North American Journal of Fisheries Management 24:1353–1362, 2004

Evaluating Stream Habitat Survey Data and Statistical Power Using an Example from Southeast Alaska

MASON D. BRYANT*

U.S. Forest Service, Pacific Northwest Research Station, 2770 Sherwood Lane, 2A, Juneau, Alaska 99801, USA

JOHN P. CAOUETTE

U.S. Forest Service, Alaska Region, Juneau, Alaska 99801, USA

BRENDA E. WRIGHT

U.S. Forest Service, Pacific Northwest Research Station, 2770 Sherwood Lane, 2A, Juneau, Alaska 999801, USA

Abstract.—Stream habitat surveys and watershed assessments have been developed and used as monitoring tools for decades. Most rely on type I error as the primary criterion, with minor consideration of statistical power and effect size. We test for statistical differences in fish habitat condition between harvested and nonharvested watersheds from habitat survey data collected in southeast Alaska. We apply statistical power analysis to judge whether nonsignificant results can be interpreted with confidence. None of the fish habitat variables we examined were significant at $\alpha = 0.05$; however, several *P*-values were less than 0.10 and consistent differences between harvested and nonharvested reaches were observed among channel types. Statistical power is low and the probability of not detecting differences is high when the effect size, scaled to the standard deviation of the measurement, is small to medium. For large effect sizes, the ability to detect differences was greater but did not exceed 85% for any measurement. Statistical power, effect size, and biological significance of the outcome are important considerations when the results are interpreted and can lend additional information to managers making decisions with data that are less than perfect.

Ecological monitoring has been bound by a scientifically conservative paradigm that is ruled by type I error. The paradigm fits well for hypothesis testing in a rigorously controlled experimental design and it is well entrenched in philosophy of science literature (Popper 1959). Attempts to assess large watersheds or effects of management activities across watersheds and landscapes are frequently based on compromises that facilitate ease of data collection and economic efficiency. In many instances, assessments and monitoring programs have not used more sophisticated statistical protocols such as some of the examples presented by McDonald (2002). In some cases monitoring programs have been based on retrospective data and often a less than perfect dataset. Managers are forced to make decisions from imperfect information and the conservative paradigm of hypothesis testing at $\alpha = 0.05$ may not be appropriate. Johnson (1999) discusses various aspects of setting *P*-values in hypothesis testing, reasons why they may be arbitrary, and some alternatives.

Large-scale landscape assessments have been developed and implemented to evaluate watershed condition over a wide geographic range. These include assessments arising from the Northwest Forest Plan for the Pacific Northwest (FEMAT 1993) and the Tongass Land Management Plan (U.S. Forest Service 1995) for southeast Alaska. Others, such as the Environmental Monitoring and Assessment Program, were developed for a broader geographic range (USEPA 2002). These assessments include attempts to identify effects of land management activities on salmon habitat and salmon abundance. The development of a set of core measures that can be used to measure fish habitat condition over a wide range of geographic conditions has been a central issue in these assessments. A set of core measures was proposed to monitor fish habitat in the Columbia River basin (U.S. Forest Service 1994). A similar approach, based in part on the assessment for the Columbia River basin, was developed for fish habitat in southeast Alaska (U.S. Forest Service 1995). Measurement of habitat variables has been examined

* Corresponding author: mdbryant@fs.fed.us

Received July 2, 2003; accepted March 18, 2004

TABLE 1.—Criteria used to screen the habitat survey data used in our analysis.

Measure	Type of information	Description of measurements	Criterion
Location	Latitude and longitude at stream mouth	Degrees	U.S. Forest Service, Region 10 Stream Habitat Survey Protocol[a]
Metadata	Units of measure	English or metric	Common definitions
Management	Timber harvested or un-harvested		Harvest units were upstream or immediately adjacent to the stream section surveyed; no criteria for time of harvest were applied.
Channel type	Physical dimensions verified	Metric	Cross-sectional procedure (Harrelson et al. 1994)
Substrate	Sizes	Metric	Wolman pebble counts (Wolman 1954)
Bank-full width	Physical dimensions	Calculated (average)	Cross-sectional procedure
Bank-full depth	Physical dimensions	Calculated (Average)	Cross-sectional procedure
Channel gradient	Slope of stream	Percent	Cross-sectional procedure
Channel bed width	Active channel	Average	Common definitions and methodology
Reach length	Total survey length	Meters	Common definitions and methodology
Pools	Number in survey	Count	Region 10 Stream Habitat Survey Protocol (meets all three criteria)
Residual pool depth	Average of all pools measured in survey	Maximum depth-pool tail depth > 0.15 m	Region 10 Stream Habitat Survey Protocol
Pool length	Amount in stream survey	Meters	Common definitions and methodology
Large wood pieces	Total pieces	Count	Region 10 Stream Habitat Survey Protocol
Key large wood pieces	Total key pieces	Count	Region 10 Stream Habitat Survey Protocol

[a] From geographical information system and global positioning system data and U.S. Geological Survey topographical maps.

in considerable detail to determine observer variability and precision (Minns et al. 1996; Bauer and Ralph 1999; Kaufmann and Larsen 1999; Kaufmann et al. 1999; Archer et al. 2004).

The issue of consistent measures of habitat appears to be reasonably well defined, and tools are available to evaluate this aspect of habitat assessment (Roper et al. 2002). Kaufmann et al. (1999) found that habitat area was relatively imprecise, was affected by stream stage, and varied among observers. However, measurements of other variables, including those in the core variables we use, were either precise or moderately precise, as described by the signal-to-noise ratio (Kaufmann et al. 1999). Among these variables are the width-to-depth ratio and residual depth. Archer et al. (2004) found the least amount of observer variability in measures of pool frequency, substrate size distribution, and counts of large wood. Ralph et al. (1994) also found that measurement of the volume and position of large wood debris was consistent among their surveys. All of these results assume that field crews are well trained.

We examine the ability of a set of core variables to measure fish habitat, using data collected in southeast Alaska to detect differences in fish habitat attributable to timber harvesting. Statistical power analysis is applied to judge whether nonsignificant results can be interpreted with confidence. We use type I and type II error rates and effect size to interpret our results in the context of environmental monitoring.

Methods

Data source.—We used data from monitoring surveys designed to assess stream habitat conditions in the Tongass National Forest. The surveys were conducted from 1994 through 1998 in 3rd- to 4th-order streams that supported anadromous salmon populations. All of the watersheds included in this study are less than 50 km long, ranging from high-gradient steep tributaries to low-gradient floodplain river systems and intertidal reaches. The surveys were conducted by trained crews and followed the methodology that has been incorporated into stream habitat survey protocols in use throughout the Tongass National Forest (U.S. Forest Service 2001). Samples of harvested and nonharvested reaches were selected from watersheds with economically viable amounts of commercially productive forest; therefore, the samples presented in this study do not represent the population of all streams in the Tongass National Forest. Each reach was surveyed once. We reviewed and screened all data to ensure quality, integrity, and consistency with the protocols (Table 1). Stream sample lengths varied considerably, ranging from 100 m to more than 6,000 m; however,

TABLE 2.—Data collection methods and equations used to calculate the eight habitat response variables from field surveys.

Habitat response variable	Equation	Data collection method
Width-to-depth ratio[a]	Bank-full width: bank-full depth	Bank-full width; bank-full depth (mean and maximum)
Total large wood pieces/meter	No. pieces/meter surveyed	Total count of large wood pieces >1 m long and 0.1 m in diameter; total length of stream surveyed
Total key pieces large wood/meter	No. key pieces/meter surveyed	Total count of key large wood pieces (key piece size based on average channel bed width of stream surveyed); total length of stream surveyed
Pools/kilometer	No. pools/kilometer surveyed	Total count of pools; total length of stream surveyed
Pool spacing	(Length of stream surveyed/channel bed width)/total number of pools	Total length of stream surveyed; channel Bed width; Total number of pools
Residual pool depth/channel bed width	Average of all pool residual depth/average channel bed width	Residual pool depth = maximum pool depth-pool tail depth; channel bed width (width of stream from bottom of bank-full to bottom of bank-full)
d50[b]	Median particle size	Measure intermediate diameter of 100 pebbles
Pool length/meter	Total pool length/total length of stream surveyed	Sum of all pool lengths; total length of stream surveyed

[a] Dunne and Leopold (1978) Rosgen (1996).
[b] Wolman (1954).

we did not observe a strong correlation between harvest status and the lengths of the stream samples. The mean sample length was 811 m and the median 470 m.

In each sample reach, the physical habitat measurements included all parameters that we used to calculate the eight core variables. The selection of these eight variables evolved from earlier monitoring efforts, including those in the Columbia River basin (U.S. Forest Service 1994) and southeast Alaska (U.S. Forest Service 1995). The eight core variables were width-to-depth ratio, total number of large wood pieces per meter, total key large wood pieces per meter, pools per kilometer, pool spacing, residual pool depth per channel bed width, median bed particle size, and pool length per meter (U.S. Forest Service 2001). The data collected to determine values for each of these variables are described in Table 2.

Stream reach samples were categorized according to the southeast Alaska Channel Type Classification System, a hierarchical classification system designed to manage the variation among streams and stream reaches (Paustian 1992). Stream reaches were separated into nine basic fluvial process groups according to such physical attributes as channel gradient, channel pattern, stream bank incision and containment, riparian plant community, bank-full width, and dominant substrate size (Paustian 1992). We collected data in seven of the nine process groups: floodplain; moderate-gradient mixed control; low-gradient contained; moderate-gradient contained; high-gradient contained; palustrine; and alluvial fan (see Paustian 1992). Harvest status of each stream reach was determined by the presence or absence of timber harvest upstream or adjacent to the stream reach sample, regardless of the age of the harvest. All watersheds were harvested with high-lead clear-cut logging between 1970 and 1990.

The final data set included 128 stream reach samples. In none of the 128 samples were all eight core habitat variables measured and recorded. However, pools per kilometer and pool spacing were recorded in 127 sample reaches; the total number of key pieces of large wood per meter had the fewest data (67 sample reaches). Samples sizes varied considerably among process groups and harvest status. High gradient contained, palustrine, and alluvial fan stream reaches that had fewer than five samples per process group and harvest status were not tested. Low- and moderate-gradient contained process groups were combined to form a single contained process group (LC&MC). We tested three process groups, floodplain, moderate-gradient mixed control, and the LC&MC group.

Statistical analysis.—Stream reach samples were sorted and grouped by process group and watershed harvest status. We used box and whisker plots to portray distributions by process group and watershed harvest status. The ''box lines'' (horizontal lines forming the box) indicate the 25th, 50th, and 75th percentiles; the ''whisker lines'' (horizontal lines outside the box) mark the 10th and 90th percentiles; and the ''+'' symbol is the arithmetic mean. When distributions were consistently different from normal and variances between groups were not equal, we applied data transformations. In general, standard assumptions for parametric statistical testing (i.e., normal distributions and homogeneous variances) appeared to be better met by using either a square root or logarithmic transformation—procedures common in biological count data (Zar 1984). All habitat variables were square root–transformed, except pool spacing and median particle size were log-transformed.

Because of the number of missing data points, a simultaneous multivariate analysis on all response variables was not feasible without dropping a substantial portion of the data. Instead, we opted to conduct a series of *t*-tests on each response variable within process groups. The null hypothesis of no difference between harvest and nonharvest groups was tested against an alternate hypothesis, a one-directional difference in the harvest groups. The direction of the alternate hypothesis was determined a priori, based on current scientific knowledge regarding relationships between past timber harvest practices and the fish habitat variables. For example, timber harvest is expected to increase the width-to-depth ratio of streams and reduce the frequency of pools. All habitat variables were tested for a decrease in the harvested group, except width-to-depth ratio and pool spacing, which were tested for an increase.

Testing for differences across all three process groups-—floodplain, moderate-gradient mixed control, and LC&MC—using an analysis of variance (ANOVA) framework would increase the statistical power of the hypothesis testing conducted in this paper. The advantage of an ANOVA framework is that information from all process groups are used simultaneously in the statistical testing. However, ANOVA requires independence among samples and homogeneous variances among process groups. Unfortunately, we could not meet these assumptions with our data, because in some instances stream reaches classified under different process groups were subunits of the same stream. Furthermore, the reach sample lengths varied considerably by process group. This variation in the size of the experimental unit (i.e., reach sample lengths) among process groups could lead to substantial differences in estimates for the variation within process groups.

Statistical power.—Sample size requirements (n) were calculated with predetermined levels for type I and type II error rates (α and β), sample variation (σ), and effect size (m_a—m_b) using statistical power equations (equation 1 obtained from Zar 1984), that is,

$$n = (t_{\alpha},_{v} + t_{\beta},_{/v})^2\ s^2\ /(m_a - m_b)^2, \qquad (1)$$

where $m_a - m_b$ equals the difference between means and $t_{\alpha},_{v} + t_{\beta},_{v}$ are values for Student's t at sample sizes equal to v.

Cohen (1988) offers three categories for the effect size d: small, medium, and large. These categories are defined as 20, 50, and 80%, respectively, of 1 standard deviation. We substituted Cohen's effect size $d = |m_a—m_b|/s$ into equation (1) to obtain equation (2):

$$n = s^2(\mathrm{t}_{\alpha},_{v} + t_{\beta},_{v})^2/d^2. \qquad (2)$$

We used equation (2) to create a table of minimum required sample sizes (n per group) associated with various combinations of statistical power parameters (i.e., $\alpha = \beta$ and Cohen's effect sizes).

Results

The overall variation of the measurements tends to be large with respect to the mean (Figure 1). The coefficients of variation (CV) range from 40% to 80%. One-way statistical *t*-tests between harvested and nonharvested groups resulted in no single *t*-test with a *P*-value lower than 0.05 (Table 3). Differences were consistent among process groups; for example, fewer pools and less LWD were observed in harvested watersheds for all process groups. *P*-values for many process groups were 0.10 or less, including pools per kilometer and residual pool depth to channel bed width in the floodplain process group, total large woody debris per meter in the mixed control process group, and included median particle size and width-to depth ratio for the LC&MC combined process group. The tests of width-to-depth ratio resulted in relatively low *P*-values for all process

→

FIGURE 1.—Box-and-whisker plots for eight core habitat measures observed for streams in harvested (H) and nonharvested (N) watershed by process group. Abbreviations are as follows: d50, median bed particle size; FP, floodplain; MC&LC, low-gradient contained and moderate-gradient contained combined; and MM, moderate-gradient mixed control. See text for additional details.

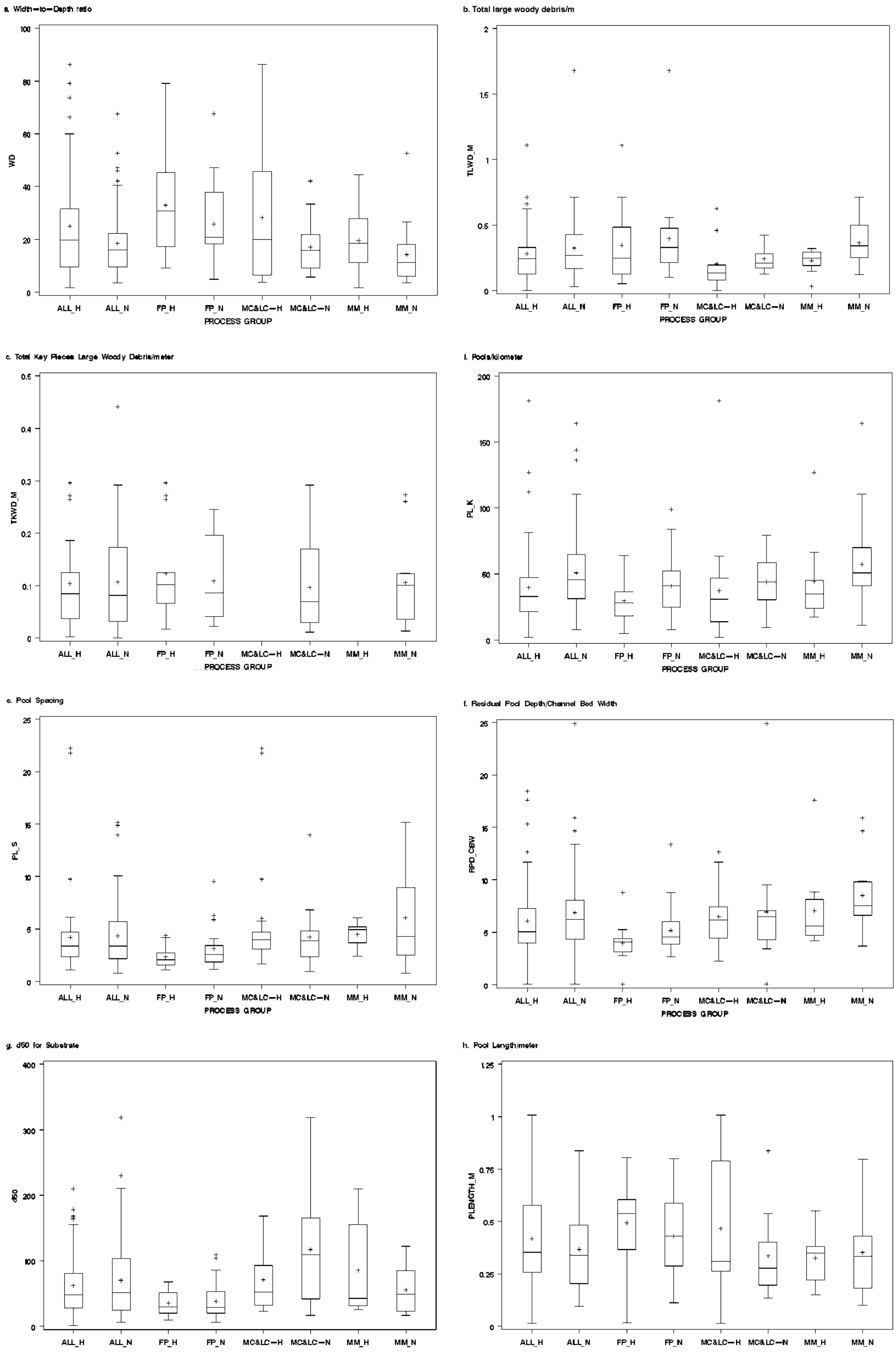
a. Width-to-Depth ratio
WD
b. Total large woody debris/m
TLWD_M
c. Total Key Pieces Large Woody Debris/meter
TKWD_M
d. Pools/kilometer
PL_K
e. Pool Spacing
PL_S
f. Residual Pool Depth/Channel Bed Width
RPD_CBW
g. d50 for Substrate
d50
h. Pool Length/meter
PLENGTH_M
ALL_H ALL_N FP_H FP_N MC&LC-H MC&LC-N MM_H MM_N
PROCESS GROUP

TABLE 3.—Results of *t*-tests for habitat variables by process group and harvest status. Habitat variables are abbreviated as follows: WD, width-to-depth ratio; TLWD/M, total number of large wood pieces per meter; TKWD/M, total number of key large wood pieces per meter; POOL/KM, number of pools per kilometer; PL SPC, pool spacing; RPD/CBW, residual pool depth per channel bed width; d50, median bed particle size; and PLNGTH/M, pool length per meter. The abbreviation N_N stands for the number of samples in the nonharvested group, the abbreviation N_H for the number of samples in the harvested group. The direction of difference between harvest groups (Dir) is indicated by upward-pointing (the value for the harvested group > that for the nonharvested group) and downward-pointing arrows (the value for the harvested group < that for the nonharvested group). Asterisks indicate significance at the 0.10 level.

Habitat variable	Transformation	Dir	Floodplain			Moderate-gradient mixed control			Low-and moderate-gradient contained		
			N_N	N_H	P	N_N	N_H	P	N_N	N_H	P
WD	Square root	↑	23	15	0.11	18	11	0.11	18	18	0.10*
TLWD/M	Square root	↓	16	17	0.39	11	9	0.07*	9	9	0.26
TKWD/M	Square root	↓	18	13	0.67	12	1	NA	10	4	0.32
POOL/KM	Square root	↓	25	19	0.06*	18	10	0.15	18	18	0.13
PL SPC	Natural logarithm	↑	25	19	0.93	18	10	0.68	18	18	0.14
RPD/CBW	Square root	↓	22	19	0.06*	13	10	0.15	17	17	0.27
d50	Natural logarithm	↓	22	17	0.56	15	10	0.15	15	14	0.09*
PLNGTH/M	Square root	↓	16	12	0.78	13	10	0.39	17	15	0.91

groups, whereas pool length per meter resulted in high *P*-values for all process groups.

The values in Table 4 show that there is relatively low power for all variables at small and medium effects sizes (20% and 50% of one standard deviation, respectively) for the statistical tests in this study. The statistical power is low (β = 0.15, 20% of one standard deviation) for width-to-depth ratio in floodplain channels; therefore, a difference will go undetected 85% of the time. At an effect size of 80% of one standard deviation the power of the statistical power is much higher (β = 0.76), but an effect will go undetected 24% of the time. The results from equation (2) show that we would need 542 samples to detect a difference at α = 0.05 and β = 0.05 for a small effect size (Table 5). In some cases, the sample sizes available with our data can detect differences for large effects (80% of one standard deviation) at α = 0.10, β = 0.10. These are differences that Cohen (1988) describes as grossly visible to the naked eye. For our data, we need 16 samples for each group to detect a medium effect, differences barely visible to the naked eye at 50% of one standard deviation (Cohen 1988) when α and β are equal to 0.25 (Table 5).

Discussion

Although the differences that were observed between harvested and nonharvested watersheds were not significant at α = 0.05, the means for the width-to-depth ratios were larger in the harvested watersheds than in the nonharvested watersheds for all process groups, and the means for total large wood per meter were smaller in all harvested process groups. Relatively low *P*-values observed for the *t*-tests of the measurements of width-to-depth ratio, total large wood per meter, pools per kilometer, and residual pool depth indicate potential differences in fish habitat resulting from timber harvest. These differences concur with other studies that compare stream habitat in watersheds with

TABLE 4.—Statistical power ($1 - \beta$) for *t*-tests (one-way; α = 0.05); 20, 50, and 80% are effect sizes expressed as a percentage of one standard deviation. See Table 3 for habitat variable abbreviations.

Habitat variable	Floodplain			Moderate-gradient control			Mixed low- and moderate-gradient contained		
	20%	50%	80%	20%	50%	80%	20%	50%	80%
WD	0.15	0.43	0.76	0.13	0.36	0.66	0.15	0.43	0.76
TLWD/M	0.14	0.40	0.72	0.11	0.29	0.53	0.16	0.25	0.46
TKWD/M	0.13	0.38	0.69				0.14	0.21	0.38
POOL/KM	0.16	0.50	0.83	0.13	0.34	0.63	0.15	0.43	0.76
PL SPC	0.16	0.50	0.83	0.13	0.34	0.63	0.15	0.43	0.76
RPD/CBW	0.15	0.46	0.80	0.12	0.31	0.57	0.14	0.42	0.74
d50	0.15	0.45	0.78	0.12	0.33	0.60	0.13	0.36	0.66
PLNGTH/M	0.13	0.36	0.66	0.12	0.31	0.57	0.14	0.40	0.72

TABLE 5.—Number of samples needed (per group) to detect significant effects for type I errors (α) of 0.05, 0.10, 0.15, 0.20, and 0.25 and type II errors (β) of 0.05, 0.10, 0.15, 0.20, and 0.25 and effect sizes of 20, 50, and 80%. The values in bold italics are the sample sizes observed in this study.

Levels of α and β	Effect size 20%	50%	80%
(α = 0.05, β = 0.05)	542	88	35
(α = 0.10, β = 0.10)	330	54	***22***
(α = 0.15, β = 0.15)	216	35	***14***
(α = 0.20, β = 0.20)	143	24	***10***
(α = 0.25, β = 0.25)	92	***16***	***7***

timber harvest and without timber harvest. Montgomery et al. (1995) and Ralph et al. (1994) found fewer pools per kilometer and less large wood per meter in harvested units than in nonharvested units. Nonetheless, we found no single statistical *t*-test with a *P*-value less than 0.05 (i.e., the null hypothesis of no difference was not rejected). However, if we are willing to increase our significance level to α = 0.10, we would conclude significant differences between management treatments for pools per kilometer, and residual pool depth per channel bed width in the floodplain process group, total number of key pieces of large wood per meter in the moderate-gradient mixed control process group, and width-to-depth ratio and median bed particle size in the combined LC&MC process group (Table 5).

Two habitat measures (large wood and pool frequency) examined in this study are of considerable biological significance for juvenile salmonids. Both were found to be lower in harvested watersheds. The low statistical power suggests our data may not detect such effects if we maintain a type I error rate of 0.05. Reed and Blaustein (1997) point to considerable risk when type I error rates are too strict. Furthermore, statistical significance and biological significance are not necessarily the same (Hayes and Steidl 1997). Biological significance may be of considerably more importance than statistical significance with respect to potential risk, for example, detecting a biologically significant decline in populations (Reed and Blaustein 1997). Ralph et al. (1994) and Montgomery et al. (1995) provide corroborative evidence of loss of large wood and pools in watersheds where riparian timber has been removed. Considerable evidence demonstrates that changes in both variables have biologically significant effects on salmon populations (Bisson et al. 1987; Meehan and Bjornn 1991). Large wood in streams generally results in habitats that support more juvenile salmonids and provide increased survival rates during critical points in their life cycle, such as during fall and winter (Bryant 1985; Beechie and Sibley 1997; Hauer et al. 1999; Solazzi et al. 2000; Roni and Quinn 2001; Sharma and Hilborn 2001; Rosenfeld and Huato 2003). Pools and pools with complex habitat created by large wood support larger numbers of several stream-rearing salmonids, including juvenile coho salmon, juvenile cutthroat trout, and bull trout (Nass et al. 1996; Healy and Lonzarich 2000; Rosenfeld et al. 2000; Sharma and Hilborn 2001).

Failure to reject the null hypothesis may lead managers to the inappropriate conclusion that no effect exists when two other important elements, type II error rate and effect size, are not explicitly discussed. Choosing values for type I and type II error rates (α and β), sample variation (σ), and effect size (m_a—m_b) is important (Fairweather 1991; Downes et al. 2002). In many environmental studies, data needed for estimating means and variance in equation (1) often do not exist. In remote areas such as southeast Alaska, obtaining these data and determining costs associated with type I and type II errors is difficult and expensive. In the absence of any information on costs associated with the type I and type II error rates, a traditional science paradigm suggests a conservative alpha, such as α = 0.05, and a more liberal beta, such as β = 0.10 or β = 0.20; however, Peterman (1990a) argues that in the absence of any information on costs, type I and type II errors should be treated equally.

Peterman (1990b) illustrates one example of biological effects and the consequences of not considering type II error rates in his examination of results in which Nickelson (1986) failed to reject the hypothesis that the marine survival of coho salmon was density-independent. Peterman (1989) points out that the probability of a type II error was at least 81%; therefore, hatchery managers could not with any degree of certainty accept the hypotheses that marine survival was density-independent. The results do not support a conclusion that managers reasonably could expect more adult coho salmon to result from an increased production of hatchery-raised coho salmon smolt without potential effects on natural populations. Other processes and relationships may explain the results: Marine upwelling, decadal ocean cycles, and changes in ocean temperature suggest more strongly the possibility of a density-dependent component in marine survival (Nickelson 1986; Beamish et al. 1997; Welch et al. 1998a, 1998b). Certainly, from a management perspective, it may pay to consider giving beta at least the same weight as alpha, because the costs of a type II error could prove disastrous and lead to both short- and long-term costs. As Fairweather (1991) points out: ''a Type 2 (II) error would most probably lead to a hopelessly false sense of security while environmental degradation continued until very extreme and undeniable damage became evident.''

Defining a meaningful biological effect size for fish habitat is challenging. Biological effect sizes often are reported across a wide range of sizes. For example, in their discussion of biological effects, Roper et al. (2002) estimated that differences of 50% in their measurements of physical habitat would be required before consequences of management could be measured. Meaningful effects, or biological significance, can be defined in controlled studies more easily than those normally encountered in monitoring programs. For example, toxicity studies that determine lethal levels establish a basis for potential biological effects are relatively straightforward and can be defined in terms such as LD50 (that is, the concentration lethal to 50% of the organisms studied). Reed and Blaustein (1997) discuss in considerable detail the problems of assigning biological significance to the analysis of population changes and population declines specifically. They point out that an 80% decline followed by 90% increase may not be biologically significant, whereas a sustained smaller decline (i.e., Cohen's small effect) would be biological significant if it led to extirpation of the population. They conclude that in these cases that managers and scientists will have to ''work by consensus'' to establish biological significance.

The ability to perceive differences provides a starting point for the consensus suggested by Reed and Blaustein (1997). The sample sizes available in our study provide enough statistical power to detect group differences that are grossly visible to the naked eye ($\alpha = \beta = 0.10$; Table 5). However, in most cases, the statistical power was insufficient to detect large differences that could be easily perceived. This is an indication that forest managers and users may be seeing the changes on the ground, but models or monitoring studies cannot detect them—presenting a potential problem for land managers and decision makers. We suggest that sample size be scaled to effect size; in our example we would recommend a sample size of 53 for each group to detect differences at a medium effect size with $\alpha = 0.10$ and $\beta = 0.10$.

The data we used for this analysis are representative of a large body of monitoring data collected from forested watersheds throughout North America. In many cases, these studies apply a conservative testing paradigm with $\alpha = 0.05$ without discussion of either type II error rates or effect sizes, leading to the false assumption that the null hypothesis can be accepted when in fact all that has occurred is a decision to set a high threshold for significance. Most scientists recognize that not rejecting the null hypothesis does not mean it should be accepted and that failing to reject the null hypothesis can be attributed to low sample sizes and low statistical power combined with high variation in measurements rather than to the absence of an effect. This is a particularly important consideration in monitoring studies where the desired outcome is often finding no effect. Accepting the null hypothesis of ''no effect'' can have important and potentially costly consequences. An important consideration in setting α is the risk of making a type II error and the biological significance and consequences that would follow. Yet, determining risk and the best values for these parameters can be as much of a philosophical or political process as it is statistical (Millard 1987). Nonetheless, all three—α, β, and effect size—deserve equal status in the development of monitoring studies and the evaluation of their results.

Acknowledgments

An unpublished manuscript prepared by Kathy Coghill provided the starting point for the development of this paper. Many people assisted with the development of protocols, training, and data collection, including Steve Paustian, Stephen Levesque, Dan Kelliher, John Hannon, Julianne Thompson, Dick Aho, Ron Medel, William Lorenz, and Ron Dunlap. Kim Frangos made a substantial contribution in developing and managing the database. Jeff Kershner and Brett Roper provided thorough and helpful reviews of an earlier version of this paper.

References

Archer, E. K., B. B. Roper, R. C. Henderson, N. Bouwes, S. C. Mellison, and J. L. Kershner. 2004. Testing common stream sampling methods: How useful are these techniques for broad-scale, long-term monitoring? U.S. Forest Service, Rocky Mountain Research Station, General Technical Report RMRS-GTR-122, Fort Collins, Colorado.

Bauer, S. B., and S. C. Ralph. 1999. Aquatic habitat indicators and their application to water quality objectives within the Clean Water Act. U.S. Environmental Protection Agency, Region 10, EPA-910-R-99-014, Seattle.

Beamish, R. J., C. M. Neville, and A. J. Cass. 1997. Production of Fraser River sockeye salmon (*Oncorhynchus nerka*) in relation to decadal-scale changes in the climate and the ocean. Canadian Journal of Fisheries and Aquatic Sciences 54:543–554.

Beechie, T. J., and T. H. Sibley. 1997. Relationships between channel characteristics, woody debris, and fish habitat in northwestern Washington streams. Transactions of the American Fisheries Society 126: 217–229.

Bisson, P. A., R. E. Bilby, M. D. Bryant, C. A. Dolloff, G. B. Grette, R. A. House, M. L. Murphy, K. V. Koski, and J. R. Sedell. 1987. Large woody debris in forested streams in the Pacific Northwest: past, present, and future. Pages 143–190 *in* E. O. Salo and T. W. Cundy, editors. Streamside management: forestry and fishery interactions. College of Forest Resources, University of Washington, Seattle.

Bryant, M. D. 1985. Changes 30 years after logging in large woody debris, and its use by salmonids. Pages 329–334 *in* R. Johnson, C. D. Ziebell, D. R. Patton, P. F. Folliott, and R. H. Hamre, editors. Riparian ecosystems and their management: reconciling conflicting uses. U.S. Forest Service, Rocky Mountain Forest and Range Experiment Station, General Technical Report RM-120, Fort Collins, Colorado.

Cohen, J. 1988. Statistical power for the behavioral sciences. Lawrence Erlbaum Associates, Hilldale, New Jersey.

Downes, B. J., L. A. Barmuta, P. G. Fairweather, D. P. Faith, M. J. Keough, P. S. Lake, B. D. Mapstone, and G. P. Quinn. 2002. Monitoring ecological impacts: concepts and practice in flowing waters. Cambridge University Press, Cambridge UK.

Dunne, T., and L. B. Leopold. 1978. Water in environmental planning. Freeman, San Francisco.

Fairweather, P. G. 1991. Statistical power and design requirements for environmental monitoring. Australian Journal of Marine and Freshwater Research 42:555–567.

FEMAT (Forest Ecosystem Management Team). 1993. Forest ecosystem management: an ecological, economic, and social assessment. FEMAT, Portland, Oregon.

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

Hauer, F., G. Poole, J. Gangemi, and C. Baxter. 1999. Large woody debris in bull trout (*Salvelinus confluentus*) spawning streams of logged and wilderness watersheds in northwest Montana. Canadian Journal of Fisheries and Aquatic Sciences 56:915–924.

Hayes, J. P., and R. J. Steidl. 1997. Statistical power analysis and amphibian population trends. Conservation Biology 11:273–275.

Healy, B., and D. Lonzarich. 2000. Microhabitat use and behavior of overwintering juvenile coho salmon in a Lake Superior tributary. Transactions of the American Fisheries Society 129:866–872.

Johnson, D. H. 1999. The insignificance of statistical testing. Journal of Wildlife Management 63:763–772.

Kaufmann, P. R., and D. P. Larsen. 1999. Power to detect regional trends in habitat characteristics. Toward ecosystem-based management: breaking down the barriers in the Columbia River basin. American Fisheries Society, Spokane, Washington.

Kaufmann, P. R., P. Levine, E. G. Robison, C. Seeliger, and D. Peck. 1999. Quantifying physical habitat in wadeable streams. U.S. Environmental Protection Agency, Environmental Monitoring and Assessment Program, EPA 620/R-99/003, Corvallis, Oregon.

McDonald, T. L. 2002. Review of environmental montitoring: survey designs. Environmental Monitoring and Assessment 1:1–16.

Meehan, W. R., and T. C. Bjornn. 1991. Salmonid distributions and life histories. Pages 47–82 *in* W. R. Meehan, editor. Influences of forest and rangeland management on salmonid fishes and their habitats. American Fisheries Society, Special Publication 19, Bethesda, Maryland.

Millard, S. P. 1987. Environmental monitoring, statistics, and the law: room for improvement. American Statistician 41:249–253.

Minns, C. K., J. R. M. Kelso, and R. G. Randall. 1996. Detecting the response of fish to habitat alterations in freshwater ecosystems. Canadian Journal of Fisheries and Aquatic Sciences 53(Supplement 1):403–414.

Montgomery, D. R., J. M. Buffington, R. D. Smith, K. M. Schmidt, and G. Pess. 1995. Pool spacing in forest channels. Water Resources Research 31: 1097–1105.

Nass, B., K. English, and H. Frith. 1996. Assessment of summer rearing habitat and juvenile coho abundance in the Kwinageese River, B.C., 1992. Department of Fisheries and Oceans, Report 2375, Prince Rupert, British Columbia Canada.

Nickelson, T. E. 1986. Influences of upwelling, ocean temperature, and smolt abundance on marine survival of coho salmon (*Oncorhynchus kisutch*) in the Oregon production area. Canadian Journal of Fisheries and Aquatic Sciences 43:527–535.

Paustian, S. J. 1992. A channel type users guide for the Tongass National Forest, southeast Alaska. U.S. Forest Service, Region 10, Technical Paper 26, Juneau, Alaska.

Peterman, R. M. 1989. Application of statistical power analysis to the Oregon coho salmon (*Oncorhynchus kisutch*) problem. Canadian Journal of Fisheries and Aquatic Sciences 46:1183–1187.

Peterman, R. M. 1990a. The importance of reporting statistical power: the forest decline and acidic deposition example. Ecology 7:2024–2027.

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

Popper, K. R. 1959. The logic of scientific discovery. Routledge, New York.

Ralph, S. C., G. C. Poole, L. L. Conquest, and R. J. Naiman. 1994. Stream channel morphology and woody debris in logged and unlogged basins of western Washington. Canadian Journal of Fisheries and Aquatic Sciences 51:37–51.

Reed, J. M., and A. R. Blaustein. 1997. Biologically significant population declines and statistical power. Conservation Biology 11(1):281–282.

Roni, P., and T. P. Quinn. 2001. Density and size of juvenile salmonids in response to placement of large woody debris in western Oregon and Washington streams. Canadian Journal of Fisheries and Aquatic Sciences 58:282–292.

Roper, B. B., J. L. Kershner, L., E. K. Archer, and N. Bouwes. 2002. An evaluation of physical stream habitat attributes used to monitor streams. Journal of American Water Resources 38:1637–1646.

Rosenfeld, J. S., and L. Huato. 2003. Relationship between large woody debris characteristics and pool formation in small coastal British Columbia streams. North American Journal of Fisheries Management 23:928–938.

Rosenfeld, J., M. Porter, and E. Parkinson. 2000. Habitat factors affecting the abundance and distribution of juvenile cutthroat trout (*Oncorhynchus clarki*) and coho salmon (*Oncorhynchus kisutch*). Canadian Journal of Fisheries and Aquatic Sciences 57:766–774.

Rosgen, D. L. 1996. Applied river morphology. Western Hydrology, Lakewood, Colorado.

Sharma, R., and R. Hilborn. 2001. Empirical relationships between watershed characteristics and coho salmon (*Oncorhynchus kisutch*) smolt abundance in 14 western Washington streams. Canadian Journal of Fisheries and Aquatic Sciences 58:1453–1463.

Solazzi, M. F., T. Nickelson, S. Johnson, and J. Rodgers. 2000. Effects of increasing winter rearing habitat on abundance of salmonids in two coastal Oregon streams. Canadian Journal of Fisheries and Aquatic Sciences 57:906–914.

U.S. Forest Service. 1994. Section 7 fish habitat monitoring protocol for the upper Columbia River basin, 6th revision. U.S. Forest Service, Portland, Oregon.

U.S. Forest Service. 1995. Synthesis: anadromous fish habitat assessment. U.S. Forest Service, Juneau, Alaska.

U.S. Forest Service. 2001. Fish and aquatic stream habitat survey. Chapter 20 *in* Aquatic ecosystem management handbook FSH 2090, R-10 amendment 2090–2001–1. U.S. Forest Service, Region 10, Juneau, Alaska.

USEPA (U.S. Environmental Protection Agency). 2002. Research Strategy Environmental Monitoring and Assessment Program. U.S. Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Report 620/R-02/002, Research Triangle Park, North Carolina.

Welch, D. W., Y. Ishida, and K. Nagasawa. 1998a. Thermal limits and ocean migrations of sockeye salmon (*Oncorhynchus nerka*): Long-term consequences of global warming. Canadian Journal of Fisheries and Aquatic Sciences 55:937–948.

Welch, D. W., B. R. Ward, B. D. Smith, and J. P. Eveson. 1998b. Influence of the 1990 ocean climate shift on British Columbia steelhead (*Oncorhynchus mykiss*) and coho salmon (*O.kisutch*) populations. North Pacific Fisheries Commission, Technical Report, Vancouver.

Wolman, M. G. 1954. A method of sampling coarse river bed material. Transactions of the American Geophysical Union 35:951–956.

Zar, J. H. 1984. Biostatistical analysis. Prentice-Hall, Englewood Cliffs, New Jersey.