Monitoring Protocols: Effectiveness Monitoring of Physical/Environmental Indicators In Tributary Habitats

Prepared by:

T. W. Hillman
A. E. Giorgi

BioAnalysts, Inc.
Boise, Idaho

Prepared for:

Bonneville Power Administration
Portland, Oregon

25 July 2002

Table of Contents

1.0 Introduction

The U.S. Army Corps of Engineers, the Bureau of Reclamation, and the Bonneville Power Administration (collectively referred to as the Action Agencies) recently prepared a draft Implementation Plan for the Federal Columbia River Power System (FCRPS) (ACOE, BR, and BPA 2001). The Implementation Plan responds to the Biological Opinions (BiOp) issued by the U.S. Fish and Wildlife Service (USFWS) and the National Marine Fisheries Service (NMFS) on the effects to listed species from operations of the federal hydropower system. The Implementation Plan outlines a blueprint that organizes collective fish recovery actions by the Action Agencies.

One important component of the Implementation Plan is the implementation of Habitat Strategies that will improve the survival of listed species within tributary habitats. The goal of the Tributary-Habitat Strategy is to improve survival by protecting and enhancing the structure and function of the aquatic ecosystem. This will be accomplished by protecting existing "high-quality" habitat, enhancing degraded habitat on a priority basis and connecting those to other properly functioning habitats, and preventing further degradation of habitat and water quality. According to the Implementation Plan, the Action Agencies are required to assess the benefits associated with implementing those actions at prescribed check-ins over the next decade and beyond.

The Implementation Plan identifies four Tiers of responses that should be monitored to ensure management actions are effective and listed populations are advancing toward recovery. Of immediate concern to the Action Agencies are responses that are likely to provide insight into action effectiveness, particularly as determined over the next decade. Once compliance has been established (Tier 4), the next response level is Tier 3, where physical/environmental and biological responses are expressed. Biological responses will typically lag behind changes in physical/environmental conditions.

The purpose of this paper is to identify protocols suitable for monitoring physical/environmental conditions at Tier 3, which are pertinent to the needs expressed in the Implementation Plan and the NMFS BiOp. Indeed, Reasonable and Prudent Alternative (RPA) 183 in the NMFS BiOp directs research on tributary mitigation actions intended to improve salmon and steelhead habitat and survival. In addition, the BiOp (Section 9.6.5.3.3) indicates that each major habitat management action should be assessed immediately to obtain enough information for a complete evaluation at the 5- and 8-year check-in points. We wrote this document to aid researchers in developing effectiveness research programs that are consistent with the mandates of the BiOp.

This monitoring plan focuses on freshwater tributary systems that comprise and drain into the Columbia and Snake rivers. Although we wrote this paper as a stand-alone document, it is actually one component of a larger Research, Monitoring, and Evaluation (RME) program. Other papers will identify protocols for monitoring biological components in tributary streams, and both biological and physical components in the mainstem, estuary, and near-shore ocean habitats. This paper is closely tied to the paper that describes effectiveness monitoring of biological indicators in tributary habitats

(NMFS in prep). In fact, research programs designed to assess the effectiveness of management actions in tributaries should measure both physical/environmental and biological indicators. The reason for this is to test relationships between the physical environment and biological responses.

The specific objectives of this report are to (1) identify an appropriate set of physical/environmental indicators or variables to be monitored, (2) recommend protocols for collecting those data, and (3) outline general guidelines for developing effectiveness monitoring designs. Before we address these objectives, we briefly survey the status of several regional monitoring programs that track habitat status and trends over broad geographic areas. This review provides context for the Action Agencies offsite mitigation efforts.

We divided this report into six major parts. The first part (Section 2.0) provides the rationale for establishing a physical/environmental-monitoring plan. Section 3.0 identifies and describes different types of monitoring, and identifies and compares various monitoring programs proposed or currently in use in the Pacific Northwest. Section 4.0 identifies and describes physical/environmental indicator variables that can be monitored as part of the Implementation Plan. Importantly, this section also presents general criteria or "interim performance standards" for each of the indicators. Most of these performance standards (PS) should be considered as working hypotheses, because no standard applies to all environments in all conditions. Section 5.0 identifies protocols for measuring each indicator, while Section 6.0 addresses issues associated with valid monitoring designs. Finally, Section 7.0 describes how this document should be used to guide effectiveness monitoring in tributary habitats. It provides the reader with a checklist of questions for designing an effectiveness monitoring study.

Throughout this document we attempted to keep discussions fairly general. Because this report discusses some issues that are quite involved, we used footnotes to define technical terms, offer further explanation, offer alternative explanations, or to describe a given topic or thought in more detail. We hope the reader will not be too distracted by the extensive use of footnotes. In some instances, however, it was necessary to provide considerable detail within the text (e.g., discussion on choosing sample sizes).

2.0 Rationale for Physical/Environmental Monitoring Program

The Action Agencies have an interest and responsibility to assess if actions funded by them under the 2000 FCRPS BiOp are effective. Given this, the RME program of the Implementation Plan focuses on the effectiveness of habitat management actions or RPAs. The Implementation Plan provides explicit language that directs the scope of monitoring activities. The three primary objectives of RME under the Implementation Plan are:

1. Identify the physical and biological responses to management actions.

2. Track the status of fish populations (e.g., Evolutionarily Significant Units) and their environment relative to required performance standards.
3. Resolve critical uncertainties in the methods and data required for the evaluation of future population performance and needed survival improvements.

The Implementation Plan, which is consistent with the NMFS and USFWS BiOps, directs the Action Agencies to develop a monitoring program to determine if RPAs achieve specified targets, goals, or objectives (hereafter referred to as performance standards or PS). As a result, the Action Agencies must develop performance measures (PM) that monitor progress toward specified PS. According to the Implementation Plan, physical/environmental PM can be placed into one of three classes: (1) preservation measures, (2) water quality/quantity measures, and (3) physical measures. As such, monitoring can involve either the enumeration of measures (e.g., number of acres of riparian habitat secured or miles of stream accessible following barrier improvement or removal), or the measurement of a change in a physical attribute (e.g., document the change in water temperature). To be useful, physical/environmental PM should have some, but preferably all of the following characteristics:

- They should reflect effects of habitat-based management actions.
- They should be measurable using established methods and technology.
- They should be readily interpretable (i.e., changes in the physical/environmental indicator should not be confounded or masked by either extraneous processes or inherent variability associated with the variable itself).

To satisfy these characteristics, the Implementation Plan requires the monitoring plan to identify a suite of physical/environmental indicator variables that are likely to respond to proposed actions (see Section 4.0).[1] In addition, the monitoring plan must describe suitable sampling designs. For example, the monitoring plan must consider the sampling frame, spatial and temporal variation, sampling methods, sampling frequency, and the use of reference (control) areas and baseline information (see Sections 5.0 and 6.0).

A separate but related matter concerns the time required for physical/environmental indicators to change in response to actions. Depending on the nature of the action, some indicators will respond rapidly (within years), while others require decades. Also, the natural variability of the physical/environmental indicators can affect our ability to detect a response. Noisy or cyclical processes may require concerted sampling effort (i.e., use of appropriate spatial references), or an extended time frame to document a habitat response, if it occurs.

Implementation of effectiveness monitoring experiments by the Action Agencies cannot occur in isolation. Proposed projects to be funded by the Action Agencies are subject to

[1] The measurement of physical/environmental indicators is no substitute for monitoring biological indicators (e.g., survival, fish condition, abundance). The NMFS BiOp clearly states that the program will monitor changes in survival associated with habitat management actions. Both physical/environmental and biological indicators need to be measured as part of a valid monitoring program.

technical review by the Independent Scientific Review Panel (ISRP) of the Northwest Power Planning Council (NPPC). The ISRP reviews and assesses the merits of habitat projects funded by BPA. The Action Agencies will rely on the ISRP to provide direction and guidance regarding proposed monitoring and evaluation projects.

3.0 Classification of Monitoring Activities

3.1 Types of Monitoring

In general, monitoring can be defined as a series of observations or measurements over time (MacDonald et al. 1991). Often the purpose of monitoring is to document changes associated with the implementation of some management action(s). In this case, monitoring is essentially an experiment. For example, one might measure water temperature in treatment and control sites several times before and after the removal to livestock from the riparian zone (removal of livestock is the management action). Here, monitoring is used to demonstrate an improvement in water quality (i.e., temperature) caused by the exclusion of livestock from the riparian zone. Importantly, monitoring is not limited to measuring temporal variability, but should describe both temporal and spatial variability (see Section 6.0). As a result, federal and state agencies have defined several types of monitoring. Below, we identify several different types of monitoring, following the definitions provided in MacDonald et al. (1991), OPSW (1999), and FISRWG (2001).

Trend Monitoring—Trend monitoring involves measurements taken at regular time intervals in order to assess the long-term trend in a particular parameter. Usually, the measurements are not taken specifically to evaluate management practices. Rather, they serve to describe changes in the parameter over time.

Status or Baseline Monitoring—Baseline monitoring is used to characterize existing or undisturbed conditions, and to establish a database for future comparisons. The intent of baseline monitoring is to capture temporal variability of the parameters of interest. There is no explicit end point at which continued baseline monitoring becomes trend monitoring.

Implementation Monitoring—This type of monitoring assesses whether activities were carried out as planned. This is generally carried out as an administrative review and does not require any parameter measurements. This type of monitoring cannot directly link management actions to physical/environmental responses, as no physical/environmental parameters are measured.

Effectiveness Monitoring—Effectiveness monitoring evaluates whether the management activities achieved the desired effect or goal. Success may be measured against "controls," "baseline conditions," or "desired future conditions." Project monitoring, a type of effectiveness monitoring, addressed the effectiveness of a particular project and the combination of measures used to protect aquatic habitat.

Validation Monitoring—Validation monitoring assesses the performance of a model or standard. It questions whether the underlying management assumptions and models are correct.

Compliance Monitoring—This type of monitoring determines whether specified criteria are being met. The criteria can be numeric or descriptive. Generally, regulations associated with individual criterion specify the location, frequency, and method of measurement.[2]

Clearly, these types of monitoring are not mutually exclusive. Usually the distinction between them is determined more by the purpose of monitoring rather than by the type and intensity of measurements. Nevertheless, MacDonald et al. (1991) broadly classified types of monitoring according to the frequency and duration of monitoring and the intensity of data analysis (Table 1).

Table 1. General characteristics of different types of monitoring (from MacDonald et al. 1991).

Type of monitoring	Frequency of measurements	Duration of monitoring	Intensity of data analysis
Trend	Low	Long	Low to moderate
Baseline or Status	Low	Short to medium	Low to moderate
Implementation	Variable	Duration of project	Low
Effectiveness	Medium to high	Short to medium	Medium
Validation	High	Medium to long	High
Compliance	Variable	Depends on project	Moderate to high

The ISRP (2001) recently identified the type and level of monitoring they expected to see in most habitat-based proposals. They stressed that for most projects, effectiveness monitoring should be the goal. That is, monitoring efforts should focus on documenting environmental or biological responses in the local vicinity of the project. They recommended that projects be less concerned with documenting changes across watersheds or sub-basins. Other specialized efforts should track conditions at those scales.

[2] This definition differs from that in the NMFS (2000) Biological Opinion. On page G-1, NMFS (2000) defines Compliance Monitoring as, "[h]ave management actions been properly implemented and maintained?" This definition is more consistent with Implementation Monitoring.

As we stated earlier, this document addresses effectiveness monitoring, or more appropriately, effectiveness research. Action 9 in the 2000 FCRPS BiOp specifies two principle motivations for effectiveness (Tier III) research:

> *"…Research, monitoring, and evaluation will provide data for resolving a wide range of uncertainties, including…establishing causal relationships between habitat (or other) attributes and population response, and assessing the effectiveness of management actions."*

The first motivation is to determine if actions are accomplishing their objectives (i.e., *assess the effectiveness of management actions*). The second is to develop mechanistic understanding of the relationships between salmon population response and RPA manipulations to guide future decisions on RPA activities (i.e., *establish causal relationships between habitat attributes and population response*). Therefore, the Action Agencies identified two levels of effectiveness monitoring: **basic effectiveness research** to estimate the effects of actions on environmental conditions and salmonid survival, and **intensive effectiveness research** to understand the mechanisms underlying environmental and survival changes.

Basic effectiveness research is designed to assess both the specific effects of individual management actions and the generic effects of classes of actions across the listed ESUs. The Action Agencies expect that most investigators will execute basic effectiveness research. Therefore, this type of effectiveness research will monitor a prescribed set of physical-environmental indicators. Their experimental designs will test hypotheses regarding the effects of management actions on physical-environmental conditions and life-stage survivals. Comparisons across classes of actions will be facilitated by measurement of a consistent set of indicators.

Intensive effectiveness research, on the other hand, is designed to investigate in detail mechanistic relationships between management actions and the environment or survival responses of salmonids. It consists of detailed ecological and ecosystem experiments. Hypotheses tested within an intensive research program address the ecological mechanisms behind the effects of management actions directly, rather than implicitly as in basic effectiveness research.

In this document, we focus on basic effectiveness research. A separate document prepared by the NMFS will address intensive effectiveness research.

3.2 Regional Monitoring Programs

When monitoring the effectiveness of management actions in tributary habitats, it is important that monitoring is consistent with other monitoring programs in the region. Clearly, a great deal of effort and money can be saved by implementing monitoring methods that are consistent across the basin. In addition, this should increase the probability of detecting a change in habitat conditions, in an inherently variable and dynamic environment.

Federal, state, and tribal agencies have instituted a variety of natural resource/ecosystem-based programs designed to monitor physical/environmental conditions in aquatic habitats. This section reviews the status of several prominent programs. This review is intended to provide information about the type, quality, and quantity of habitat data being sampled and archived by resource agencies in the Pacific Northwest. These programs include:

- Northwest Forest Plan (NFP)
- PACFISH/INFISH
- Interior Columbia Basin Ecosystem Management Project (ICBEMP)
- Oregon Plan for Salmon and Watersheds (OPSW)
- Salmon and Steelhead Habitat Inventory and Assessment Program (SSHIAP)
- The Northwest Power Planning Council Sub-Basin Planning Program
- Environmental Monitoring and Assessment Program (EMAP)

Northwest Forest Plan:

The NFP is based on the record of decision regarding the ESA-listing of the spotted owl as endangered. Actions under the plan affect the U.S. Forest Service (USFS) and Bureau of Land Management (BLM). Geographically, the affected area covers the range of the northern spotted owl as well as the marbled murrelet. In general this encompasses the area west of the Cascade crest, from Northern California through Washington. Thus, there is limited overlap with the Columbia Basin as treated under the Implementation Plan. The aquatic and riparian ecosystems are fundamental components treated under the NFP.

Under the NFP, Reeves et al. (2001) developed the Aquatic and Riparian Effectiveness Monitoring Plan (AREMP). The monitoring plan is intended to characterize the ecological condition of watersheds and aquatic ecosystems. It will describe present conditions, track trends in condition over time, and report on the effectiveness of the NFP. Its focus is to describe status and subsequent trends in ecosystem condition. In its current form the monitoring plan references preferred measures and protocols for monitoring indicators of interest. The plan also describes sampling designs and analytical approaches in more detail than other plans reviewed thus far.

Physical/environmental indicators identified in the AREMP include (Reeves et al. 2001):

- Road/stream crossing density
- Channel connectivity, sinuosity, pool depth and frequency
- Structural complexity (e.g., LWD, boulders)
- Substrate composition
- Water quality
- Water quantity

Data were collected at a pilot scale in 2000 and 2001 to test and tighten protocols and logistics. The intent is to conduct more systematic sampling in 2002 and beyond. Based on information presented in Reeves et al. (2001), it is unclear to what extent a final data-management system has been adopted for full implementation.

The AREMP provides an excellent model for the Action Agencies to consider for application in the Columbia Basin. Unfortunately, the geographical overlap of the NFP and the Implementation Plan is very limited. Therefore, the Action Agencies cannot rely heavily on the data collected and compiled under the NFP.

PACFISH/INFISH:

In 1994, the USFS and BLM developed ecosystem-based, aquatic habitat and riparian-area management strategies for Pacific salmon and other anadromous species (PACFISH) and an Inland Native Fish Strategy (INFISH) for resident fish species outside the anadromous areas. Geographically, the program covers most of the Columbia Basin east of the Cascade Crest, and focuses on lands managed by the two federal agencies.

Those agencies and other cooperating agencies are actively engaged in a tributary habitat-monitoring program, which they refer to as effectiveness monitoring. The goal is to determine whether management practices implemented under PACFISH/INFISH are effective in maintaining or restoring the function of aquatic systems. Their monitoring plan can be viewed on their website (www.fs.fed.us/biology/fishecology). The version we reviewed was published in June 2001. Mr. Jeff Kershner with the USFS provided additional information.

Their plan is quite detailed and prescribes protocols for monitoring an assortment of habitat indicators (Appendix A). Those indicators are consistent with the "properly functioning condition" (PFC) indicators identified by NMFS (1996) and many of the environmental attributes incorporated into the Ecosystem Diagnosis and Treatment (EDT) model. Data are being collected systematically under the plan, including information gathered from existing reports, measuring indicators in the field, and through the use of remote sensing. At this juncture three, eight-person crews are collecting data in 160 6-HUC watersheds throughout the Columbia Basin. At the end of this field season, they will have surveyed about 450 6-HUC watersheds. The data are entered into a database (Access) that is maintained by the two agencies. The data are readily available to other agencies, and will be available on the web sometime in 2002.

Interior Columbia Basin Ecosystem Management Project:

According to the information we received from Mr. Carl Pence with the USFS, the following is a characterization of monitoring activities under this federal program involving both the USFS and BLM. While a final EIS was published, the Record of Decision (ROD) is still pending. Therefore, ICBEMP is not an established program. No physical/environmental habitat data are being sampled region-wide under ICBEMP. A general monitoring plan has been developed, but without a final ROD, monitoring will not proceed.

Oregon Plan for Salmon and Watersheds:

The Oregon Plan for Salmon and Watersheds has a statewide focus. It has two components: (1) the Coastal Salmon Restoration Initiative and (2) the Healthy Streams Partnership (HSP). With regard to habitat monitoring activities, the HSP is most immediately relevant to assessing tributary habitat status and trends. As part of the HSP, the technical guidebook, "Water Quality Monitoring" was published in 1999. That document identifies monitoring protocols for measuring water quality indicators. Those indicators include temperature, dissolved oxygen, pH, conductivity, turbidity, pollutants (toxic chemicals), and macroinvertebrates. In addition to these indicators, the Plan has provisions for monitoring physical habitat characteristics and conducting watershed assessments.[3]

Data are currently being collected as part of the program. Water quality information is being collected at various locations by a number of different groups statewide. The program classifies data into one of three levels according to data quality. Level A is the highest quality data and is appropriate for assessing compliance with water quality standards; Level C is the lowest quality data and is suitable for educational purposes but not rigorous resource quality assessments. Level B is intermediate in quality and is suitable for screening information or providing early warnings. For Implementation Plan application, Level A data seem most appropriate.

Geographically, the overlap with the Implementation Plan area is minimal. The emphasis so far has been on coastal areas, Willamette River, and the lower Columbia. As a consequence, there may be little information that is applicable in establishing aquatic habitat status with respect to the needs of the Implementation Plan. Nevertheless, some tributaries may be well described. Perhaps that information could be integrated into a larger information management system.

Salmon and Steelhead Habitat Inventory and Assessment Program:

The Salmon and Steelhead Habitat Inventory and Assessment Program (SSHIAP) is an information system that characterizes freshwater and estuarine habitat conditions and fish stock distribution in Washington State. The Washington Department of Fish and Wildlife and the Northwest Indian Fisheries Commission maintain the system jointly. Data describing habitat and fish stock conditions are consolidated in an Access database. Data will be displayed via a GIS system linked to the main database. The type of information contained in the database includes: fish distribution, barriers, and a variety of indicators that describe habitat condition.

[3] We encourage investigators to review the Oregon Plan documents. They include discussions on sampling design components and have experience with monitoring juvenile fish.

SSHIAP provides an infrastructure for organizing and compiling environmental and fish information collected by a variety of jurisdictions, agencies, and fisheries projects statewide. SSHIAP technicians instruct biologists in other organizations on the use of the system, which is voluntary. Thus far, most of the information entered into the system is from the western side of the state. The state is organized into 62 water resource inventory areas (WRIA). In a recent accounting, 25 of those areas had completed computer-based maps and 10 contained information on habitat condition and fish stocks distribution.

SSHIAP recently completed a survey of protocols for collecting habitat data (Johnson et al. 2001). Their objective was to provide guidance to diverse groups collecting environmental data. As the program develops, the intent is that data collected as part of sub-basin planning will be incorporated into the information system and will be available to the NMFS Technical Recovery Team.

The amount of information compiled by the SSHIAP for the geographic area covered under the Implementation Plan is limited. Their website gives an overview of the status of information compiled. As of 6 September 2001, the lower Columbia system is sparsely characterized and information from the upper Columbia reach has not yet been incorporated into the SSHIAP information system.

Northwest Power Planning Council Sub-Basin Planning:

An important part of sub-basin planning under the NPPC Fish and Wildlife Program is monitoring habitat condition. Habitat attributes are the foundation of the EDT model, which is used in many sub-basin assessments. NPPC staff developed a conceptual design for monitoring and evaluation (Bisbal 2001). But, because they are in the process of developing a monitoring plan based on those guidelines, there is no systematic habitat monitoring currently being conducted. They do intend to implement monitoring within the next few years. Presumably the monitoring plan will identify preferred monitoring protocols.

Environmental Monitoring and Assessment Program:

The Environmental Monitoring and Assessment Program (EMAP) is an EPA research project to develop tools to monitor and assess the status and trends of ecological resources across the nation. The ultimate goal is to translate environmental monitoring data from different temporal and spatial scales into assessments of ecological condition. The objectives are to advance ecological and risk assessment, guide national monitoring efforts, and improve the understanding of ecosystem dynamics. Although the scope of this program encompasses more than freshwater aquatic systems, it adequately treats freshwater ecosystems.

The information management system is a central feature of the system. The information management plan describes the system in detail and can be referenced through the EMAP web page. EPA researchers, other federal agencies, and state agencies can use EMAP.

Various agencies could benefit by adopting the EMAP infrastructure for collating and archiving environmental information. The program offers more than merely a database (Oracle). It specifies sampling guidelines, statistical designs, QA/QC procedures, supporting analytical tools, and establishes standards and protocols for collecting and managing environmental data. The system is a template and functional model for groups to use in assembling an integrated database for characterizing ecological status and trends for various ecosystems at locales across the country.

A branch of the program, Regional EMAP (R-EMAP), was established to assist agencies at the local level in the application of EMAP. R-EMAP uses EMAP's statistical design and indicators at smaller spatial and temporal scales. Table 2 lists R-EMAP projects and their status in Region 10, an area that embraces the Columbia Basin. There are efforts already in place that may offer the Action Agencies an opportunity to learn first hand the strengths and limitations of EMAP with respect to Implementation Plan objectives.

Table 2. List of R-EMAP Projects in Region 10.

Fiscal Year	**Project**	**Status**
FY94 - 96	Ecological Condition of Streams in the Coast Range Ecoregion of Oregon and Washington.	Completed
FY97	Ecological Condition of Upper Chehalis Basin Streams. Draft 2/5/01.	Field work completed, reporting continuing
FY97 - 98	Oregon Upper Deschutes River Basin	Field work completed, reporting continuing
FY99	Washington & Oregon Western Cascades Ecoregion	Field work completed, reporting continuing
FY00 - 03	Funds committed to the Western Pilot thru FY03	

An important issue is that the program is not systematically monitoring aquatic habitat condition across the Columbia Basin. Thus, there is currently no centralized data set that captures the existing information being collected basin-wide. However, PACFISH/INFISH, OPSW, and the NFP indicate that data being collected under their programs could be incorporated into EMAP. This would appear to be advantageous for the Action Agencies, if this diverse array of information could be integrated into a centralized system.

Synopsis:

There are several programs in place that are actively collecting and compiling data that describe the physical/environmental conditions of tributaries in the Columbia Basin. This network of information should be useful in characterizing the status of aquatic habitat within most of the area of interest under the Implementation Plan. However, no one program provides the needed geographic coverage or the suite of desired indicators critical to the needs of the Action Agencies. Nevertheless, several of these programs collectively (e.g., PACFISH/INFISH, NFP, and OPSW) may satisfy the needs of broad-scale habitat status monitoring expressed in the FCRPS BiOp. Indeed, programs like PACFISH/INFISH monitoring respond directly to other anadromous fish BiOps and should apply to some of the habitat monitoring needs under the FCRPS BiOp.

This document makes a distinction between habitat monitoring suitable for status and trends, and monitoring needed to evaluate the effectiveness of individual projects funded by the Action Agencies. The programs described above are more appropriate for monitoring status and trends. Therefore, the remainder of this document will focus on developing strategies and protocols needed to evaluate the effectiveness of projects funded by the Action Agencies.

4.0 Physical/Environmental Attributes

4.1 Indicator Variables

As noted above, the Action Agencies are directed to identify physical/environmental PM[4] that will be used to assess progress toward specified PS. These PM consist of a suite of "indicator" variables that should be sensitive to the proposed RPAs. Physical/environmental indicator variables have been identified to meet various purposes including assessment of fisheries production, identifying limiting factors, assessing effects of various land uses, and evaluating habitat improvement activities. For the purpose of evaluating habitat strategies, we believe physical/environmental indicators should possess the following characteristics:

- They should relate quantitatively with salmonid production.
- They should be sensitive to land-use activities or stresses.
- They should be consistent with other regional monitoring programs.
- They should lend themselves to reliable measurement.

Because the Implementation Plan and BiOp call for PS for each indicator variable, another characteristic is that one should be able to identify reasonable numeric PS for each variable. This, however, is a difficult task and presently is not possible for all indicators. Therefore, we identified narrative PS for those indicators that have no numeric standard.

We reviewed the literature (e.g., Bjornn and Reiser 1991; Spence et al. 1996; Gregory and Bisson 1997; and Bauer and Ralph 1999) and several regional monitoring programs (see Section 3.2) to identify a suite of physical/environmental indicator variables that fit the above criteria. Although we found a large number of indicator variables that could be used to monitor physical/environmental conditions, we believe many of those identified by the NMFS (1996) (Appendix B) and USFWS (1998) (Appendix C) as important

[4] In the context of habitat-based projects, PM can be used interchangeably with several terms that are often used in monitoring programs including: habitat indicators, variables, attributes, metrics, and parameters.

attributes of "properly functioning condition" closely fit our criteria. Indeed, the NMFS and USFWS use these indicators to evaluate the effects of land-management activities for conferencing, consultations, and permits under the ESA.

The physical/environmental indicators identified by the NMFS and USFWS are consistent with many indicators used in other monitoring programs. For example, indicators used by the NMFS are a mix of PACFISH (Appendix A) and other habitat indicators. The habitat component of the USFWS matrix varies only slightly from the NMFS matrix and is used to assess habitat conditions within bull trout streams. These indicators are also consistent with "key" parameters used in the Ecosystem Diagnosis and Treatment model. Recent analyses by Mobrand Biometrics indicated that certain 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).

We combined the NMFS and USFWS matrices to establish a list of pathways, general indicators, and specific indicators that should be monitored by the Action Agencies as part of basic effectiveness research (Table 3). This resulted in six pathways that address water quality, habitat access, habitat quality, channel conditions, flow and hydrology, and watershed conditions. Each of these pathways consists of one or more general indicators. For example, water temperature, sediment/turbidity, and contaminants/nutrients are the three general indicators of water quality. In turn, each general indicator is described by several specific indicators, which are measurable and linked directly to PS. We identified a total of 27 specific indicator variables (Table 3)[5]. In sum, these pathways and their associated indicators address watershed process and "input" variables (e.g., artificial physical barriers, road density, and disturbance) as well as "outcome" variables (e.g., temperature, sediment, woody debris, pools, riparian habitat, etc.), as required by the Implementation Plan and BiOp.

Both the Implementation Plan and BiOp require the Action Agencies to identify PS for physical/environmental conditions in tributaries within the Columbia River basin. This plan identifies interim PS for each of the 27 specific indicators (Table 3). Because these PS provide regional targets or goals, they will not be attainable in all environmental settings. For example, a large woody debris PS of >20 pieces per mile cannot be attained in a non-forested stream and may not be attainable in some forested streams (see Fox 2001). Therefore, this plan identifies "interim" PS, which can be modified to reflect conditions in a specific watershed or stream reach based on local geology, topography, climate, and potential vegetation. In fact, this plan encourages managers and investigators to modify the PS according to the potential of the environmental setting in which they implement management activities. Thus, the interim PS identified in this document could be viewed as working hypotheses. Indeed, it is anticipated that as the

[5] The reader should understand that this is simply a minimum set of physical/environmental indicators that should be included in effectiveness monitoring studies. Researchers may want to include additional indicators as part of their study.

effectiveness monitoring program matures, the results produced will demonstrate the actual appropriateness and usefulness of PS like those in the matrix of pathways and indicators. As a general guideline, we recommend that "specific" standards be based on historical information if available, or information from similar watersheds that are mostly undisturbed. In the absence of "specific" PS, "interim" PS can be used to guide monitoring.

The interim PS that we identify in this report are closely associated with physical/environmental conditions considered by the NMFS (1996) and USFWS (1998) as functioning properly. We believe these PS more likely set inapplicable target values across large geographic areas than contribute to incremental habitat deterioration. This strategy is consistent with the ESA, which requires the federal regulatory agencies to use a very conservative approach to protect habitat for endangered species.

What follows is a brief description of each indicator variable. We define each indicator, describe its response to land use or management activities[6], and identify its corresponding interim PS. Section 5.0 identifies recommended methods for measuring each indicator. We have not attempted to convert all measurements to either English or metric systems. Readers will find both in this section. 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), and OPSW (1999).

---

[6]Although physical/environmental indicators can be affected by natural events such as landslides, storms, floods, and fires, our discussion here focuses on effects of various land uses on indicators. Effectiveness research, however, must be able to separate effects associated with natural events and land-use activities.

Table 3. Classes, interim performance standards (PS), and protocols for monitoring physical/environmental factors in tributary habitat (Tier 3 Monitoring). Classification follows the NMFS (1996) and USFWS (1998) matrices of pathways and indicators for evaluating effects of human activities on salmonid habitat.

Pathway	General Indicators	Specific Indicators	Interim PS	Recommended Protocols
Water quality	Water temperature	MDMT	Salmon and Steelhead: Spawning: June-Sept 17.5ºC Sept-May 14.5ºC Rearing 17.5ºC Migration 17.5ºC Adult holding 17.5ºC	Schuett-Hames, et al. (1999a); Zaroban (2000)
		MWMT	Bull Trout: Incubation 2-5ºC Rearing 4-10ºC Spawning 1-9ºC Salmon and Steelhead: Spawning: June-Sept 15ºC Sept-May 12ºC Rearing 15ºC Migration 15ºC Adult holding 15ºC	Schuett-Hames et al. (1999a); Zaroban (2000)

Table 3. Continued.

Pathway	General Indicators	Specific Indicators	Interim PS	Recommended Protocols
Water quality	Sediment/Turbidity	Turbidity	Acute <70 NTU Chronic <50 NTU For streams that naturally exceed these standards: Turbidity should not exceed natural baseline levels at the 95% CL.	OPSW (1999)
		Depth fines	Fines (<0.85 mm) within spawning gravels <12%.	Platts et al. (1983); Schuett-Hames et al. (1999b)
	Contaminants/Nutrients	Metals/Pollutants	Not to exceed EPA (1986, 1999) general criteria.	APHA, AWWA, and WEF (1999)
		Ph	6.5-9.0	OPSW (1999)
		DO	One-day minimum ≥8.0 mg/L Seven-day mean ≥9.5 mg/L	OPSW (1999)
		Nitrogen	Nitrate ≤10 mg/L Nitrite ≤0.06 mg/L Ammonia = see EPA (1986)	OPSW (1999)
		Phosphorus	Phosphates: Lake/Res ≤0.025 mg/L Streams to Lake/Res ≤0.050 mg/L Other streams ≤0.10 mg/L	OPSW (1999)

Table 3. Continued.

Pathway	General Indicators	Specific Indicators	Interim PS	Recommended Protocols
Habitat Access	Artificial Physical Barriers	Road crossing (culverts)	General: Upstream and downstream passage is possible at all flows. Numeric: Connectance = 1	Parker (2000); WDFW (2000)
		Diversion dams	General: Upstream and downstream passage is possible at all flows and no fish end in irrigation systems. Numeric: Connectance = 1	Bain and Stevenson (1999); WDFW (2000)
		Fishways	General: Upstream and downstream passage is possible at all flows. Numeric: Connectance = 1	WDFW (2000)
Habitat Quality	Substrate	Dominant substrate	Gravels or small cobbles make up >50% of the bed materials in spawning areas.	Bevenger and King (1995); Bunte and Abt (2001)
		Embeddedness	Embeddedness in spawning and rearing areas is <20%.	MacDonald et al. (1991)
	LWD	Pieces per mile	LWD in forested streams is >20 pieces/mile and an adequate source of LWD is available for recruitment.	Overton et al. (1997); BURPTAC (1999)

Table 3. Continued.

Pathway	General Indicators	Specific Indicators	Interim PS	Recommended Protocols
Habitat Quality	Pools	Pools per mile	Pool frequency: Channel width — No. pools/mile 5 ft — 184 10 ft — 96 15 ft — 70 20 ft — 56 25 ft — 47 50 ft — 26 75 ft — 23 100 ft — 18 125 ft — 14 150 ft — 12 200 ft — 9	Overton et al. (1997); Platts et al. (1983)
		Pool quality	Maximum pool diameter exceeds the mean stream width by >10% and residual pool depth >1 m. Maximum pool diameter exceeds the mean stream width by >10%, residual pool depth is 0.6-1.0 m, and the pool has abundant fish cover.	Platts et al. (1983)
	Off-channel habitat	Off-channel habitat	For channels with gradients <3%, several backwater areas with cover and low-energy off-channel habitats are present.	WFPB (1995); Reeves et al. (2001)

Table 3. Concluded.

Pathway	General Indicators	Specific Indicators	Interim PS	Recommended Protocols
Channel Condition	Width/Depth Ratio	W/D	W/D < 10	BURPTAC (1999)
	Streambank Condition	Bank stability	Banks along >80% of any stream reach are >90% stable.	Platts et al. (1987); BURPTAC (1999)
Flow/Hydrology	Streamflows	Change in peak flow	Peak flows within a watershed do not exceed natural baseline conditions at the 95% level.	Bain and Stevenson (1999); MacDonald et al. (1991)
		Change in base flow	Base flows within a watershed do not exceed natural baseline conditions at the 95% level.	Bain and Stevenson (1999); MacDonald et al. (1991)
		Change in timing of flow	Timing of major flow events (peak and base flows) within a watershed does not exceed natural baseline conditions at the 95% level.	Bain and Stevenson (1999)
Watershed Condition	Road Density	Watershed road density	Road density <2 miles/mile^2	WFC (1998); Reeves et al. (2001)
		Riparian-road index	RRI = 0.00	WFC (1998)
	Disturbance	Equivalent clearcut area	Disturbance is <15% ECA of watershed with no disturbance concentrated in unstable areas, refugia, or riparian areas.	USFS (1974); King (1989)
	Riparian Habitat	Percent vegetation altered	<20% of riparian vegetation is altered.	Platts et al. (1987)

Water Quality

Water Temperature:

Definition—Water temperature is the net result of a variety of energy transfer processes, including radiation inputs, evaporation, convection, conduction, and advection. Water temperature reflects both the seasonal change in net radiation and the daily changes in air temperature. Stream characteristics such as velocity, depth, canopy cover, and groundwater inflow can modify patterns of energy input and output. Typically, peak daily temperatures occur in the late afternoon, while daily minima occur just before dawn. Seasonal patterns of stream temperature generally track patterns of incoming solar radiation, but with a lag of one to two months.

We selected two temperature metrics that can serve as specific indicators of water temperature: maximum daily maximum temperature (MDMT) and maximum weekly maximum temperature (MWMT) (Table 3). 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.

Response to Activities—Any activity that affects the energy transfer process can alter stream temperatures. For example, removal of riparian vegetation can increase the incident solar radiation and hence increase maximum summer water temperatures. Beschta et al. (1987) showed that complete removal of the forest canopy increases maximum daily stream temperatures in the summer by 3-8°C, although daily summer minima increased by only 1-2°C. Other activities that can alter steam temperatures include transportation systems,[7] urban and industrial development, grazing and other agricultural activities, water withdrawals, mining, diversions and dams, and point-source thermal inputs. In general, any activity or combination of activities that affect stream flow, velocity, depth, or canopy cover can alter the temperature regime of streams.

Performance Standards—The EPA, state agencies, and tribes have established temperature criteria for coldwater fishes. They developed these criteria to meet site-specific requirements for successful migration, spawning, egg incubation, fry and juvenile rearing, and adult rearing by various species or classes of coldwater fishes. This plan adopts those criteria as PS for water temperatures in tributary habitat (Table 3).

- For bull trout, the EPA and USFWS have established incubation, rearing, and spawning MWMT temperature criteria of 2-5ºC, 4-10ºC, and 1-9ºC,

[7] We define transportation systems as multilane paved highways, unpaved secondary roads, and railroad corridors.

respectively. Spawning migration corridors for bull trout should not exceed 15ºC MWMT.

States are currently reviewing and revising temperature criteria for salmon and steelhead. For example, the Washington Department of Ecology (WDOE) recently proposed a salmon and steelhead spawning criterion of 15ºC MWMT, with MDMT no greater than 17.5ºC during June through mid-September (Hicks 2000). From mid-September through May, WDOE proposes a criterion for spawning of 12ºC MWMT, with MDMT no greater than 14.5ºC. For waters used for rearing, migration, or holding by salmon and steelhead, WDOE proposes a 15ºC MWMT, with MDMT no greater than 17.5ºC.

- Until temperature criteria are finalized, interim PS for water temperatures for salmon and steelhead should follow Washington State temperature criteria.[8]

Sediment and Turbidity:

Definition—Sediment is fragmented material from weathered rocks and organic material that is suspended in, transported by, and eventually deposited by water or air. Therefore, sediment encompasses two overlapping areas of interest: (1) sediment transport and (2) sediment deposition. Sediment transport is a function of stream flows and the rate and size of the sediment supply. Sediment transport usually increases logarithmically with stream flow. Sediment moves either in suspension within the water column as suspended load or by bouncing or rolling along the stream bottom as bedload. Typically, particles >1.0 mm in diameter are transported as bedload, while particles <0.1 mm in diameter are transported as suspended load. Depending on intensity and duration, suspended sediment can be deleterious to fish. Bedload can also be detrimental to fish because it can scour redds, reduce food supply, and alter channel morphology.

During low-flow conditions, fine sediments (sand, silt, and clay) are deposited on the steambed. These fine sediments can fill the interstitial spaces between larger particles and block intergravel water flow that supports benthic macroinvertebrates, small fish, some fish spawning, and egg incubation.

We identified two sediment-related specific indicators: turbidity and depth fines (Table 3). 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. Depth fines refer to the amount of fine sediment (<0.85 mm) within the streambed. Depth fines are usually estimated to a depth of 6-12 inches within spawning gravels.

[8] Currently, the public is reviewing WDOE proposed temperature criteria. Should the WDOE criteria change as a result of public comments, the Action Agencies may adopt revised criteria as PS for water temperatures for salmon and steelhead in tributary habitats.

Response to Activities—Most land-use activities can affect the amount of fine sediment transported to streams. These activities include urban and industrial development, timber harvest, channelization, dams, transportation systems, mining, and agriculture. Activities that remove vegetation along the stream channel have a tendency to increase the recruitment of fine sediments to streams. In forested areas, for example, road building and road maintenance are generally the primary sources of fine sediments. Key factors that affect recruitment of fine sediment to streams from land-use activities include: (1) intensity of the disturbance, (2) the areal extent of disturbance, (3) the proximity of the disturbance to the channel system, and (4) storm events experienced during the periods when the site is most sensitive to erosion (Swanson et al. 1987).

Performance Standards—We identified general PS for both turbidity and depth fines (Table 3). The Action Agencies understand that there are periods when a stream is relatively turbid (e.g., during storms and episodes of snowmelt). These ephemeral high concentrations appear to have little effect on juvenile and adult salmonids (Bjornn and Reiser 1991). Therefore, we selected two PS for turbidity; an acute (short-term) standard and a chronic (long-term) standard.

- The interim PS for turbidity are: (1) an acute (short-term) standard of <70 NTU (nephelometric turbidity unit) and (2) a chronic (long-term) standard of <50 NTU (based on information in Bjornn and Reiser (1991) and MacDonald et al. (1991)).

The acute standard would apply during the period of storms and episodes of snowmelt. The chronic standard would apply at times other than during periods of storms and snowmelt. For streams that naturally exceed these standards (e.g., glacial-fed streams), the PS does not allow turbidity to exceed natural baseline levels[9] at the 95% confidence level.

Fine sediments (<0.85 mm) within spawning gravels can affect the survival of embryos and alevins by reducing the circulation of water through the redd (Bjornn and Reiser 1991). Indeed, there is a direct relationship between the survival of embryos and alevins and the percentage of fines within redds.

- The interim PS for depth fines (<0.85 mm) within spawning gravels is <12% fines.

Both the NMFS (1996) and USFWS (1998) consider spawning gravels with <12% fines as functioning properly.

Contaminants and Nutrients:

[9] Natural baseline conditions can be assessed from historical (pre-disturbance) data. If historical data do not exist, baseline levels can be assessed in similar, mostly undisturbed watersheds.

Definition—Water pollution can be defined as any physical or chemical change in surface water (or groundwater) that harms living organisms or makes water unfit for beneficial uses. Water pollution is generally characterized as originating from either "point" or "nonpoint" sources. Point-source pollution is associated with a particular site on a stream and typically involves a known quantity and type of pollutant that can be controlled at the site. An example is effluent from a factory outlet (end-of-pipe discharge) delivered directly to a stream. In contrast, nonpoint-source pollution typically results from multiple contaminant sources in the vicinity where water quality is impaired. An example of nonpoint-source pollution is the input of petroleum products (e.g., oils and greases) along the course of a stream that parallels a transportation system and flows through an urban area.

We identified five specific indicators associated with contaminants and nutrients: metals and pollutants, pH, dissolved oxygen (DO), nitrogen, and phosphorus (Table 3). Most of these indicators are commonly measured because of their sensitivity to municipal and industrial pollution and their importance in aquatic ecosystems. Pollutants can include a large suite of factors such as pesticides, insecticides, herbicides, fertilizers, detergents and cleaning solvents, bacteria from fecal wastes, road salts and other chemicals from surface treatments, petroleum products, and water-soluble radioactive isotopes. Earlier we discussed thermal pollution and fine sediments. The most dangerous heavy metals include lead, mercury, arsenic, cadmium, tin, chromium, zinc, and copper. It is not necessary to measure all these factors within a given monitoring program. We recommend that only those metals or pollutants that are known to be a concern in a specific project area be monitored.

We include pH and DO because these parameters are often included in water quality monitoring programs (e.g., OPSW 1999). 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 added base or acid, 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 parts per million (ppm) or 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.

We identified 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.

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.

<u>Response to Activities</u>—Land-use activities such as mining, agriculture, urban and industrial development, forestry, dams, and transportation systems can alter the water quality of streams. Mining is generally responsible for elevated levels of heavy metals in streams, although transportation systems also contribute metals to streams. Activities such as urban and industrial development, forestry, and agriculture can increase concentrations of pesticides, herbicides, insecticides, fertilizers, and bacteria in streams. Transportation systems and urban and industrial development can increase inputs of petroleum products, road salts, and other chemicals from surface treatments. Although excessive amounts of metals and pollutants can directly kill biota, at lower concentrations they can affect biota indirectly by reducing production, growth, and fecundity, or by affecting behavior (e.g., avoidance behavior).

Activities such as forestry, acid rain (from burning fossil fuels), urban and industrial development, and mining can alter the pH of aquatic ecosystems. Forest management activities can indirectly affect pH by introducing large amounts of organic debris, which can increase the concentration of organic acids, oxygen demand, and CO_2 inputs. Fertilizers from sewage or industrial discharge, failing septic systems, and agricultural and urban runoff can increase plant growth, which alters pH. Hard-rock mining is the activity most likely to substantially alter the pH of aquatic systems. Water draining from mine tailings, settling ponds, and adits can be highly acidic. A reduction in pH exacerbates the problems associated with heavy metals by increasing their solubility and hence their mobility and rate of biological uptake.

Several activities can either directly or indirectly affect the concentration of DO within water. For example, activities such as agriculture, forestry, urban and industrial development, and transportation systems can increase the input of organic compounds, which reduce DO by increasing BOD. Dams can indirectly reduce DO in reservoirs by increasing water temperatures and trapping organic sediments, and can directly increase DO in the tailrace during periods of spill. Low DO in streams is commonly associated with major point sources such as pulp mills or municipal-waste treatment facilities.

Activities such as urban and industrial development, agriculture, transportation systems (burning of fossil fuels), and forestry can affect nutrient concentrations in streams. Inadequate human waste disposal is the primary factor responsible for increasing concentrations of nitrogen and phosphorus in streams. Inadequate human-waste disposal can result from dispersed recreation, septic tanks, and municipal wastewater treatment plants. Livestock represent another potentially important source of nutrient contamination. Excessive nutrient loading leads to "eutrophication" of streams and lakes. Eutrophication is an excessive growth of aquatic plants, which can affect the structure and function of aquatic ecosystems by altering daily fluctuations in pH and DO.

<u>Performance Standards</u>—The EPA (EPA 1986; EPA 1999) established general national criteria for most contaminants and nutrients. We adopted these criteria as PS for contaminants and nutrients.

For most chemicals (pesticides, insecticides, herbicides, oils and greases, etc.), the EPA has recommended maximum allowable mean concentrations over a 24-hr period. These concentrations vary according to the size of the stream and the designated uses of the water body (EPA 1986; EPA 1999). The maximum allowable mean concentrations are based on a combination of the acute toxicity as defined by the LC-50 and a safety factor. For heavy metals, criteria may vary depending on hardness (ppm of $CaCO_3$), temperature (ºC), and pH. The EPA water quality criteria manual (EPA 1986; EPA 1999) should be consulted for maximum allowable concentrations of chemicals and metals in water bodies.

The EPA has established a pH range of 6.5 to 9.0 as the criteria necessary to protect freshwater aquatic life (EPA 1999). For DO, the 1-day minimum and 7-day mean concentration should be 8.0 and 9.5 mg/L, respectively (EPA 1986). These criteria apply to waters containing salmonid populations and are based on the assumption that intergravel DO is about 3 mg/L less than the DO concentration in surface water. This makes the 1-day minimum and 7-day mean intergravel DO concentration 5.0 and 6.5 mg/L, respectively.

The national drinking water standard for nitrate-nitrogen is 10 mg/L (EPA 1987). A standard for nitrite-nitrogen has not been established because nitrite is such a transient form. However, the EPA (1986) believes that nitrite-nitrogen at or below 0.06 mg/L should protect salmonids. Water bodies with high nitrite concentrations are likely to be highly polluted and not meet existing standards for other indicators such as DO. Criteria concentrations for ammonia vary depending on temperature and pH. The EPA (1986) established criteria for salmonids for the pH range of 6.5-9.0 and temperatures of 0-30ºC. We adopted these criteria as PS for ammonia (see EPA 1986).

The EPA (1986) has made some general recommendations regarding the maximum concentration of phosphorus in streams and lakes. To prevent eutrophication, total phosphates as phosphorus (PO_4^{-3}-P) should not exceed 0.025 mg/L for any lake or reservoir. Where streams enter into reservoirs or lakes, the concentration of total phosphates as phosphorus should not exceed 0.050 mg/L; total phosphorus concentrations should not exceed 0.10 mg/L for streams that do not flow into reservoirs or lakes.

- The PS for contaminants and nutrients follow EPA general criteria. The standards for some areas within the Columbia River basin could deviate from EPA criteria if the states or tribes have justified other more appropriate criteria.

Habitat Access

Artificial Physical Barriers:

Definition—A physical barrier is any physical obstruction that prevents or impedes the upstream and/or downstream passage of fish. Here, we are concerned only with "artificial" or "man-made" physical barriers. In this context, the definition excludes natural barriers such as falls, cascades, log-jams, and beaver dams, and also thermal and chemical barriers. The Water-Quality Pathway deals with the latter two. Flows suitable for passage are considered under the Flow/Hydrology Pathway. Therefore, the focus here is on passage features such as culverts, dams, and fishways.

We identified three specific indicators associated with artificial physical barriers: road crossings (culverts), dams, and fishways (Table 3). Roads and highways are common in the Columbia 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. 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).

Response to Activities—Activities that affect fish passage include transportation systems, urban and industrial development, agriculture, and dams. These activities result in the construction of physical obstructions that can block fish passage. Large hydroelectric facilities are self evident, but other structures like small diversion dams, culverts, and unscreened water-withdrawal systems may impede fish migration during certain time periods. Unscreened diversions are common in the Columbia Basin.

Performance Standards—The presence of artificial (man-made) physical barriers can limit the distribution and abundance of salmonids within a watershed. We selected both a general and numeric PS for man-made physical barriers within watersheds (Table 3).

- The general PS for man-made physical barriers (road crossings, dams, and fishways) is that upstream and downstream fish passage is possible at all flows and no fish end in irrigation systems. This standard applies to any given man-made physical barrier within a watershed.

A numeric PS based on "connectance" can be applied to the watershed. We define connectance as a ratio, which is calculated as the number of stream miles "currently" accessible (i.e., connected) within a watershed, divided by the number of stream miles "possibly" accessible if man-made physical barriers were removed or made passable. Importantly, connectance is a comparison of "current" and "possible" connections under existing conditions. It is not concerned with areas that historically blocked fish passage but presently are passable. It is concerned with man-made structures that under existing conditions block fish passage but can be removed or made passable. Because some man-made structures will not be passable (e.g., Grand Coulee Dam), connectance treats these artificial barriers as permanent barriers. That is, connectance would be calculated for areas downstream from such permanent barriers.

- For a given watershed, the PS for connectance should be "1". Therefore, each of several man-made physical barriers would need to meet the general PS before the numeric standard for the watershed is achieved.

Habitat Quality

Substrate:

Definition—Substrate refers to the bottom material of a water body. Substrate composition serves three important functions: (1) it determines the roughness of the stream channel, (2) provides micro-conditions needed by fish and macroinvertebrates, and (3) provides clues to local and watershed influences on stream habitat quality. Disturbances caused by various land uses can alter surface water runoff and sedimentation rates, and these processes are reflected in the size composition of surface substrate.

We identified two specific indicators of substrate: dominant substrate and embeddedness. 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.

Response to Activities—Most land-use activities can alter the substrate composition of streambeds. The primary land uses include channelization, urban and industrial development, transportation systems, dams, mining, forestry, and agriculture. These activities tend to increase erosion and sediment delivery rates. Most of the material reaching the stream as a result of land-use activities will be fine sediments (sand-sized or smaller). For example, streams in heavily roaded and logged watersheds tend to have higher embeddedness than undisturbed or partially disturbed watersheds. The deposition of fines in streams has a series of adverse effects on aquatic biota.

Performance Standards—Substrate composition and embeddedness can affect salmonid spawning and rearing habitat (Bjornn and Reiser 1991). Spawning areas with high percentages of gravels or small cobbles and low levels of fine sediments tend to correlate with higher survival of embryos and alevins. Areas with high percentages of clean (i.e., lack fine sediments) cobbles or larger substrates provide concealment cover for juvenile salmonids during winter.

- The interim PS for substrate are: (1) gravels or small cobbles make up >50% of the bed materials in spawning areas and (2) embeddedness in spawning and rearing areas is <20% (Table 3).

These standards comport with conditions described as properly functioning by the NMFS (1996) and USFWS (1998).

Large Woody Debris:

Definition—Large woody debris (LWD) consists of large pieces of relatively stable woody material located within the bankfull channel and appearing to influence bankfull flows. LWD is also referred to as large organic debris (LOD) and coarse woody debris (CWD). On the east-side of the Cascade Mountains, LWD is defined as any log with a diameter greater than 30 cm (1 ft) and a length greater than 10.6 m (35 ft).[10] In coastal streams, LWD is any log with a diameter and length greater than 60 cm (2 ft) and 15 m (50 ft), respectively. 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.

[10] This definition is from NMFS (1996) and is used to define properly functioning condition. Other reports offer different dimensions. For example, Armantrout (1998) and BURPTAC (1999) defined LWD as any piece with a diameter >10 cm and a length >1 m. Sedell et al. (1988) and Schuett-Hames et al. (1994) defined LWD as any piece with a diameter >10 cm and a length >2 m, while Overton et al. (1997) define it as any piece with a diameter >10 cm and a length >3 m or two-thirds of the wetted stream width.

Several factors affect the amount of LWD in streams. Stream size is an important factor, with smaller streams usually containing more wood than larger systems. Amount of LWD in streams is positively related to the density of trees in the riparian zone. Streambed characteristics also influence LWD amount, as streams with boulder or bedrock substrates typically contain only about half the LWD compared to streams with finer substrates. Stream size also influences the size of LWD in the channel as well as the amount. Generally, the average size of LWD in a stream channel increases with increasing stream size. Finally, catastrophic events, such as major windstorms, fires, or landslides, can affect the amount and location of LWD in some streams.

We selected number of pieces per stream mile as the one specific indicator of LWD in streams (Table 3).

<u>Response to Activities</u>—Nearly all land-use activities can affect the amount of LWD in streams. Activities such as urban and industrial development, channelization, transportation systems, mining, dams, timber harvest, agriculture, and recreation can directly and indirectly affect recruitment of LWD to streams. Historically, wood was removed from streams to improve navigation, reduce flooding hazards, reduce bank and bed scour, and to provide upstream passage for anadromous fish. Presently, the removal of trees from riparian areas is the primary factor affecting the amount of LWD in streams. A reduction in the amount and mean size of LWD in streams can lead to a decrease in pool frequency and size, and a decrease in the amount of spawning gravels and finer organic matter retained by LWD.

<u>Performance Standards</u>—LWD affects stream systems and their biota in a number of ways. For example, large wood can influence channel morphology by affecting meandering, sediment transport, storage of organic debris (including carcasses), bank stability, channel width, pool formation, and distribution of gravel bars. LWD is perhaps the most important source of habitat and cover for salmonid populations in streams. LWD increases habitat complexity, ensuring that cover and suitable habitat is available over a wide range of flow and climatic conditions. Generally, there appears to be a direct relationship between the amount of LWD and salmonid production; there are no studies that indicate an upper end to this relationship.

- The interim PS for LWD in forested streams is >20 pieces per mile and an adequate source of woody debris available for both short and long-term recruitment (Table 3).[11]

[11] We encourage the reader to review the thesis by Martin Fox (2001). His work looks at the quantities and volumes of instream wood in forested basins within Washington. He established wood targets for forested streams based on channel size, ecoregion, geomorphology, and disturbance regimes. His research can be used to develop more specific targets for LWD. Another useful reference is Appendix A in Bauer and Ralph (1999). Those authors reveal the relationship between basin size and numbers of LWD in the upper Middle Fork Salmon basin.

This standard comports with conditions described as properly functioning by the NMFS (1996) and USFWS (1998).

Pool Habitat:

Definition—Pools are sections of the stream channel that have a concave profile along the longitudinal axis of the stream, or are areas of the stream channel that would contain water even if there were no flow. This means that the maximum depth of pools is deeper than the average thalweg depth, and water velocities at low flows often are lower than the mean velocity. Pools are usually classified by the process that created the pool (e.g., dammed or scour pools). Dammed pools are formed by downstream damming action, while scour pools are formed by erosion action when flowing water impinges against and is diverted by a streambank or channel obstruction.

We identified two specific indicators associated with pool habitat: number of pools per mile and pool quality (Table 3). Pool quality refers to the ability of a pool to support the growth and survival of fish. Pool size (diameter and depth) and the amount and quality of cover determines overall pool quality. Pool cover is any material or condition that conceals or protects fish from predators or competitors and may consist of logs, organic debris, overhanging vegetation, cobble, boulders, undercut banks, or water depth.

Response to Activities—Land-use activities that affect sediment dynamics, flow characteristics, or LWD recruitment can affect the number and quality of pools in a stream. Activities such as urban and industrial development, transportation systems, channelization, dams, mining, agriculture, and timber harvest can reduce the number and quality of pools in a stream. Dammed pools and backwater pools are particularly susceptible to activities that increase the input of fine sediments to stream channels. Activities that reduce the input of LWD also affect dammed pools. Similarly, a change in the size or frequency of peak flows will alter the ability of the stream to transport coarse sediment, and this may alter pool quality.

Performance Standards—Pools are an important morphological feature in streams and an essential type of fish habitat. Pools are typically needed to provide habitat for different species and age classes of fish. In general, dammed or backwater pools provide important winter habitat for salmonids, while scour pools with overhanging banks offer cover and foraging stations for salmonids during summer. In small streams, pools provide the majority of the summer rearing habitat.

- Because the number of pools per stream mile is a function of channel width, we identified interim PS for different channel widths (Table 3). PS range from 184 pools/mile for streams with channel widths of five feet or less to 18 pools/mile for channels wider than 100 ft. These standards comport with conditions described as properly functioning by the NMFS (1996).

In addition to pool frequency, pools should maintain adequate cover for fish. Indeed, there is a close relationship between high quality pools and fish abundance (Platts et al. 1983).

- Because pool quality is a function of both pool size and cover, the interim PS for pool quality are: (1) the maximum pool diameter exceeds the average stream width by 10% or more and residual pool depth[12] is >1 m deep, or (2) the maximum pool diameter exceeds the average stream width by 10% or more, the residual pool depth is between 0.6-1.0 m deep, and the pool has abundant[13] fish cover (Table 3).

Off-Channel Habitat:

Definition—Off-channel habitat consists of side-channels, backwater areas, alcoves or sidepools, 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. Backwater areas are secondary channels in which the inlet becomes blocked but the outlet remains connected to the main channel. Alcoves are deep areas along the shoreline of wide and shallow stream segments. 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 overbank 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.

We identified the presence of off-channel habitat as a specific indicator of habitat quality (Table 3). This standard is specific to channels with gradients <3% (WFPB 1995).

Response to Activities—Activities the alter streamflows and sediment dynamics, or disturb banks and riparian habitat can affect the presence of off-channel habitat. For example, urban and industrial development, mining, dams, and transportation systems can bury off-channel habitat. These land uses and others (e.g., timber harvest, channelization, and agriculture) can also reduce off-channel habitat by altering the connection between the off-channel habitat and the main

---

[12] Residual pool depth is the difference between the maximum pool depth and the pool crest outlet depth. This measure is independent of streamflow at time of measurement and is sensitive to land-management alterations.
[13] Abundant fish cover means that the pool has excellent instream cover and most of the perimeter of the pool has overhead cover (Platts et al. 1983).

channel and by increasing fine sediments that are deposited in low-energy off-channel areas. Activities that remove vegetation from off-channel areas can increase water temperatures and aquatic plant growth, which together can reduce DO concentrations.

Performance Standards—Off-channel habitats provide critical rearing habitat for juvenile salmonids that emerge during spring high-flow conditions. They also provide winter-rearing habitat for salmonids, especially coho salmon.

- The general PS requires the presence of several[14] backwater areas with cover and low-energy off-channel habitats (Table 3). This standard applies only to channels with gradients less than 3% (WFPB 1995).

This standard comports with conditions described as properly functioning by the NMFS (1996) and USFWS (1998).

Channel Condition

Width/Depth Ratio:

Definition—The width/depth ratio is an index of the cross-section shape of a stream channel at bankful level. The ratio is a sensitive 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.

We selected the width/depth ratio as a specific indicator of channel condition. The ratio is expressed as bankful width (geomorphic term) divided by the mean cross-section depth.

Response to Activities—Activities that increase recruitment of sediment to channels or affect streamflows can alter the width/depth ratio. For example, transportation systems, mining, timber harvest, and agriculture can increase the amount of sediment delivered to the stream channel. Usually an increase in coarse sediment will lead to an accumulation of sediment in the deeper portions of the stream channel. If the runoff remains unchanged, an unconfined channel generally responds by increasing its width. Thus, changes in the width/depth ratio can be used as an indicator of change in the relative balance between the sediment load and the sediment transport capacity. Activities that remove riparian vegetation can decrease bank and channel stability and thereby initiate bank

[14] At this time there is no specific standard for the number of different types of off-channel habitats needed to meet properly functioning condition. We suggest that the standard be specific to each channel type and not deviate by more than 10% from the number of off-channel habitats that occurred naturally within the channel (or within a suitable reference channel).

erosion and channel widening, which increases the width/depth ratio. Also, activities that increase the size of peak flows can increase the width/depth ratio.

Performance Standards—An increase in the width/depth ratio can have adverse effects on the biological community of streams. For example, a decrease in depth tends to reduce the number of pools, and this will reduce habitat for fish. An increase in channel width will lead to an increase in net solar radiation and higher summer water temperatures. The combination of shallower pools and increased solar radiation can affect the suitability of stream habitat for salmonids.

- The interim PS is a width/depth ratio of <10 (Table 3).

This standard is consistent with the definition of properly functioning condition (NMFS 1996; USFWS 1998).

Streambank Condition:

Definition—The streambank is defined as the ground that borders a channel above the streambed and below the level of rooted vegetation. It is the portion of the channel cross-section that restricts lateral movement of water during normal streamflow. Streambank condition refers to the stability or alteration of the banks. Under optimum conditions, banks are well vegetated and stable. Well-vegetated banks are usually stable regardless of bank undercutting, which provides excellent cover for fish. They show virtually no evidence of alteration. On the other hand, vertical, eroded, or laid-back banks are unstable and heavily altered. Unstable banks provide little cover for fish.

We selected streambank stability as the one specific indicator of streambank condition (Table 3). 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.

Response to Activities—Certain land uses, such as agriculture (especially livestock grazing), mining, transportation systems, urban and industrial development, and timber harvest can reduce the stability of streambanks. These activities tend to alter or remove streamside vegetation, which can reduce the stability of banks. In contrast, land uses that harden banks with riprap or concrete tend to increase bank stability. However, this form of channelization is unnatural and is often used because other land uses removed bank-stabilizing vegetation. Additionally, some forms of bank-hardening structures reduce cover for fish (e.g., concrete).

Performance Standards—Bank stability is an indicator of channel condition and can directly affect the quality of fish habitat. Unstable banks contribute sediment to the stream channel by slumps and erosion. Because all the material from an eroding streambank is delivered directly into the channel, the adverse effects of bank instability can be much greater than the adverse effects of a comparable area of eroding hill slope. Not only do eroding banks contribute fine sediments to the channel, bank erosion tends to increase stream width and decrease stream depth, which in concert decreases the suitability of habitat for salmonids. Generally, eroding banks provide little cover for fish. In addition, eroding banks support little or no riparian vegetation. Thus, the loss of fish cover along an eroding streambank will be exacerbated by the reduction in riparian cover.

- The interim PS for streambank condition requires that the banks along 80% or more of any stream reach are >90% stable[15] (Table 3).

This standard comports with conditions considered by the USFWS (1998) as functioning adequately.

Flows and Hydrology

Streamflows:

Definition—Streamflow is a measure of the volume of water passing a given point per unit of time. The basic unit of streamflow in the U.S. is the cubic foot per second (cfs). In simple terms, it is a volume of 1 ft^3 of water flowing through a 1-ft^2 plane in one second (the cubic foot refers to volume, and can be any dimensions). The formula for the unit of flow is: $Q = A \times V$, where Q is flow (cfs), A is the area through which the water is flowing (ft^2), and V is the velocity (ft/s). Streamflows plotted over time form a hydrograph. Hydrographs show the pattern of streamflows that occur over a season (seasonal flow patterns) or over a year (annual flow patterns). Because the shape of the hydrograph (i.e., how quickly streamflows rise and fall, timing of peaks, and the magnitude of peak and base flows) can affect stream biota, hydrograph characteristics are useful in the classification of streams for biological purposes (Hawkes 1975).

We identified three specific indicators of streamflows: change in peak flow, change in base flow, and change in timing of flow (Table 3). 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.

[15] The 80% of any reach represents a quantitative factor, while the 90% stability represents a qualitative factor.

Response to Activities—Nearly all land-use activities can affect streamflow patterns within a watershed. Urban and industrial development and transportation systems increase impervious surfaces, which increase surface runoff (increases peak flows) and reduces subsurface lateral flow (reduces base flows). Urban and industrial development can further affect base-flow conditions of effluent (gaining) streams by removing water from the aquifer. Livestock grazing can compact soils and thus reduce infiltration rates and soil moisture storage capacity (increases peak flows and reduces base flows). By removing forest canopy, timber harvest reduces rain and snow interception (increase peak flows) and also reduces rates of evapotranspiration (increases base flows). Dams can affect streamflow patterns in several ways, depending on the purpose of the dam. For example, diversion dams tend to reduce base-flow conditions by diverting water during low-flow periods. Hydroelectric facilities, on the other hand, can reduce peak flows by storing water during spring runoff and increase base flows by releasing water during low-flow conditions.

Performance Standards—Peak flows have important effects on stream channel morphology and streambed particle size. Because higher flows move larger particles, peak flows determine the stable particle size in the streambed. Large stable particles provide important habitat for invertebrates and small fish (especially juvenile salmonids). The size of peak flows also determines the stability of LWD and the rate of bank erosion, which affect the quality of habitat for fish. The majority of sediment transport occurs during peak flows, as sediment transport capacity increases logarithmically with discharge. The relationship between sediment load and transport capacity affects the distribution of habitat types, channel morphology, and streambed particle size.

- The interim PS for peak flow requires that peak flows within a watershed do not exceed natural baseline conditions (see footnote 5) at the 95% level (Table 3).

Base-flow conditions are important primarily for maintaining aquatic habitat and adequate migration corridors. An increase in low flows will increase the wetted perimeter and water depth, and thereby provide more habitat for fish. Increased flows will also reduce the magnitude of any temperature increase because of land-use activities.

- The interim PS for base flow requires that base flows within a watershed do not exceed natural baseline conditions (see footnote 5) at the 95% level (Table 3).

The timing of major flow events (e.g., peak flows and base flows) can affect habitat quality and the abundance, distribution, and behavior of fish in streams. For example, downstream migration of smolts is in part related to increases in streamflows. Higher streamflows usually correspond with greater water velocities, which carry smolts to the estuary.

- The interim PS for timing of streamflows requires that the timing of major flow events (peak and base flows) do not exceed natural baseline conditions (see footnote 5) at the 95% level (Table 3).

Watershed Conditions

Road Density:

Definition—A road is any open way for the passage of vehicles or trains. We identified 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 (miles) within a watershed divided by the area of the watershed (miles^2). The RRI is expressed as the total mileage of roads within riparian areas divided by the total number of stream miles 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.

Response to Activities—The transportation system is the most apparent land-use activity affecting road density. However, other activities associated with roads include urban and industrial development, timber harvest, agriculture and livestock grazing, mining, and recreation (e.g., fishing). All these activities tend to increase the density of roads within a watershed.

Performance Standards—Roads contribute more sediment to streams than any other land-use activity (Meehan 1991). Most problems are with older roads that are located in sensitive terrain and roads that have been essentially abandoned, but are not adequately configured for long-term drainage (Quigley and Arbelbide 1997). Increasing road density tends to correlate with declining stream habitat conditions and aquatic integrity. That is, roads directly affect natural sediment and hydrologic regimes by altering streamflow, sediment loading, sediment transport and deposition, channel morphology, channel stability, substrate composition, stream temperatures, water quality, and riparian conditions within a watershed. Indeed, anadromous salmonids are less likely to use moderate to highly roaded areas for spawning and rearing (Quigley and Arbelbide 1997). Roads also provide access, and the activities that accompany access magnify their negative effects on stream habitat and biota (e.g., poaching and the stocking of non-native fish).

- The interim PS for roads are: (1) road density within a watershed remain less than 2 miles/miles^2 and (2) RRI = 0.00 (Table 3).

Watershed Disturbance:

Definition—Disturbance can be defined as any relatively discrete event in time that is characterized by a frequency, intensity, and severity outside a predictable range, and that disrupts ecosystem, community, or population structure and changes resources or the physical environment (Resh et al. 1988). Here, we are most concerned with disturbances to vegetation within the watershed. These disturbances can affect water yield[16] and sediment recruitment to streams.

We selected "equivalent clearcut area" (ECA) as the single indicator of watershed disturbance. ECA is defined as the area of a watershed that has been disturbed by timber harvest, roads, and fires, with an adjustment factor to account for the hydrologic recovery that results from forest regeneration (USFS 1974; King 1989). The adjustment is based on regeneration (size of trees) and elevation.

Response to Activities—The three primary activities that affect ECA include timber harvest (clearcuts and partial cuts), roads, and fires. These activities reduce the density of mature trees within a watershed and therefore affect streamflows and sediment dynamics in streams. For example, roads intersect surface and subsurface flows, and their ditches capture and concentrate small streams and overland flow. This tends to increase surface delivery of water to the main stream. In addition, roads can convert groundwater seepage into surface flow, significantly increasing the rate of water movement to the main stream and effectively increasing stream drainage density. Timber harvest and fires, on the other hand, remove the forest canopy and change the hydrologic behavior of an area by altering interception, transpiration, and snowmelt processes. These changes usually result in increased peak flows and an increase in annual water yield. As the forest regenerates, the forest canopy develops and reestablishes interception and transpiration processes.

Performance Standards—Trees play an important role in governing runoff rates within a watershed. The crowns of mature trees intercept about 30% of annual precipitation, which is then lost to evaporation. Trees also control snowmelt by casting shade on the snowpack and reducing wind speeds at the snow surface. Trees withdraw water from the soil through transpiration. In riparian areas, trees and shrubs rely on groundwater that supplies streams during low-flow periods. Because tree removal affects streamflows and sediment dynamics, it can also affect the quality of habitat for salmonids.

- The interim PS for watershed disturbance is <15% ECA of the watershed with no disturbance concentrated in unstable areas, refugia, or riparian areas (Table 3).

This standard comports with conditions considered properly functioning by the NMFS (1996) and USFWS (1998).

[16] Water yield is the total outflow from a watershed through surface runoff or subsurface aquifers within a given time period.

Riparian Habitat:

<u>Definition</u>—In the broad sense, riparian habitat includes the ecosystems adjacent to a river or stream. More specifically, riparian habitat refers to banks and floodplains on water bodies where sufficient soil moisture supports the growth of mesic vegetation (i.e., vegetation that requires a moderate amount of moisture). For convenience, we define riparian areas 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. In most ecoregions, one can find a wide variety of vegetation types on the streambanks and floodplains, including coniferous and deciduous trees, grasses, shrubs, forbs, ferns, and mosses.

We identified percent altered vegetation as the one specific indicator of riparian habitat (Table 3). Percent altered refers to the percentage of riparian vegetation along the stream channel that has been removed or altered by disturbance (includes both land-use activities and natural disturbances such as fires, floods, etc.).

<u>Response to Activities</u>—Any land-use activity within the riparian zone can affect the structure and function of riparian vegetation. The high productivity of riparian areas often results in a more intensive exploitation of riparian resources. In many areas the largest trees are in the floodplains and alluvial valleys, and the riparian zones have been more heavily logged because the trees were readily accessible. Grazing pressure is usually higher in riparian areas because of more shade, surface water for drinking, and more succulent vegetation there. Transportation systems (e.g., highways, roads, and railways) are often located within riparian areas. In addition, urban areas are usually located adjacent to streams and rivers. Riparian areas also tend to be the focus of recreational activities such as camping and fishing.

<u>Performance Standards</u>—The type and amount of riparian vegetation is an important controlling factor for stream temperatures and bank erosion, and both temperature and bank erosion can be related directly to the quality of fish habitat. The riparian zone also plays an important role in defining channel morphology and creating fish rearing habitat through the input of LWD. Finally, the riparian zone can control the amount of sediment and nutrients reaching the stream channel from up-slope sources.

- The interim PS for riparian habitat is <20% of the riparian vegetation can be altered (Table 3).

This standard comports with watershed conditions described by the NMFS (1996) and USFWS (1998) as functioning properly.

4.2 Project-Specific Indicators

By incorporating the matrices developed by the NMFS and USFWS, we have effectively reduced the total array of possible variables into a small suite of indicators that should be sensitive to management activities. However, we do not recommend that all these variables be measured for each monitoring project. Rather, we propose that the investigator measure only those indicators that are linked directly to the proposed action. In other words, the most useful indicators are likely to be those that represent the first links of the cause-effect chain. 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 management action and the effect.

Tables 4a and 4b identify indicators likely to be affected by a particular land-use or management activity. Table 4a identifies indicators most likely affected by different land uses, while Table 4b identifies indicators that are directly related (sensitive) to specific activities identified in the NMFS (2000) Biological Opinion. The intent of the tables is to provide an initial screen or filter for selecting the most appropriate monitoring indicators for a particular situation. The "usefulness" of an indicator is based on: (1) sensitivity of an indicator to the specified activity; (2) importance of the indicator to the overall health of the aquatic ecosystem; and (3) the cost of measurement and data analysis, including consideration of the sampling frequency and time needed to detect change. It should be understood that these general guidelines cannot specify which indicators are most appropriate under all conditions. Nevertheless, they provide a qualitative indication as to the monitoring indicators most likely to be useful most of the time.

Table 4a. Ranking of the usefulness of physical/environmental indicators to monitoring effects of different activities in tributary habitat within the Columbia River basin. 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.

Pathway	Specific indicators	Land-use activities							
		Urban/industrial development	**Transportation Systems**	**Channelization**	**Dams**	**Mining**	**Timber harvest**	**Agriculture**	**Grazing**
Water quality	MDMT	1-2	3	3	1-2	1-2	1-2	2	2
	MWMT	1-2	3	3	1-2	1-2	1-2	2	2
	Turbidity	1-2	2	3	1-2	1-2	2	1-2	1-2
	Depth fines	1-2	2	2	1-2	2-3	2	1-2	1-2
	Metals/pollutants	1	2-3	3	3	1-2	3	1-2	3
	pH	1-2	3	3	3	1-2	3	3	3
	DO	1-2	2-3	3	1-2	3	3	2	2
	Nitrogen	1-2	3	3	3	3	3	1-2	2
	Phosphorus	1-2	3	3	3	3	3	1-2	1
Habitat Access	Road crossings	1	1	3	3	2	1-2	3	3
	Diversion dams	3	3	3	1	3	3	3	3
	Fishways	3	3	3	1-2	3	3	3	3
Habitat Quality	Dominant substrate	1-2	2	2	2	2	2	1-2	1-2
	Embeddedness	1-2	2	2	1-2	2	2	1-2	1-2
	LWD	1-2	3	2-3	3	3	1-2	3	3
	Pool frequency	1-2	2	2	1-2	2	2	2	2
	Pool quality	1-2	2	2	1-2	2	2	2	1-2
	Off-channel habitat	1-2	1-2	1-2	2	2	2	2	2
Channel Condition	Width/depth	1-2	2	1	1-2	3	2	1-2	1
	Bank stability	1-2	2	1	2	3	2	1-2	1
Flow/Hydrology	Change in peak Q	1	3	3	1	3	3	3	3
	Change in base Q	1	3	3	1	3	2	2	2
	Change in Q timing	1-2	3	3	1	3	3	3	3
Watershed Condition	Road density	1	1	3	3	2	1-2	3	3
	Riparian-road index	1	1	3	3	2-3	1-2	3	3
	Equivalent clearcut	3	1-2	3	3	3	1	3	3
	Percent veg altered	1	2	3	3	3	1-2	1	1

Table 4b. Ranking of the usefulness of physical/environmental indicators to monitoring effects of different activities identified in the NMFS (2000) Biological Opinion. 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.

Pathway	Specific indicators	Actions identified in NMFS (2000) Biop							
		Irrigation screens	**Blockage removal**	**Sediment reduction**	**Improve water quality**	**Nutrient enrichment**	**Restore instream flows**	**Restore riparian function**	**Restore stream complexity**
Water quality	MDMT	3	2	3	1	2	1-2	1-2	3
	MWMT	3	2	3	1	2	1-2	1-2	3
	Turbidity	3	1-2	1	1	1	1-2	2	3
	Depth fines	3	1-2	1	1-2	2	2	2	2
	Metals/pollutants	3	3	2	1	2	3	3	3
	pH	3	3	3	1	1	3	2-3	3
	DO	3	2	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
	Dams	1-2	1	3	3	3	3	3	3
	Fishways	2-3	1-2	3	3	3	3	3	3
Habitat Quality	Dominant substrate	3	2	1	3	3	1-2	3	1-2
	Embeddedness	3	1-2	1	1-2	3	1-2	2	1-2
	LWD	3	3	3	3	3	3	1	1
	Pool frequency	3	1-2	2	3	3	1-2	1-2	1
	Pool quality	3	1-2	1	2	3	1	1-2	1
	Off-channel habitat	3	2	2	3	3	1	1-2	1
Channel Condition	Width/depth	3	1-2	1-2	3	3	1-2	1-2	1
	Bank stability	3	2	1-2	3	3	2	1	1
Flow/Hydrology	Change in peak Q	3	1-2	3	3	3	1	2	1-2
	Change in base Q	3	1-2	2	3	3	1	2	1-2
	Change in Q timing	3	1-2	3	3	3	1	2	1-2
Watershed Condition	Road density	3	3	1-2	3	3	3	3	2
	Riparian-road index	3	3	1-2	3	3	3	1	2
	Equivalent clearcut	3	3	1-2	3	3	3	2	2
	Percent veg altered	3	3	1-2	3	3	3	1	1-2

To demonstrate the use of Tables 4a and 4b, we offer three different examples of management actions that may be implemented in the Columbia River basin.

Provide Fish Passage at Irrigation Diversions:

Salmonid resources are affected by irrigation throughout the Columbia River basin. Diversions of irrigation water throughout the Columbia, Snake, and Salmon rivers can either block adults or send juveniles into fields to die. Diversions in the upper Salmon River basin remained unscreened until the 1950s. Even today, irrigation water withdrawal continues to reduce accessibility of habitat (NMFS 2000; USFWS 2000).

If one were interested in improving habitat access, possible management actions would be to provide passage at these irrigation diversions and screen intake structures. Because the objectives of the management actions are to ensure passage and prevent entrainment, the physical/environmental indicator most directly related to the action is dam (diversion) passage (Tables 4a and b). This requires the researcher to take physical measurements of the diversion and the stream, and if necessary, run a hydraulic model to assess the barrier status of the diversion. One would take these measurements following the protocols identified in Section 5.0. In this case, it is not necessary to monitor other physical/environmental pathways, such as water quality, habitat quality, channel condition, and watershed condition.[17]

Removal of Metals in Streams:

The Columbia Basin has several abandoned mines, many of which have created pollution problems. In some anadromous streams, acid and metal pollution from abandoned mines has persisted over decades after mining ceased (Spence et al. 1996). In Panther Creek, for example, reductions in salmon and steelhead runs closely correlate with the amount of mining in the drainage. Indeed, continuous leaching of toxic materials from the Blackbird Mine eliminated the large chinook salmon run in Panther Creek. Cleanup efforts have reduced toxic materials in Panther Creek and the stream now supports increasing numbers of salmon and steelhead.

If the objective of the management action is to remove elevated levels of metals from a stream, the physical/environmental indicators most directly related to the action are metals and pH (Table 4a). These water quality indicators would be monitored according to the procedures outlined in Section 5.0. One would not

[17] In some situations, the diversion may reduce streamflows downstream from the diversion to a level that precludes fish passage. In addition, return water from the diversion may elevate temperatures and suspended sediment loads. In this case, one would measure indicators such as baseflow, temperature (MDMT and MWMT), and turbidity in addition to assessing barrier status. In all cases, an understanding of the diversion and its effects on the aquatic habitat is needed to identify which indicators should be monitored.

need to monitor other indicators unless the mining activities also affected other pathways, such as habitat quality (e.g., high levels of fine sediments), channel condition (e.g., bank stability), and/or watershed condition (e.g., altered riparian vegetation). If mining altered other pathways, the investigator would need to monitor the appropriate indicators associated with each pathway. An understanding of the effects of mining in the monitoring area is needed before one can select the appropriate indicators.

Removal of Livestock Grazing:

Extensive deterioration of riverine-riparian habitats in the Columbia River basin began with severe overgrazing in the late nineteenth century (Platts 1990). Approximately 90% of all rangeland in the West is grazed (720 million acres). Livestock graze about 8 million acres of private and state land in Idaho. In Oregon, Washington, and Idaho combined, about 41% of the total land is grazed. Improper grazing damages riverine-riparian habitats, decreasing fish productivity. Over twenty years ago Behnke (1977) noted that the best opportunity to increase fish populations in western North America is to improve riparian habitats adversely modified by livestock grazing. His comments are still true today.

If the objective is to improve habitat conditions in a stream damaged by livestock grazing, removal of livestock from riparian areas is the most logical management action. Because grazing can alter most pathways, one should monitor several indicator variables. Indicators most directly related to the action include depth fines, phosphorus, embeddedness, pool quality, width/depth ratio, bank stability, and percent vegetation altered (Table 4a). Again, an understanding of the effects of grazing on the stream in the monitoring area is needed to select the appropriate suite of indicator variables.

4.3 Classification (Stratification) Variables

Because responses of indicator variables to management actions may vary depending on both large-scale (e.g., ecoregion, geology, province) and smaller-scale (e.g., channel characteristics and riparian vegetation type) characteristics, this plan proposes that investigators measure descriptive landscape and aquatic habitat variables following a hierarchical classification system. 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).

We propose a classification system that incorporates the entire spectrum of processes influencing stream features and recognize 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 5). The State of Montana used a similar system to identify reference areas for assessing mining damages in the Clark Fork River (Hillman et al. 1995). 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. Below we define each classification variable. Section 5.0 identifies recommended methods for measuring each variable.

Table 5. List of classification (stratification) variables and protocols for measuring physical/environmental characteristics of streams and their basins.

Pathway	General characteristics	Specific characteristics	Recommended Protocol
Regional Setting	Ecoregion	Bailey's classification	Bain and Stevenson (1999)
		Omernik's classification	Bain and Stevenson (1999)
	Physiographic Province	Province	Bain and Stevenson (1999)
	Geology	Geologic districts	Overton et al. (1997)
Drainage Basin	Geomorphic Features	Basin area	Bain and Stevenson (1999)
		Basin relief	Bain and Stevenson (1999)
		Drainage density	Bain and Stevenson (1999)
Valley Segment	Valley Characteristics	Valley bottom type	Cupp (1989a,b); Naiman et al. (1992)
		Valley bottom width	Naiman et al. (1992)
		Valley bottom gradient	Naiman et al. (1992)
		Valley containment	Bisson and Montgomery (1996)
Channel Segment	Channel Characteristics	Elevation	Overton et al. (1997)
		Channel type (Rosgen)	Rosgen (1996)
		Bed-form type	Bisson and Montgomery (1996)
		Channel gradient	Overton et al. (1997)
	Riparian Vegetation	Cover group	Overton et al. (1997)
		Community type	Overton et al. (1997)

Regional Setting

At the regional scale, climate and geology are primarily responsible for setting the stage on which factors that operate at more local scales and shorter time frames act to shape channel conditions. Identifying the regional setting makes it easier to group similar habitats and recognize local variability (Whittier et al. 1988; NRC 1992). Therefore, for physical/environmental variables, it is appropriate to select a top tier that is stratified primarily on ecoregion, geology, and physiographic province (Table 5).

Ecoregion:

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, we selected the two most commonly used ecoregion systems, Bailey (1978) and Omernik (1987). Bailey's approach uses macroclimate[18] 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 macrofeatures 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) responded to a need for regionalizing water resource management and 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). We recommend that both systems be used to classify ecoregions in project areas.

[18] Macroclimate is the climate that lies just beyond the local modifying irregularities of landform (e.g., orographic effects) and vegetation, and is interpreted as having an overriding effect on the composition and productivity of ecosystems from region to region (Bailey 1998).

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, 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).

Geologic Districts:

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.

Drainage Basin

Drainage basins are nested within ecoregions and geologic districts. A drainage basin is an area of the landscape occupied by a surface stream or water body together with all the tributary streams, surface, and subsurface water flows. The boundary or "rim" of the drainage basin is the drainage divide, and follows the highest points between two drainage basins. The boundaries of a drainage basin are often used to explain biogeographic distributions of fish. Geomorphic features of the basin are important in predicting flood patterns, estimating sediment yield, and predicting water availability and quality. It is therefore important to understand the geomorphic setting of a stream in its basin.

Geomorphic Features:

We identified three important geomorphic features of drainage basins: basin area, basin relief, and drainage density (Table 5). Basin area (a.k.a. drainage area or catchment area) is the total land area, 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 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. The last geomorphic feature, drainage density, is an index of the length of stream per unit area of basin and is calculated as the drainage area divided by the total stream length. 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).

Valley Segment

The stream valley is nested within the basin and is defined as the elongated and low areas of the landscape. It exists between higher terrain features such as uplands, hills, and mountains where runoff and sediment transport occur through downslope convergence. Within the hierarchy of spatial scales, the valley segment represents the largest physical subdivision that land-use activities can directly alter. As such, it is important to include it as a level in our hierarchical classification. Furthermore, valley segments within a watershed influence the distribution and abundance of aquatic biota by governing the characteristics of water flow and the capacity of streams to store sediment and transform organic material (Hynes 1970).

Valley Features:

We selected four important features of the valley segment: valley bottom type, valley bottom width, valley bottom gradient, and valley confinement (Table 5). 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.[19] Valley bottom width is the ratio of the valley bottom[20] width to active channel width. Valley gradient is the slope or the change in vertical elevation per unit of horizontal valley distance. Valley gradient is typically measured in lengths of about 300 m (1000 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).

Channel Segment

[19] Table 7.3 in Naiman et al. (1992) identifies and describes various valley bottom types.
[20] Valley bottom is defined as the essentially flat area adjacent to the stream channel.

Channel segments are governed by valley and basin characteristics. They are generally classified into various types based on valley form, channel width, average depth and velocity, mean discharge, gradient, channel roughness, sediment load and sizes, channel entrenchment, sinuosity, and other attributes (Rosgen 1996). They form and are maintained by the interaction of streamflow and sediment regimes in a process that yields consistent average channel shape and size. Because land-use activities can alter the interaction of streamflow and sediment regimes, channel morphology may change in response to human disturbance.

Channel Characteristics:

We identified four important characteristics of the channel segment: elevation, channel gradient, channel type, and bed-form type (Table 5). Elevation 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.[21] 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. We recommend that all studies include at least Level I (geomorphic characterization) classification. Depending on the objectives of the monitoring program, additional levels of classification may be necessary.

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. 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.

---

[21] Indices include entrenchment, gradient, width/depth ratio, sinuosity, and dominant channel material.

Riparian Vegetation:

Because riparian vegetation has an important influence on stream morphology and aquatic biota, we identified two characteristics of riparian vegetation: riparian cover group and riparian community type (Table 5). Riparian cover group refers to the dominant vegetative cover type (Overton et al. 1997). The classification consists of two cover groups, wooded and meadow. Wooded riparian areas are characterized by streamside or upslope tree stands that have the potential to supply LWD to the stream channel. Meadow riparian areas are characterized by streamside or floodplain grasses, forbs, or shrubs (including willows) that have little potential to contribute LWD to the stream channel. Riparian community type is a repeated and defined assemblage of riparian plant species. It requires knowledge of plant classification.

5.0 Recommended Methods for Measuring Physical/Environmental Attributes

There are several publications that describe methods for measuring physical/environmental attributes. Not surprisingly, there can be several different methods for measuring the same variable. For example, channel substrate can be described using surface visual analysis, peddle 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. In this document, we identify only one method that should be used to measure indicator and classification variables. The reason for this is to create a standard protocol that will allow comparison of measures across basins and projects. If different methods are used to measure an indicator, then results cannot be compared across studies.[22]

Below we recommend methods for measuring each physical/environmental indicator and classification variable. Our selection of methods relied heavily on the work of Johnson et al. (2001). Johnson et al. (2001) reviewed 112 documents that described 429 sampling protocols for collecting physical and biological data on 48 different attributes. They used selection criteria combined with a scientific peer-review process to recommend a subset of protocols for use across the Pacific Northwest. Their selection criteria consisted of: (1) a review of the protocol elements, (2) the accessibility and practicability to workers with diverse training, (3) applicability across the different environments, (4) listing of tools and implements needed, and (5) type of data generated. The document was published in final form after extensive review by more than 30 peers. Our discussion here is restricted to identifying protocols described in Johnson et al. (2001) for measuring

[22] One of the goals of the Implementation Plan is to develop a regional database. The complexity of that database is greatly simplified if variables are measured and reported using the same methods.

indicator and classification variables, leaving the rest of Johnson's work for the interested reader.

5.1 Indicator Variables

Table 3 identifies the recommended protocols for measuring physical/environmental indicators. We refer the reader to those documents for detailed descriptions of methods and measuring instruments. Here we provide only a general description of recommended methods and instruments.

Water Quality

Water Temperature:

We recommend the use of data loggers for measuring MWMT and MDMT. Both Schuett-Hames et al. (1999a) and Zaroban (2000) describe 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. These manuals provide standard methods for conducting temperature-monitoring studies associated with land-management activities and for characterizing temperature regimes throughout a watershed.

Sediment and Turbidity:

We recommend that investigators measure turbidity with a portable turbidimeter (calibrated on the nephelometric turbidity method) following protocols described in Chapter 11 in OPSW (1999). This guidebook provides a standardized method for measuring turbidity, including criteria for selecting monitoring sites, data quality guidelines, equipment, field measurement procedures, and methods to store and analyze turbidity data.

We recommend that investigators measure depth fines with McNeil core samplers following methods described in Platts et al. (1983) and Schuett-Hames et al. (1999b). We recommend the volumetric method for processing samples sorted via a standard set of sieves (Schuett-Hames et al. 1999b). The volumetric method measures the millimeters of water displaced by particles of different size classes.

Contaminants and Nutrients:

Measurement of metals and pollutants should follow procedures described in APHA, AWWA, and WEF (1999). This document identifies standard methods for the examination of water and wastewater. It describes quality assurance procedures; methods for development and evaluation; collections and preservation of samples; laboratory apparatus, reagents, and techniques; and safety.

For other water quality factors, such as pH, dissolved oxygen, nitrogen, and phosphorus, we recommend that the investigator follow the procedures described in OPSW (1999). The guidebook provides a standardized method for measuring pH (Chapter 8), DO (Winkler Titration Method—Chapter 7), nitrogen and phosphorus (Chapter 10), including criteria for selecting monitoring sites, data quality guidelines, equipment, field measurement procedures, and methods to store and analyze water quality data.

Habitat Access:

Artificial Physical 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. The methods outlined in this manual assess connectivity of fish habitats on a watershed scale. We recommend that investigators follow the methods outlined in these two documents.

Habitat Quality

Substrate:

There are two specific indicators of substrate, dominant substrate and embeddedness. We recommend that investigators use pebble counts to identify dominant substrate. Bevenger and King (1995) describe the pebble count method. We recommend that investigators use a modified version of this method. Rather than pick a substrate particle at toepoint, we recommend the use of a 60 x 60-cm sampling frame (see Bunte and Abt 2001). The sampling frame is intended to reduce operator influence on the selection of particle sizes. In field tests, the sampling frame produced slightly coarser size distributions than the traditional heel-to-toe walk. The sampling frame also produced more similar sampling results between two investigators than heel-to-toe walks.

We recommend the method described in MacDonald et al. (1991) for measuring embeddedness. The method involves the use of a 60-cm diameter hoop as the basic sample unit. The use of hoops rather than individual particles as the basic sampling unit substantially increases the number of particles that must be measured, but reduces the variability among sample units. This makes it easier to detect change and results in an embeddedness value that more closely represents the condition of the stream reach.

Large Woody Debris:

We recommend that the investigator follow the methods described in Overton et al. (1997) and BURPTAC (1999) for estimating the number of pieces of large woody debris in forested streams. These guidelines describe procedures for dealing with single pieces and aggregates.

Pool Habitat:

We recommend that the investigator follow the methods described in Platts et al. (1983) and Overton et al. (1997) for estimating pool frequency. Overton et al. (1997) provide a good description of the various types of pools and how to identify them. Platts et al. (1983) describe methods for estimating pool quality (see their Table 1). We added residual pool depth to the criteria in Platts et al. (1983). Residual pool depth is the difference between the maximum pool depth and the pool crest outlet depth (Overton et al. 1997 describe methods for measuring these two depths). Residual pool depth is independent of streamflow at time of measurement and is sensitive to land-management actions.

Off-Channel Habitat:

WFPB (1995) describes methods for identifying and rating off-channel habitat quality. These habitats provide important winter rearing habitat for fish and apply to channels with gradients less than 3%.

Channel Condition

Width/Depth Ratio:

We recommend that the investigator follow the protocol described in BURPTAC (1999) for estimating width/depth ratios, with one exception. Rather than measure wetted width and wetted depth, we recommend that the investigator measure mean bankful width and mean bankful depth. BURPTAC (1999) describes methods for estimating W/D rations in both single channels and split channels.

Streambank Condition:

We recommend that investigators use methods in Platts et al. (1987) and BURPTAC (1999) for estimating streambank stability. These methods apply to both the left and right banks of the water body.

Flows and Hydrology

Streamflows:

We recommend that investigators use USGS flow data where available to assess changes in peak, base, and timing of flows. For those streams with no USGS stream-gauge data, we recommend methods described in Chapter 14 in Bain and Stevenson (1999) for measuring stream flows. We recommend that velocities be measured with a calibrated water-velocity meter rather than the float method.

Watershed Conditions

Road Density:

WFC (1998) and Reeves et al. (2001) describe methods for calculating road density within watersheds. The index is simply the total length of roads within a watershed divided by the area of the watershed. WFC (1998) describes a method for estimating riparian-road index, which is expressed as the total mileage of roads within riparian areas divided by the total number of stream miles within the watershed. WFC (1998) defines the riparian areas as those areas falling within the federal buffer zones.

Watershed Disturbance:

Equivalent clearcut area is the metric that describes watershed disturbance. This metric is defined as the area of a watershed that has been disturbed by timber harvest, roads, and fires. Methods for calculating ECA are outlined in USFS (1974) and King (1989).

Riparian Habitat:

Percent altered vegetation is the specific indicator of riparian condition. This metric describes the percentage of riparian vegetation along the stream channel that has been removed or altered by disturbance. We recommend that investigators follow the methods described in Platts et al. (1987) for estimating percent altered vegetation.

5.2 Classification (Stratification) Variables

Table 5 identifies the recommended protocols for measuring classification variables. We refer the reader to those documents for detailed descriptions of methods and measuring instruments. Here we provide only a general description of recommended methods and instruments.

Regional Setting

Ecoregions:

We recommend that investigators use the methods described in Chapter 3 in Bain and Stevenson (1999) for describing ecoregions. Until we better understand the relationships between fish abundance and distribution and the two classes of ecoregions, we recommend that investigators use both classifications (Bailey's and Omernik's). Published maps of ecoregions are available to assist with classification work.

Physiographic Province:

Delineation of physiographic provinces is based on topography and, to a lesser extent, climate, which governs the processes that shape the landscape. Chapter 3 in Bain and Stevenson (1999) outline methods for describing physiographic provinces. Physiographic maps are available to aid classification work.

Geology:

Geologic districts are areas of similar rock types or parent materials. Overton et al. (1997) outline methods for identifying gross geologies. Published geology maps aid in the classification of rock types.

Drainage Basin

Geomorphic Features:

Geomorphic features include the characterization of basin area, basin relief, and drainage density. Chapter 4 in Bain and Stevenson (1999) outlines standard methods for estimating these parameters. We recommend that investigators use USGS topographic maps (1:24,000 scale) and GIS to estimate these parameters.

Valley Segment

Valley Characteristics:

Important valley characteristics include valley bottom type, valley bottom width, valley bottom gradient, and valley confinement. We recommend that investigators follow the methods of Cupp (1989a,b) and Naiman et al. (1992) to describe valley bottom types. Naiman et al. (1992) describes 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.

Channel Segment

Channel Characteristics:

Important characteristics of channel segments include elevation, channel gradient, channel type, and bed-form type. Overton et al. (1997) describe methods for measuring elevation and channel gradient. Bisson and Montgomery (1996) describe in detail the method for identifying channel bed-form types. We recommend that description of channel type follow the techniques of Rosgen

(1996). We recommend that all studies 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.

Riparian Vegetation:

Riparian cover group and community classification are the two descriptors of riparian vegetation. We recommend that investigators used the methods described in Overton et al. (1997) to assess cover group and riparian community classification.

6.0 Monitoring Designs

In this section we tackle the difficult task of describing valid designs for monitoring management actions or RPAs in tributary habitats. It is not our intent to outline possible designs for all types of management actions. Rather, this section briefly describes the minimal requirements of valid monitoring designs. For a more detailed discussion of experimental designs, we refer the reader to Box et al. (1978), Green (1979), Keppel (1982), Mead (1988), Hairston (1989), or Manly (1992, 2001). Any one of these sources will provide the reader with a more detailed discussion of sampling designs than what we can offer in this section.

Basic effectiveness research requires the use of valid experimental protocols. As we indicated earlier, the intent of effectiveness research is to demonstrate causal relationships. That is, we want to know if the habitat action resulted in a desired change in the physical/environmental and biological indicators. To establish cause-effect relationships, the investigator must follow a few simple steps. Initially, the investigator must define the problem (which can be stated as a question), select a management action to fix the problem, and state objectives and testable hypotheses. Implementation of a satisfactory management action requires the investigator to have knowledge of the conditions existing before the actions are implemented. A common failure in developing a valid effectiveness monitoring study is an inadequate understanding of current conditions, potential conditions, and limiting factors. With an understanding of initial conditions, the investigator can select appropriate management actions. This, in turn, leads to identification of study objectives and testable hypotheses. A hypothesis should, at the very least, contain one or more overt or implied predictions.[23]

Following the initial steps, the investigator then identifies dependent (indicator) and independent variables, selects a sample, selects reliable and valid measuring instruments, and selects a valid statistical design.[24] These steps logically follow the objectives and

[23] An investigator would not implement a management action if there were no expectation of some result, however nebulous, and that expectation constitutes an implied prediction.
[24] Not all management actions or RPAs will require a rigorous "statistical" design. For example, identifying and fixing migration barriers does not require a statistical design. However, the majority of actions will require some elements of statistical design to monitor the effectiveness of the action.

hypotheses of the study. Indeed, the objectives and hypotheses dictate the selection of variables, samples, and the statistical design. Earlier we identified and described variables (indicators) and reliable measuring instruments for monitoring physical/environmental conditions in tributary habitats (see Sections 4 and 5). The remainder of this report focuses on sampling methods and statistical designs.

We divided this section into two parts. The first part deals with sampling and provides definitions, methods of selecting a sample, and a general overview of sample size considerations and measurement error and bias. This part leans heavily upon the work of Scheaffer et al. (1990) to which the reader is referred for a much fuller treatment.[25] The second part, statistical design, describes briefly different types of research designs that can be used to measure physical/environmental conditions in tributary habitats. Here the discussion draws on statistical theory, which is an inherent component of nearly all effectiveness research programs. Although many fear statistics, its importance is obvious because much of what is learned about the stream environment is based on numerical data. Indeed, statistics provide the scientific basis and procedures for studying numerical data and making inferences about a population. Our approach is non-mathematical, which means that those looking for an in-depth discussion of statistics will not find it here.

6.1 Sampling

Definitions

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*.[26] All the possible sampling units available within the area (population) constitute the *sampling frame*.[27] 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. 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,

[25] For a more advanced treatment of sampling see Cochran (1977) or Thompson (1992).
[26] 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.
[27] 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.

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., 100-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.

Methods of Selecting a Sample

Selection of a sample is a crucial step in monitoring physical/environmental conditions in streams. The "goodness" of the sample determines the generalizability of the results. Because monitoring studies usually require a large amount of time and money, non-representative results are wasteful. Therefore, it is important to select a method that increases the degree to which the selected sample represents the population. We describe the five most commonly used methods for monitoring physical/environmental conditions: random sampling, stratified sampling, systematic sampling, cluster sampling, and multi-stage sampling. The reader should consult Scheaffer et al. (1990) for information on how to calculate means, variances, and standard errors for each sampling method.

> **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.[28] 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 the each choice is equally likely.

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

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.[29] 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 select 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 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

[29] 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.

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 districts (primary sampling units) and then divide each geological district 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 districts), 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.

Choosing Sample Sizes

We now address the question, "to have a high probability of detecting a management (treatment) effect at least as large as the 'biologically' significant one, what sample size should the investigator use?" This is one of the most important questions of an effectiveness research study. If the sample is too small, the results of the study may not be generalizable 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 we consider these, it is appropriate to discuss the factors that one needs to consider when selecting a total sample size.

The total sample size for survey or descriptive studies[30] depends upon the population size (total number of units in the sampling frame), population 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.[31] 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. 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), and Methodologist's Toolchest, which is produced by Idea Works (1997).[32]

Effectiveness research 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 research (i.e., statistical significance indicates that the management action did what it was suppose to do).[33] Therefore, when selecting a total sample size, 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.[34] What follows is a brief description of each of these factors. First, however, we describe briefly the errors of inference.

> **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

---

[30] Descriptive or survey studies (a.k.a. mensurative studies) describe population characteristics based on a sample of the population (see Section 6.2). These studies describe current or existing conditions.

[31] 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.

[32] The use of trade or firm names in this paper is for reader information only and does not imply endorsement by the Action Agencies of any product or service.

[33] As we 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.

[34] 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, we include them in our discussion on variability and in Section 6.2.

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 quickly 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 ($1-\beta$) 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.

<u>Significance Level</u>—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.

<u>Effect size</u>—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.[35] For example, a practical significant effect size of interest might be

[35] 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

the difference between the maximum acceptable percentage of fine sediments in spawning gravels (Interim PS is <12% fines; see Table 3) 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 general, sample size is inversely related to effect size. That is, 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 we 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

---

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.

variability using methods described in Scheaffer et al. (1990).[36] In general, the greater the variability the larger the sample size needed to detect a significant difference.

<u>Statistical Power</u>—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. Sample size is directly proportional to statistical power. 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 ($\beta = 0.20$). This comports with Peterman (1990) and Green (1994), who suggest that fisheries researchers should like 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 research studies. Because of the time and cost of sampling 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 to reduce sample size (Lipsey 1990). In addition, it is unwise to reduce statistical power ($1-\beta$), 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

[36] 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, we do 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.

the perceived consequences of Type I and Type II errors. We recommend that the investigator 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 the PS (or 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,[37] stratification, or covariate analysis). Earlier we identified sensitive indicator variables (Section 4.0) and reliable methods for measuring those variables (Section 5.0). Later we discuss ways to reduce sampling error and bias, and describe various techniques of experimental design.

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 and Methodologist's Toolchest. We suggest that the investigator use the method that meets their particular needs.

Measurement Error and Bias

Measurements and estimates are never perfect. Indeed, most physical/environmental 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. In this section we identify and describe the various types of errors. We also describe ways to minimize these errors.

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

[37] Although unreplicated random block designs are useful methods of reducing variability, we do not recommend them for monitoring physical/environmental 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).

result that displays no systematic *bias*[38] 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.[39]

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, we divided systematic errors into two general classes: those that occur because of inadequate procedures and those that occur during data processing. We consider each of these 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 physical/environmental 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

---

[38] *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.

[39] 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.

imperceptibly as equipment ages or observers change their perspectives (especially true of "visual" measurements). Failure to calibrate equipment introduces error, as does demanding more 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. We offer the following guidelines for achieving this goal.

(1) Physical/environmental measures based on counts (e.g., LWD frequency)

- 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) Physical/environmental measures based on visual estimates (e.g., 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) Physical/environmental measures based on instruments (e.g., 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 physical/environmental indicators

- Use modern GPS technology 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.

(5) Handling of physical/environmental 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 we leave 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). 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. 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.[40] 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.

6.2 Statistical Designs

In this document we define "statistical design" as the logical structure of effectiveness research studies. It does not necessarily mean that all effectiveness research studies require rigorous statistical analysis. Rather, it implies that all effectiveness research studies, regardless of the objectives, must be designed with a logical structure that reduces the likelihood that rival hypotheses are correct.[41] Our purpose in this section is four-fold. First, we will describe the various classes of designs that can be used to monitor physical/environmental variables. Second, we identify the minimum requirements of effectiveness research designs. Third, we identify several designs that can be used to monitor physical/environmental conditions. Finally, we introduce the concept of bioequivalence. This section is not exhaustive, but it should provide the reader with the minimum information necessary to design a valid effectiveness research

[40] Another reason that outliers should be treated carefully is because they can invalidate standard statistical inference procedures. Outliers tend to affect assumptions of variability and normality.
[41] 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.

study. The following discussions draw heavily on the work of Hairston (1989), Hicks et al. (1999), Krebs (1999), and Manly (1992, 2001).

Classification of Monitoring Designs

One can classify monitoring designs based on the "validity" of the monitoring studies. 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 research 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, the investigator should select a design strong in both internal and external validity. This is not always possible, however, and the investigator should seek to have internal validity when possible. Without internal validity the data are difficult to interpret because of the confounding effects of uncontrolled variables.

Different monitoring designs have different capabilities for guarding against sources of invalidity. For our purposes here, we identify two different study paradigms: (1) *mensurative* (a.k.a. descriptive or observational studies, unplanned experiments) and (2) *manipulative* (a.k.a. causal-comparative/experimental studies, true, planned experiments) studies (Table 6). In the first case, data are collected to test hypotheses or answer questions concerning the current status of the population or setting (e.g., percent native riparian vegetation) with a certain precision and accuracy in summary statistics. Often the investigator collects data by observing a process that may not be well understood. For example, monitoring the quality of pools in a stream reach describes the current condition and trend of pools in the reach, but does not identify factors that affect the quality of pools in the reach. That is, mensurative studies cannot tell us with a high degree of certainty why pool quality in a reach increases or decreases. Put another way, conclusions about causation from mensurative studies are not necessarily wrong. The problem is that there is little assurance that they are right. Because these studies usually cannot infer causal relationships, they have low internal validity.[42]

Manipulative ("kick-it-and-see") studies, on the other hand, attempt to establish cause-effect relationships. In manipulative studies, the alleged "cause," the management activity or characteristic believed to make a difference, is referred to as a treatment; the more general term for "cause" is independent variable. The difference, or "effect," which is determined to occur or not occur is referred to as the indicator (dependent) variable.

[42] Correlational studies are a special type of mensurative research that attempt to assess whether a relationship exists between two or more variables. Correlational studies are descriptive in that they cannot conclude that one variable is the cause of another; there may be a third factor that "causes" both of the related variables. Correlational studies do not establish cause-effect relationships (for an alternate view see Shipley 2000), although they may indicate fruitful avenues of inquiry for testing of cause-and-effect.

Thus, a study that investigates a cause-effect relationship investigates the effect of an independent variable on an indicator variable. In a manipulative study the investigator manipulates at least one independent variable and observes or measures the effect on one or more indicator variables. Manipulative studies require both treatment and control (reference) groups. Because of the direct manipulation of independent variables, these studies are the only ones than can truly establish cause-effect relationships. Unlike mensurative studies, manipulative studies have high internal validity.

The two study paradigms (mensurative and manipulative) provide managers with very different types of information. Because no treatment or management activity is implemented in mensurative studies, these studies are not recommended for effectiveness research. Mensurative studies are more appropriate for status and trend monitoring. Therefore, we will not consider these in any detail in this document. Manipulative studies, on the other hand, involve assigning some treatment or management activity to an experimental unit. These studies are most appropriate for effectiveness research. The remainder of this section will focus on manipulative studies for effectiveness research.

Table 6. General comparison of characteristics, types of designs, and examples of mensurative and manipulative studies (modified from Hicks et al. 1999).

	Mensurative studies	**Manipulative studies**
Descriptions	Observational studies Descriptive studies Unplanned studies	True experiments Causal-comparative studies Planned studies
Characteristics	Finite sampling No randomization Inferences limited to study area	Random assignment of treatments Inferences to study protocol Controls
Types of designs[1]	BACI in impact assessment Before-after accident ANOVA designs without randomization Retrospective studies	BACI with random assignment Paired design with random assignment ANOVA designs with randomization
Examples	Effects of legacy timber harvest on stream habitat Effects of grazing on stream habitat Effects of historical mining on stream habitat Measure of stream temperature upstream and downstream from disturbed area	Study of controlled and randomized grazing Study of controlled and randomized riparian-timber harvest Study of controlled and randomized treatments of LWD in streams

[1] In many cases the same statistical analysis can be used with either mensurative or manipulative studies. However, the validity of any inferences will depend on the type of study.

Minimum Requirements of Effectiveness Designs

Although it might seem superfluous to describe the requirements of effectiveness research designs, they have been violated enough that all should be warned of errors. In general, the more complex the study, the more complex the requirements, but the minimum requirements include *randomization*, *replication*, and *controls*. We assume that the investigator has a good understanding of existing conditions in the study area.

> **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.[43] 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 6.1 on selection of sample size). 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.
>
> 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

[43] 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.

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.[44] 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 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 a mountain stream. 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.

[44] 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.

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 we 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 (McNeal 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 effectiveness 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, we recommend that investigators consult with a professional statistician during the development of effectiveness research studies.

Controls—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.[45] 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

[45] Lee (1993, pg 205) offers a quote 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.)"

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.

In this section we pointed out that the minimum requirements of effectiveness research include randomization, replication, and controls. We have also noted that in some instances effectiveness research 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. We recommend that the reader see Cook and Campbell (1979) for a detailed discussion of quasi-experimental studies.

Types of 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 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. In this section we describe valid designs that could be used to monitor the effectiveness of management actions in tributary habitats. This discussion is not exhaustive, nor is it complete and should not be considered the final authority on designs for effectiveness research. Our intent is merely to introduce the reader to a few simple designs.

Recall from our earlier discussion that a study is valid if observed results are directly related to the manipulation of the independent variable (internal validity) and they can be generalized to situations outside the study area (external validity). If an invalid study design is used, one cannot determine whether the lack of response (positive or negative) in indicator variables resulted from adding an unnecessary treatment (e.g., new management action), or because the research design was unable to identify a true treatment effect. For example, stream-habitat condition may respond favorably or unfavorably to some treatments, but because habitat indicators were not sampled consistently in treatment and control sites for a sufficient period of time (under-replicated and hence low statistical power), the analysis may not identify significant treatment effects.

Effectiveness research studies should therefore be designed to reduce sources of external and internal invalidity. *Internal validity* refers to the condition that observed differences on the indicator variable are a direct result of manipulation of the independent variable, not some other nuisance variable. In other words, the outcome of the study is the result of the management action, not something else. If someone can offer an alternative explanation (i.e., a more reasonable hypothesis supported by the data) for the observed results, the study was not internally valid.

External validity concerns the extent to which the results of the study can be generalized, or applied, to environments outside the study area. In other words, the results of the study can be expected in other settings, at other times, as long as the conditions are similar to those of the study area. With some thought, one can see that it becomes difficult to design a study with both high internal and external validity.[46] Because the intent of effectiveness research is to demonstrate a treatment effect, the study should err on the side of internal validity. Below we identify some 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 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, then 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.

[46] Studies with high internal validity tend to have low external validity. In the same way, studies with high external validity tend to have lower internal validity.

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, proper use of randomization, replication, and controls is all that is needed to reduce the threats to internal and external validity. The remainder of this section will describe two general types of designs that can be used to monitor the effectiveness of management actions in tributary habitats.

BACI Designs:

One experimental design that considers changes in indicator variables within treatment and control sites both before and after treatment is the BACI or Before-After-Control-Impact design (Stewart-Oaten et al. 1986, 1992; Smith et al. 1993). This type of design is also known as CTP or Control-Treatment Paired design (Skalski and Robson 1992), or Comparative Interrupted Time Series design (Manly 1992). Although names differ, these designs are essentially the same. That is, they require data collected simultaneously at both treatment and control 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 site.[47] The pretreatment sampling serves to evaluate success of the pairings and establishes the 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

[47] The use of several test and control 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 sites. These “Beyond BACI” designs and their analyses are described in more detail in Underwood (1996).

likely to occur if similar climatic and environmental conditions equally influence sampling units. Precision of the design can be improved further if treatment and control stream reaches are paired according to a hierarchical classification approach (see Section 5.2). Thus, indicator variables in stream reaches with similar climate, geology, and geomorphology (geomorphic guilds) should track each other more closely than those in reaches with only similar climates.

It is important that control and treatment 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. 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 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). 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 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.

ANOVA Designs:

Although BACI designs are powerful tools for assessing cause-effect relationships, there are other effectiveness research designs that can be used to assess treatment effects. For example, factorial experiments, split-plot experiments, Latin squares, nested, and repeated-measures designs with randomization (Keppel 1982; Mead 1988; Manly 1992) are robust ANOVA-based methods for assessing cause-effect relationships. Some RPAs may include the need to test a variety of risky or innovative management actions. The case of the riparian forest is an example where several indicators can be used to prescribe management actions. Management actions in riparian areas could be tested using a variety of direct indicators (e.g., percent altered riparian vegetation) and several indirect indicators (e.g., LWD or bank stability) (see Tables 4a and b). In any case, ANOVA is a valid approach for testing the effectiveness of management actions.

Usually, during field studies the investigator will not be able to experimentally control all possible variables. However, the investigator may be able to measure extraneous variables[48] during the study and then use analysis of covariance to remove their influence on the indicator variable. Analysis of covariance is an ANOVA-based design that allows the investigator to compare group means on an indicator variable, after these group means have been adjusted for differences between the groups on some extraneous (covariate) variable. Because analysis of covariance is a combination of regression analysis with an ANOVA, the indicator variable must be related to the covariates. Analysis of covariance is used to: (1) increase precision (power) in an experiment, (2) control for extraneous variables, and (3) compare regressions within several groups (Dowdy and Wearden 1983). For example, suppose an investigator is interested in comparing the effectiveness of different riparian management actions on the number of pieces of LWD in stream channels. Suppose also that the number of pieces of LWD in a channel is related to the amount of LWD stored in the channel at the beginning of the study. That is, the more LWD in the stream channel, the more likely recruited LWD will be retained. In this example, the indicator variable is number of pieces of LWD at the *end* of the study and the covariate is the number of pieces at the *beginning* of the study. Analysis of covariance would adjust the means of the indicator (number of pieces of LWD) on the basis of the covariate means, and then compare these adjusted means to assess differences among treatment groups. Thus, the researcher is able to "statistically" control any initial differences that may be present and that may confound differences among treatment groups.

In the above example, the investigator used the same instrument to measure the covariate and the indicator variable. That is, both the covariate and indicator variable were the same variable (number of pieces of LWD). In contrast, it is possible for the covariate and the indicator to be different variables. Here, the covariate and indicator variables are obtained by different measurements. For example, an investigator may be interested in assessing the effects of different riparian management actions on stream temperatures (indicator variable). However, extraneous variables such as elevation, stream size, aspect, and azimuth affect stream temperatures independent of the treatment. If the investigator is unable to control for these extraneous factors in the study design (e.g., through stratification), the investigator may be able to remove their effects statistically. That is, if these variables are related to temperature but not to each other, analysis of

[48] Extraneous variables are also known as nuisance variables, rival variables, or covariates. The effects of extraneous variables can confound effectiveness monitoring studies.

covariance may be used to adjust mean temperature scores. In this example, the indicator (stream temperature) and covariates (elevation, stream size, aspect, and azimuth) are not the same variables. Although this is a simple example of the use of analysis of covariance, covariance can also be used in factorial analysis, repeated measures analysis, and even multivariate analysis. We caution that analysis of covariance is no substitute for good experimental design.

As with BACI designs, randomization should be used whenever possible. ANOVA designs can be analyzed properly when treatments are not randomly assigned to units; however, valid subsampling must occur within the units. Valid statistical inferences can be drawn to the sites or units, but the authority of a randomized design is not there to "prove" cause-effect relationships.

The Concept of Bioequivalence

Most monitoring studies in the Columbia Basin will involve evaluation of the effectiveness of management actions on disturbed lands. The overall goal will be to restore[49] a site until the distribution or level of indicator variables measured on the site is "equal" to the distribution or level of the indicators on an undisturbed control site, or until certain PS are achieved. For example, one could monitor the effectiveness of reclamation of a stream that has been damaged from surface mining. The approach is to measure appropriate indicator variables (from Table 4a and b) in the "treatment" area (randomly selected sites in the mine-damaged area) and control area (randomly selected sites within undisturbed reference streams that are similar to the damaged stream but lack mining damages). One could use a BACI or ANOVA design to monitor the effectiveness of the restoration activities. In this example, the investigator would select subsamples (not true statistical replicates) and sample sizes using methods described in Section 6.1, measure indicators with standard methods described in Section 5.1, and compare treatment and control sites using statistical methods.[50]

Classical null hypothesis tests[51] may not be appropriate in situations such as deciding whether a damaged site has been restored. Rather, the hypothesis that is tested should be that the site is still damaged (i.e., guilty until proven innocent). For example, the classical null hypothesis could be

[49] We define "restoration" as a process that involves management decisions and manipulation to enhance the rate of recovery (after Davis et al. 1984). We believe the goal of restoration should be to reestablish an ecosystem's ability to maintain its function and organization without continued human intervention. It does not mandate returning to some arbitrary prior state. Indeed, restoration to a previous condition often is impossible or even undesirable ecologically.

[50] The reader should recognize that this study is an example of a quasi-experiment because selection of sites for "treatment" and "control" is not by a random process (i.e., restoration actions or treatments are not randomly assigned to sites). Therefore, statistical inferences are limited to the specific sites under study. Subjective inferences can be extended beyond the specific area if enough replications of the study produce similar effects. However, statistical inferences beyond the study sites are not possible.

[51] The classical null hypothesis is simply a statement of "no difference."

H_0: The mean for an indicator on the disturbed (treated) site and the mean for the indicator on the control site are equal

and the alternative

H_1: The mean for an indicator on the disturbed (treated) site and the mean for the indicator on the control site are different.

One would assume that the site had been restored if the null hypothesis is not rejected. However, failing to reject the null hypothesis is not considered as scientific proof that the null hypothesis is true. A difference may exist, but high variation due to inadequate replication or imprecise measurements may yield data for which the null hypothesis is not rejected. Considering statistical power provides one way to design the experiment to be precise enough to detect effects of ecological significance, but it is an indirect way to reach the conclusion that a treatment has no effect. On the other hand, an experiment may be too precise. A large experiment could reject the null hypothesis even if the effect of the treatment is extremely small. In this case the results are statistically significant, but not ecologically significant.

A better approach would be to test whether treated and control sites do *not* differ by an important amount, or if PS have been achieved. This approach is based on the concept of "bioequivalence" (McDonald and Erickson 1994; Underwood 1998; Manly 2001). That is, a damaged site might be considered bioequivalent to a control site if the mean of the indicator on the damaged site is "equivalent" to the mean of the indicator on the control site. One must of course specify the limits within which the treated sites will be considered equivalent to the control sites, because the difference between sites will never be exactly zero. In the example of our mine-damaged area, one could specify equivalence as the mean of the indicator variable on the treated site falling within, say, 80% of the mean of the indicator on the control site. Here, bioequivalence can be examined by testing the null hypothesis

H_0: The mean of an indicator on treated sites is at least 80% of the mean of the indicator at control sites ($\mu_t \leq 0.8\mu_c$).

against the alternative hypothesis

H_1: The mean of the indicator at treated sites is larger than 80% of the mean of the indicator on control sites ($\mu_t > 0.8\mu_c$).

Note that now the null hypothesis is that the sites are not equivalent. The data have to provide evidence that this is not true before the investigator can declare the sites equivalent. That is, the area is assumed damaged unless the data suggest otherwise.

Tests for bioequivalence appear to be useful in monitoring the effectiveness of management actions in tributary habitats and in adaptive management. One practical problem, however, is that the investigator must define bioequivalence before conducting the study, and all parties should agree on the definition. This could get quite involved because a different definition for bioequivalence may attend each indicator variable. We hope this will not preclude the use of tests for bioequivalence.

6.3 Synopsis

Section 6 discusses nothing new. All of it has been discussed in much great detail elsewhere. What we have tried to do is summarize the important ingredients of valid effectiveness research. We outlined methods for selecting unbiased samples, discussed the importance of choosing appropriate sample sizes, and identified errors and biases that can sneak into monitoring studies. We also identified the minimum requirements of valid study designs, identified common threats to validity, and described two general types of monitoring designs that can be used to assess the effectiveness of management actions in tributary streams. We believe that all investigators should have a general understanding of these principles.

There is no substitute for a sound sampling program. Indeed, the success of effectiveness research depends on a sound sampling program. Control of the sampling program must be maintained throughout the period of study. Changes in the sampling program should be avoided (e.g., changing sampling techniques, adding or dropping sampling sites, or changing sampling periods).[52] In addition, it is important not to overemphasize statistical significance (as opposed to practical significance), because statistical significance is tied to sample size, α-level, magnitude of the treatment, magnitude of temporal variance, and temporal covariance between control and treatment sites through time. In fact, the approach that we advocate requires the investigator to define practical significance before conducting the study. This avoids a statistically significant result that is not necessarily significant practically, or finding a significant result merely by collecting more samples.

In most cases we believe that conclusions from effectiveness research should be "design-based" rather than "model-based." The former uses randomization when collecting data, while the latter uses the randomness inherent in the assumed model. The advantage of the design-based approach, which we introduced in Section 6.1,[53] is that valid inferences are possible and justified by the design of the study and the way one collects data. The investigator can defend conclusions provided that there is agreement about which indicators should be measured, the procedures used to do the measuring, and the design protocol. It may be better, therefore, to use point estimates and confidence intervals rather than strict tests of significance. Understanding discussions about significance may be easier if results are presented in terms of confidence intervals. Nevertheless, results of valid monitoring studies should be used to evaluate management actions or RPAs.

---

[52] There are times when events beyond the investigator's control reduce the validity of the study. Dyer (2002) refers to these events as "Demonic Intrusion" or DI-. DI- includes any seemingly malicious, but apparently random, interference of an unanticipated and unpredictable nature that negatively affects the operation and completion of the study and the proper evaluation of data. It is usually dismissed as "just bad luck," which also implies paranormal causation. The investigator has little control over DI-, but does have control over purposeful changes in the sampling program.

[53] All the sampling methods that we covered in Section 6.1 are design-based, because this is how the classical theory for sampling finite populations developed.

7.0 Applications

The preceding sections serve notice that considerable care must be put into the appropriate methods and logic structure of monitoring the effectiveness of management actions or RPAs in tributary habitats. It is our intent in this section to distill the information presented in this document into a concise outline that the investigator can follow to develop a statistically valid effectiveness research plan. For convenience, we offer this summary as a checklist of elements that force the investigator to consider all aspects of valid effectiveness research. Although these elements or statements are generic, the investigator must address each statement in order to demonstrate complete understanding of the research problem and effectiveness monitoring.

A. Problem Statement and Overarching Issues:

- Describe the physical/environmental problem that needs to be improved or corrected.
- Describe the current environmental conditions at the project site.
- Describe the factors that contribute to current conditions (e.g., roads in the riparian zone increase stream siltation).
- Identify the management action(s) (treatments) that will improve existing conditions.
- Describe the goal or purpose of the management action(s).
- Identify the hypotheses to be tested.
- Identify the independent variables in the study.

B. Statistical Design:

- Identify and describe the statistical design to be used (e.g., BACI).
- Describe how treatments (management actions) and controls will be assigned to sampling units (random assignment).
- Demonstrate whether the study includes "true" replicates or subsamples.
- Describe how temporal and spatial controls will be used and how many of each type will be sampled.
- Describe the independence of treatment and control sites (are control sites completely unaffected by management actions).
- Explain how variables will be co-varied in the experiment.
- List and describe the potential threats to internal and external validity and how these threats will be addressed.
- If this is a pilot test, explain why it is needed.
- Describe the descriptive and inferential statistics that will be used to analyze the data and how precision of statistical estimates will be calculated.

C. Sampling Design:

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

D. Measurements:

- Identify the indicator (dependent) variables that will be measured.
- Describe the methods and instruments that will be used to measure the indicators.
- Describe the precision of the measuring instrument.
- Describe the possible effects of the measuring instrument on the sampling unit (e.g., core sampling for sediment may affect local sediment conditions). If the instrument affects the sampling unit, describe how the study will deal with this.
- Describe the steps that will be taken to minimize systematic errors.
- Describe a QA/QC plan, if any.
- Describe the minimum sampling frequency for field measurements.

E. Results:

- Explain how the results of this study yield will information relevant to management decisions.
- Describe how the study will provide useful results within the five- to ten-year timeframe identified in the Biological Opinion (NMFS 2000).

Although these elements are general in nature, they should be considered when designing a monitoring plan to assess the effectiveness of any management action or RPA, regardless of how simple the proposed action may be. Even a plan as simple as monitoring the effectiveness of irrigation screens requires careful consideration of all elements in the checklist. In some cases, the investigator may not be able to address all statements with a high degree of certainty because adequate information does not exist. For example, the investigator may lack information on population variability, effect size, or "practical significance," which can make it very 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. This document provides the investigator with information that can be used to address all elements in the checklist. Although the checklist is specific to monitoring physical/environmental conditions in tributary habitats, it applies equally well to monitoring biological conditions in tributary habitats.

8.0 Acknowledgments

The Bonneville Power Administration funded this work. This paper benefited from the helpful discussions of R. Beaty (BPA), D. Daley (BPA), J. Geiselman (BPA), C. Jordan (NMFS), S. Katz (NMFS), M. Newsom (BR), and C. Paulsen (Paulsen Environmental Research). This paper is a contribution from the Action Effectiveness Research Group.

9.0 References

ACOE, BR, and BPA (U.S. Army Corps of Engineers, Bureau of Reclamation, and the Bonneville Power Administration). 2001. Endangered Species Act implementation plan for the federal Columbia River power system. Draft report. U.S. Army Corps of Engineers, Bureau of Reclamation, and the Bonneville Power Administration, Portland, OR.

APHA, AWWA, and WEF (American Public Health Association, American Water Works Association, and Water Environment Federation). 1999. Standard methods for the examination of water and wastewater. 20th Edition. American Public Health Association, Washington, D.C.

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.

Bauer, S. B. and S. C. Ralph. 1999. Aquatic habitat indicators and their application to water quality objectives within the Clean Water Act. EPA-910-R-99-014, U.S. Environmental Protection Agency, Seattle, WA.

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.

Behnke, R. J. 1977. Fish faunal changes associated with land-use and water development. Great Plains-Rocky Mountain Geographical Journal 6:133-136.

Beschta, R. L., R. E. Bilby, G. W. Brown, L. B. Holtby, and T. D. Hofstra. 1987. Stream temperature and aquatic habitat: fisheries and forestry interactions. Pages 191-232 *in*: E. O. Salo and T. W. Cundy, editors. Streamside management: forestry and fishery interactions. University of Washington Institute of Forest Resources, Contribution No. 57, Seattle, 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.

Bisbal, G. A. 2001. Conceptual design of monitoring and evaluation plans for fish and wildlife in the Columbia River ecosystem. Environmental Management 28:433-453.

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.

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

Box, G. E. P., W. G. Hunter, and J. S. Hunter. 1978. Statistics for experimenters: an introduction to design, data analysis, and model building. John Wiley and Sons, Inc., New York, NY.

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

Bunte, K. and S. R. Abt. 2001. Sampling frame for improved pebble count accuracy in coarse gravel-bed streams. Journal of the American Water Resources Association 37:1001-1014.

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.

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

Cochran, W. G. 1977. Sampling techniques. Third edition. John Wiley and Sons, Inc., New York, NY.

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.

Davis, W. P., D. E. Hoss, G. I. Scott, and P. F. Sheridan. 1984. Fisheries resource impacts from spills of oil or hazardous substances. Pages 157-172 *in*: J. Cairns, Jr. and A. L. Buikema, Jr., editors. Restoration of habitats impacted by oil spills. Butterworth, Boston, MA.

Doppelt, B., M. Scurlock, C. Frissell, and J. Karr. 1993. Entering the watershed, a new approach to save America's river ecosystem. Island Press, Washington, D.C.

Dowdy, S. and S. Wearden. 1983. Statistics for research. John Wiley and Sons, Inc., New York, NY.

Dyer, A. 2002. Detecting and quantifying unpredictable events in ecological research. Bulletin of the Ecological Society of America 83:96-97.

EPA (Environmental Protection Agency). 1986. Quality criteria for water, 1986. USDC, Environmental Protection Agency, EPA 440/5-86-001, Washington, D.C.

EPA (Environmental Protection Agency). 1999. National recommended water quality criteria-correction. USDC, Environmental Protection Agency, EPA 822-Z-99-001, Washington, D.C.

FISRWG (Federal Interagency Stream Restoration Working Group). 2001. Stream corridor restoration: principles, processes, and practices. Prepared by 15 Federal agencies of the U.S. Government, GPO Item No. 0120-A; SuDocs No. A 57.6/2:EN3/PT.653, Washington, D.C.

Fox, M. J. 2001. A new look at the quantities and volumes of instream wood in forested basins within Washington State. Master's thesis. University of Washington, Seattle, WA.

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.

Frissell, C. A., W. J. Liss, and D. Bayles. 1993. An integrated, biophysical strategy for ecological restoration of large watersheds. Pages 449-456 *in*: D. F. Potts and N. E. Spangenberg, editors. Changing roles in water resources management and policy. American Water Resources Association, Bethesda, MD.

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.

Hawkes, H. A. 1975. River zonation and classification. Pages 312-374 *in*: B. A. Whitton, editor. River ecology. Blackwell Scientific, Oxford.

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.

Hicks, M. 2000. Evaluating standards for protecting aquatic life in Washington's surface water quality standards, temperature criteria. Draft discussion paper and literature summary. Washington Department of Ecology Publication Number 00-10-070, Olympia, WA.

Hillman, T. W., D. W. Chapman, T. S. Hardin, S. E. Jensen, and W. S. Platts. 1995. Assessment of injury to fish populations: Clark Fork River NPL sites, Montana, January 1995. In: J. Lipton, et al. Aquatic resources injury assessment report, upper Clark Fork River Basin and appendices. RCG/Hagler, Bailly, Inc., Boulder, CO.

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.

Hynes, H. B. N. 1970. The ecology of running waters. University of Toronto Press.

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

ISRP (Independent Scientific Review Panel). 2001. Preliminary review of fiscal year 2002 project proposals for the Mountain Snake and Blue Mountain Provinces.

Independent Scientific Review Panel for the Northwest Power Planning Council, ISRP 2001-9, Portland, OR.

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.

Keppel, G. 1982. Design and analysis: a researcher's handbook. Second edition. Prentice-Hall, Inc., Englewood Cliffs, NJ.

King, J. G. 1989. Streamflow responses to road building and harvesting: a comparison with the equivalent clearcut area procedure. USDA Forest Service, Research Paper INT-401, Ogden, UT.

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

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

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.

McDonald, L. L. and W. P. Erickson. 1994. Testing for bioequivalence in field studies: has a disturbed site been adequately reclaimed? Pages 183-197 *in*: D. J. Fletcher and B. F. J. Manly, editors. Statistics in ecology and environmental monitoring. University of Otago Press, Dunedin.

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.

Mead, R. 1988. The design of experiments: statistical principles for practical application. Cambridge University Press, New York, NY.

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.

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.

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. Endangered Species Act – Section 7 Consultation, biological opinion, reinitiation of consultation on operation of the Federal Columbia River Power System, including the juvenile fish transportation program, and 19 Bureau of Reclamation projects in the Columbia Basin. National Marine Fisheries Service, Northwest Region, Seattle, WA.

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

ODEQ (Oregon Department of Environmental Quality). 1995. 1992-1994 water quality standards review. Final issue papers. Oregon Department of Environmental Quality, Portland, OR.

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.

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.

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.

Peck, D. V., J. M. Lazorchak, and D. J. Klemm, editors. Unpublished draft. Environmental monitoring and assessment program—surface waters: western pilot study field operations manual for wadeable streams. U.S. Environmental Protection Agency, Washington, D.C.

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

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.

Platts, W. S. 1990. Fish, wildlife and livestock: protection of riparian areas. Western Wildlands 16:16-19.

Quigley, T. M. and S. J. Arbelbide, editors. 1997. An assessment of ecosystem components in the interior Columbia Basin and portions of the Klamath and Great Basins. USDA Forest Service and USDI Bureau of Land Management, General Technical Report PNW-GTR-405, Portland, OR.

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.

Resh, V. H., A. V. Brown, A. P. Covich, M. E. Gurtz, H. W. Li, G. W. Minshall, S. R. Reice, A. L. Sheldon, J. B. Wallace, and R. C. Wissmar. 1988. The role of disturbance in stream ecology. Journal of the North American Benthological Society 7:433-455.

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

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.

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.

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.

Sedell, J. R., G. H. Reeves, R. R. Hauer, J. A. Stanford, and C. P. Hawkins. 1990. Role of refugia in recovery from disturbances: modern fragmented and disconnected river systems. Environmental Management 14:711-724.

Shipley, B. 2000. Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference. Cambridge University Press, New York, NY.

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.

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.

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.

Swanson, F. J., L. E. Benda, S. H. Duncan, G. E. Grant, W. F. Megahan, L. M. Reid, and R. R. Ziemer. 1987. Mass failures and other processes of sediment production in Pacific Northwest forest landscapes. Pages 9-38 *in*: E. O. Salo and T. W. Cundy, editors. Streamside management: forestry and fishery interactions. University of Washington Institute of Forest Resources, Contribution No. 57, Seattle, WA.

Thompson, S. K. 1992. Sampling. John Wiley and Sons, Inc., New York, NY.

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.

Underwood, A. J. 1998. Design, implementation, and analysis of ecological and environmental experiments: pitfalls in the maintenance of logical structure. Pages 325-349 *in*: W. J. Resetarits, Jr. and J. Bernardo, editors. Experimental ecology: issues and perspectives. Oxford University Press, New York, NY.

USFS (U.S. Forest Service). 1974. Forest hydrology part II—hydrologic effects of vegetation manipulation. USDA Forest Service, Missoula, MT.

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.

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.

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.

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.

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.

Zaroban, D. W. 2000. Protocol for placement and retrieval of temperature data loggers in Idaho streams. Idaho Division of Environmental Quality, Boise, ID.

Appendix A. Final indicator selection summary, showing relationship to stressors, a composite usability ranking, and an indication of how the data will be gathered. Reported in the June 2001 draft (PACFISH/INFISH 2001)

Indicator	Direct/Indirect[1]	Usability	Data Collection[2]
Land Use History and Current Management (upland and riparian)			
equivalent road acres	D	high	all*
road density - hydrologically connected	D	high	all*
# of culverts and stream crossings	D	high	office, field*
culvert failure rate	D	high	office, field
mining history/extent	D	?	office, field
forest condition: fire frequency, harvest	D	med/high	office, field
roads: landslide frequency, size, location	D	med	office, field*
Riparian/Floodplain Habitat			
Bank material – soil type, comp., infilt.	I	high	field*
fragmentation of riparian veg - high contrast	I	high	rm, field
seral stage / structural complexity of riparian	I	high	rm, field
floodplain interactions/connectivity	I	med/high	field
effective ground cover	D	high	field*
In-channel/Community Integrity			
invertebrate community structure	I	med/high	field*
water quality - direct measures	I	med	field*
water temperature - direct measures	I	high	field*
frequency, distribution, arrangement of LWD	I	high	field*
cross section mapping	I	high	field*
width to depth ratio, frequency of large pools, longitudinal profiles, residual pool depth, bank angles, % undercut bank, substrate comp., bank stability	I	high	field, rm*

[1] Direct (D) or indirect (I) measure of a stressor

[2] Remote sensing(rm) = aerial photos, maps, infra-red, and satellite imagery; office = information on file in Forest offices or that can be gathered through library research; field = requires field data collection; all=all three of these techniques are used.

* Data is quantitative, measured and not estimated

Appendix B. Matrix of pathways and indicators for streams east of the Cascade Mountains (developed by the NMFS 1996).

Pathway	Indicators	Existing condition		
		Properly functioning	At risk	Not properly functioning
Water Quality	Temperature	10-13.9 C	13.9-15.6 C (spawning) 13.9-17.8 C (migrate/rearing)	>15.6 C (spawning) >17.8 C (migrate/rearing)
	Sediment/turbidity	<12% fines (<0.85 mm) in gravel and turbidity low	12-20% fines and turbidity moderate	>20% fines and turbidity high
	Chemical contamination/ nutrients	Low levels of chemical contamination from land-use sources, no excessive nutrients, no CWA 303d designated reaches	Moderate levels of chemical contamination from land-use sources, some excess nutrients, one CWA 303d designated reach	High levels of chemical contamination from land-use sources, high levels of excess nutrients, more than one CWA 303d designated reach
Habitat Access	Physical barriers	Any man-made barriers present in watershed allow upstream and downstream fish passage at all flows	Any man-made barriers present in watershed do not allow upstream or downstream fish passage at base (low) flows	Any man-made barriers present in watershed do not allow upstream or downstream fish passage at a range of flows
Habitat Elements	Substrate	Dominant substrate is gravel or cobble (interstitial spaces clear), or embeddedness <20%	Gravel and cobble is subdominant, or if dominant, embeddedness 20-30%	Bedrock, sand, silt, or small gravel dominant, or if gravel and cobble dominant, embeddedness >30%
	Large woody debris	>20 pieces/mile >12" diameter >35 ft length; and adequate sources of woody debris recruitment in riparian areas	Currently meets standards for properly functioning, but lacks potential sources from riparian areas of woody debris recruitment to maintain that standard	Does not meet standards for properly functioning and lacks potential large woody debris recruitment
	Pool frequency: Channel width — No. pools/mile 5 ft — 184 10 ft — 96 15 ft — 70 20 ft — 56 25 ft — 47 50 ft — 26 75 ft — 23 100 ft — 18	Meets pool frequency standards and large woody debris recruitment standards for properly functioning habitat	Meets pool frequency standards but large woody debris recruitment is inadequate to maintain pools over time	Does not meet pool frequency standards

Appendix B. Continued.

Pathway	Indicators	Existing condition: Properly functioning	At risk	Not properly functioning
Habitat Elements (cont.)	Pool quality	Pools >1 meter deep (holding pools) with good cover and cool water, minor reduction of pool volume by fine sediment	Few deeper pools (>1 m) present or inadequate cover/temperature, moderate reduction of pool volume by fine sediment	No deep pools (>1 m) and inadequate cover/temperature, major reduction of pool volume by fine sediment
	Off-channel habitat	Backwaters with cover, and low energy off-channel areas (ponds)	Some backwaters and high energy side channels	Few or no backwaters, no off-channel ponds
	Refugia (important remnant habitat for sensitive aquatic species)	Habitat refugia exist and are adequately buffered (e.g., by intact riparian reserves); existing refugia are sufficient in size, number, and connectivity to maintain viable populations or sub-populations	Habitat refugia exist but are not adequately buffered (e.g., by intact riparian reserves); existing refugia are insufficient in size, number, and connectivity to maintain viable populations or sub-populations	Adequate habitat refugia do not exist
Channel Condition and Dynamics	Width/depth ratio	<10	10-12	>12
	Streambank condition	>90% stable; i.e., on average, less than 10% of banks are actively eroding	80-90% stable	<80% stable
	Floodplain connectivity	Off-channel areas are frequently hydrologically linked to main channel; overbank flows occur and maintain wetland functions, riparian vegetation and succession	Reduced linkage of wetland, floodplains and riparian areas to main channel; overbank flows are reduced relative to historic frequency, as evidenced by moderate degradation of wetland function, riparian vegetation/succession	Severe reduction in hydrologic connectivity between off-channel, wetland, floodplain and riparian areas; wetland extent drastically reduced and riparian vegetation/succession altered significantly

Appendix B. Concluded.

Pathway	Indicators	Existing condition		
		Properly functioning	At risk	Not properly functioning
Flow/Hydrology	Change in peak/base flows	Watershed hydrograph indicates peak flow, base flow, and flow timing characteristics comparable to an undisturbed watershed of similar size, geology, and geography	Some evidence of altered peak flow, baseflow, or flow timing relative to an undisturbed watershed of similar size, geology, and geography	Pronounced changed in peak flow, baseflow, or flow timing relative to an undisturbed watershed of similar size, geology, and geography
	Increase in drainage network	Zero or minimum increases in drainage network density due to roads	Moderate increases in drainage network density due to roads (e.g., about 5%)	Significant increases in drainage network density due to roads (e.g., about 20-25%)
Watershed Conditions	Road density and location	<2 mi/mi^2, no valley bottom roads	2-3 mi/mi^2, some valley bottom roads	>3 mi/mi^2, many valley bottom roads
	Disturbance history	<15% ECA (entire watershed) with no concentration of disturbance in unstable or potential unstable areas, refugia, or riparian areas	<15% ECA (entire watershed) but disturbance concentrated in unstable or potential unstable areas, refugia, or riparian areas	>15% ECA (entire watershed) and disturbance concentrated in unstable or potential unstable areas, refugia, or riparian areas
	Riparian conservation areas (RHCA – PACFISH and INFISH)	Riparian conservations areas provide adequate shade, large woody debris recruitment, and habitat protection and connectivity in all subwatersheds, and buffers include known refugia for sensitive aquatic species (>80% intact), and/or for grazing impacts: percent similarity of riparian vegetation to the potential natural community/ composition >50%	Moderate loss of connectivity or function (shade, LWD recruitment, etc.) of riparian conservation areas, or incomplete protection of habitats and refugia for sensitive aquatic species (about 70-80% intact), and/or for grazing impacts: percent similarity of riparian vegetation to the potential natural community/composition 25-50% or better	Riparian conservation areas are fragmented, poorly connected, or provide inadequate protection of habitats and refugia for sensitive aquatic species (<70% intact), and/or for grazing impacts: percent similarity of riparian vegetation to the potential natural community/composition <25%

Appendix C. Matrix of physical/environmental pathways and indicators for east-side steams developed by the USFWS (1998).

Pathway	Indicators	Existing condition		
		Functioning adequately	Functioning at risk	At unacceptable risk
Water Quality	Temperature	MWMT in a reach during the following life history stages: Incubation 2-5C Rearing 4-12C Spawning 4-9C Temperatures do not exceed 15C in areas used by adults during the local spawning migration	MWMT in a reach during the following life history stages: Incubation <2 or 6C Rearing <4 or 13-15C Spawning <4 or 10C Temperatures in areas used by adults during the local spawning migration sometimes exceeds 15C	MWMT in a reach during the following life history stages: Incubation <1 or >6C Rearing >15C Spawning <4 or >10C Temperatures in areas used by adults during the local spawning migration regularly exceed 15C
	Sediment (in areas of spawning and incubation)	<12% fines (<0.85 mm) in gravel	12-20% fines in gravel	>20% fines in gravel
	Chemical contamination/ nutrients	Low levels of chemical contamination from land-use sources, no excessive nutrients, no CWA 303d designated reaches	Moderate levels of chemical contamination from land-use sources, some excess nutrients, one CWA 303d designated reach	High levels of chemical contamination from land-use sources, high levels of excess nutrients, more than one CWA 303d designated reach
Habitat Access	Physical barriers	Any man-made barriers present in watershed allow upstream and downstream fish passage at all flows	Any man-made barriers present in watershed do not allow upstream or downstream fish passage at base (low) flows	Any man-made barriers present in watershed do not allow upstream or downstream fish passage at a range of flows
Habitat Elements	Substrate embeddedness in rearing areas	Reach embeddedness <20%	Reach embeddedness 20-30%	Reach embeddedness >30%
	Large woody material	>20 pieces/mile >12" diameter >35 ft length; and adequate sources of woody debris available for both long and short-tern recruitment	Currently levels are being maintained at minimum levels desired fro "functioning adequately," but potential sources for long term woody debris recruitment are lacking to maintain these minimum values	Current levels are not at those desired values for "functioning adequately," and potential sources of woody debris for short and/or long term recruitment are lacking

Appendix C. Continued.

Pathway	Indicators	Existing condition		
		Functioning adequately	Functioning at risk	At unacceptable risk
Habitat Elements (cont.)	Pool frequency and quality	Channel width / No. pools/mile 0-5 ft 39 5-10 ft 60 10-15 ft 48 15-20 ft 39 20-30 ft 23 30-35 ft 18 35-40 ft 10 40-65 ft 9 65-100 ft 4 Pools have good cover and cool water and only minor reduction of pool volume by fine sediment	Pool frequency is similar to values in "functioning adequately," but pools have inadequate cover/temperature, and/or there has been a moderate reduction of pool volume by fine sediment	Pool frequency is considerably lower than values for "functioning adequately," also cover/temperature is inadequate, and there has been a major reduction of pool volume by fine sediment
	Large pools (in adult holding, juvenile rearing, and overwintering reaches where streams are >3m in wetted width at base flow)	Each reach has many large pools >1 m deep	Reaches have few large pools (>1 m) present	Reaches have no deep pools (>1 m)
	Off-channel habitat	Watershed has many ponds, oxbows, backwaters, and other off-channel areas with cover, and side-channels are low energy areas	Watershed has some ponds, oxbows, backwaters, and other off-channel areas with cover, but side-channels are generally high energy areas	Watershed has few or no ponds, oxbows, backwaters, or other off-channel areas
	Refugia (important remnant habitat for sensitive aquatic species)	Habitats capable of supporting strong and significant populations are protected and are well distributed and connected for all life stages and forms of the species	Habitats capable of supporting strong and significant populations are insufficient in size, number and connectivity to maintain all life stages and forms of the species	Adequate habitat refugia do not exist
Channel Condition and Dynamics	Wetted width/maximum depth ratio in scour pools in a reach	<10	11-20	>20
	Streambank condition	>80% of any stream reach has >90% stability	50-80% of any stream reach has >90% stability	<50% of any stream reach has >90% stability

Appendix C. Continued.

Pathway	Indicators	Existing condition Functioning adequately	Existing condition Functioning at risk	Existing condition At unacceptable risk
Channel condition and dynamics (cont.)	Floodplain connectivity	Off-channel areas are frequently hydrologically linked to main channel; overbank flows occur and maintain wetland functions, riparian vegetation and succession	Reduced linkage of wetland, floodplains and riparian areas to main channel; overbank flows are reduced relative to historic frequency, as evidenced by moderate degradation of wetland function, riparian vegetation/succession	Severe reduction in hydrologic connectivity between off-channel, wetland, floodplain and riparian areas; wetland extent drastically reduced and riparian vegetation/succession altered significantly
Flow/Hydrology	Change in peak/base flows	Watershed hydrograph indicates peak flow, base flow, and flow timing characteristics comparable to an undisturbed watershed of similar size, geology, and geography	Some evidence of altered peak flow, baseflow, or flow timing relative to an undisturbed watershed of similar size, geology, and geography	Pronounced changed in peak flow, baseflow, or flow timing relative to an undisturbed watershed of similar size, geology, and geography
	Increase in drainage network	Zero or minimum increases in active channel length correlated with human-caused disturbance	Low to moderate increases in active channel length correlated with human caused disturbance	Greater than moderate increase in active channel length correlated with human caused disturbance
Watershed Conditions	Road density and location	<1 mi/mi^2	1-2.4 mi/mi^2	>2.4 mi/mi^2
	Disturbance history	<15% ECA of entire watershed with no concentration of disturbance in unstable or potential unstable areas, refugia, or riparian areas	<15% ECA of entire watershed but disturbance concentrated in unstable or potential unstable areas, refugia, or riparian areas	>15% ECA of entire watershed and disturbance concentrated in unstable or potential unstable areas, refugia, or riparian areas

Appendix C. Concluded.

Pathway	Indicators	Existing condition		
		Functioning adequately	Functioning at risk	At unacceptable risk
Watershed Conditions (cont.)	Riparian conservation areas (RHCA – PACFISH and INFISH)	The riparian conservation areas provide adequate shade, large woody debris recruitment, and habitat protection and connectivity in subwatersheds, and buffers include known refugia for sensitive aquatic species (>80% intact), and adequately buffer impacts on rangelands; percent similarity of riparian vegetation to the potential natural community/ composition >50%	Moderate loss of connectivity or function (shade, LWD recruitment, etc.) of riparian conservation areas, or incomplete protection of habitats and refugia for sensitive aquatic species (about 70-80% intact), and adequate buffer impacts on rangelands: percent similarity of riparian vegetation to the potential natural community/composition 25-50% or better	Riparian conservation areas are fragmented, poorly connected, or provide inadequate protection of habitats for sensitive aquatic species (<70% intact, refugia does not occur), and adequately buffer impacts on rangelands: percent similarity of riparian vegetation to the potential natural community/composition <25%
	Disturbance regime	Environmental disturbance is short lived; predictable hydrograph, high quality habitat and watershed complexity providing refuge and rearing space for all life stages or multiple life-history forms. Natural processes are stable.	Scour events, debris torrents, or catastrophic fires are localized events that occur in several minor parts of the watershed. Resiliency of habitat to recover from environmental disturbances is moderate.	Frequent flood or drought producing highly variable and unpredictable flows, scour events, debris torrents, or high probability of catastrophic fire exists throughout a major part of the watershed. The channel is simplified, providing little hydraulic complexity in the form of pools or side channels. Natural processes are unstable.