. River conservation and management. John Wiley and Sons, New York, NY.

Nawa, R. K., C. A. Frissell, and W. J. Liss. 1988. Life history and persistence of anadromous fish stocks in relation to stream habitats and watershed classification. Annual progress report to Oregon Department of Fish and Wildlife, Portland, OR.

NCSS (Number Cruncher Statistical Systems). 2000. Pass 2000 power analysis and sample size for Windows. NCSS, Kaysville, UT.

NMFS (National Marine Fisheries Service). 1996. Making Endangered Species Act determinations of effect for individual or grouped actions at the watershed scale. The National Marine Fisheries Service Environmental and Technical Services Division, Habitat Conservation Branch, Seattle, WA.

NMFS (National Marine Fisheries Service). 2000. Guidelines for electrofishing waters containing salmonids listed under the Endangered Species Act. National Marine Fisheries Service, Portland, OR.

NRC (National Research Council). 1992. Restoration of aquatic ecosystems: science, technology, and public policy. National Academy Press, Washington, D.C.

Omernik, J. M. 1987. Aquatic ecoregions of the conterminous United States. Annals of the Association of American Geographers 77:118-125.

OPSW (Oregon Plan for Salmon and Watersheds). 1999. Water quality monitoring, technical guide book. Version 2.0. Corvallis, OR. Web link: http://www.oweb.state.or.us/publications/index.shtml

Overton, C. K., S. P. Wollrab, B. C. Roberts, and M. A. Radko. 1997. R1/R4 (Northern/Intermountain Regions) fish and fish habitat standard inventory procedures handbook. USDA Forest Service General Technical Report INT-GTR-346, Ogden, UT.

Overton, W. S., D. White, and D. L. Stevens. 1990. Design report for EMAP environmental monitoring and assessment program. U.S. Environmental Protection Agency, EPA/600/3-91/053, Corvallis, OR.

Parker, M. A. 2000. Fish passage – culvert inspection procedures. Watershed Restoration Technical Circular No. 11. Ministry of Environment, Lands and Parks and Ministry of Forest, British Columbia.

Parker, R. A. and N. G. Berman. 2003. Sample size: more than calculations. The American Statistician 57:166-170.

Parmenter, A. W., A. Hansen, R. E. Kennedy, W. Cohen, U. Langener, R. Lawrence, B. Maxwell, A. Gallant, and R. Aspinall. 2003. Land use and land cover in the greater Yellowstone ecosystem: 1975-1995. Ecological Applications 13:687-703.

Paulsen, C., S. Katz , T. Hillman, A. Giorgi, C. Jordan, M. Newsom, and J. Geiselman, 2002 Guidelines for Action Effectiveness Research Proposals for FCRPS Offsite Mitigation Habitat Measures. Bonneville Power Administration, U.S. Bureau of Reclamation, U.S. Army Corps of Engineers, and NOAA Fisheries, Portland, OR. Web link:

http://www.efw.bpa.gov/cgiin/FW/welcome.cgi?ViewMode=ExternalView.

Peck, D. V., J. M. Lazorchak, and D. J. Klemm. 2001. Environmental monitoring and assessment program—surface waters: western pilot study field operations manual for wadeable streams. Draft Report. EPA/XXX/X-XX/XXX, U.S. Environmental Protection Agency, Washington, D.C. Web link: http://www.epa.gov/emap/html/pubs/docs/groupdocs/surfwatr/field/ewwsm01.html *[Although this draft document states that it should not be cited or quoted, some of the material in the report is an important improvement to Lazorchak et al. (1998). By not citing the document, it may give the appearance that I improved some of the methods outlined in the Lazorchak et al. report. To avoid this, I feel it necessary to offer credit where credit is due.]*

Peterman, R. M. 1990. Statistical power analysis can improve fisheries research and management. Canadian Journal of Fisheries and Aquatic Sciences 47:2-15.

Peterson, J. T., R. F. Thurow, and J. W. Guzevich. 2004. An evaluation of multipass electrofishing for estimating the abundance of stream-dwelling salmonids. Transactions of the American Fisheries Society 133:462-475.

Platts, W. S., W. F. Megahan, and G. W. Minshall. 1983. Methods for evaluating stream, riparian, and biotic conditions. USDA Forest Service General Technical Report INT-138, Ogden, UT.

Platts, W. S. and twelve others. 1987. Methods for evaluating riparian habitats with applications to management. USDA Forest Service General Technical Report INT-221, Ogden, UT.

Reeves, G. H., and nine others. 2001. Aquatic and riparian effectiveness monitoring plan for the Northwest Forest Plan. USDA Forest Service, Pacific Northwest Research Station, Corvallis, OR.

Reynolds, J. B. 1996. Electrofishing. Pages 221-253 *in*: B. R. Murphy and D. W. Willis, editors. Fisheries techniques, 2nd edition. American Fisheries Society, Bethesda, MD.

Reynolds, L., A. T. Herlihy, P. R. Kaufmann, S. V. Gregory, and R. M. Hughes. 2003. Electrofishing effort requirements for assessing species richness and biotic integrity in Western Oregon streams. North American Journal of Fisheries Management 23:450-461.

Roni, P., editor. 2005. Monitoring stream and watershed restoration. American Fisheries Society, Bethesda, MD.

Roni, P., L. Weitkamp, and J. Scordino. 1999. Identification of essential fish habitat for salmon in the Pacific Northwest: initial efforts, information needs, and future direction. American Fisheries Society Symposium 22:93-107.

Roper, B. B., J. L. Kershner, and R. C. Henderson. 2003. The value of using permanent sites when evaluating stream attributes at the reach scale. Journal of Freshwater Ecology 18:585-592.

Rosgen, D. 1996. Applied river morphology. Wildland Hydrology, Pagosa Springs, CO.

Royce, W. F. 1996. Introduction to the practice of fishery science. Revised edition. Academic Press, New York, NY.

Scheaffer, R. L., W. Mendenhall, and L. Ott. 1990. Elementary survey sampling. Fourth edition. PWS-KENT Publishing Company, Boston, MA.

Schuett-Hames, D., J. Ward, M. Fox, A. Pleus, and J. Light. 1994. Large woody debris survey module. Section 5 *in*: D. Schuett-Hames, A. Pleus, L. Bullchild, and S. Hall, editors. Ambient monitoring program manual. Timber-Fish-Wildlife TFW-AM9-94-001, Northwest Indian Fisheries Commission, Olympia, WA. Web link: http://www.nwifc.wa.gov/TFW/documents.asp

Schuett-Hames, D., A. E. Pleus, E. Rashin, and J. Matthews. 1999a. Method manual for the stream temperature survey. Timber-Fish-Wildlife TFW-AM9-99-005, Northwest Indian Fisheries Commission, Olympia, WA. Web link: http://www.nwifc.wa.gov/TFW/documents.asp

Schuett-Hames, D., R. Conrad, A. Pleus, and M. McHenry. 1999b. Method manual for the salmonid spawning gravel composition survey. Timber-Fish-Wildlife TFW-AM9-99-006, Northwest Indian Fisheries Commission, Olympia, WA. Web link: http://www.nwifc.wa.gov/TFW/documents.asp

Simonson, T., J. Lyons, and P. Kanehl. 1994. Quantifying fish habitat in streams: transect spacing, sample size, and a proposed framework. North American Journal of Fisheries Management 14:607-615.

Skalski, J. R. and D. S. Robson. 1992. Techniques for wildlife investigations, design and analysis of capture data. Academic Press, New York, NY.

Smith, E. P., D. R. Orvos, and J. Cairns. 1993. Impact assessment using the before-after-control-impact (BACI) model: concerns and comments. Canadian Journal of Fisheries and Aquatic Sciences 50:627-637.

Smith, R. L. and T. M. Smith. 2001. Ecology and field biology. Sixth edition. Benjamin Cummings, New York, N.Y.

Spence, B. C., G. A. Lomnicky, R. M. Hughes, and R. P. Novitzki. 1996. An ecosystem approach to salmonid conservation. TR-4501-96-6057, Management Technology, Corvallis, OR.

Stevens, D. L. 1997. Variable density grid-based sampling designs for continuous spatial populations. Environmetrics 8:167-195.

Stevens, D. L. 2002. Sampling design and statistical analysis methods for the integrated biolgocial and physical monitoring of Oregon streams. Report No. OPSW-ODFW-2002-07, The Oregon Plan for Salmon and Watersheds, Oregon Department of Fish and Wildlife, Corvallis, OR.

Stevens, D. L. and A. R. Olsen. 1999. Spatially restricted surveys over time for aquatic resources. Journal of Agricultural, Biological, and Environmental Statistics 4:415-428.

Stevens, D. L. and N. S. Urquhart. 2000. Response designs and support regions in sampling continuous domains. Environmetrics 11:13-41.

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

Stewart-Oaten, A., J. R. Bence, and C. W. Osenberg. 1992. Assessing effects of unreplicated perturbations: no simple solutions. Ecology 73:1396-1404.

Strahler, A. N. 1952. Hypsometric (area-altitude) analysis of erosional topography. Bulletin of the Geological Society of America 63:1117-1142.

Strange, R. J. 1996. Field examination of fishes. Pages 433-446 *in*: B. R. Murphy and D. W. Willis, editors. Fisheries techniques, 2nd edition. American Fisheries Society, Bethesda, MD.

Thompson, W. L., G. C. White, and C. Gowan. 1998. Monitoring vertebrate populations. Academic Press, Inc., New York, N.Y.

Thurow, R. F. 1994. Underwater methods for study of salmonids in the Intermountain West. USDA Forest Service, General Technical Report INT-GTR-307, Ogden, UT.

Thurow, R. F., J. T. Peterson, C. A. Larsen, and J. W. Guzevich. 2004. Development of bull trout sampling efficiency models. Final report to U.S. Fish and Wildlife Service, Aquatic Resources Division, Lacey, WA.

Thurow, R. F., J. T. Peterson, and J. W. Guzevich. 2006. Utility of validation of day and night snorkel counts for estimating bull trout abundance in first- to third-order streams. North American Journal of Fisheries Management 26:117-132.

Underwood, A. J. 1994. Things environmental scientists (and statisticians) need to know to receive (and give) better statistical advice. Pages 33-61 *in*: D. J. Fletcher and B. F. J. Manly, editors. Statistics in ecology and environmental monitoring. University of Otago Press, Dunedin.

Underwood, A. J. 1996. Spatial and temporal problems with monitoring. Pages 182-204 *in*: G. Petts, and P. Calow, editors. River restoration. Blackwell Science, Cambridge, MA.

USFWS (U.S. Fish and Wildlife Service). 2000. Biological Opinion, effects to listed species from operations of the Federal Columbia River Power System. U.S. Fish and Wildlife Service, Regions 1 and 6, Portland, OR.

Van Deventer, J. S. and W. S. Platts. 1989. Microcomputer software system for generating population statistics from electrofishing data—user's guide for MicroFish 3.0. USDA Forest Service General Technical Report INT-254, Ogden, UT.

WDFW (Washington Department of Fish and Wildlife). 2000. Fish passage barrier and surface water diversion screening assessment and prioritization manual. Washington Department of Fish and Wildlife Habitat Program, Environmental Restoration Division, Olympia, WA. Web link: http://www.wa.gov/wdfw/hab/engineer/fishbarr.htm

Wertz, W. A. and J. F. Arnold. 1972. Land systems inventory. U.S. Forest Service, Intermountain Region, Ogden, UT.

WFC (World Forestry Center). 1998. Pilot study report, Umpqua land exchange project. World Forestry Center, Portland, OR. Web link: http://www.or.blm.gov/umpqua/documents.htm

WFPB (Washington Forest Practices Board). 1995. Washington forest practices board manual: standard methodology for conducting watershed analysis under Chapter 222-22 WAC. Version 3.0. Washington Forest Practices Board, Olympia, WA. Web link: http://www.dnr.wa.gov/forestpractices/watershedanalysis/

Whittier, T. R., R. M. Hughes, and D. P. Larsen. 1988. Correspondence between ecoregions and spatial patterns in stream ecosystems in Oregon. Canadian Journal of Fisheries and Aquatic Sciences 45:1264-1278.

Williams, G. P. 1978. Bank-full discharge of rivers. Water Resources Research 14:1141-1154.

Wipfli, M. S. and D. P. Gregovich. 2002. Export of invertebrates and detritus from fishless headwater streams in southeastern Alaska: implications for downstream salmonid production. Freshwater Biology 47:957-969.

Wolman, M. G. and L. B. Leopold. 1957. River floodplains: some observations on their formation. U.S.G.S. Professional Paper 282-C.

WSRFB (Washington Salmon Recovery Funding Board). 2003. Draft 5/16/2003 monitoring and evaluation strategy for habitat restoration and acquisition projects. Washington Salmon Recovery Funding Board, Olympia, WA. Web link: http://www.iac.wa.gov/srfb/docs.htm

Zaroban, D. W. 2000. Protocol for placement and retrieval of temperature data loggers in Idaho streams. Idaho Division of Environmental Quality, Boise, ID. Web link: http://www.deq.state.id.us/water/tlp.htm

APPENDIX A—WENATCHEE BASIN

APPENDIX B—ENTIAT BASIN

APPENDIX C—METHOW BASIN

APPENDIX D—OKANOGAN BASIN

APPENDIX E—CHELAN RIVER AND SMALL TRIBUTARIES