Use of decision analysis to design a
habitat restoration experiment

by

Ian James Parnell

B.Sc., Simon Fraser University, 1990

RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF RESOURCE MANAGEMENT

in the

School of Resource and Environmental Management

Report No. 276

SIMON FRASER UNIVERSITY

August 2002

Approval

Name:	Ian James Parnell
Degree:	Master of Resource Management
Report number:	276
Title of Research Project:	Use of decision analysis to design a habitat restoration experiment
Supervisory Committee:	

Dr. Randall Peterman
Senior Supervisor
Professor
School of Resource and Environmental Management
Simon Fraser University

Dr. Michael Bradford
Committee Member
Adjunct Professor
School of Resource and Environmental Management,
Simon Fraser University

Date Approved: ______________________________

Abstract

Recovery plans for endangered salmon stocks often include aggressive restoration of freshwater spawning and rearing habitat. However, there is large uncertainty about its effectiveness for increasing freshwater survival rates compared to cheaper, passive, actions that focus on habitat protection. Experimental implementation of restoration projects could reduce uncertainty and improve future recovery decisions, but optimal designs should balance statistical requirements for high power against the social costs associated with uncertain outcomes. I used decision analysis to design an example experiment for testing the relative effectiveness of aggressive and passive habitat actions for increasing the egg-to-parr survival rate of spring chinook salmon (*Oncorhynchus tshawytscha*). This approach not only accounted for the costs of experimenting, but also the magnitude of costs for different outcomes and their probability of occurrence. I ranked the candidate designs using an objective of minimizing expected total cost to society and found that the most expensive, highest-power design was optimal. This choice was robust to a wide range of assumptions, but primarily depended upon the high social costs incurred under outcomes where stocks went extinct. These results are different from other research that shows less powerful experiments can be optimal.

Acknowledgements

I thank my supervisory committee, Randall Peterman and Mike Bradford, for their advice, expertise, encouragement, and patience. I owe thanks to the members of the Fisheries Research Group for providing helpful suggestions during the early stages of this project. I thank the fisheries biologists of the Idaho Department of Fish and Game for providing data and information on Idaho salmon research and recovery programs: in particular Charlie Petrosky, Judy Hall-Griswold and Doug Nemeth. Finally, I thank my colleagues at ESSA Technologies Ltd., Dave Marmorek, Calvin Peters and Clint Alexander, who, through our shared work experiences, have helped shape my thinking during the latter stages of this project.

Table of Contents

List of Tables

List of Figures

Introduction

Recent precipitous declines in the abundance of many stocks of Pacific salmon (e.g., Slaney et al. 1996; Schaller et al. 1999) are attributed partly to the degradation of their freshwater spawning and rearing habitat (NAP 1996). Consequently, the restoration of freshwater spawning and rearing habitat is a commonly used action to achieve recovery of such populations. Habitat restoration actions fall into two general categories: “passive”, or “aggressive” (e.g., NAP 1996). Passive actions include government legislation and regulations designed to protect salmon stream habitat (e.g., British Columbia’s Fish Protection Act) and rely on natural processes to restore habitat to its pristine state. Because natural processes can operate on decadal time scales (e.g., Roni et al. 2002) management agencies may also use “aggressive” restoration actions to manipulate freshwater habitat directly to speed its recovery (e.g., British Columbia’s Watershed Restoration Program (WRP) and Fisheries and Oceans Canada’s Habitat Restoration and Salmonid Enhancement Program (HRSEP)).

Both types of restoration actions incur social costs - costs incurred by society as a result of government actions. These costs include lost economic opportunity due to new regulations as well as the cost of projects funded with public money. The costs of aggressive actions include the costs of passive actions and thus are always greater. For example, regional habitat protection regulations will also apply to those watersheds where aggressive actions take place, so the costs of aggressive actions will be added to the costs of ongoing passive actions. The costs of aggressive actions also have a higher profile than passive costs. While the costs of passive actions may be spread widely over various stakeholder groups, aggressive actions are funded through budgets allocated to specific resource sectors (i.e. fisheries) and can make up a high proportion of those budgets. This inflicts additional opportunity costs.-.the net benefit forgone because the resources providing the service can no longer be used in their next most beneficial use (Tietenberg 1992). For example, implementing and properly monitoring an aggressive habitat restoration program may use up funds that could have been used for an alternative, and perhaps more successful, recovery program.

Aggressive habitat actions are often justified on the assumption that they will restore spawning and rearing habitat, and consequently the salmon populations that depend on it, more quickly than passive habitat actions (e.g., Slaney 2000). This is usually just a hypothesis; there is limited evidence that the application of aggressive restoration actions is generally successful at increasing production of the freshwater lifestages of salmon (e.g., Roni et al. 2002). Using an experimental approach when implementing aggressive habitat restoration projects to deliberately test this hypothesis could reduce uncertainty about their future effectiveness and the benefits for both salmon and society (e.g., MacGregor et al. 2002). Proper experimental design contributes to this goal in at least two ways: (1) it increases the probability of detecting true effects of some specified magnitude (i.e. statistical power, Peterman 1990); and (2), it increases the strength of inferences about results of actions by reducing the confounding of management actions with uncontrolled environmental processes (Green 1979).

Statistically powerful experiments are not always economically optimal due to the costs of experimenting and the potential costs and benefits of decisions based on the outcome of the experiment and their probability of occurrence (Walters and Collie 1988; MacGregor et al. 2002; Keeley and Walters 1994). The net value of experimenting will depend on who bears the biological, social, and economic costs of experimental errors of inference and their probability of occurrence, factors that should be considered explicitly prior to the initiation of an experimental management program. *A priori* statistical power analysis (Peterman 1990) and decision analysis (Clemen 1996, Peterman and Anderson 1999) are useful tools for assessing the relative value of different experimental designs in terms of both social and scientific objectives (Peterman 1990; Peterman and Antcliffe 1993; MacGregor et al. 2002) and both have been applied to the design of resource management experiments (MacGregor et al. 2002, Keeley and Walters 1994, Walters and Green 1997, McAllister and Peterman 1992a,b).

One area where such considerations are especially relevant is the Columbia River basin where salmon stocks have declined sharply since the development of the Columbia River hydrosystem, leading to listing many stocks under the United States' Endangered Species Act (ESA) (Schaller et al. 1999). The Northwest Power Planning Council's Fish and Wildlife program spends millions of dollars annually to help recover threatened salmon stocks (e.g., BPA 2001). Recent modeling analyses have provided contradictory advice, finding that either the breaching of certain dams (Peters and Marmorek 2001), or off-site mitigation efforts (e.g., habitat restoration) in combination with improved downstream migration conditions for smolts (e.g., transportation around dams) (Kareiva et al. 2000) will be the best option for recovery. Prior to deciding whether to breach dams, the National Marine Fisheries Service (NMFS) opted to first try and achieve recovery through a combination of “reasonable and prudent actions”, including freshwater habitat restoration, with periodic evaluation of their effectiveness at 3-, 5- and 8-years (NMFS 2000). The large amount of money being spent annually on recovery as well as the implications of dam breaching in terms of lost electric power and impaired river transportation indicate that the outcomes of these decisions have high social costs, and that these costs may be distributed asymmetrically across stakeholder groups (e.g., those who value the existence of the salmon vs. those who rely on economical hydroelectic power). Therefore, it is important to design restoration experiments that can provide good estimates of effectiveness and that also consider the costs of the uncertain outcomes of these experiments.

Here I show an example of how to design a Before-After-Control-Impact-Paired series (BACIP) (Stewart-Oaten et al. 1986) management experiment for comparing the relative effectiveness of passive and aggressive habitat restoration actions, roughly in the time-frame of the evaluation period set by NMFS (2000). I set this experiment in the context of increasing the egg-to-parr survival rate of ESA-listed stocks of spring chinook salmon (*Oncorhyncus tschwaytscha*) in the Middle Fork Salmon River basin, Idaho by reducing stream sedimentation. The Middle Fork Salmon River (a river in the Columbia River basin) is in the Idaho Batholith (Andrews and Everson 1988), a geographic area of granitic soils particularly susceptible to erosion (Platts et al. 1989, Rhodes et al. 1994). Many of the land-use practices there, such as livestock grazing, can increase sediment input to salmon streams (e.g., Meehan 1991). Increased fine sediment reduces the quality and quantity of juvenile rearing habitat by covering redds and suffocating incubating eggs, entombing alevin, removing habitat for the benthic organisms that are food for juvenile salmon, and filling interstitial cobble spaces and pools where juvenile chinook hide and overwinter (Rhodes et al. 1994). Various state, federal and tribal management agencies have implemented both passive and aggressive restoration actions in this area to address severe sediment problems believed limiting for the production of juvenile chinook. An example of a passive action is the United States Forest Service's (USFS) revised grazing plans for the Marsh Creek and Bear Valley/Elk Creek watersheds (Beamesderfer et al. 1997). An example of an aggressive action is the USFS's and Shoshone-Bannock Tribe's fencing and re-vegetation program in the Bear Valley/Elk Creek watershed (Andrews and Everson 1988).

Because there is as yet no coordinated approach to experimental evaluation of habitat restoration activities across the Columbia River basin, I assume this experiment takes place in isolation of other activities throughout the Columbia basin and that managers can afford to monitor at most two watersheds. I assume that they have pre-existing baseline data available to them from other monitoring programs that will allow implementation of a BACIP experiment without preliminary baseline monitoring. I also assume that if the results of the experiment indicate that aggressive treatment is better than the passive treatment, this will trigger the release of funds allowing wider application of the aggressive treatment to other candidate watersheds.

My example has three major components: experimental design, a priori statistical power analysis, and decision analysis. I will identify the rank order of alternative experimental designs based on two objectives: 1) a social objective of minimizing the expected total cost to society and 2) a statistical objective of being the quickest to achieve an acceptably high level of statistical power (i.e. ≥ 0.8). I calculate outcomes in terms of expected costs because it is difficult to estimate the intangible benefits to society of enhancement for endangered salmon stocks (e.g., Loomis and White 1996). Depending on whether the more expensive, higher power designs reduce costs to society more than their additional cost to implement, the rank order of designs may differ for these two objectives.

Methods

Experimental design

I broke the experimental design into several logical components to facilitate description. The first component describes the purpose of the experiment and covers the experimental objective, treatment and management hypothesis. The second component covers the statistical requirements including the biological measurements of outcomes and the BACIP monitoring framework. The third component combines elements of the first two into a model of the costs of experimenting. A specific experimental design is a single combination of the number of years of post-treatment monitoring, the level of statistical significance used for hypothesis testing, and the type of monitoring program used to estimate biological outcomes (i.e. a change in the egg-to-parr survival rate).

Experimental objective, treatment and management hypothesis

The experimental objective is to compare the relative effectiveness of passive land use regulations and a form of aggressive habitat manipulation for reducing stream sedimentation and increasing the egg-to-parr survival rate of juvenile chinook salmon. The treatment consists of applying aggressive sediment control actions (e.g., road deactivation) to one stream, while continuing to manage the other under an existing passive regime that relies on land-use restrictions (e.g., grazing management) to reduce sediment input. The management hypothesis is that the aggressive restoration action will increase the egg-to-parr survival rate of spring chinook salmon more quickly than the passive restoration action.

Index of egg-to-parr survival rate and BACIP monitoring framework

Index of egg-to-parr survival rate

I used parr density/spawner abundance (P/S) as an index of the egg-to-parr survival rate. Developing this index is more expensive than either a parr density or spawner abundance index, but it accounts for the effect of spawner abundance on parr density, is linked closely to freshwater rearing conditions, and can respond to changes in the first year after treatment. This will reduce confounding compared to just using parr abundance, or spawner abundance alone, decrease response time, and improve inferences about the effect of habitat restoration actions. These advantages may offset the higher data collection costs.

I assumed that fry emigration is minimal and not related to habitat quality so that P/S is an index of both habitat quality and the egg-to-parr survival rate. This is a necessary assumption because summer parr surveys index parr populations after egg-to-parr survival and fry emigration. If there is significant, habitat related fry emigration, the egg-to-parr survival and fry emigration rates will be confounded (i.e. a high emigration will be perceived as low egg-to-parr survival rate). However, the results of an Idaho Department of Fish and Game study suggest that fry emigration is not significant at the observed low spawner abundances for the stocks whose data I use in this analysis; a survey of mainstem Middle Fork Salmon River rearing areas found parr densities only 3-13% of those observed in the sample areas of the tributary streams (Scully et al. 1990).

I developed the P/S index using parr density and spawner abundance data collected for ESA-listed spring chinook stocks in tributary streams of the Middle Fork Salmon River, Idaho (Bear Valley/Elk Creek, Marsh Creek, and Sulphur Creek, Table 1, top box). Bear Valley/Elk Creek is impacted by sediment while Marsh Creek and Sulphur Creek are pristine spring chinook spawning and rearing habitat (Hall-Griswold and Petrosky 1996). I used parr density data collected in C-channel habitat (Rosgen 1985) by the Idaho Department of Fish and Game’s (IDFG) General Parr Monitoring (GPM) program (Hall-Griswold and Petrosky 1996) from 1985-1996. C-channel habitat is the preferred habitat of juvenile chinook and under the low seeding levels observed from 1985 to 1996, summer parr could be expected to concentrate there, making the data representative of the true parr distribution in the sampled streams. The spawner abundance data are derived from annual fall redd counts conducted by IDFG for these same streams, expanded to an estimate of total annual spawner abundance by adjusting for stream length (Beamesderfer et al. 1997).

Table 1. Summary of biological data. The top box presents summary statistics, sampling information and estimated egg-to-parr survival rates for Bear Valley/Elk Creek (BVC), Marsh Creek (MCR) and Sulphur Creek (SCR), tributaries of the Middle Fork Salmon River, Idaho. “Ln(P/S)” is the natural log of the ratio of parr density per 100m^2 and spawner abundance. “x” is mean , s^2 is sample variance, s is sample standard deviation, n is the number of annual Ln(P/S) data points for each stream, and CV is the coefficient of variation (the ratio of the standard error of the mean to the mean). “Stream sections sampled/year” is the range in number of stream sections sampled each year to estimate parr density for each stream. Egg-to-parr survival rate is estimated in a separate analysis. The middle box presents the correlations of the annual Ln(P/S) for each stream. The bottom box presents the summary statistics for the mean of the paired differences in Ln(P/S) for the two possible BACIP pairings under the assumption that Bear Valley/Elk Creek is the stream impacted by sedimentation.

Stream specific biological data[1]			
	BVC Ln(P/S)	**MCR Ln(P/S)**	**SCR Ln(P/S)**
x	-5.75	-2.27	-2.92
s^2	0.62	0.76	2.33
s	0.73	0.95	1.47
n	11	10	8
CV	0.04	0.13	0.18
Stream sections sampled/year	6 to 11	3 to 7	1 to 2
Egg-to-Parr survival rate[2]	1.2%	21.8%	11.9%

Correlation results, Ln(P/S)		
	BVC	**MCR**
MCR	0.34	
SCR	0.64	0.17

Mean and variance of the baseline paired differences ($D_{i,j}$)		
	$D_{BVC-SCR}$	**$D_{BVC-MCR}$**
x	-2.71	-3.38
s^2	1.34	0.93
s	1.16	0.96
n	8	10
CV	0.15	0.09

1 Parr density data from IDFG GPM database (Hall-Griswold and Petrosky 1996). Spawner abundance data from Beamesderfer et al. 1997.
2 I.J. Parnell, unpublished data.

The parr data showed a strong linear relationship with spawner abundance at the low-seeding levels in the data set (correlations ranged from 0.72 to 0.74, I.J. Parnell. unpublished data), but this linear relationship may not hold if juvenile populations increase substantially following successful treatment. A curvilinear Beverton-Holt egg-to-smolt relationship is commonly assumed for the chinook stocks of the Salmon River (e.g., Bjornn 1978, Bowles and Leitzinger 1991). Therefore, an important analytical decision was whether to model density-dependent effects explicitly. If not accounted for, these effects could confound results; increases in the P/S index could be interpreted as positive effects of habitat restoration but might actually reflect a density-dependent increase in the egg-to-parr survival rate under declining spawner abundance. If density-dependent effects were important, then statistical tests of change in the index of survival rate (P/S) would need to focus on changes in the parameters for models of density-dependent egg-to-parr survival. Alternatively, if density-dependent effects were not important during the experiment, I could use the simple P/S index. To resolve this, I asked two questions: 1) “Do the data indicate that the egg-to-parr survival rate is density-dependent?”, and 2) “Are density-dependent effects likely to become important over experimental periods in the range of those specified by NMFS 2000 (i.e. 3-, 5- and 8-years)?”

I found that a density-dependent model of parr production fit the data no better than a density-independent linear model of parr production (I.J. Parnell, unpublished data). Modeling the effects of recovery for a stock parameterized with the Middle Fork Salmon data showed that even under an unrealistic assumption of instant recovery in egg-to-parr survival rate from that of a degraded stream to that of a pristine one, it took 39 years for the spawning stock to rebuild to a range of abundances where density-dependent effects on the egg-to-parr survival rate would be important (I.J. Parnell, unpublished data). This is not surprising because these stocks are likely held well below their carrying capacity by the dam-related mortality they experience during their downstream migration through the Columbia River hydrosystem to the Pacific ocean (e.g., Deriso et al. 2001). Therefore, density-dependent effects would not be important for low-abundance stocks like those whose data I use here over experimental periods in the range of 3-8 years.

Based on these results, I assumed the following linear model of parr production (in numbers of parr produced per year per 100 m^2 of C-channel rearing habitat) for the rest of the analysis:

$$P_{i,j,k} = a_k \times S_{i,j,k} \times e^{e_{i,j,k}} \times e^{n_{i,j,k}} \tag{1}$$

where P is parr density, S is spawner abundance, a is productivity (a function of the density-independent egg-to-parr survival rate and fecundity), i is the index of year, j is the index of period (before [B] or after [A] treatment), k is the index of treatment (i.e. aggressive = a; passive= p), $\varepsilon_{i,j,k}$ is normally distributed process error (Bradford 1995), and $\nu_{i,j,k}$ is normally distributed measurement error. To produce the index used for statistical tests of change due to restoration I rearranged equation 1 and log-transformed it to normalize the errors, an important consideration for commonly used tests of significance.

$$Ln\left(\frac{P}{S}\right)_{i,j,k} = Ln(a)_k + e_{i,j,k} + n_{i,j,k} \tag{2}$$

BACIP Monitoring Framework

I used a Before-After-Control-Impact-Paired series (BACIP) design (Stewart-Oaten et al. 1986) as the monitoring framework. This design pseudo-replicates samples in time; more years of sampling increase statistical power. The paired BACIP differences for each year before and after treatment were estimated as:

(3) $$\mathrm{D}_{\mathrm{i,j}} = \mathrm{Ln}\left(\frac{\mathrm{P}}{\mathrm{S}}\right)_{i,j,a} - \mathrm{Ln}\left(\frac{\mathrm{P}}{\mathrm{S}}\right)_{\mathrm{i,j,p}} + (\varepsilon_{i,j,a} - \varepsilon_{i,j,p}) + (\nu_{i,j,a} - \nu_{i,j,p})$$

and the average $\mathrm{D}_{i,j}$, $\overline{\mathrm{D}}_j$, estimated the true average differences, D_j, across years for each period (i.e., Before ($\overline{\mathrm{D}}_B$) and After ($\overline{\mathrm{D}}_A$) treatment). I used the average and variance of the paired differences in Ln(P/S) for the baseline data to estimate the true average and variance for D_B and also for D_A in the absence of any treatment effect (Table 1, bottom box). I used a two-tailed t-test to test for a difference in the Before and After means because I was not certain of the direction of response in Ln(P/S) to aggressive habitat restoration relative to the passive action, even though I expected it to increase egg-to-parr survival rate more quickly. The statistical hypothesis for the two-sample two-tail t test was:

(3a) $\mathrm{H}_{\mathbf{o}} : \overline{\mathrm{D}}_{\mathbf{B}} = \overline{\mathrm{D}}_{\mathbf{A}}$, or $\overline{\mathrm{D}}_{\mathbf{B}} - \overline{\mathrm{D}}_{\mathbf{A}} = 0$

(3b) $\mathrm{H}_{\mathbf{A}} : \overline{\mathrm{D}}_{\mathbf{B}} \neq \overline{\mathrm{D}}_{\mathbf{A}}$, or $\overline{\mathrm{D}}_{\mathbf{B}} - \overline{\mathrm{D}}_{\mathbf{A}} \neq 0$ i.e. there is some difference, either positive or negative.

At the end the experiment, I assumed a decision was made based on the outcome of the t-test to either apply the aggressive treatment to a wider range of candidate streams, or to stop application of the aggressive treatment and focus on other restoration methods. I assumed the following decision rule: if the null hypothesis was rejected, the treatment effect was in the direction expected for a faster increase in the egg-to-parr survival rate for the aggressively treated stream relative to the passive one. If the t-test and BACIP assumptions were met, I concluded that the aggressive habitat restoration action was more effective than the passive restoration action and should be applied more widely.

Pairing Treatment and Control sites

A major assumption of the BACIP design is that pairing treatment and control streams that show high temporal correlation in Ln(P/S) will decrease variance in D_j and increase statistical power (Stewart-Oaten et al. 1986). Chinook stocks of the Snake River basin, including the three stocks whose data I used in this analysis, show a high degree of correlation in their temporal patterns of spawner abundance (Walters et al. 1989, Botsford and Paulsen 2000). However, covariance in spawner abundance data does not necessarily justify pairing of treatment-control streams based on the performance measure used here, Ln(P/S), because factors driving covariation in spawner abundance (S) could occur in one or more of the many life stages that occur between eggs and resulting adult spawners (e.g., Korman and Higgins 1997). Therefore, I evaluated covariation in the egg-to-parr life stage by first fitting the transformed model of parr production (eq. 2) to the data for each of the three candidate streams to remove the spawner effect. I then correlated the residuals to assess the degree of similarity between them following the example of Peterman et al. (1998) (I.J. Parnell, unpublished data). Sulphur Creek was selected to represent the “control” (passively restored) stream because it had the highest correlation with Bear

Valley/Elk Creek (correlation of 0.64 vs. 0.34 for Marsh Creek vs. Bear Valley/Elk Creek, Table 1, middle box).

Model of the costs of experimenting

Cost of Experimenting

The costs of experimenting included four basic components: implementation of aggressive treatment, maintenance of the treatment, monitoring in the aggressively treated and passively treated streams, and analysis of data at the end of the experimental period. Implementation costs include project management, design, and application of treatment (Table 2). Application of treatment includes the one-time costs of materials, labor, and equipment. Maintenance costs cover the annual cost of maintaining treatment (e.g., inspection and repair of roads and culverts). The monitoring component covers the costs of collecting parr density and spawner abundance data using summer snorkel counts and redd counts respectively. Although not explicit to this analysis, I also included the cost of monitoring the physical response of the system to treatment (e.g., %sand and cobble embeddedness indices) because this would also be necessary information for making inference about sediment reduction actions. The costs of analysis are incurred at the end of the experimental period. A general model combines these components in terms of present economic value:

$$(4) \quad C_{E,i} = (n_{pairs} \times C_{imp}) + (a \quad \times \ddot{e}^{n_A}) + \sum_{t=1}^{n_A} [(2 \quad \times m \quad \times n_{pairs} + n_{pairs} \times C_{main})_{n_A} \times \ddot{e}^{t}]$$

where $C_{E,i}$ is the total cost of experiment i, n_{pairs} is the number of paired treatment-control watersheds (one here), n_A is the total length of the experiment after treatment in years, C_{imp} is the implementation cost, C_{main} is the annual treatment maintenance cost, m is annual monitoring cost per stream, a is the cost of analysis at the end of the experimental period, and λ is the discount factor ($1/[1+r]$) used to convert costs to present value, where r is the discount rate. r was set to 5% for base-case calculations and was varied in sensitivity analyses. Equation 4 is an adaptation of cost models presented by Keeley and Walters (1994) and Antcliffe (1992). A summary of these symbols and other symbols discussed later is presented in Table 3.

An important consideration in the design of management experiments is to identify the conditions that might lead to choosing a more expensive, higher-power design over a less expensive, lower-power design (e.g., MacGregor et al. 2002). To explore this, I modeled both lower- and higher-cost BACIP-type monitoring programs. The lower-cost program combines a low-sampling-intensity parr sampling design similar to the IDFG GPM program (Hall-Griswold and Petrosky 1996) and the “Index stock spawner abundance program” described by Beamsderfer et al. (1997). The higher-cost sampling program combines a higher sampling intensity parr sampling program similar to that used in the IDFG Idaho Supplementation Studies (ISS) program (Bowles and Leitzinger 1991) and a spawner abundance program assumed to increase the precision and accuracy of spawner abundance estimates relative to that used for the low-cost program. I assumed that the more extensive ISS-type program would be preferred, given sufficient funding. The low-cost program is the base-case. The detailed components of these costs are described below (Table 2).

To estimate costs for each of the lower- and higher-power monitoring designs, I made the annual monitoring cost (*m)* a function of the number of stream sections sampled in parr monitoring and the cost of estimating spawner abundance:

$$m = C_{\mathbf{sediment}} + C_{\mathbf{spawner}} + \left[C_{\mathbf{base}} + \left(C_{\mathbf{section}} \times S_{\mathbf{sampled}}\right)\right] \tag{5}$$

where for each stream C_{sediment} is the cost of sediment sampling, held constant across all designs, C_{spawner} is the cost of estimating spawner abundance, C_{base} is a base travel cost associated with parr sampling that is incurred regardless of the number of stream sections sampled, C_{section} is the average sampling cost per stream section, and S_{sampled} is the number of stream sections sampled during parr surveys. I used equation 5 to shift from lower- to higher-cost monitoring by increasing S_{sampled} from the average cost observed under the GPM monitoring program for the Middle Fork Salmon River streams to the average cost observed for an ISS-type monitoring program that provides more precise estimates of mean parr abundance. I also increased the cost of estimating spawner abundance (C_{spawner}) (Table 2).

Table 2. Estimated costs for components of experimental design.

Cost component	Cost ($)	Source
One time costs		
Initialization:		
Design	$23,000	Adapted from Andrews 1988
Implementation	$205,000	
Analysis	$50,000	Estimate
Annual costs		
Project management	$25,247	Adapted from Andrews 1988
Maintenance	$4,000	Adapted from Andrews 1988
Monitoring:		
<u>Parr density</u>		
Low-sampling-intensity		
Bear Valley/Elk Creek	$2,290	D. Nemeth, IDFG pers. comm.,1998.
Sulfur Creek	$1,527	
High-sampling-intensity		
Bear Valley/Elk Creek	$8,439	ISS estimate
Sulfur Creek	$5,490	ISS estimate
<u>Spawner abundance</u>		
Inexpensive	$2,290	Estimate
Expensive	$6,870	Estimate (3 x inexpensive)
Sediment sampling	$20,413	Adapted from Andrews 1988

Table 3. Summary of base-case parameter values. For those parameters varied in sensitivity analyses, the range is noted in the "Definitions" column.

Symbol	Base-case value	Definition
n_B	8 years	Baseline monitoring period, constant.
n_A	6 and 12 years	Post-treatment experimental period over which the costs of experimenting are accumulated.
n_w	10	Number of candidate watersheds to which the aggressive treatment is applied when the inference is that it is better than passive treatment.
T	20 years	Management period (Post-treatment + Post-Inference)
$T\text{-}n_A$	Calculated	Duration of the post-inference period over which the costs of outcomes are accumulated.
α	0.05, 0.1, 0.2	Level of statistical significance, i.e. probability of Type I error
β	Function of s^2, Δ, α, and n_A	Probability of Type II error
$1-\beta$		Statistical power
Δ	Instant: Δ =1.37 Trend: At $n_A = 6$, $\Delta = 1.7$ At $n_A = 12$, $\Delta = 3$	Effect size; natural-log of the net increase in egg-to-parr survival rate between the aggressively and passively treated streams. Implemented either as an instantaneous net 4-fold increase, or as a time-trend of a 4-fold increase over 5 years.
$\overline{D}_{\mathbf{B}}$	-2.71	Mean of paired differences in Ln(P/S) over the baseline period for BVC and SCR.
s_B^2	1.34	Variance of the paired differences in the baseline period.
$\overline{D}_A$	$\overline{D}_{\mathbf{B}}$ +Ln(Δ)	Mean of paired differences in Ln(P/S) over the experimental period.
s_A^2	Low \$: $s_A^2 = s_B^2$ High \$: CV = 15%	Variance of paired differences over the experimental period. Equal to baseline variance under lower-cost monitoring (Low \$). Coefficient of variation (CV) maintained at 15% for higher-cost monitoring (High \$).
P_{Ho}	0.50	Probability of the null hypothesis (varied from 0 to 1 in sensitivity analyses)
r	5%	Discount rate (%) (varied from 0% to 10% in sensitivity analysis)

A priori statistical power analysis

I used a priori power analysis to calculate the statistical power of different experimental designs to detect some specified effect size using the analytical methods of Cohen (1988):

(6) $$\hat{a} = Z_{1-b} = \frac{d(n_{Total} - 1)\sqrt{2n_{Total}}}{2(n_{Total} - 1) + 1.21(Z_{1-a} - 1.06)} - Z_{1-a}$$

where β is the probability of Type II error, Z_{1-b} is the percentile of the unit normal curve which estimates power, Z_{1-a} is the percentile of the unit normal curve for the significance criterion (for two-tail tests a = α/2), d is the Standardized Effect Size (derived below), and n_{Total} is sample size ($n_{Total} = n_B + n_A$, where n_B is the baseline number of years of monitoring). Statistical power is 1-β. I included modifications recommended by Cohen (1988) to account for unequal sample sizes between the Before and After periods, a necessary requirement because the length of the post-treatment period (n_A) varies between experimental designs. I also accounted for unequal population variance between the $D_{i,j}$ in the Before and After periods. Statistical power is a function of four components: (1) the level of statistical significance (α), (2) post-treatment sample size (n_A) (3) the effect size of interest (Δ), and (4) sample variance (s^2).

Level of statistical significance (a)

The level of statistical significance (α) is commonly set to 0.05. However, this is often too stringent for environmental monitoring (Mapstone 1995), especially when monitoring low-abundance stocks at higher risk of extinction from even small further declines (e.g., Rhodes et al 1994). I therefore also set α to 0.1 and 0.2.

Sample size (n_A)

I considered two specific base-case post-treatment periods (n_A) of 6 and 12 years. The baseline (pre-treatment) sample size (n_B) remains constant at 8 years because it is based on pre-existing data (Table 1). Six years provides a reasonable number of parr/spawner data points while twelve years would provide the same number of data points for R/S should further ancillary analyses be required. These periods are in the range of the 3-, 5-, and 8-year “check-ins” proposed by NMFS (2000) for evaluation of the effectiveness of proposed “Reasonable and Prudent Alternatives” to dam breaching for recovering ESA-listed salmon stocks.

Effect size (D)

The effect size is related to the change in the mean of the paired difference in pre-treatment Ln(P/S), $\overline{D}_B$,that is important to detect from a decision-maker’s point of view, or based on other judgments. Connecting a change in $\overline{D}_B$ to an effect size of biological interest required three steps: (1) determining the change in the egg-to-parr survival rate (Δ) that is important to detect from a recovery perspective, (2) expressing it as a change in $\overline{D}_B$, and (3) expressing this change as the standardized effect size (d) used to calculate statistical power.

Step1: Biological Effect Size (Δ)

I assumed a net 4-fold increase in survival for the biological effect size (a multiplicative increase (Δ) in the egg-to-parr survival rate), based on the following considerations. First, I estimated the maximum possible increase in the survival rate as the difference between the estimated egg-to-parr survival rate for an impacted stream (i.e., 1.2% in Bear Valley/Elk Creek) and a pristine stream (i.e., 21.8% in Marsh Creek), an approximate 17-fold increase (Table 1, I.J. Parnell, unpublished data). Second, I estimated the expected magnitude and timing of an increase in survival under passive treatment. Rhodes et al. (1994) cite an observed 10-fold increase in survival-to-emergence, an index of egg-to-fry survival rate, over 15 years under a passive form of restoration (cessation of logging) after a massive sediment influx to the South Fork Salmon River, Idaho. A 10-fold increase from the estimated current egg-to-parr survival rate in Bear Valley/Elk Creek is about 12.3%. To justify wider application of the aggressive treatment, the increase in survival for the aggressively treated stream must be higher than that achieved under passive treatment. I assumed managers would want to achieve close to the 17-fold maximum increase in the egg-to-parr survival rate, or a net 4.8-fold increase in survival relative to the passively treated stream over 15 years. I assumed that a slightly more conservative net 4-fold increase would be satisfactory.

I applied this biological effect size using two different scenarios. For base-case conditions, I assumed that the 4-fold increase was instantaneous and constant across all n_A. I then tested the sensitivity of base-case results to a more realistic but slower trend of a net 4-fold increase over 15 years in the egg-to-parr survival rate of the aggressively treated stream. Under this approach there was a

1.7-fold increase in egg-to-parr survival rate at n_A = 6 and a 3-fold increase at n_A = 12.

Step 2: BACIP effect size (Ln(Δ) = $\overline{D}_B$ - $\overline{D}_A$)

I converted increases in the egg-to-parr survival rate to changes in the BACIP statistic as follows:

$$(7) \qquad \dot{D} = \overline{D_B} - \overline{D_A} = \overline{D_B} - \left[\overline{D_B} + \mathrm{Ln}(\Delta)\right] = \mathrm{Ln}(\Delta)$$

where $\dot{D}$ is the average difference between periods, $\overline{D}_B$ is estimated from the data and Δ is the multiplicative change in the egg-to-parr survival rate derived above.

Step 3: Standardized Effect Size (d)

The standardized effect size, d (Cohen 1988), is,

$$(8) \qquad d = \frac{|\dot{D}|}{s_{pooled}}$$

where $\dot{D}$ is the BACIP statistic estimated using equation 7 and s_{pooled} is the pooled standard deviation estimated by equation 9.

$$(9) \qquad s_{pooled} = \sqrt{\frac{s_B^2 \times (n_B - 1) + s_A^2 \times (n_A - 1)}{n_A + n_B - 2}}$$

s^2 and n are the sample variance and sample size for $\overline{D}_B$ and $\overline{D}_A$ as indicated by subscripts. s^2 for $\overline{D}_B$ is estimated from the baseline data and s^2 for $\overline{D}_A$ is estimated as described next.

Sample variance (s^2)

I estimated the base-case variance of $\overline{D}_B$ for the Before period (s_B^2) from the paired-differences in Ln(P/S) between the treatment and control streams (Table 1). The variance in $\overline{D}_A$ for the After period (s_A^2) was either equal to s_B^2 under lower-cost monitoring, or adjusted downward to maintain the Before period CV of 15% (Table 1, bottom box) under higher-cost monitoring. I assumed that the correlation between the two streams did not change in the After period and that the higher cost monitoring program reduced s_A^2 by reducing the measurement error component of equation 1.

Decision analysis

Formal decision analysis has eight basic components (e.g., Peterman and Anderson 1999). The decision tree (Figure 1) is a graphical summary of the decision framework.

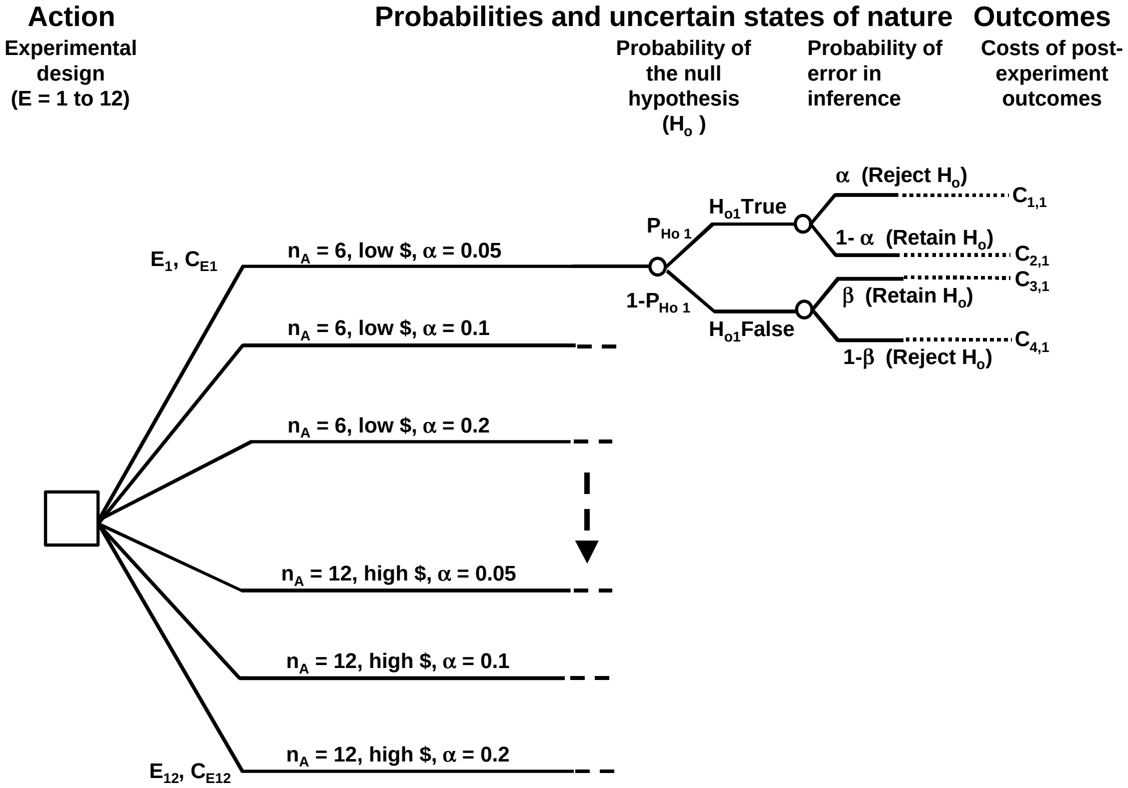

Figure 1. Decision tree for calculating the expected total costs of alternative management actions (i.e. BACIP experiments) for determining whether an aggressive habitat restoration action is better than a passive action. Management actions emerging from the (square) decision node are alternative experimental designs; each action is a single combination of post-treatment monitoring period (n_A), monitoring method (high \$ or low \$), and level of statistical significance (α) and is associated with a specific cost, C_E. Uncertain states of nature (emerging from the circles) are (1) either the null hypotheses is true that there is no difference between aggressive and passive actions, or the null is false and the alternative hypothesis is true that the aggressive action is better, and (2) the errors in inference associated with the null (Type I error) or the alternative (Type II error) hypothesis, or the correct inferences. Probabilities for each of these alternative states of nature are symbolized, respectively, as P_{Ho}, $1-P_{Ho}$, α, β, $1-\alpha$, and $1-\beta$. The C symbols at the terminal branches of the decision tree (far right) represent the total post-experiment cost associated with each management action and state of nature. The dashed lines plus arrow indicate that some designs are not shown, but Table 4 lists the full set.

The management objective: The primary objective for the decision analysis is to select the experimental design that minimizes social costs, as indexed by the expected total cost of an experimental design. Because this objective may lead to a different chosen action than a statistically based selection of experimental designs, I also examined an alternative objective to choose the design that most quickly achieves an acceptable statistical power (i.e. ≥ 0.8) to detect treatment effects.

The alternative management actions: There are twelve alternative management actions (i.e. BACIP experimental designs) representing alternative combinations of sample size (years of post-treatment sampling), monitoring methods (low or high cost), and level of statistical significance (α) (Table 4). The alternative management actions are represented by the lines leaving the square node in Figure 1.

Table 4. Alternative management actions (BACIP experimental designs) considered in this analysis. Sampling cost is the cost of the annual monitoring program used to estimate Ln(P/S), α is the level of statistical significance, and n_A is the number of years of post-treatment monitoring.

Design Number	**Sampling cost**	α	**Sample size (n_A)**
1	Low $	0.05	6
2	Low $	0.1	6
3	Low $	0.2	6
4	High $	0.05	6
5	High $	0.1	6
6	High $	0.2	6
7	Low $	0.05	12
8	Low $	0.1	12
9	Low $	0.2	12
10	High $	0.05	12
11	High $	0.1	12
12	High $	0.2	12

The uncertain states of nature: This analysis is set in the context of classical statistics hypothesis tests so the uncertain states of nature are (1) the null hypothesis of no difference in D between the Before and After periods, or the alternative hypothesis that there is a difference; and (2) a Type I or II error in inference, or a correct inference was drawn. There are four possible states represented by the lines leaving the circles in Figure 1.

The probability of each state of nature: The probabilities for the two possible states of the first category of uncertainty (P_{Ho} and $1-P_{Ho}$) are not known prior to the experiment; therefore, I set P_{Ho} to 0.5 for base-case runs and varied it in sensitivity analyses. There are four states for the second category of uncertainty (Figure 1, Table 5). If the null hypothesis is the true state of nature, the outcome of the statistical test will be either a Type I error in inference (committed with probability equal to α) or the correct inference (with probability equal to$1-\alpha$) that there is no difference between the aggressive and passive treatment. When the alternative hypothesis is the true state of nature, the outcome of the statistical test will be either a Type II error in inference (committed with probability equal to β), or the correct inference (with probability equal to $1-\beta$, or statistical power) that the aggressive treatment is better than the passive treatment. I set α to 0.05, 0.1 and 0.2 and calculated β (eq. 6) for each case.

Table 5. Four possible outcomes for classical inference. Adapted from Peterman 1990.

States of nature	**Decision**	
	Reject H_o	**Retain H_o**
H_o True	Type I error (α) Cost = $C_{1,i}$	Correct ($1-\alpha$) Cost = $C_{2,i}$
H_o False	Correct ($1-\beta$) power Cost = $C_{4,i}$	Type II error (β) Cost = $C_{3,i}$

The model of outcomes expresses results in terms of the expected total cost of an experimental design for society, $E(C_T)$:

(10) $$E(C_{T,i}) = C_{E,i} + E(C_{O,i})$$

where $C_{E,i}$ is the cost of experimental design i (eq. 4) and $E(C_{O,i})$ is the weighted sum of the four possible costs of outcomes, each weighted by its probability of occurrence.

(11) $$E(C_{O,i}) = P_{Ho} \times \left[\alpha \times C_{1,i} + (1-\alpha) \times C_{2,i}\right] + (1 - P_{Ho}) \times \left[\beta \times C_{3,i} + (1-\beta) \times C_{4,i}\right]$$

If the null hypothesis is the true state of nature (i.e. the aggressive action is not better than the passive action), then the costs of outcomes are either $C_{1,i}$ (Type I error) or $C_{2,i}$ (correct) (Table 5). $C_{1,i}$ included the unjustified costs of expanding the aggressive treatment to more watersheds. I assumed that both outcomes led to stock extinction because expanded treatment under a Type I error is ineffective and passive actions continued under the correct decision are too slow to stem declines in population. Therefore, $C_{1,i}$ and $C_{2,i}$ also included the cost of extinction as an index of the lost existence value of the fish. However, it is possible that the correct decision here would result in the initiation of research actions that might be successful in preventing stock extinction, which would reduce $C_{2,i}$.

If the alternative hypothesis is the true state of nature (i.e. the aggressive action works better than the passive action), the costs of outcomes will be either $C_{3,i}$ (Type II error) or $C_{4,I}$ (correct, power) (Table 5). $C_{3,i}$ also included the lost existence value of the fish (cost of extinction) for the same reason as for $C_{1,I}$, as well as the cost of implementing recovery actions that were unsuccessful (were too late) when the Type II error was discovered. $C_{4,i}$ included the cost of justifiably expanding the aggressive treatment to candidate watersheds. Although not considered in this analysis, the correct decision here is also associated with potential future benefits such as revenue from future salmon harvests.

I used representative costs to estimate the order-of-magnitude for $C_{1,i}$ to $C_{4,i}$ in the context of the ESA-listed stocks whose data I used for this analysis (Table 6). I calculated treatment expansion costs using a modified version of the experimental cost model (eq. 4).

$$(12) \qquad C_{Expand} = \left(n_w \times C_{\text{imp}} \times l^{T-n_A}\right) + \sum_{t=n_A+1}^{T} \left[\left\{n_w \times \left(m + C_{\text{maint}}\right)\right\}_{n_A} \times l^{t}\right]$$

where n_W is the number of candidate watersheds that treatment is applied to (10 for this analysis), C_{imp} is as for eq. 4 except that is does not include annual project management costs, *m* is the annual cost of monitoring each watershed, T is the duration of the management period (20 years for this analysis) (Figure 2), $T-n_A$ is the duration of the post-experimental period, C_{maint} is as for eq. 4, and λ is the discount factor. The number of candidate watersheds (n_W), or universe of inference (Walters and Green 1997) are all those watersheds in the region of the Middle Fork Salmon River where sedimentation has been identified as a production constraint for spring chinook that could conceivably be addressed through aggressive restoration actions. During Columbia Basin system planning, sedimentation was identified as a production constraint for spring chinook in 34 subbasins of the Salmon River watershed (IDFG 1990), of which the Middle Fork Salmon is a tributary. Many of these watersheds are nested within others also listed as being sediment impacted and so could likely be treated at the same time, therefore 10 seems a reasonable base-case value for n_W. A management period (T) of 20 years falls in the range of management periods that have been considered by other researchers exploring the optimal design of salmon enhancement experiments (e.g., MacGregor et al. 2002, 15-20 years; Keeley and Walters 1994, 30 years). I assumed managers would continue to allocate funds to monitoring of all treated watersheds because they are aware of the potential for errors in inference. I assumed no further periodic analysis costs, though such analysis would be required to process monitoring data and evaluate stock status. I did not consider possible economies of scale that might reduce implementation and monitoring costs over a larger number of watersheds.

The cost of implementing ineffective alternative actions upon detecting a Type II error was incurred as a discounted lump sum at the end of the management period.

$$C_{fix} = C_{newprogram} \times \lambda^{T} \tag{13}$$

$C_{newprogram}$ was similar in magnitude to that of the existing recovery programs, (e.g., the BPA 2001 budget noted above), but set slightly higher ($50,000,000) because I assumed that the urgency of trying to reverse the Type II error for an ESA-listed stock would justify massive spending.

I represented the cost of extinction of the fish as the discounted lump sum of the annual sunk cost of recovery actions over the duration of the management period (T).

(14) $$C_{extinction} = T \times C_{sunk} \times \lambda^{T}$$

I estimated C_{sunk} as the FY 2001 Fish and Wildlife program budget for the Salmon River region ($ 31,387,793) (BPA 2001). Only about 22% of the full annual budget ($ 6,976,744) is specifically allocated to habitat restoration actions, the rest goes to other recovery related research, however, I assumed the full level of funding was an index of the value of these stocks to society, so it served as a useful proxy for the intangible costs such as the existence value of the fish. These costs can be very high for endangered species (e.g., Osler et al. 1991). Because these cost are hard to quantify, I did sensitivity analyses on the relative magnitude of the costs of Type I and Type II errors, and the relative magnitude of experimental costs ($C_{E,i}$) and the costs of outcomes ($E(C_{o,i})$).

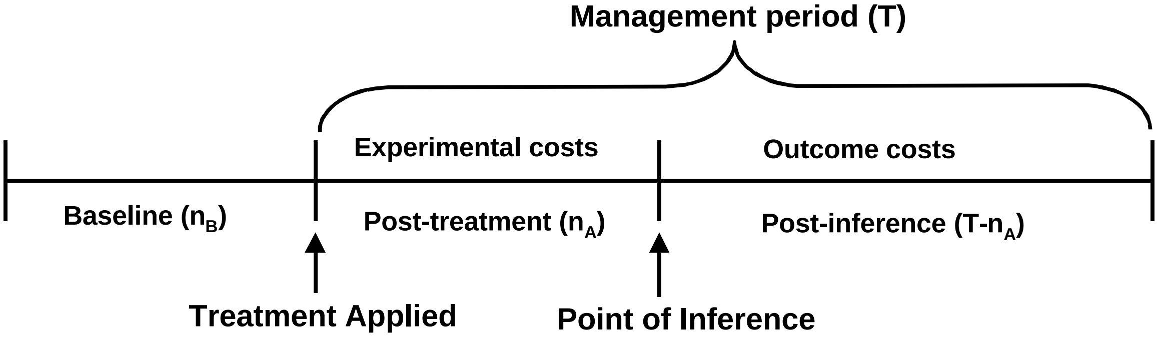

Figure 2. Time line for calculation of costs. $E(C_T)$ includes costs only over the experimental management period during which active experimentation and analysis occurs. The management period (T) includes the Post-treatment period during which the costs experimenting are accumulated and the Post-inference period during which the costs of outcomes are accumulated. The costs of experimenting include the cost of treatment, monitoring, and maintenance over the experimental period and the cost of analysis at the Point of Inference. The costs of outcomes include the cost of expanding treatment to other watersheds if it is deemed successful, the correction of Type II errors, and the cost of extinction. T is the duration in years of the experimental management period, n_B is the duration of the baseline period, n_A is the duration of the post-treatment period and T-n_A is the duration of the post-inference period.

Table 6. The costs of outcomes associated with post-inference decisions.

State of Nature 1	State of Nature 2 (inferred)	Cost	$ value
Null Hypothesis True (aggressive treatment no better than passive treatment)	Type I error(C_{O1})	Expansion of aggressive treatment to candidate streams (C_{expand}) +	Costs outlined in Table 1 accumulated over the post-inference period (n_P) and multiplied by the number of treated systems (ten).
		Existence value of fish ($C_{exticntion}$)	$ 31,387,793*T, incurred once at the end of the management period (T).
	Correct (C_{O2})	Existence value of fish ($C_{extinction}$)	As above
Null Hypothesis False (aggressive treatment better than passive treatment)	Type II error (C_{O3})	Unsuccessful recovery actions implemented to correct Type II error. (C_{fix}) +	$50,000,000, incurred once at the end of the management period (T).
		Existence value of fish ($C_{extinction}$)	As above
	Correct (C_{O4})	Expansion of aggressive restoration treatment to candidate streams (C_{expand})	As above

<u>Ranking alternative management actions:</u> I ranked the experimental designs according to the primary objective of minimizing the expected total cost of the experiment ($E(C_T)$) and also according to the alternative objective of most quickly achieving an acceptably high statistical power (i.e. >= 0.8).

<u>Sensitivity analyses:</u> I performed sensitivity analyses to explore how robust the optimal base-case design choice (i.e. recommended experimental design) was to the following uncertainties: the temporal response of egg-to-parr survival

rate in the aggressively treated stream (instantaneous vs. time-trended), the probability of the null hypothesis (P_{Ho}), the discount rate (*r*), the relative magnitude of the costs of Type I and Type II errors, and the relative magnitude of the cost of experimenting and the costs of outcomes.

Results and Discussion

Base-case Results

For the primary objective of minimizing the expected total cost to society, the optimal design was #12, which was composed of 12 years of post-treatment monitoring, the higher-cost monitoring program, and a level of significance of 0.2 (n_A=12, High \$, α = 0.2) (Table 7). For the secondary objective of most quickly achieving acceptable statistical power (>= 0.80), the optimal design was #6 (n_A=6, High \$, α= 0.2) (Table 7, shaded row). There was a tradeoff between the primary and secondary objectives in terms of time (12 vs. 6 years), $E(C_T)$ (\$121.41 vs. \$130.40 million) and statistical power (1.00 vs. 0.94). The secondary objective experiment cost more because even though it was only 6 years long, it had lower power (higher β) and the high Type II error costs thus contributed more to $E(C_T)$ (eq. 11). It cost \$121 million to treat one stream because $E(C_T)$ was composed of more than just monitoring and restoration costs, it also included the cost of extinction of the fish, which was very high (Table 6). Although higher-power experiments removed this cost from the Type II component of $E(C_T)$, it did not affect the probability of incurring it under P_{Ho}, where it was incurred for both a Type I error and a correct decision.

Table 7. Base-case ranking of experimental designs for the primary objective of minimizing $E(C_T)$. The shaded row indicates the design selected under the secondary objective, i.e. the shortest experimental period that achieves statistical power ≥ 0.8. "#" is the design number in Table 2. "Rank" indicates the rank of the experimental design under the primary ranking criterion of minimum expected total cost to society. "Monitoring $" refers to higher (High $) or lower (Low $) cost monitoring programs. "α" is the level of statistical significance used for hypothesis tests. "n_A" is the number of post-treatment years, or the experimental period. "$E(C_T)$" is the expected total cost to society of an experimental design. "Power" is statistical power.

#	Rank	Monitoring $	α	n_A	$E(C_T)$ $ x 10^6	Power
12	1	High $	0.2	12	121.41	1.00
11	2	High $	0.1	12	122.29	0.99
10	3	High $	0.05	12	124.52	0.97
6	4	High $	0.2	6	130.40	0.94
9	5	Low $	0.2	12	133.26	0.90
5	6	High $	0.1	6	140.94	0.86
8	7	Low $	0.1	12	145.49	0.81
3	8	Low $	0.2	6	146.63	0.81
4	9	High $	0.05	6	155.39	0.75
7	10	Low $	0.05	12	160.51	0.69
2	11	Low $	0.1	6	165.31	0.67
1	12	Low $	0.05	6	184.76	0.53

Sensitivity of the optimal design to a trend in effect size

I calculated base-case results using an unrealistic assumption of an immediate net 4-fold increase in the egg-to-parr survival rate in the treated stream for both n_A (6 and 12 years). To explore the sensitivity of the optimal design choice to this assumption, I calculated $E(C_T)$ and statistical power for the case where the increase in egg-to-parr survival rate followed a trend over time (Table 8). While there were some differences in the rank order of these results for the lower ranked base-case designs, design 12 was still optimal. With the time-trend in increasing egg-to-parr survival rate, no design met the secondary objective, but the highest power was achieved by the best design for the primary objective (design 12). Thus, the optimal decision was insensitive to a trend in effect size for the primary objective, but sensitive for the secondary objective.

The overall costs of the experiments were higher for a trend because power was generally lower, so the high cost of a making a Type II error (C_3) contributed more to $E(C_T)$. The contribution of the costs of outcomes associated with a true null hypothesis (Type I error, C_1 and correct, C_2) did not change. Power was lower with a trend in productivity because the mean post-treatment difference in Ln(P/S) ($\overline{D}_A$) was calculated around an increasing trend from lower to higher values of $D_{i,A}$, rather than around a constant mean difference in the egg-to-parr survival rate, as for the base-case. Thus, mean D_A under a trend was lower than under the instant-increase (base-case) scenario for a given experimental period (n_A). A lower $\overline{D}_A$ implies a lower effect size (Δ) and thus lower power. For this analysis, without a trend in effect size, the post-treatment differences reflected a constant net 4-fold increase in the egg-to-parr survival rate for both $n_A = 6$ and 12. However, with the trend in effect size, there was only a net 1.4-fold increase at $n_A = 6$ and a net 1.8-fold increase at $n_A = 12$, even though the actual final increase achieved was 1.7-fold and 3-fold for $n_A = 6$ and 12 respectively.

Table 8. Ranking of experimental designs for a trend in effect size. No design met the secondary objective of having the shortest experimental period where statistical power meets or exceeds 0.8. “Base-case Rank” indicates the rank of the design for under base-case conditions (Table 7). Other column headings are the same as in Table 7.

#	Base-case Rank	Monitoring $	α	n_A	$E(C_T)$ \$ x 10^6	Power
12	1	High $	0.2	12	165.24	0.66
11	2	High $	0.1	12	185.63	0.51
9	5	Low $	0.2	12	194.52	0.44
10	3	High $	0.05	12	203.49	0.37
8	7	Low $	0.1	12	213.47	0.29
6	4	High $	0.2	6	221.33	0.26
3	8	Low $	0.2	6	226.71	0.22
7	10	Low $	0.05	12	227.01	0.19
5	6	High $	0.1	6	235.72	0.15
2	11	Low $	0.1	6	239.39	0.12
4	9	High $	0.05	6	244.41	0.09
1	12	Low $	0.05	6	246.73	0.07

Sensitivity of the optimal design to the assumed P_{Ho}

Varying the probability of the null hypothesis (P_{Ho}) shifts the relative contribution of the Type I and Type II cost components to expected total cost (eq. 11), but does not affect statistical power, which remains constant for a given design. I explored the sensitivity of the optimal design choice under the base-case parameter set to P_{Ho} by varying it between 0 and 1.0, which corresponded to the expectation that the aggressive action was certain to work better than the passive action to zero probability that it would. At the extremes of this range, the composition of the expected outcome cost component of $E(C_T)$ shifted from either all Type I ($P_{HO} = 1$) to all Type II ($P_{HO} = 0$) related costs. I evaluated sensitivity by looking at how these different values of P_{Ho} affected the ranking of the four top-ranked base-case designs (Figure 3). Design 12 remained optimal over a wide range in P_{Ho}, but lower power designs became optimal for $P_{Ho} \geq 0.9$; as P_{Ho} increased, first design 11 and then design 10 became optimal.

$E(C_T)$ for all four designs decreased and diverged from one another as the certainty that habitat treatment works increased (P_{Ho} approaches 0). $E(C_T)$ decreased because the contribution of the costs associated with a true null hypothesis decreased. The costs of the different designs diverged as certainty increased, with higher power designs having the lowest costs. This occurred because their lower Type II error probabilities weighted Type II error costs less and thus contributed the least to $E(C_T)$. $E(C_T)$ increased and converged as the certainty that habitat treatment worked decreased (P_{Ho} approached 1) because the costs associated with the true null hypothesis made up an increasing proportion of $E(C_T)$ (Figure 1). Under very low certainty ($P_{Ho} > 0.9$), low-α designs became optimal because they gave the least weight to Type I error costs and thus contributed the least to $E(C_T)$.

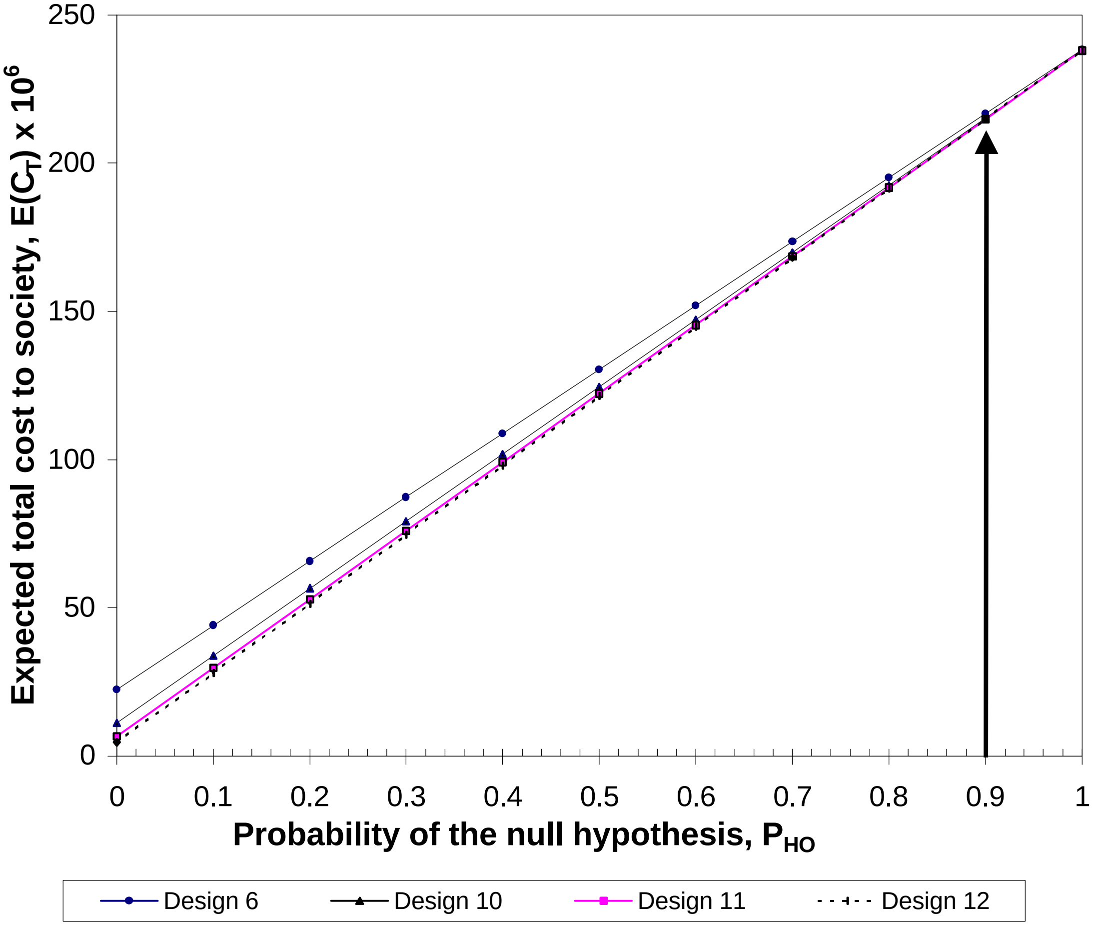

Figure 3. Sensitivity of the optimal design choice to the probability of the null hypothesis (P_{Ho}). P_{Ho} is the probability that there is no difference in effectiveness between aggressive and passive actions (P_{Ho}). The lines represent the four top-ranked base-case designs (Table 7). At any value of P_{Ho}, the design with the lowest value of $E(C_T)$ is selected under the primary objective of minimizing the expected total cost to society. The arrow indicates the approximate value of P_{Ho} at which the optimal design switched from design 12 to the lower-power design 11.

Sensitivity of the optimal design to the assumed discount rate (*r*)

I explored the sensitivity of the ranking of the four top-ranked base-case designs (Table 7) to discount rates (*r*) that ranged from 0 to 10% (Figure 4). Higher values of *r* imply that future conditions are less important for decision-making in the present, while lower values imply that future conditions are relatively more important. $E(C_T)$ declined as *r* increased, but design 12 remained optimal across all values of *r* explored here. Thus, there was no tradeoff between present and future interests for this range in *r*. There was very little difference in $E(C_T)$ for designs 10, 11 and 12 because they all have the same costs of outcomes (Table 9). $E(C_T)$ for design 6 was always highest because it had higher costs of outcomes (Table 9). The difference in $E(C_T)$ between all four designs generally narrowed as *r* increased because when higher value was placed on the present value of costs, the benefits of higher power experiments (in terms of lower expected cost) were less able to offset the large costs incurred for outcomes associated with a true null hypothesis (Figure 1, cost of a Type I error, C_1, and a correct inference, C_2).

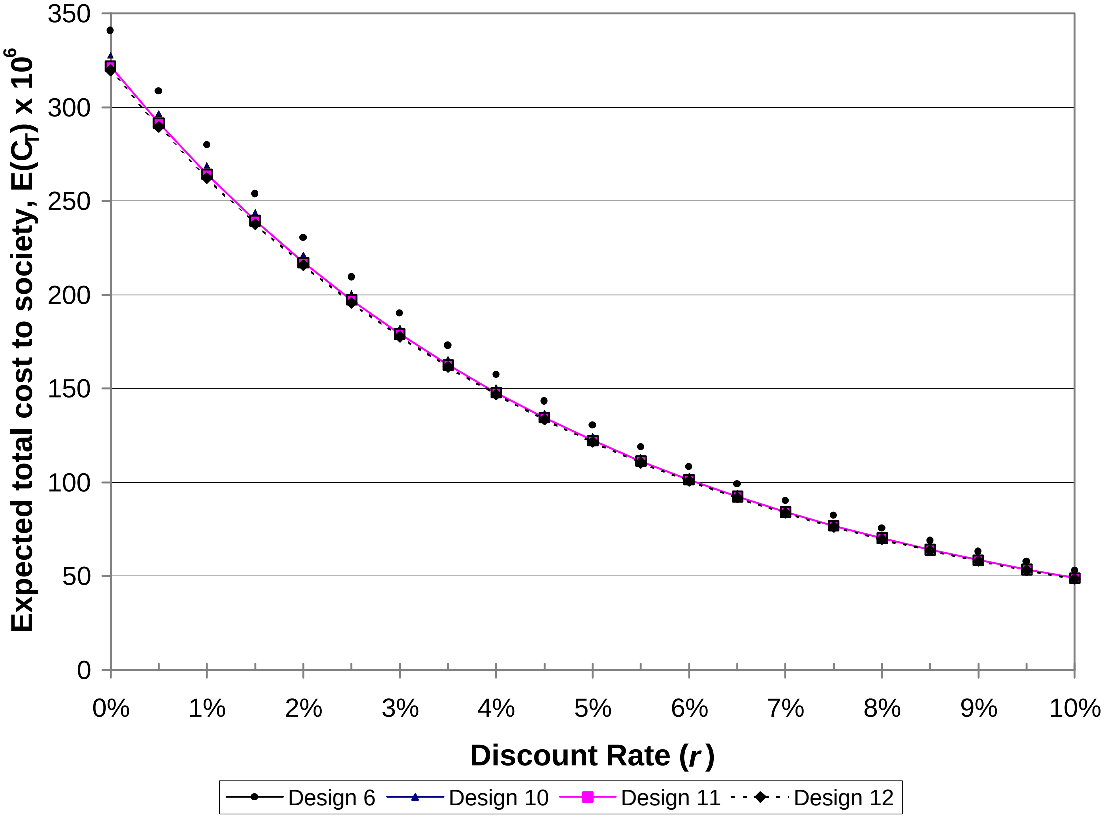

Figure 4. Sensitivity of the optimal design choice to the assumed discount rate (*r*).The lines represent the four top-ranked base-case designs (Table 7). At any value of *r*, the design with the lowest value of E(C_T) is selected under the primary objective of minimizing the expected total cost to society. The optimal design choice does not change over the range in *r* explored here.

Sensitivity of the optimal design to the ratio of Type II and Type I costs

I estimated the costs of outcomes for this analysis (Table 6), but the actual costs associated with a Type II error are difficult to determine, especially when they cannot be easily be quantified in terms of dollars alone, as is the case for ESA-listed species (e.g., Loomis and White 1996). To assess how uncertainty about the cost of a Type II error affected the ranking of experimental designs under the primary objective, I varied the relative magnitude of the costs of Type II and Type I errors. Since the costs of outcomes changed with respect to n_A (Table 9), I kept the costs of a Type I error constant for a particular experimental period and varied the costs of a Type II error around them from 0.05 to 10 times their magnitude. Design 12 remained optimal over most of this range, but the lower-power design 11 ($n_A = 12$, High \$, $\alpha = 0.1$) became optimal at a ratio of the Type II to Type I costs of about 0.15 (Figure 5). This occurred because below a ratio of 0.15, the benefits of the higher-power design 12 ($n_A = 12$, High \$, $\alpha = 0.2$), in terms of its higher power reducing Type II costs, no longer offset the higher costs of a Type I error incurred through its higher probability of a Type I error relative to the lower-power design 11 ($n_A = 12$, High \$, $\alpha = 0.1$). That is, a lower probability of Type I error (α) became more important than a lower probability of making a Type II error because it reduced the contribution of the large Type I costs to $E(C_T)$.

The slopes of the four lines in Figure 5 are quite different because for each design, only the costs of a Type II error changed. Thus the effective slope of each line became equal to $(1-P_{Ho})*(\beta)$ (eq. 11). P_{Ho} is constant for all four designs, so lower power (higher β) designs had higher slopes and thus steeper lines.

Table 9. The costs of outcomes for the base-case parameter set. Costs of outcomes and the ratio of the costs of Type II and Type I errors are shown for the two experimental periods (n_A) and the higher- and lower-cost monitoring designs for the four outcome of Table 5. Costs are in millions of dollars; the ratio of costs for Type II and Type I errors is dimensionless. The numbers prefixed by “D” are the designs associated with each combination of monitoring cost and n_A (e.g., “D12” is design 12 of Table 4).

States of Nature		**Costs of Outcomes ($\$ \times 10^6$)**	
		n_A = 6	**n_A = 12**
Higher Cost Monitoring		**D4, D5, D6**	**D10, D11, D12**
P_{Ho} True	**Reject H_o (Type I error)**	241.32	239.32
	Retain H_o (correct)	236.59	236.59
P_{Ho} False	**Retain H_o (Type II error)**	273.91	264.44
	Reject H_o (correct)	4.73	2.72
	Ratio of costs for Type II and I errors (Type II/Type I)	1.14	1.10
Lower Cost Monitoring		**D1, D2, D3**	**D7, D8, D9**
P_{Ho} True	**Reject H_o (Type I error)**	240.54	238.94
	Retain H_o (correct)	236.59	236.59
P_{Ho} False	**Retain H_o (Type II error)**	273.91	264.44
	Reject H_o (correct)	3.95	2.34
	Ratio of costs for Type II and I errors (Type II/Type I)	1.14	1.11

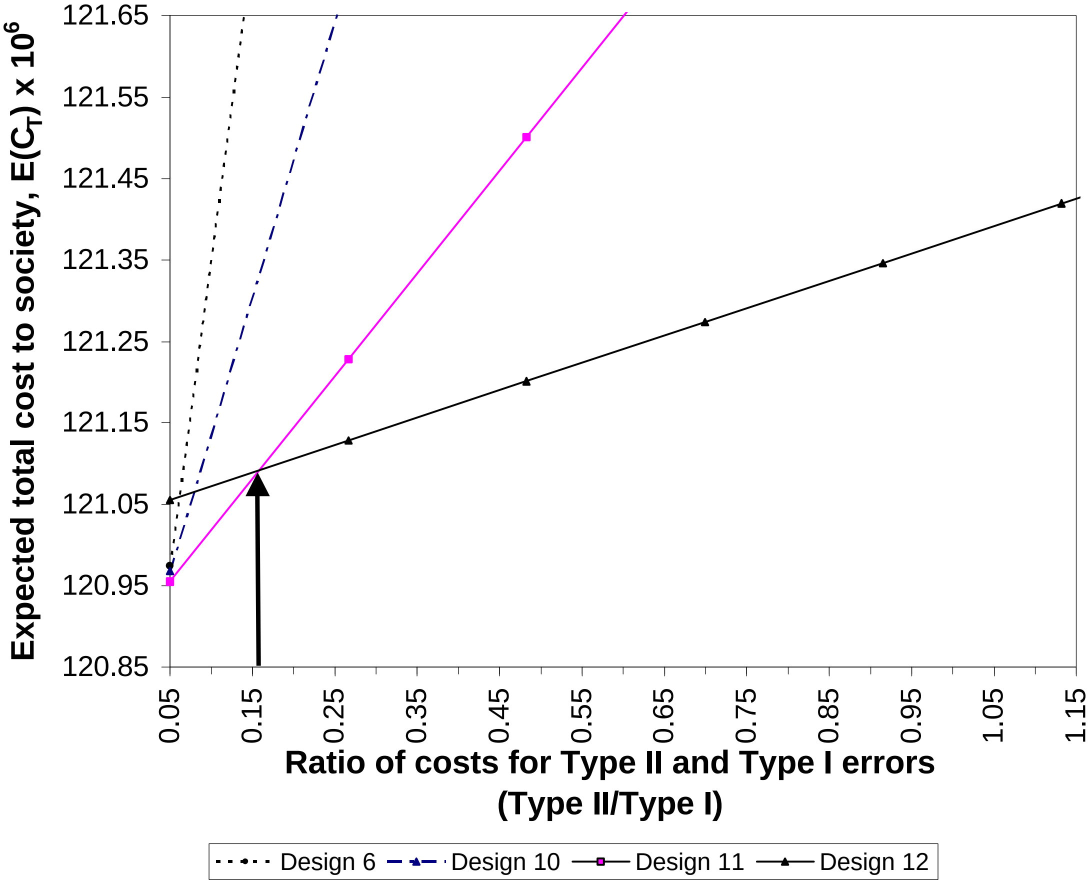

Figure 5. Sensitivity of the optimal design choice to the ratio of Type II and Type I costs. The lines represent the four top-ranked base-case designs (Table 7). The vertical arrow indicates the ratio at which the optimal design switches from the highest power design, #12, to the lower power design 11 (at a ratio of approximately 0.15).The base-case Type II/Type I cost ratio is 1.10 for designs 10-12 and 1.14 for design 6 (Table 7).

Sensitivity of the optimal design to the relative magnitude of the costs of experimenting and the costs of outcomes

For the base-case parameter set (Table 3), minimum $E(C_T)$ is reached when the number of experimental years (n_A) is highest, which also gives the highest statistical power in all cases. This is because the large costs of outcomes (Table 9) relative to the costs of experimenting (Table 10) require the highest power design to reach the minimum $E(C_T)$ for the two conditions of n_A explored here. However, if conditions were such that the costs of experimenting were a more substantial proportion of the costs of outcomes, the rank order of designs relative to the base-case results might change such that a lower power design becomes optimal. This might occur if my example costs of outcomes severely overestimate true costs of outcomes (e.g., inflated cost of extinction).

Table 10. The costs of experimenting for the base-case parameter set. Costs are shown for the two experimental periods (n_A) of 6 and 12 years for the higher and lower cost monitoring programs. These costs apply to the designs in the bottom row of the table.

	Cost of Experimenting	
Cost of Monitoring	**$n_A = 6$**	**$n_A = 12$**
Higher	$750,568	$1,111,307
Lower	$652,748	$940,493
Design #	D1-D6	D7-D12

I explored the sensitivity of the ranking of the four top-ranked base-case designs (Table 7) to cost conditions that brought the costs of experimenting (C_E) and the costs of outcomes (C_1-C_4) closer in magnitude (Figure 6). I only varied the cost of extinction component of the costs of outcomes since existence value depends on future biological conditions and social values and so is probably the most uncertain cost component. I assumed that the costs of experimenting, the costs of expansion, and the cost of correcting Type II errors were certain, though the latter cost could also depend heavily on future social values.

Design 12 remained optimal over a wide range in the cost of extinction. The second ranked design of the four compared, design 11, only became optimal when the cost of extinction was about 4.2% of its base-case value ($ 9.86 million vs. $ 236.6 million for the base-case). The only difference between designs 12 and 11 is the level of statistical significance ($\alpha = 0.2$ vs. $\alpha = 0.1$), which leads to a small difference in statistical power (power = 1 vs. power = 0.99). When the cost of extinction is at 4.2% of its base-case value, the benefits of the base-case optimal design, in terms of the lower costs incurred under a false null hypothesis (H_o False, Figure 1) relative to the lower-power design 11, no longer offset the higher costs it incurs under a true null hypothesis (Ho True, Figure 1) with its higher probability of incurring a Type I error. The lines for designs 10, 11 and 12 in Figure 5 were similar because their base-case costs of outcomes were identical (Table 9) and their statistical power was very similar (Table 7). The line for design 6 diverged from the lines for the other designs as the cost of extinction decreased because it had higher base-case costs of outcomes (Table 9), and the lowest statistical power of the four designs (Table 7).

Although this analysis shows that the optimal design could change when the cost of outcomes became less important, in practical terms, there was no real difference between designs 10, 11, and 12 below 4.2% (left of the vertical line in Figure 6). They differed by less than 3%, and it is unlikely that the existence value of these stocks could be so severely overestimated. Thus, the selection of design 12 was robust to the magnitude of the costs of outcomes relative to the costs of experimenting.

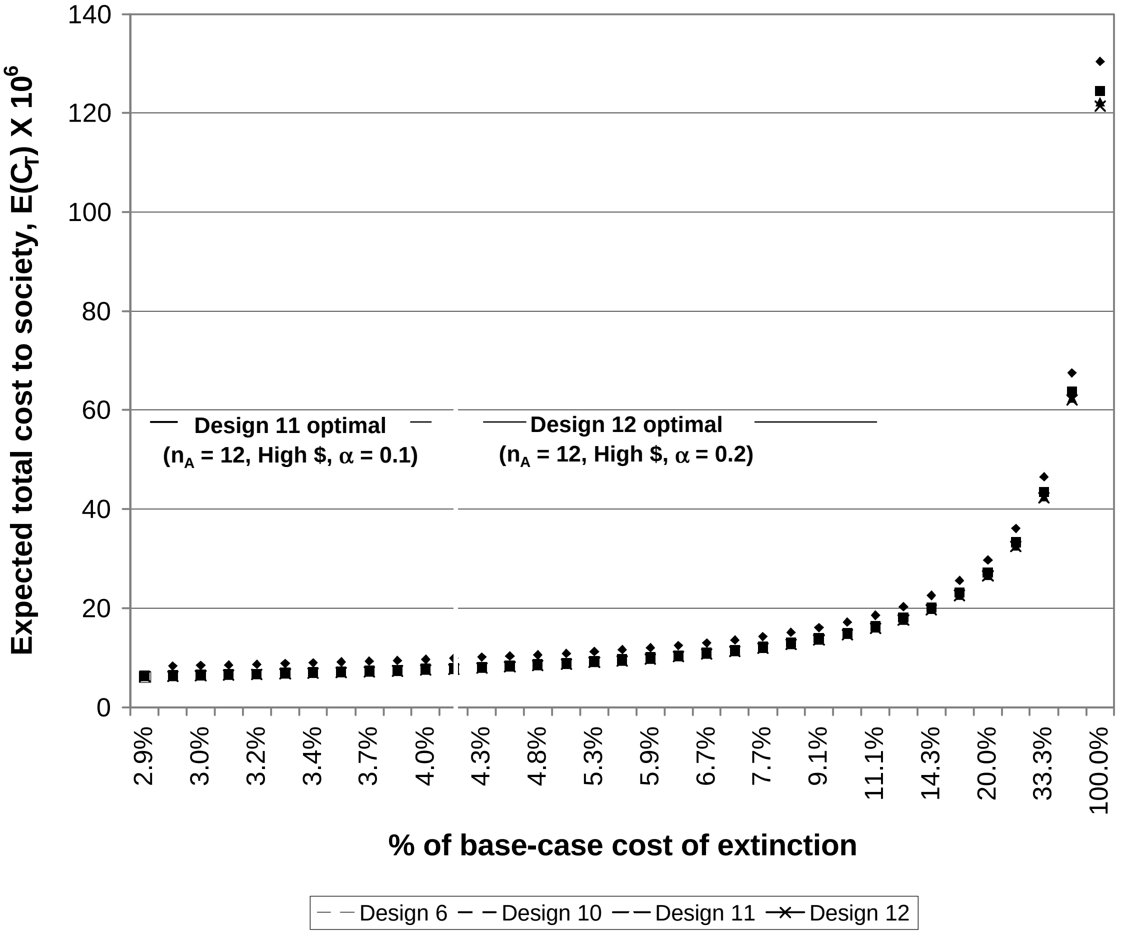

Figure 6. Sensitivity of the optimal design choice to the relative magnitude of the costs of experimenting and the costs of outcomes. The lines represent the four top-ranked base-case designs (Table 7). The costs of experimenting were held constant at their base-case values (Table 10). The costs of outcomes were varied by changing the cost of extinction. The vertical line indicates the point below which the optimal design choice switched from design 12 to design 11. At that point the cost of extinction and the expected (weighted average) costs of outcomes for the base-case optimal design were 4.2% and 5.5% of their base-case values respectively.

Sensitivity of the optimal design to the number of candidate watersheds (n_W) and the duration of the management period (T)

I initially assumed a single value for both the number of candidate watersheds (n_W) to which the aggressive treatment was applied after the experiment and the duration of the management period (T). However, the sensitivity analyses showed that the optimal design choice was driven by the high costs of outcomes, which depend on both n_W and T. Therefore, I briefly explored the sensitivity of the optimal design to these parameters.

I found that the optimal design (design 12) was robust to a wide range in n_W. Increasing n_W increased the cost of expanding treatment, and thus increased the need to avoid making a Type I error. This resulted in the lower-power (lower α) design 11 becoming optimal at $n_W = 73$, well above the maximum number of candidate watersheds (34) for the Salmon River watershed as estimated from IDFG (1990). Reducing n_W decreased the costs of a Type I error, making them less important relative to the cost of a Type II error. Thus, for $n_W < 10$, the highest power design (design 12) remained optimal.

I also found that the optimal design choice was robust to the duration of the management period (T). Increasing T increased the costs of outcomes, but also increased the influence of the discount rate (*r*), which offset those increased costs. This effect dropped the expected total cost to society $E(C_T)$ for the optimal design below its base-case value for $T > 21$ years. Within the base-case cost framework, increasing T increased the cost of a Type I error more than the cost of a Type II error and the benefits of higher power designs no longer offset the growing cost of a Type I error at $T \approx 80$ years. At that point the lower power (lower α) design 11 became optimal. The value of T where the optimal design switched to the lower power design decreased as r increased. For $r = 10\%$, the switch occurred at $T \approx 55$ years. Both of these values of T were greater than what might be considered a reasonable period for evaluating recovery actions for endangered species. Decreasing the duration of T did not affect the optimal design because r had less influence on the costs of outcomes over shorter periods. Thus, the Type II costs remained higher than Type I costs for $T<20$ years and design 12, the highest power design, remained optimal.

Sensitivity of the optimal design choice to variance in $\overline{D}_A$

I did not explicitly explore the sensitivity of the base-case optimal design to uncertainty about the level of post-treatment variance, other than to have higher and lower-variance monitoring designs. Variance is likely poorly estimated because there were few baseline data points ($n_B = 8$) and the parr data were collected from parr populations generated at low spawner abundances (i.e., density-independent egg-to-parr survival rate). As stocks recover and density dependent effects become important, variance in Ln(P/S) can be expected to change. However, the robustness of the optimal design to assumptions about effect size, α, and the higher-cost lower variance or lower-cost higher variance monitoring designs suggests that it would also be robust to increased post-treatment variance, though $E(C_T)$ would be higher because of lower power.

General Discussion

Primary and secondary objectives

The design that best met the primary management objective of minimizing expected total costs ($E(C_T)$) was design 12 (n_A = 12, High \$, α = 0.2), which also had the highest statistical power over the range of designs that I considered (Table 7). The optimal design was robust to assumptions about the temporal response of wide range of assumptions about effect size (i.e. trend vs. no trend), although the magnitude of $E(C_T)$ and power varied in those different cases (e.g., Table 7 vs. Table 8). It was also robust to the assumed discount rate (*r*), the number of candidate watersheds (n_w), the duration of the management period (T), and most likely to post-treatment variance, $\overline{D}_A$. The optimal design was less robust to assumptions about the probability of the null hypothesis (P_{Ho}), the ratio of the Type II to Type I error costs, and the relative magnitude of the cost of experimenting and the costs of outcomes, where in each case the optimal decision switched to a lower power design at some value. The switch in the optimal design choice to the next highest power design (design 11 in each case) did not affect the cost of experimenting (Table 10) because only the level of statistical significance changed between designs. This indicates that the controlling factor for the switch in each case was the pattern of change in the costs of making a Type I error relative to the costs of making a Type II error.

The design that best met the secondary objective of most quickly reaching an acceptably high statistical power (i.e. ≥ 0.8) was not robust to the assumption of a trend in effect size. Under base-case conditions, design 6 was the optimal design choice under this secondary ranking criterion (Table 7), but with a trend in effect size, no design was optimal (Table 8). However, the optimal design under the primary objective had the highest power and lowest expected total cost.

These sensitivity analyses show that the optimal design changed under certain conditions for reasons that were both logical and consistent with the decision framework. More importantly, they showed that the optimal design choice was robust over reasonable ranges for assumptions. For this example then it appears that it is worth spending more time and money to do monitoring well.

Factors worth further consideration

The sensitivity analyses highlight two factors worth further consideration. First, although the probability of the null hypothesis cannot be known prior to the experiment, if managers believe the probability of the aggressive restoration action not working better than the passive action could be as high as 0.9, it becomes important to either not experiment at all and to turn to other recovery options, or to select experimental designs that minimize the probability of making a Type I error. Given the widespread application of aggressive restoration techniques, it seems unlikely that managers would believe the probability of H_o could be as high as 0.9. A second and more important consideration is the structure for the costs of outcomes, in particular, the very influential and large cost of extinction. In the context of this analysis, it seems unlikely that the costs of Type II errors could be 15% of the costs of Type I errors when both outcomes lead to stock extinction and incur that cost. Similarly, it also seems unlikely that for endangered salmon stocks the cost of extinction could fall below 4.2% of that estimated in this analysis. However, because this cost is so influential and because it includes costs such as existence value that are difficult to estimate, it would be important to consider carefully the magnitude of the cost of extinction before the final selection of the experimental design, and how it might be incurred under different outcomes. Contingent valuation methods could be applied to estimate these existence values in terms of society's willingness to pay for recovering endangered salmon stocks (Loomis and White 1996).

Contingent valuation of the existence value of salmon has been done before in the Columbia River basin and the results provide an interesting contrast to my estimates of the cost of existence. Olsen et al. (1991) conducted an existence valuation study to estimate the willingness-to-pay and willingness-to-accept of users and non-users (existence value only and some probability of future use) for a doubling of Columbia River salmon stocks. Their estimates (in 1996 dollars) ranged from US $42,415,000 per year for existence value only to US $110,943,000 per year for users, over the whole Columbia River basin. These values imply that a lower-power design than design 12 would be optimal could. Although Olsen et al.'s (1991) estimates are for the whole Columbia River basin, their existence value estimate is already lower than the value for the cost of extinction at which the optimal design choice for this analysis switched to the lower power (lower α) design 11 (US $49 million). Additionally, the Mountain Snake region is only a small area of the Columbia basin, adjusting Olsen et al.'s (1991) existence value downward to reflect this would imply that an even lower power design than design 11 could be optimal. Thus, my cost of extinction either severely overestimates existence value, or the annual BPA expenditure I used to represent existence value is confounded with other values such as use values.

An alternative explanation for the difference between the magnitude of Olsen et al.'s (1991) existence costs and those I used for this analysis is that they reflect existence values at different scales of society. Olsen et al (1991) derived their costs by surveying residents of the Pacific Northwest. However, the BPA (2001) budget costs for the Mountain Snake region that I used to estimate the cost of extinction are driven in part by the requirements of the federal Endangered Species Act, and thus reflect the value held for endangered salmon at the broader scale of the entire population of the United States.

Tradeoffs between objectives

Within the context of the primary decision objective there was no tradeoff between social value and statistical power; the lowest cost occurred for the highest statistical power (Table 7). However, there was a tradeoff between the primary and secondary objectives. For the primary objective, social costs were minimized at $121.41 million for an experimental period of 12 years and power of close to 1.0 (Table 7). For the secondary objective, an acceptably high level of statistical power (0.94) was achieved in 6 years at an $E(C_T)$ of $130.40 million (Table 7). Although results would be achieved sooner for the secondary objective relative to the primary objective, which may be desirable when trying to evaluate recovery efforts for rapidly declining stocks, the higher probability of making a Type II error brings additional expected social costs of $8.99 million.

In fact, for the base-case, there is less than a 10% difference in $E(C_T)$ between the top and fifth ranked designs (12 and 9 respectively) a difference of \$11.85 million (Table 7). The difference in power between these designs is 0.1 (power = 1 for design 12 and 0.9 for design 9). Given the uncertainties not addressed by this analysis, the top five designs may be effectively equal with respect to the primary objective of minimizing $E(C_T)$, and other objectives that I have not considered may play a larger role in decision making.

One such objective alluded to above is minimizing the probability of extinction. Minimizing the probability of extinction is likely to be an objective for experiments that explore recovery actions for endangered species. In this analysis, I have assumed that managers are risk-neutral and base their decisions about choice of experimental design solely on the stated primary objective of choosing the experimental design that minimizes social cost. Under this assumption, the optimal design has a duration of 12-years. However, for endangered stocks, longer experiments will be associated with a higher probability of extinction. Under these circumstances, a decision-maker may be risk-seeking with respect to the primary objective, that is be willing to accept higher social costs (accept a higher probability of a Type II error) in order to get experimental results more quickly and minimize the probability of extinction. For my base-case results (Table 7), the increase in the probability of Type II error is not substantial because the optimal design under the secondary objective also has very high power (design 6, power = 0.94). Recall, however, that for a trend in effect size, no design achieved a statistical power > 0.8 (Table 8). Thus, if habitat recovery followed trend like that modeled in this analysis, the risk-seeking manager would have to accept much higher social costs, and higher probability of Type II error for shorter duration experimental periods.

This example illustrates that the results of my analysis could change if more objectives, such as minimizing the probability of extinction, were considered. Performance measures for additional objectives could be included and multi-attribute utility analysis techniques (Keeney and Raiffa 1976) used to facilitate tradeoff analyses and the elicitation of stakeholder values. This would strengthen the decision process; therefore, including more objectives would be a useful extension of this analysis.

Comparison to the experimental valuation approach of Walters and Green 1997

Walters and Green (1997) defined a valuation framework for the selection of optimal experimental designs that consisted of four general components: (1) universe of inference, (2) treatment options, (3) impact hypotheses and baseline policy option, and (4) value measures. My decision framework is really a special case of their general approach, with some important differences with respect to the use of a baseline policy, the assignment of probabilities to uncertain states of nature, and the definition of “optimal” experimental design.

Walters and Green (1997) recommend identifying the baseline management policy that would be applied in the absence of experimenting. I did not do this explicitly, but such a baseline non-experimental policy could be continuing to rely on passive habitat restoration actions to recover endangered stocks. The baseline total cost to society in this case would not be a weighted sum over uncertain outcomes, but only the discounted cost of extinction over the management period (T) ($236,594,581 for T = 20), which is $115.18 million more than the expected cost of the optimal experimental design under the primary objective. In fact, the cost of the baseline policy is larger than all 12 designs considered in my analysis (Table 7).

Walters and Green (1997) also recommend using a range of hypotheses of about the response of the experimental system to experimental actions (the effect size), each linked through models to a specific set of future biological and socio-economic benefits and costs. Thus, there can be many branches to the uncertain state of nature (effect size). Bayesian techniques would be used to calculate the posterior probability of each hypothesis and weight their outcomes in the calculation of the expected value over all hypotheses. In contrast, my analysis considered only the probability of detecting a single pre-specified effect size using classical statistical techniques for which there are only four uncertain states of nature (Figure 1). This could bias $E(C_T)$ because it does not include the costs associated with small differences from the effect size of interest. This may be especially influential when there are large differences in value for the costs of Type I and Type II errors (MacGregor et al 2002), but that is not the case here (Table 9).

Walters and Green's (1997) definition of optimal design differs from that used for this analysis. Walters and Green (1997) note that optimum designs for their valuation equation will occur where the rate of increase in long-term value with the number of experiment units and/or number of years of monitoring (the design variables) just balances the rate of loss in short-term value. Thus, they discuss a “global” optimum across those design variables for a specific parameter set. My cost function (eq. 10) is similar in structure to their valuation equation and for a single set of parameters (single experimental design) will also produce an optimum (minimum $E(C_T)$) when the rate of increase in experimental costs (C_E) balances the rate of decline in the expected (weighted average) costs of outcomes ($E(C_O)$), both rates with respect to the number of years of experimental monitoring (n_A). However, in this analysis, I only evaluated $E(C_T)$ at two points ($n_A = 6$ and 12) for six discrete design categories (combinations of α and monitoring cost). Thus, the “optimal” design in this case is only optimal with respect to this set of 12 discrete designs.

Comparison with results from of other research

My results contrast with those of others besides Walters and Green (1997). Keeley and Walters (1994) and MacGregor et al. (2002) found that optimal experimental designs can occur at levels of statistical power considerably less than 0.8. However, my sensitivity analyses showed that the optimal base-case design under the primary objective could switch from a higher-power to a lower-power design when the cost of extinction and consequently the expected (weighted average) costs of outcomes became closer in magnitude to the costs of experimenting. A switch from the base-case optimal design 12 to the lower-power design 11 occurred when the cost of extinction was roughly 4.2% of its base-case value (Figure 6). At this point the benefits of the base-case optimal design, in terms of reduced social costs relative to lower-cost lower-power design

(design 6), no longer offset the higher costs for a Type I error that it incurred under a true null hypothesis. 4.2% of the base-case cost of extinction for design 12 was $9,858,108. The expected (weighted average) total costs were about 5.5% of their base-case value ($6.61 million vs. $122.29 million). The total cost of experimenting (capital costs + monitoring costs + maintenance costs + analysis costs) was $1,111,307 (Table 10). The ratio of experimental to expected (weighted average) costs of outcomes was about 0.17, a 19-fold increase from the ratio of 0.009 for base-case conditions.

For their optimal designs, MacGregor et al. (2002) and Keeley and Walters (1994) also appear to have high ratios of experimental to expected (weighted average) costs of outcomes. Using as an example MacGregor et al.'s (2002) Scenario F with an optimal monitoring design of 9 systems for 2 years and a high-cost monitoring program (at CDN $80,000 per system per year) and per-treated system capital costs of CDN $91,525, the total costs of experimenting would be CDN $2,263,725. The expected net present value (ENPV) for Scenario F was CDN $672,560. Since ENPV includes benefits less the costs of experimenting, I assumed that a crude analogy of the weighted costs of outcomes that I use ($E(C_o)$) is the sum of the ENPV for Scenario F and its costs of experimenting (CDN $2,936,285). The ratio of the cost of experimenting and crude expected (weighted average) outcomes for Scenario F was 0.77, much higher than the ratio of 0.009 for my optimal design under base-case conditions.

Similarly, I replicated Keeley and Walters' (1994) approach and found for their base-case cost conditions an optimal design of 8 streams (4 treatment control pairs) and 4 years of monitoring. The costs of experimenting (the sum of capital costs, monitoring, and maintenance costs) for this design were CDN $18,000,000 while the ENPV (less experimental costs) was CDN $62,196,392. In this case, I was able to calculate ENPV separately from the costs of experimenting. The ratio of the costs of experimenting to ENPV was approximately 0.29, again much higher than the ratio for my optimal design of 0.009. Thus, the costs of experimenting for both MacGregor et al (2002) and Keeley and Walters (1994), make up a larger proportions of the expected (weighted average) costs of outcomes than for my base-case result (ratio of 0.009), but are similar in proportion to that for which my base case design switched to a lower power design during sensitivity analyses (0.17%). Indeed, my analysis showed that lower power designs will be optimal too as the costs of the experimenting begin to make up a larger proportion of the expected (weighted average) costs of outcomes.

These conditions would be more likely to occur over the experimental periods I considered in this analysis for net-value models that consider both benefits and costs. This is because the benefits that accrue under the different outcomes will help offset their costs and reduce the overall magnitude of the expected value of outcomes relative to the magnitude of the costs of the experimenting. For example, there could be future benefits from fishery openings on these populations. Such benefits would reduce the magnitude of costs when the aggressive action was better than the passive action by offsetting some of the expansion costs associated with the correct decision (power). This would increase the rate of decline in $E(C_O)$ with respect to n_A so that lower power, experiments would be more likely to be optimal within the range of experimental periods considered here (6 to12 years). Similarly, a correct decision when the aggressive action was not better than the passive action (occurring with probability $1-\alpha$) could also be associated with a delayed accumulation of benefits if future actions helped recover the stocks.

I could not use a net value (benefit minus cost) approach because it is difficult to estimate the intangible benefits the existence of endangered species hold for society. While such values can be roughly inferred for the present and past from money spent on recovery actions, future benefits are less clear. They cannot be indexed by the commercial value of the fish because there is no tribal, recreational, or commercial harvest of ESA-listed stocks. Under the current low spawner abundance conditions, it is not likely that these stocks could recover to abundances able to support harvest during the management period used for this analysis even if the aggressive habitat restoration action was successful. However, if appropriate, a benefit-cost model could more quickly bring the expected value of outcomes closer in magnitude to the cost of experimenting, making lower-power and lower-cost experiments optimal.

Tendency to overestimate the expected total cost to society

My base-case optimal result would overestimate the expected total cost of the experiment if the future cost structure changed. For example, this could happen if the value that society places on the existence of wild salmon declines in the future. I assumed a constant annual application of the full 2001 Salmon

River research budget (about $ 31 million/year, Table 6) over the management period and applied this as a cost under decisions where stocks went extinct, which implies an enormous social value associated with preserving wild salmon stocks. If this value instead declined in the future (e.g., a weakened Endangered Species Act, or a critical need for cheap electricity), the expected (weighted average) costs of outcomes would decline, bringing them closer in magnitude to the costs of experimenting. In that situation, lower power designs would more likely become optimal.

Utility of decision analysis

Decision analysis was useful for determining an optimal BACIP experimental design based on an index that incorporated both biological uncertainty and socio-economic costs. It provided a framework for exploring quantitatively the robustness of the base-case results to explicit assumptions about the components of statistical power, the costs of experimenting, and the costs of outcomes. The results of these sensitivity analyses highlighted important factors that should be considered further. This example framework could be easily adapted and applied to more complex BACIP decision problems incorporating more detailed biological and statistical models, a broader range of objectives, as well as socio-economic models with a more refined structure for the costs of experimenting and the costs of outcomes.

References

Andrews, J. 1988. Anadromous fish habitat enhancement for the Middle Fork and Upper Salmon River, Annual Report 1988. U.S. Department of Energy, Bonneville Power Administration, Division of Fish and Wildlife, Portland, Oregon. Project 84-24. 31 p.

Andrews, J. and L. Everson. 1988. Middle Fork and Upper Salmon River Habitat Improvement Implementation Plan FY 1988-1992. U.S. Department of Energy, Bonneville Power Administration, Division of Fish and Wildlife, Portland, Oregon. Project 84-24. 33 p.

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

Beamesderfer, R. C. P., H. A. Schaller, M. P. Zimmerman, C. E. Petrosky, O. P. Langness, and L. LaVoy. 1997. Spawner-recruit data for spring and summer chinook salmon populations in Idaho, Oregon and Washington. In Plan for Analyzing and Testing Hypotheses (PATH): retrospective and prospective analyses of spring/summer chinook reviewed in FY 1997. Edited by D.M. Marmorek and C.N. Peters, ESSA Technologies Ltd., Vancouver, B.C. Available from Bonneville Power Administration, Portland, Oregon (http:// www.efw.bpa.gov/Environment/PATH/reports/1997retro/toc.htm).

Bjornn, T. C. 1978. Survival, production, and yield of trout and chinook salmon in the Lemhi River, Idaho. *Project F-49-R*, University of Idaho.

Botsford, L. W., and C. M. Paulsen. 2000. Assessing covariability among populations in the presence of intraseries correlation: Columbia River spring-summer chinook salmon (*Oncorhyncus tshawytscha*) stocks. Can. J. Fish. Aquat. Sci. 57: 616-627.

Bowles, E., and E. Leitzinger. 1991. Salmon supplementation studies in Idaho rivers (Idaho Supplementation Studies). Experimental Design. U.S. Department of Energy, Bonneville Power Administration, Division of Fish and Wildlife, Project 89-098.

BPA (Bonneville Power Administration) 2002. Fish and Wildlife Program Budget Tracking Report 4th Quarter, Fiscal Year 2001 Final. (http://www.efw.bpa.gov/portal/ColumbiaBasin/FishAndWildlifeProjects/BPA/Fiscal/02Feb04Mth/rptQR2001Q4.pdf).

Bradford, M. J. 1995. Comparative review of Pacific salmon survival rates. Can. J. Fish. Aquat. Sci. 52: 1327-1338.

Clemen, R. T. 1996. Making hard decisions: an introduction to decision analysis. 2nd ed., Duxbury Press, Wadsworth Publishing Co., Belmont, Ca.

Cohen, J. 1988. Statistical power analysis for the behavioural sciences. 2nd Edition., Lawrence Erlbaum Associates., Hillsdale, N.J.

Deriso, R. B., D. R. Marmorek, and I. J. Parnell. 2001. Retrospective patterns of differential mortality and common year-effects experienced by spring and summer chinook salmon (*Oncorhynchus tshawytscha*) of the Columbia River. Can. J. Fish. Aquat. Sci. 58: 2419-2430.

Green, R. H. 1979. Sampling design and statistical methods for environmental biologists, John Wiley & Sons, Toronto.

Hall-Griswold, J. A., and C. E. Petrosky. 1996. Idaho Habitat/Natural Production Monitoring Part I Annual Report 1995.U.S. Department of Energy, Bonneville Power Administration, Environment, Fish and Wildlife, Portland, Oregon. Project 91-73

IDFG (Idaho Department of Fish and Game) 1990. Salmon River Subbasin Salmon and Steelhead production plan. September 1 1990. Columbia Basin System Planning. Lead agency: Idaho Department of Fish and Game. Co-Writers: Nez Pierce Tribe of Idaho and Shoshone-Bannock Tribes of Fort Hall. Funds provided by the Northwest Power Planning Council and the Agencies and Indian Tribes of the Columbia Basin Fish and Wildlife Authority.

Kareiva, P., M. Marvier, and McClure, M. M. 2000. Recovery and management options for spring/summer Chinook Salmon in the Columbia River Basin. Science 290: 977-979

Keeley, E. R., and C. J. Walters. 1994. The British Columbia watershed restoration program: summary of the experimental design, monitoring, and restoration techniques workshop. Watershed Restoration Management Report Number 1, British Columbia Ministry of Environment, Lands and Parks, and Ministry of Forests, Victoria, British Columbia.

Keeney, R.L. and H. Raiffa. 1976. Decisions and multiple objectives: preferences and value trade-offs. John Wiley and Sons, New York.

Korman, J., and P. S. Higgins. 1997. Utility of escapement time series data for monitoring the response of salmon populations to habitat alteration. Can. J. Fish. Aquat. Sci. 54: 2058-2067.

Loomis, J. B., and D. S. White. 1996. Economic values of increasingly rare and endangered fish. Fisheries 21: 6-11.

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

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

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

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

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

NAP (National Academy Press) 1996. Upstream: Salmon and society in the Pacific Northwest, National Academy Press, Washington, DC. 472 p.

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. NMFS, Portland, Oreg.

Olsen, D., J. Richards, and R. D. Scott. 1991. Existence and sport values for doubling the size of Columbia River Basin salmon and steelhead runs. Rivers 2: 44-56.

Peterman, R. M. 1990. Statistical power analysis can improve fisheries research and management. Can. J. Fish. Aquat Sci. 47: 2-15.

Peterman, R. M., and J. L. Anderson. 1999. Decision analysis: A method for taking uncertainties into account in risk-based decision making. Human Ecol. Risk Assess. 5: 231-244.

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

Peterman, R. M., B. J. Pyper, M. F. Lapointe, M. D. Adkison, and C. J. Walters. 1998. Patterns of covariation in survival rates of British Columbia and

Alaskan sockeye salmon (*Oncorhynchus nerka*) stocks. Can. J. Fish. Aquat. Sci. 55: 2503-2517.

Peters, C.N. and D.R. Marmorek. 2001. Application of decision analysis to evaluate recovery actions for threatened Snake River spring and summer chinook salmon (*Oncorhynchus tshawytscha*). Can. J. Fish. Aquat. Sci. 58: 2431-2446.

Platts, W. S., R. J. Torquemada, M. L. McHenry, and C. K. Graham. 1989. Changes in salmon spawning and rearing habitat from increased delivery of fine sediment to the South Fork Salmon River, Idaho. Trans. Am. Fish Soc. 118: 274-283.

Rhodes, J. J., D. A. McCullough, and F. A. Espinosa. 1994. A coarse screening process for potential application in ESA consultations, Columbia River Intertribal Fish Commission. Technical Report 94-4.

Roni, P., T. J. Beechie, R. E. Bilby, F. E. Leonetti, M. M. Pollock, and G. R. Pess. 2002. A review of stream restoration techniques and a hierarchical strategy for prioritizing restoration in Pacific Northwest watersheds. North. Am. J. Fish. Manag. 22: 1-20.

Rosgen, D. L. 1985. A stream classification system. *In* Riparian ecosystems and their management: reconciling conflicting uses. *Edited by* R. R. Johnson, C. D. Ziebell, D. R. Palton, P. F. Ffolliot, and R. H. Hamre, eds., USDA Forest Service General Technical Report RM-120, Fort Collins, Colo. pp. 91-95.

Schaller, H. A., C. E. Petrosky, and O. P. Langness. 1999. Contrasting patterns of productivity and survival rates for stream-type chinook (*Oncorhynchus tshawytscha*) populations of the Snake and Columbia rivers. Can. J. Fish. Aquat. Sci. 56(6): 1031-1045.

Scully, R. J., E. J. Leitzinger, and C. E. Petrosky. 1990. Idaho habitat evaluation for off-site mitigation record. Part I in Idaho Department of Fish and Game. 1990. Idaho habitat evaluation for off-site mitigation record, Annual Report 1988. Department of Energy, Bonneville Power Adminstration, Division of Fish and Wildlife, Portland, Oregon. Project 83-7.

Slaney, P. A. 2000. Turning the tide: towards recovery of water quality and wild fish habitat in BC's forested watersheds. *Foreword in* Annual compendium of aquatic rehabilitation projects for the watershed restoration program 1999-2000, Watershed Restoration Project Report No. 18. *Edited by* D. Underhill. Watershed Restoration Program, Ministry of Environment, Lands and Parks, Victoria, British Columbia.

Slaney, T. L., K. D. Hyatt, T. G. Northcote, and R. J. Fielden. 1996. Status of anadromous salmon and trout in British Columbia and Yukon. Fisheries 21: 20-35.

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

Tietenberg, T. 1992. Environmental and Natural Resource Economics. Third Edition. HarperCollins Publishers Inc. New York, NY.

Walters, C. J., and J. S. Collie. 1988. Is research on environmental factors useful to fisheries management? Can. J. Fish Aquat. Sci. 45: 1848-1854.

Walters, C. J., J. S. Collie, and T. Webb. 1989. Experimental designs for estimating transient responses to habitat alteration: is it practical to control for environmental interactions? *In* Proceedings of the National Workshop on Effects of Habitat Alteration on Salmonid Stocks. *Edited by* C. D. Levings, L. B. Holtby, and M. A. Henderson. Can. Spec. Publ. Fish. Aquat. Sci. No. 105. pp. 13-20.

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