NORTH AMERICAN
JOURNAL OF FISHERIES MANAGEMENT

Volume 12 Winter 1992 Number 1

North American Journal of Fisheries Management 12:1–18, 1992

Experimental Design in the Management of Fisheries: A Review

MURDOCH K. MCALLISTER[1] AND RANDALL M. PETERMAN

School of Resource and Environmental Management, Simon Fraser University
Burnaby, British Columbia V5A 1S6, Canada

Abstract.—Despite the accumulating theoretical interest in experimental management, there are few practical applications of it. Because most fisheries management plans lack rigorous experimental design, managers often face controversy when results appear consistent with several alternative mechanisms or when results yield little information about causes of fish population dynamics. We provide a synthesis of the problems of experimental design in fisheries science and management, and we show how these problems can be solved to generate better information and better decisions, especially when combined with proper statistical practices and formal decision analysis. Reasons why experimental management is currently rare are (1) lack of unfamiliarity of management agencies with designing management actions to yield information, (2) logistical constraints placed on replication by fish migratory patterns, (3) resistance by fishermen who fear experimentation will lower incomes, or (4) risk of stock collapse. However, recent applications show that these constraints can be overcome if (1) experimentation is done on a small scale, such as on small stocks, (2) fishermen are compensated when cuts in fishing effort are required, (3) fishermen believe that experimentation is in their best interest (e.g., when it is likely to increase harvests), or (4) control populations are held in reserve in case overharvesting occurs in a treated population. Several quantitative decision analyses that include uncertainty show that experimentation sometimes has much higher expected economic value than not experimenting.

Even though most fisheries biologists are trained to do field and laboratory experiments using principles of experimental design, approaches to fisheries management typically have been nonexperimental (Larkin 1972; Walters 1986). Consequently, unexpected results of management actions often lead to confusion rather than to learning, because the management manipulations were not designed to distinguish rigorously between alternative explanations of observations. In other words, we are seldom able to use surprising results as tests of understanding.

Many controversies exist over whether fish stocks are being underharvested or overharvested and whether management actions or environmental causes have led to the collapse of stocks (e.g., Glantz and Thompson 1981). One root cause of these controversies is the large uncertainty in our scientific understanding of managed systems. Part of this uncertainty results from the most common approach to fisheries management, called passive adaptive management (Walters and Hilborn 1976), in which historical data are periodically fitted to a model, new parameters are estimated, and new controls on fishing practices are identified. This form of incremental management often prevents scientists from resolving uncertainties about alternative hypotheses concerning productivity or optimal harvest for a stock because it tends to limit the range of observed stock abundances and it

[1] Present address: School of Fisheries, WH-10, University of Washington, Seattle, Washington 98195, USA.

generates only small contrasts in fishing patterns over time (Walters and Collie 1989).

Several other factors contribute to uncertainty in our scientific understanding of managed systems. Research scientists are continually reducing uncertainty in some aspects of fisheries biology (e.g., somatic growth, natural mortality, migration), but the inability to manipulate systems over large areas and to control perturbations caused by fishing impedes understanding of the dynamics (e.g., recruitment processes) of harvested fish populations. Also, results of small-scale field experiments may differ from those of large-scale experiments because of community interactions (Peterman 1991). Partly for these reasons and partly because managers rarely have designed and implemented policies to test alternative biological hypotheses, we have few unambiguous understandings of complex interactions among components of fish communities. Thus, theories about such interactions and about responses of communities to disturbances remain based primarily on incomplete and indirect evidence (Larkin 1972; Walters and Hilborn 1978; Mercer 1982; Walters 1986; Colby et al. 1987; Sainsbury 1988).

These sources of uncertainty in our understanding of fish population dynamics have led several authors to advocate an experimental approach to management (Larkin 1972; Loftus 1976; Walters and Hilborn 1976, 1978; Holt 1977; Holling 1978; Silvert 1978; Ludwig and Hilborn 1983; Peterman and Routledge 1983; Lee and Lawrence 1986; Walters 1986; Sainsbury 1988; Walters et al. 1988; Buckley 1989; Collie et al. 1990; McAllister 1990). Walters' (1986) book, devoted entirely to this topic, presents compelling arguments for applying a rigorous, quantitative approach. Experimental management promotes a scientific basis for management yet does not require action to be dependent on extensive past studies; instead, actions can be based on limited current data as long as uncertainties are recognized openly and experiments are designed to generate new information to resolve them. For example, proponents of the experimental approach advocate that management actions be treated as deliberate experiments to test hypotheses about components of population dynamics of fish stocks and to identify optimal management regimes (e.g., Walters and Hilborn 1978; Peterman and Routledge 1983). This manipulative approach is essential to most resource management sciences, and it allows fisheries managers to learn from their past actions (Larkin 1972). The main goal of this approach is to reduce uncertainty in our understanding of the dynamic components of harvested fish populations so the populations can be managed more effectively.

An implicit assumption in experimental management is that information gained by experimentation will be used to meet management objectives. The approaches to analyzing this information include (1) informal comparisons of treatment and control populations, and (2) formal quantitative updating of parameter estimates, testing of alternative models of population dynamics, and choosing future management actions based on expected economic yield and on the actions' ability to yield the most information about remaining uncertainties. The second approach was called active adaptive management by Walters and Hilborn (1978), and it constitutes a special case of the broader category of experimental management. Active adaptive management provides a framework for learning as well as for systematically choosing and evaluating management actions as they are carried out (Lee and Lawrence 1986; Walters 1986).

Although many authors have advocated experimental management in various situations, relatively few management agencies have tried the approach. A significant drawback to active adaptive management—as the method is portrayed by Walters and Hilborn (1976), Sainsbury (1988), and others—is that it may not be readily understandable to fisheries managers and the fishing industry because it involves relatively sophisticated quantitative methods (see Walters 1986). Moreover, experimental management is only now gaining wide acceptance, and most fisheries managers are simply not yet familiar with its potential benefits. Additional reasons for the paucity of applications of experimental management involve social, economic, and biological constraints. For example, large reductions in fishing effort could generate large, informative contrasts (ranges) in catch and effort data that might allow analysts to more clearly identify the true underlying shape of density- or effort-dependent relationships on which successful management depends. However, such reductions in effort may be resisted by fishermen because of the possibility of lowered incomes. In contrast, large increases in effort may produce an unacceptable risk of stock collapse. Spatial replication of management actions to control for spatial variation in responses may be impossible if different stocks move widely across potential spatial replicates. Nevertheless, examples described below illustrate that experimental management can be applied even with these constraints.

In this paper, we attempt to make the case for a more experimental approach to fisheries management by bridging the gap between the complex quantitative methods of active adaptive management and the general conceptual approach to experimental management. First, we outline some general features of experimental design highlighted by Hurlbert (1984), and we add statistical power, Bayesian analysis, and formal decision analysis criteria to these concepts. Second, we demonstrate how various management actions and field research projects with improper experimental design have led to inconclusive results. We suggest improvements to the designs in each case. Third, we discuss the heuristic potential of some proposed and implemented experimental designs. Fourth, we review evidence from analyses of decision options showing that, in some fisheries, an experimental approach to management may generate higher economic value than current management strategies generate. Finally, we discuss the applicability of the experimental approach, as well as its constraints and how to overcome them.

General Design of Experiments

Hurlbert's Features of Experimental Design

Hurlbert (1984) identified the components of an experiment as (1) hypothesis, (2) experimental design, (3) execution of experiment, (4) data analysis, and (5) interpretation of results. In this paper, we are primarily concerned with experimental design. Advocates of experimental management emphasize that before management actions are taken, hypotheses should be clearly stated, possible biological models should be described mathematically, and experimental designs should be carefully chosen (Walters and Hilborn 1978; Walters 1986; Sainsbury 1988). Good experimental design is crucial to distinguish among alternative hypotheses (Hurlbert 1984).

Hurlbert (1984) stated that experiments based on simple methods such as comparing means with t-tests or analyses of variance should include (1) control units (e.g., populations) against which one or more treatment units are to be compared; (2) treatments that control for effects of the procedure and for temporal changes in experimental units; (3) replication in time and space, which controls for stochastic factors among replicates that are inherent in the experimental material or introduced by the experimenter; (4) interspersion of differently treated units in time or space to control for regular spatial or temporal gradients in properties of experimental units; (5) randomization in allocation of different treatments to experimental units to control for bias; and (6) statistically independent experimental units so that the responses in one experimental unit are not related to responses in other units except insofar as they share the same type of treatment and there is a treatment effect. Note, however, that some of these criteria are not necessary for some types of statistical analyses (e.g., replication is not necessary if responses to different levels of a treatment are determined by regression).

Statistical Power

The concept of statistical power is also important for the design of field and management experiments (Dixon and Massey 1969; Pearson and Hartley 1976; Green 1979, 1989; Vaughan and Van Winkle 1982; Peterman and Routledge 1983; Toft and Shea 1983; Millard and Lettenmaier 1986; Gerrodette 1987; Holt et al. 1987; Peterman 1989, 1990). Statistical power reflects the ability of an experiment to detect an effect that actually exists. The power of an experiment is equal to $1 - \beta$, where β is the probability of committing a type II error. A type II error occurs when an experiment fails to reject the null hypothesis, the hypothesis stating that there is no effect, even though an effect exists. For a given experimental situation, β is inversely related to α, the probability of committing a type I error, whereby the null hypothesis is incorrectly rejected. The power of an experiment may be increased by increasing the differences between the levels of treatments, increasing the number of experimental units, increasing the duration of the experiment, reducing sample variance caused by measurement error, or increasing the α level. Cohen (1988) and Peterman (1990) discussed statistical power further.

Until recently, most ecologists focused on avoiding the costs of type I errors (Toft and Shea 1983; Peterman 1990). However, type II errors may also be very costly, as when one incorrectly concludes that stock productivity is not declining under current management regulations (Allen 1980; Toft and Shea 1983; Green 1984; Peterman 1990). Because type II errors can be costly, managers should design experimental or management actions so that true responses can be detected with high probability (high power).

Bayesian Criteria

Bayesian analysis is another means to assess the statistical performance of an experimental design

(Edwards et al. 1963; Berger 1985). It offers an alternative to classical or "frequentist" statistical inference, of which statistical power analysis is a part. As is well known, by using data to estimate a test statistic, the classical approach addresses the question "Given that the null hypothesis is true, what is the probability of obtaining a test statistic as extreme as or more extreme than the one observed?" If the probability (i.e., *P* value) is less than the significance level chosen (usually 0.05), the null hypothesis is rejected.

In contrast, the Bayesian approach allows the scientist to address the question "What is the probability that a hypothesis is true, given the data?"—and thereby directly address the likelihood of different hypotheses being true—whereas the classical approach only indirectly provides information concerning whether some effect exists (Berger and Berry 1988; Reckhow 1990). In addition, the Bayesian approach facilitates quantification of scientists' subjective input in experimental design and statistical analysis by incorporating into the analysis the scientists' prior expectations of the likelihood of alternative hypotheses (known as prior probabilities). Bayesian methods can also be used to design and evaluate the results of large-scale field experiments for which no replication is possible (Walters 1986; Parma and Deriso 1990; Reckhow 1990). Hence, Bayesian statistical analysis has been advocated for ecological field experiments (Carpenter 1990; Reckhow 1990), and it is also starting to receive attention in fisheries science (Walters 1986; Sainsbury 1988; Parma and Deriso 1990).

To evaluate the statistical performance of an experimental design within the Bayesian framework, the analyst computes the expected rate at which the assumed-true hypothesis can be identified when there are a number of alternative hypotheses. In separate simulation analyses, each alternative hypothesis is treated as if it is true. The most informative design then becomes the one that, for example, maximizes the expected rate of improvement in the probability that the assumed-correct hypothesis is true (summed across all of the alternative hypotheses) (e.g., Parma and Deriso 1990).

However, there is a continuing controversy over the benefits and disadvantages of the Bayesian approach compared with the traditional approach (Howson and Urbach 1989 summarize many of the arguments). In spite of this, Bayesian statistics appear to be gaining wider acceptance among ecologists (e.g., Reckhow 1990).

Decision Theoretic Criteria

In addition to statistical criteria, decision theoretic criteria can be used for experimental design when components of management decisions can be identified and costs and benefits of different outcomes can be estimated. In this context analysts identify alternative hypotheses about the system of interest and several alternative management policies, usually ranging from experimental to nonexperimental. The optimal experimental design or policy option is the one that yields the highest value of the objective function—for instance, the highest expected economic value. The expected value of a policy option incorporates the alternative hypotheses of interest, the odds placed on each alternative hypothesis, and the economic value (often in terms of catch biomass) of an outcome that is expected under each alternative hypothesis. The expected value of a policy option is thus the sum of products of the probability level placed on each hypothesis and the discounted value yielded under it, summed across all of the alternative hypotheses (including the null hypothesis) (Raiffa 1968; Walters 1977, 1986; Parma and Deriso 1990). Utility functions are sometimes used to convert economic and nonmonetary measures of performance into units that reflect the relative value of each option (Keeney and Raiffa 1976; Walters 1977).

Within this framework of decision theory, the classical approach to hypothesis testing is not directly useful because it does not generate probabilities that a given alternative state of nature is true (Berger and Berry 1988). In contrast the Bayesian approach gives these probabilities directly; it is thus sometimes used in computing the expected values of alternative policy options (Walters 1986; Sainsbury 1988; Parma and Deriso 1990).

But clearly both statistical and decision theoretic criteria are important in assessing the appropriateness of experimental design in different management situations. When sensitivity analyses of a decision analysis show that uncertainties regarding utility or net economic benefits drastically influence the choice of experimental design, statistical criteria provide important additional information. Statistical criteria are also important in situations where decision processes are so complex that they preclude decision analysis (C. J. Walters, University of British Columbia, personal communication) or where economic or utility values cannot be estimated. For more discussion and details on methods of Bayesian, frequentist, and other approaches relevant to statistical inference in ecological and fisheries experiments, see Walters (1986), Carpenter (1990), and Reckhow (1990). For descriptions and discussion of decision theoretic techniques, see Raiffa (1968), Lindley (1985), and papers referenced by Walters and Holling (1990).

Problems with Past Experimental Designs and Possible Improvements

Some readers of our examples below may conclude that we have not adequately recognized the biological, logistical, or budgetary constraints faced by researchers and managers. On the contrary, we recognize these constraints, but one of our purposes is to raise the current level of awareness about the benefits of applying principles of experimental design to management so that researchers and managers can take full advantage of opportunities that arise if such constraints are relaxed. We respect past attempts at experimental management and are merely trying to constructively identify additional elements of design that could be added if similar situations arise again. We emphasize that because of biological, logistical, or budgetary constraints, some of the improvements we recommend do not create ideal experimental designs with all the appropriate features. However, our recommended improvements should at least lead to stronger tests of hypotheses, reduce the number of alternative explanations of the resulting observations, and thereby reduce uncertainty concerning the effects of past and current approaches to management.

Lack of Temporal and Spatial Replication

In situations where treatments are not replicated, interpretations of results are often limited. However, application of replicate treatments to separate experimental units enables researchers to (1) test whether treatments produce repeatable effects in situations with similar conditions, (2) distinguish treatment effects from natural variation, (3) estimate the variance in estimated parameters, and (4) test for significance. Given the situations often faced by resource managers (no replication and one or two levels of a treatment), results cannot easily be generalized to other cases. In many fisheries situations, managers are interested only in one fish stock, or they can undertake manipulations only on the entire set of managed stocks or only at one site. Thus, managers frequently initiate one treatment (e.g., a new regulation) at a time rather than conducting two or more contrasting treatments simultaneously on different populations, and manipulations often are initiated without spatial replication or simultaneous controls. Because fisheries systems are typified by spatial variability and chance events, ideal experimental conditions may not hold, and this use of unreplicated, uncontrolled treatments exacerbates difficulties in subsequent interpretation of results.

If there is any replication, it is usually temporal. Stewart-Oaten et al. (1986) argued that when there is only one control and one treatment site, measurements of responses should be taken repeatedly both before and after the manipulation begins. Without this replication in time, there is no control for temporal variability and error in the measured responses. Repeated measurements in time, however, are no substitute for replicated treatments, especially when only one stock is treated and the results will be applied to more than one fish stock in the future. Replication in time by itself does not control for, or enable measurement of, the magnitude of spatial variability or the interactions between spatial units and treatments (Green 1987). These problems of replication are exemplified below. See Walters (1986) and Walters et al. (1988, 1989) for quantitative methods (e.g., cluster analysis) for choosing spatial replicates in management situations.

In the following cases, the first and second examples show temporally but not spatially replicated manipulations, and the third example shows manipulations that were neither spatially nor temporally replicated.

Example 1.—In the 1930s the International Pacific Halibut Commission (IPHC) imposed several regulations to conserve stocks of Pacific halibut *Hippoglossus stenolepis*. For example, in 1932, in order to protect immature Pacific halibut, the IPHC closed halibut fishing year-round in two areas, one in southeastern Alaska and the other in British Columbia. These closures were maintained until the 1960s (Skud 1985). In addition, seasonal quotas and closures were established elsewhere in the 1930s and have continued since then.

The IPHC gained international recognition for rebuilding Pacific halibut stocks between 1930 and 1960. However, some scientists (Burkenroad 1948; Ketchen 1956; Fukuda 1962) have questioned IPHC's role in the recovery of the stocks. For example, Ketchen (1956) concluded that historical data are also consistent with climatic trends as a cause of recovery of Pacific halibut stocks since the 1930s. Furthermore, it is unclear how well each of the different regulations worked because they were all initiated at approximately the same time.

Recommended improvements: In the Pacific halibut fishery, spatial replication of contrasting fishery regulations is not practical because of the wide-ranging migratory behavior of halibut. Yet the effects of different regulations could have been distinguished better if the regulations had been imposed sequentially rather than simultaneously (Stewart-Oaten et al. 1986). Seasonal closures and catch quotas could have been imposed at non-overlapping times in an attempt to yield strong contrasts in abundance among sets of age-classes. For example, the effect on age-class abundances of a large quota imposed over 10 years could have been compared with that of a small quota in a subsequent period. This procedure would not have eliminated the possibility that changing marine conditions could confound the results, but one could guard against this by later reintroducing a large quota. Parma and Deriso (1990) reported a more comprehensive analysis of experimental options for Pacific halibut; it was based on a Bayesian approach that compared the expected performances of actively adaptive, passively adaptive, and nonadaptive harvest policies in discriminating between alternative hypotheses about mechanisms that could affect recruitment of Pacific halibut.

Example 2.—Between 1903 and 1936, 10 fish hatcheries in British Columbia propagated sockeye salmon *Oncorhynchus nerka* (White 1988). During the 1920s biologists questioned whether hatchery propagation of sockeye salmon actually resulted in higher freshwater survival rates than did natural propagation (Foerster 1938). Foerster (1938) compared the egg-to-smolt survival rates of natural and two artificial methods of propagation, but only at Cultus Lake, British Columbia. The different treatments were alternated between years for 12 years. Because Foerster found no significant difference in survival rates, the British Columbia Minister of Fisheries closed all 10 British Columbia sockeye salmon hatcheries in 1936 (Foerster 1968). As a result, hatcheries have not been used subsequently in British Columbia to produce sockeye salmon, with the exception of two small pilot projects (White 1988).

However, Foerster's experimental design, which employed only one hatchery and its associated wild population, gave insufficient evidence to conclude that all hatcheries in British Columbia were not more efficient at propagation of sockeye salmon than were natural means. Of course, Foerster (1968) recognized this insufficiency, and he reported that spawning conditions at Cultus Lake were atypical relative to conditions at other hatchery locations: sockeye salmon at Cultus Lake spawned along gravel inshore areas of the lake where seepage occurred, but at other hatchery locations wild populations of sockeye salmon spawned in tributary streams. Thus, egg-to-smolt survival rates at Cultus Lake were probably not representative of those at other hatchery locations (White 1988), and the findings of this study should not have been generalized to other sockeye salmon hatcheries by the Minister of Fisheries.

Recommended improvements: A design that could have permitted generalization would have included spatial replication. Hatcheries in several different river systems could have been included in the experiment. As in Foerster's (1938) study, hatchery propagation could have been alternated with artificial propagation between years, and sockeye salmon responses to each propagation method could have been monitored in each river system. However, because of the potential variation in the performance among hatcheries, perhaps it would have been even better to decide whether to maintain or close hatcheries based on their individual performance rather than on the average performance of a subset of all of the hatcheries (Foerster 1968).

Example 3.—Bilton et al. (1982) studied how size and time at release of hatchery-raised smolts of coho salmon *Oncorhynchus kisutch* affected returns. The study took place over one season at one study site. Groups of small, medium, and large smolts were tagged and released at four times between April and July. The rates of return of jack (2-year-old) and adult (3-year-old) coho salmon were then monitored. Bilton et al. (1982) determined optimal sizes and times of release by using biological and economic (benefit–cost) criteria. However, they could not estimate interannual variability in times and smolt sizes optimum for release because they had data from only 1 year. Furthermore, they could not estimate the magnitude of among-stock variability because tests were performed only at one site. Thus, they could not make generalizations applying to other years or to other stocks at other sites.

Recommended improvements: The experimental design could have been improved by replicating the same design in other years at the same site to enable estimation of temporal variation in optimal time and size of release at the study site (Stewart-Oaten et al. 1986). To develop a model to predict times and fish sizes optimal for release of hatchery-raised coho salmon from other hatcheries and in other years (and thus generate the highest sustainable economic yield), researchers could use an experimental design similar to that described by Bilton et al. (1982) but at several different hatcheries over several years. It is likely that times and fish sizes optimal for release varied between years and among sites because of variation in such features as genetic stock of coho salmon, availability of food, and risk of predation in release waters. These potentially important variables would have to be carefully and systematically monitored and recorded at all experimental sites. The effect of variables such as genetic stock on fish sizes and times that are optimal for release could be tested by replicating treatments within and across sites; such an approach incorporates a control for interaction effects between site and treatment (Green 1979, 1987).

Thus, manipulations that are not replicated in time or space may provide only limited information. As shown above, replication in time but not in space may be uninformative, especially when rare events occur between years or when changes in environmental conditions occur simultaneously with the imposed regulations. Experiments with both temporal and spatial replication are less likely to be confounded by these events and will allow the probability levels in classical hypothesis tests to be calculated after the experiment (Hurlbert 1984). Even with Bayesian criteria, such experiments also facilitate better discrimination among alternative hypotheses than do manipulations with only temporal replication (Walters 1986; Sainsbury 1988; Carpenter 1990; Reckhow 1990).

Lack of Contrasting Treatments

Absence of informative ranges for stock abundances and fishing patterns over time characterizes many fisheries, especially conservatively managed ones (Walters 1986; Walters and Collie 1989). This pattern is apparent, for example, in management of whales (Holt 1977; Allen 1980), of Pacific ocean perch *Sebastes alutus* (Walters and Collie 1989) and sockeye salmon in British Columbia (Walters and Hilborn 1976; Collie et al. 1990), and of some multispecies communities (Tyler et al. 1982; Sainsbury 1988). The authors cited above noted that ranges of fishing effort and stock abundance in these fisheries are not wide enough to identify clearly the best model of stock dynamics or to provide reasonable estimates of model parameters or maximum sustainable yield (MSY).

Managers often are reluctant to increase effort to test their understanding of stock dynamics for fear of overharvesting (Walters 1986; Sainsbury 1988; Walters and Collie 1989) or to decrease effort because of social or economic constraints. Yet the inability of mathematical and statistical methods to eliminate uncertainty associated with existing data often indicates a need for "adaptive probing"—that is, fishing effort needs to be significantly changed to generate informative contrasts in the data. This approach thus would decrease variance estimates in model parameters (Walters and Hilborn 1976; Ludwig and Hilborn 1983; Walters and Collie 1989).

Example.—Dwindling spawning escapements of chinook salmon *Oncorhynchus tshawytscha* in the Fraser River led managers to attempt to reduce the terminal gill-net harvest by imposing a maximum gill-net mesh size in 1981. It was expected that smaller nets would catch sockeye salmon and pink salmon *O. gorbuscha* effectively but not the larger chinook salmon. Indeed, the terminal gill-net catch of chinook salmon decreased significantly (K. Wilson, Canada Department of Fisheries and Oceans, personal communication), and replication over several years produced results consistent with the hypothesis that the new regulation was effective. However, adult chinook salmon abundance continued to decrease after the new regulation was imposed (Walters and Riddell 1986). Therefore, one cannot definitively conclude that the new mesh regulations worked, because the decrease in abundance may have accounted for some or all of the decrease in catch of chinook salmon (Wilson, personal communication).

Recommended improvements: Managers could have more accurately estimated the effectiveness of the mesh-size regulation in reducing the by-catch of chinook salmon by applying simultaneous contrasting treatments in which a portion of the fleet used large-mesh nets and the remainder used small-mesh nets. Managers could then have compared the catch per unit effort (CPUE) for chinook salmon between the two types of nets. If CPUE for the two types of nets had decreased by about the same amount after the new regulations were imposed, then it would have been clear that the overall decline in stock abundance was a major contributor to the decrease in CPUE. In contrast, if the CPUE for boats with small-mesh nets had decreased much more than that for boats with large-mesh nets, then the evidence would have been stronger that the change in mesh size had a larger effect than did the decrease in stock abundance. Statistical power analysis (Peterman 1990) could also have been used to estimate the number of boats to be equipped with large-mesh nets, so that the effect (if any) of the new regulation on by-catch of chinook salmon could be detected.

Lack of Statistical Independence

Pseudoreplication occurs when researchers use inferential statistics to test for treatment effects in data from experiments in which treatments were not replicated or replicates were not statistically independent (Hurlbert 1984). Green (1987) emphasized that a typical result of pseudoreplication is that the variance in error terms used in statistical tests is underestimated. This inaccuracy can occur if the variance among repeated samples of a single treatment—rather than properly replicated treatments—is used in the tests, or if treatment in one experimental replicate affects measured responses to treatment in other replicates and the replicates are treated as independent from one another in statistical tests.

Example.—Lack of independence among experimental units could also make it difficult for managers to interpret responses to treatments. In the Great Central Lake fertilization (enrichment) project in British Columbia, the returns of adult sockeye salmon to the nearby, unfertilized Sproat Lake, which is in the same river system, were monitored as a control. Addition of nutrients to the treated lake was expected to increase its primary and secondary productivity and thereby increase the production of sockeye salmon rearing in Great Central Lake. However, adult returns to both lakes increased substantially following treatment of Great Central Lake. LeBrasseur et al. (1978) attributed the unexpected increase in adult returns to Sproat Lake to (1) reduction in depensatory mortality factors (e.g., from predators feeding on smolts migrating from both lakes) as a result of substantially increased smolt output from Great Central Lake, (2) straying of Great Central Lake adults into Sproat Lake, or (3) unplanned simultaneous cultural eutrophication in Sproat Lake, which could have increased the food supply for juveniles. The first two possible interactions make Sproat Lake an unsuitable control in this experiment.

Lack of independence between spatial units is a universal problem for spatially replicated designs in fisheries management. Strong interactions between adjacent fish stocks from migration, shared predators, or competition for a shared food source may violate their statistical independence, so managers should be extremely careful in choosing replicates to minimize this problem.

Lack of Interspersion

Where there are several experimental units or stocks with different treatments, the use of interspersion is critical to avoid confounding results with spatial differences in properties of experimental units (Hurlbert 1984). These differences may represent an initial condition, or they may arise from spatially distinct disturbances that occur once an experiment is in progress (Hurlbert 1984). If the differences occur along a gradient, and if experimental units with different treatments are arrayed along that gradient, it may be difficult to distinguish treatment effects from other sources of variation. Spatially interspersing the different treatments across experimental units may reduce the potential for such sources of confusion. Parallel arguments also apply to the need for temporal interspersion, whereby a stock is exposed to treatment A for some period, treatment B for the next period, and then to treatment A again. This pattern reduces the possibility that the apparent initial response to treatment B could be confused with some temporal change in a variable that happened to occur simultaneously with initiation of treatment B.

Example.—As noted above, the by-catch of chinook salmon in the Fraser River decreased after the mesh size of gill nets was changed in 1981, but it was unclear whether this decrease was due to changing selectivity of nets or whether, as suggested by other data (Walters and Riddell 1986), chinook salmon abundance had simply decreased. If the simultaneous contrasting-treatment design recommended above proved unworkable, it might be possible to eliminate one of these explanations by monitoring total stock abundance and by interspersing contrasting treatments in time. After several years with the new, smaller-mesh nets, one could allow the old-style, large-mesh gill nets again for some period, which would create interspersion in time of the two treatments—a period of years of fishing with large-mesh gill nets followed by a period with small-mesh gill nets, and then back to large-mesh nets. If CPUE was high, low, and high in those respective periods, and if stock abundance was constant or showed a monotonic trend across the three periods, this would indicate that the change in mesh size was effective. But if the CPUE pattern was high, low, and low, it would suggest that decreasing fish abundance was mainly responsible for the decrease in by-catch in the second period and that the change in mesh size was not very important.

Lack of Randomization

Randomization is critical when there is potential bias in selecting replicates for experimentation from a population of potential replicates and in assigning different treatments to experimental units (Hurlbert 1984; Green 1987). Because of the many constraints on applying experimental designs in fisheries management, randomization is rarely possible. The following example represents a case of inadequate controls but also demonstrates how randomization can be useful.

Example.—The experimental design of the large-scale Lake Enrichment Program for sockeye salmon in British Columbia lacks proper controls because response variables have not been monitored in untreated (unfertilized) lakes to the same extent as in the 17 treated lakes (Hyatt and Stockner 1985). Ideally, controls and treatments should have been randomly assigned to all lakes considered suitable for treatment to avoid bias that could unduly influence the results. Lakes not chosen for enrichment may not serve as valid controls precisely for the reasons they were not chosen in the first place. For example, currently untreated coastal lakes that rear sockeye salmon generally are nonoligotrophic, much smaller than treated lakes, or fry-recruitment limited (K. Hyatt, Canada Department of Fisheries and Oceans, personal communication). Thus, rearing conditions may differ significantly between treated and untreated lakes, and the untreated lakes might be unsuitable as controls for statistical tests of the effects of lake enrichment on biomass of adult sockeye salmon. Given various assumptions, empirically based simulation modeling could be used to evaluate the economic value and information payoffs of randomization of treatments and controls and value of a monitoring program in more sockeye salmon rearing lakes.

Incomplete Measurement of Responses

In fisheries research and management, the response variables that are monitored are sometimes either uninformative or monitored inadequately (Walters 1986). Attributes of a good biological response variable include (1) linkage to the questions and hypothesis tested, (2) relevance to regulations and management as well as sensitivity to them, (3) economic value, (4) tendency to change in some consistent manner with respect to the type and level of treatment, and (5) ability to be estimated from sampling that is quick, precise, and minimal in cost. The choice of response variables and the spatial and temporal pattern in which they are measured also affect the probability of an experiment correctly detecting a response (Peterman 1990).

Example 1.—To determine whether hatcheries are effective, Foerster (1938) estimated egg-to-smolt survival rates for artificially and naturally produced sockeye salmon at Cultus Lake, British Columbia. However, because the ultimate goal of hatcheries is to produce adult sockeye salmon, measurement of egg-to-smolt survival rates may have been insufficient. Egg-to-adult survival rates would have been the appropriate response variable because density-dependent responses after smolts left the lake could have occurred, thereby decreasing the benefits from the hatchery (Peterman 1987, 1991; Guthrie and Peterman 1988). An even better design would have involved measurement of both survival rates to create a higher-power test.

Example 2.—In the British Columbia Lake Enrichment Program, scientists have not consistently measured primary and secondary production, sockeye salmon smolt sizes, and adult returns in all 17 lakes that were fertilized (Hyatt and Stockner 1985). Trophic responses and smolt sizes in geographically remote lakes are monitored less frequently than in more accessible ones, yet processes of cultural eutrophication could depend on accessibility. Thus, researchers may not be obtaining a representative picture of the effects of lake enrichment. Moreover, emphasis has been on the most readily measured responses—primary and secondary production and sizes of smolts (Hyatt and Stockner 1985). These responses provide only indirect evidence for the effect of lake enrichment on biomass of adult returns.

Monitoring of adult returns in a larger number of treated lakes and also in untreated control lakes could improve the scientific basis for lake enrichment. Currently, adult returns are monitored in only four treated lakes because of mixed-stock fisheries, remote location of spawning sites, and poor documentation of procedures to generate historical escapement estimates (Hyatt and Stockner 1985). With errors as large as 25% in estimated abundance (Hyatt and Stockner 1985) and large interannual variability in returns, it may take many years to detect possible responses in the abundance of the returns, as was found in Allen's (1980) whale example (see below). The potential net economic value of these additional monitoring efforts could be evaluated through quantitative analysis to determine if they are worth pursuing.

Low Statistical Power

Example.—White (1988) found that Foerster's (1938) Cultus Lake hatchery investigation had low statistical power (i.e., low probability of rejecting the null hypothesis of no difference between freshwater survival rates of hatchery and wild fish). Although Foerster (1938) failed to find a significant difference in survival rates, the low power of his experimental design meant that he had at least a 64% chance of making a type II error, if in fact the mean observed survival rates reflected the true values. Statistical power analysis was not known in Foerster's time, and it remains unclear whether artificial propagation significantly improved egg-to-smolt survival rates of sockeye salmon at Cultus Lake. The potentially high cost of a type II error is therefore possibly important; British Columbia may have forgone economic benefits of hatchery propagation of sockeye salmon for many years. The power of the test could have been improved by increasing the duration of the study and the number of hatchery sites employed. This example shows that in the absence of other criteria, unless the power of an experiment is expected to be high, experimentation may not be worthwhile.

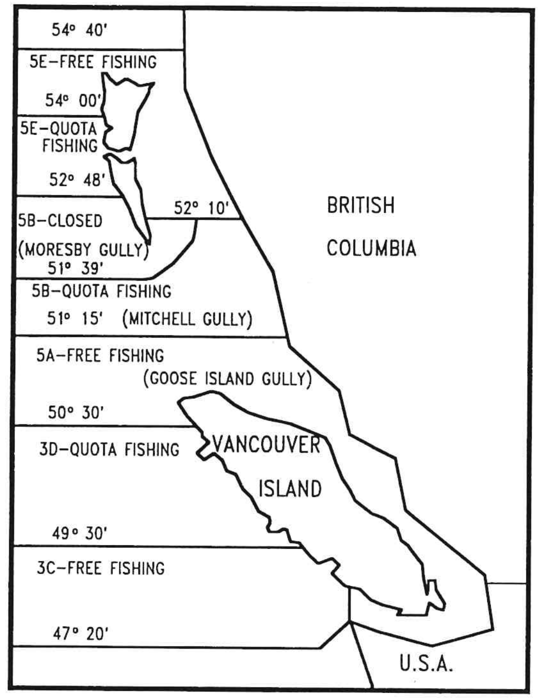

FIGURE 1.—Proposed arrangements of closed, unrestricted, and quota fishing areas for a slope rockfish experimental management program off the British Columbia coast (redrawn from Walters and Collie 1989). Areas 5E, 5B, etc., are statistical areas used by the Canadian Department of Fisheries and Oceans.

Examples of Better Experimental Management Designs

This section contains examples of management programs that address some of the experimental design problems described above. Again, these are not ideal experimental designs, but at least they help improve our knowledge of managed fish populations by eliminating some of the possible confounding interpretations.

Designs That Include Interspersion

Example.—Walters and Collie (1989) used interspersion in their proposed experimental harvest strategy for Pacific ocean perch in Canada. To obtain better estimates of stock sizes and MSYs, fishermen agreed in principle to three types of experimental fishing zones: unrestricted, quota, and no fishing (Figure 1). In the proposed design, these three different treatments are interspersed across seven fishing zones from north to south, each of which contains a distinct substock and is identified as an experimental unit. The one zone chosen for no fishing is bordered by two chosen for quota fishing to buffer the effects of unrestricted fishing in other areas and to simplify enforcement of the closure. The interspersion may help to distinguish differences in responses to quota and unrestricted fishing from latitudinal gradients that may exist in productivity and stock abundance. The arrangement also permits pairwise comparisons between adjacent quota and unrestricted fishing areas (Walters and Collie 1989).

Designs That Include Randomization

Example.—Bergh et al. (1990) evaluated two potential trawl gear experiments designed to test for an effect on catch value per hour of towing different types of trawl cod ends (diamond-mesh and knotless square-mesh nets with different mesh sizes). One experiment involved randomly assigning a single type of cod end to each vessel for each trip. The other involved a randomized block design in which each vessel was to use all types of cod ends during each fishing trip. In a complete randomized block, all types of cod ends would be used in a series of successive tows. The order of use of these cod ends would be chosen randomly—that is, if there were four types of cod ends, the first three would be picked randomly without replacement, and the fourth would be dictated by the type not picked in the first three random choices. Randomization, which occurred after the site and target species were chosen, prevented bias, for example, in the order of the choice of net types with respect to each site and species. Furthermore, the randomized block design required between 10 and 25% of the number of trips that the unblocked experimental design required to yield acceptably high statistical power in tests for the effect of gear type on catch value. The randomized block design has already been executed and has provided higher-power tests of hypotheses than originally expected (E. Pikitch, University of Washington, personal communication). The results could be instrumental in the redesign of mesh-size regulations for the demersal fishery along the Oregon and Washington coast.

Designs That Include Statistical Power Analysis

Example 1.—Peterman and Routledge (1983) showed that evaluating statistical power is essential to distinguish experimental designs that will yield valuable information from those that will not. Managers of Oregon coho salmon were uncertain whether the relationship between abundance of adult hatchery-produced coho salmon and abundance of smolts released by hatcheries is linear or shows density dependence. This is because marine survival rates are highly variable and smolt abundances vary over a relatively narrow range. Experimental management options included repeated releases of moderate or large numbers of coho smolts over varying periods. Peterman and Routledge (1983) showed that both the number of smolts released per year and the duration of the experiment critically affected the ability of an experiment to indicate whether survival rates of coho salmon in the ocean are density dependent. These estimations of power of alternative designs provide useful information to managers in their choices among alternatives where experiments and type II errors are very costly. For example, if marine survival rate is density dependent but managers assume density independence, releases of large numbers of smolts may generate low if not negative net economic benefits because costs increase with smolt releases (Peterman 1989, 1991).

Example 2.—Statistical power is also used to evaluate experiments to estimate trends in stock abundance and optimal harvest rate (Allen 1980; Gerrodette 1987; Peterman and Bradford 1987; Peterman 1990). Allen (1980) suggested that managers may learn most about optimal harvest rates of whale populations by harvesting different populations at different rates and observing responses. He evaluated whether this strategy was practical by estimating the time that would elapse before the rate of decrease in whale abundance would be significantly different from zero. Estimations were made with different α levels and levels of imprecision in the method of estimating whale abundance. Allen found that the magnitude of depletion in whale abundance that occurred before a significant decrease in abundance could be detected increased with a lower α level, a higher rate of decrease in abundance, and more imprecision in abundance estimates. In one example with a low α level (0.05), high rates of change per year (20%), and high imprecision (coefficient of variation, 30%), whale populations would be depleted severely (down to 26% of initial levels) before a significant decrease could be detected with greater than 80% probability (after 7 years). Allen (1980) emphasized the need to (1) carefully design management experiments to have high probability of producing useful results, (2) decide in advance the acceptable level of accuracy that would be sought in the results, and (3) estimate abundance in a way that minimizes the coefficient of variation.

Designs That Control for Temporal and Spatial Variability

In the following five examples, the cited authors have proposed harvest strategies designed to generate informative contrasts in the data, control for spatial variability and effects of concurrent environmental changes, and provide fail-safe plans in case of overfishing.

Example 1.—As noted above, the purpose of Walters and Collie's (1989) proposed experimental harvest strategy for Pacific ocean perch is to better estimate stock sizes and MSYs by establishing informative contrasts in the fishing effort. Survey schemes were proposed to monitor responses of each of the seven substocks in the replicated areas designated for unregulated harvest, quota management, and complete closure of fishing (Figure 1). Contingency plans were proposed to detect and respond to collapse of stocks in the unregulated zones.

Example 2.—In response to the uncertainty about the stock dynamics of whale species, some managers in the 1970s advocated a temporary moratorium on harvest. However, Holt (1977) argued that little will be learned about sustainable yields from whale stocks unless attempts are made to significantly change stock abundances. Thus, Holt (1977) proposed an experimental harvest regime in which different stocks of a whale species would be subject to differing quotas set by optimistic and pessimistic estimates of the state and productivity of the stocks. These quotas would be maintained for several years so that responses to them could be monitored. Stocks with small amounts of overlap in habitat were suggested for this experiment so that any stocks that suffered collapse could recover by gradual immigration from uncollapsed stocks.

Example 3.—Tyler et al. (1982) proposed two different experimental fishing regimes for multispecies fisheries. In the first, the objective was to find the highest sustainable effort or catch level that would maintain the viability of all species in each multispecies assemblage. The managed communities would consist of a group of similar yet spatially distinct assemblages of ecologically interdependent species. A replicate would consist of an individual assemblage. The management plan would explore a range of constant effort levels from zero to levels thought to exceed MSY. The mix of the catch would be set by fishermen and processors and would remain constant within different assemblages and over time. Areas with zero effort would serve as refugia so that migrants from them could replenish areas subject to high efforts in case of depletion.

Example 4.—The objective of the second regime of Tyler et al. (1982) was to find out how much an assemblage can be perturbed and still be able to recover its original community structure. The plan called for perturbation of one of three groups of assemblages at a time for 2 years. Different harvest efforts (some at zero for safety) would be applied to the assemblages in one group. After 2 years, the next group would be fished for 2 years and the first group would not be fished for 4 years to permit an observable recovery. Thus, each group would be fished for 2 years at a time then not fished for 4 years, and this pattern would be repeated (Figure 2). Research cruises would monitor changes in structure of the assemblage. Effort levels that threatened the presence of some key species in the assemblage would be reduced. Light effort levels would be increased toward intermediate levels that did not preclude recovery of assemblage species.

One problem common to the four examples above is that migration between units would vi-

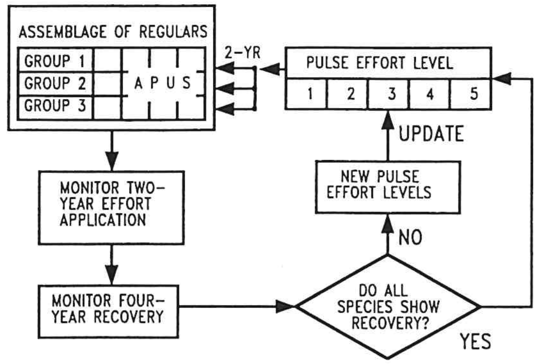

FIGURE 2.—Schematic for experimental or adaptive management of multispecies assemblages in transition states (redrawn from Tyler et al. 1982). APUs = assemblage production units.

olate their statistical independence (see below) and hence reduce the power of the experiments to detect the true response. This deficiency must be weighed against the benefit of having a fail-safe source of fish if the target stock is unexpectedly depleted.

Example 5.—Sainsbury (1988) outlined an experimental harvest strategy in a multispecies fishery on Australia's west coast. Considerable uncertainty existed about the ecological dynamics of the fishery, which involved four major demersal species, and about the harvest methods and intensity that could maximize economic yield. Extensive simulation modeling based on a Bayesian approach, plus consultation among scientists, managers, and fishermen, resulted in an experimental design in which two adjacent areas have been closed to trawl fishing for 5 years each, one starting in 1985 and the other in 1987. At the end of the closure, trap fisheries will be initiated for 5 years in each of the two areas (Figure 3). In two additional areas trawling has been maintained. The resulting design, which has spatially and temporally replicated contrasting treatments (two areas with trawling and two with trap fishing), should enable estimation of spatial and temporal variability in the responses to the management actions (Walters et al. 1988). However, the design does not control for spatial gradients in the properties of the spatial units because it does not intersperse treatments spatially (Figure 3) (see section above on interspersion).

Example 6.—In contrast to the above proposals for adaptive probing, the following experiment is proposed to test a specific hypothesis rigorously. Mean adult body weight has declined in most British Columbia pink salmon stocks in the past few

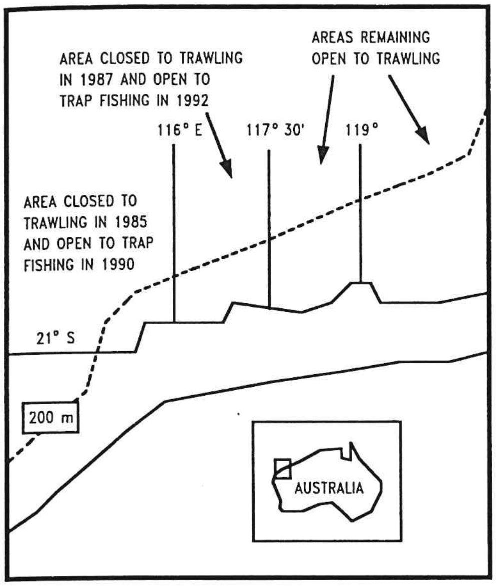

FIGURE 3.—Spatial and temporal arrangement of trawling, closed, and trap-fishing areas in the experimental demersal fishery of the northwest coast of Australia (redrawn from Sainsbury 1988).

decades (Figure 4; Ricker et al. 1978). Ricker et al. (1978) hypothesized that these trends have resulted from size-selective fishing gear. Gill nets and troll gear catch larger-than-average individuals and allow smaller individuals to escape (Ricker et al. 1978). If growth rate is heritable (Falconer 1981), as has been shown for other teleosts (Roff 1984) and for pink salmon reared under artificial conditions (Beacham and Murray 1988), the trend in body size may be an evolutionary response to repeated size-selective removal of larger-than-average fish by fishing gear. However, Ricker et al. (1978) did not exclude density-dependent or oceanographic explanations for the observed change in mean adult weight (Healey 1986). Man-

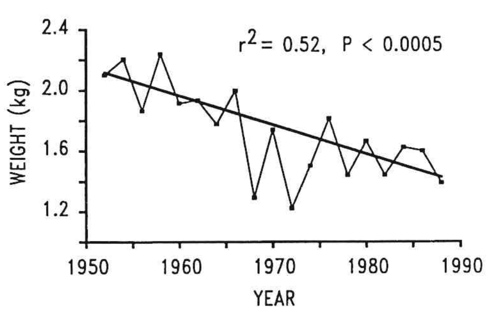

FIGURE 4.—The decrease in mean adult body weight for even-year pink salmon in statistical area 6 off the central British Columbia coast (from McAllister 1990).

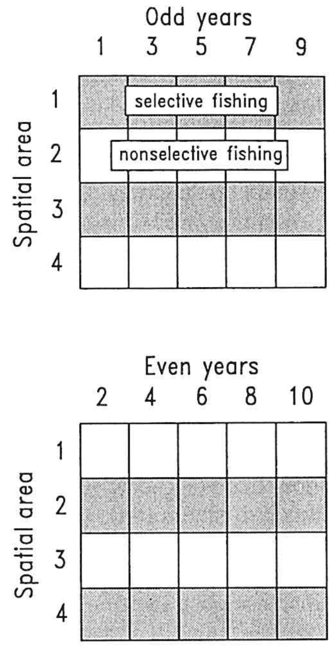

FIGURE 5.—An experimental harvest strategy to test the gear selectivity hypothesis for British Columbia pink salmon (from McAllister 1990). The fixed 2-year life cycle of pink salmon permits odd- and even-year populations to be treated as separate experimental units.

agers could rigorously test this gear selectivity hypothesis by conducting size-selective fishing with gill-net gear and nonselective fishing with seine gear in areas having separate stocks of pink salmon (Figure 5) (McAllister 1990). The trends in mean weight would be monitored in these areas. By applying different treatments simultaneously among the fishery areas, managers may be able to distinguish possible gear effects from environmental effects. Environmental effects would be present in both selectively and nonselectively fished areas, whereas gear effects would be present only in selectively fished areas. Furthermore, replication of treatments would enable managers to compute the variance and probability level of the estimate of heritability of body size (McAllister 1990).

Staircase Design to Deal with Time–Treatment Interactions

Example.—In some situations managers are prevented from distinguishing possible effects of their actions because of possible time–treatment interactions (Walters et al. 1988, 1989). Treated units may respond differently than untreated units to environmental changes. In British Columbia and Washington, several hatcheries began production of chinook and coho salmon over a decade ago. Results from coded wire tag surveys have shown substantial declines in survival rates of hatchery-produced but not of wild chinook and coho salmon (Walters and Riddell 1986). According to one hypothesis, hatchery declines represent a transient response (whose magnitude changes with time) in which high initial successes are followed by moderate sustainable production after adjustment to diseases, predation, and other limiting factors (Walters et al. 1988). However, declines have coincided with a warming trend in ocean temperatures (Walters et al. 1988). So, according to another hypothesis, hatchery-produced and wild fish are affected differently by transient environmental effects (e.g., hatchery-produced fish may be more sensitive than wild fish to a warming trend, perhaps because of differences in growth rates of hatchery and wild fish), and production levels will increase once the trend reverses (Walters et al. 1988).

To control for such time–treatment interactions, Walters et al. (1988) developed the staircase design in which treatments are initiated sequentially in separate experimental units. Monitoring would take place before initiation of treatments to detect time trends and variability in the properties to be measured that might be unique to each experimental unit. (However, it is questionable whether such time trends could be correctly assumed to continue during the experiment.) Treatment of the second experimental unit would begin 2 years after the first, for example. Treatment of the third unit would begin 3 years after the second, and so on (Figure 6). Designs can include relatively few control areas and widely spread starting times for treatments (Walters et al. 1988).

Designs in which treatments all start at the same time do not enable researchers to distinguish between transient responses to treatments and time–treatment interactions. In contrast, in a staircase design the transient responses that result from the management action (e.g., relaxation of selection for competitive ability in hatcheries) will occur at the same length of time after initiation in each experimental unit and, because the times of initiation of treatments are sequential, transient responses will not all occur in the same calendar year. On the other hand, time–treatment interactions that affect treated units independently of the time of initiation (e.g., oceanographic effects on survival rate) will show effects in the same year across all treated units and will not be apparent in control units (Figure 6). Time–treatment interactions that affect treated units only at certain periods after initiation can also be distinguished: these responses will appear only in units in which treatments are initiated at the same time; they will not appear in other units in which treatments are started at other times. Replication at each time of initiation will be required to distinguish these interaction effects from possible trends unique to each experimental unit.

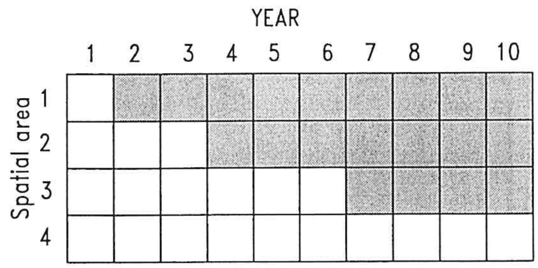

FIGURE 6.—An example of temporal and spatial arrangement of treatments in a staircase experimental design (redrawn from Walters et al. 1988). Shaded boxes show treated time–area combinations. Clear boxes show nontreated time–area combinations.

Evidence from Decision Analyses That Include Uncertainty

The most convincing arguments in favor of taking an experimental approach to fisheries management come from case studies where researchers evaluated economic benefits of a variety of management options, including status quo and experimental strategies. The most sophisticated of the analyses include (1) various biological uncertainties in the form of alternative hypotheses about the true production dynamics, weighted by their probabilities of occurrence; (2) several management options, some of which help discriminate between alternative biological hypotheses; and (3) analyses of the decision options across these hypotheses (e.g., Walters and Hilborn 1976; Walters 1977; Sainsbury 1988; McAllister 1990; Parma and Deriso 1990; Welch and Noakes 1990). For example, Sainsbury (1988) calculated that, given the uncertainty about the extent of interaction among species in the northwestern Australian groundfish community, an experimental fishing strategy (Figure 3) is likely to generate as much as 1.99 times the current value of the fishery. This result depends on the duration of the experimental trial and the assumed-true biological situation. Also, this experimental fishing strategy has a higher probability of enabling discrimination between alternative ecological models and their parameters than do other possible management policies. Similarly, McAllister (1990) estimated through extensive modeling and sensitivity analyses that a block-designed fishing experiment on pink salmon in British Columbia (Figure 5) has a high probability of detecting significant heritability in salmon body weight, which in combination with size-selective fishing might be the cause of the large decrease in body weight of fish. This experiment can also generate as much as a 45% increase in present economic value of the catch. Finally, Welch and Noakes (1990) showed that, regardless of the uncertainty concerning the degree of interaction among year-classes of Adams River sockeye salmon in British Columbia, an experiment with equal escapement targets each year (as opposed to the current wide range of escapement targets in each group of 4 years) could increase the net present value of the fishery by between Can$675 million and $3 billion, depending on the true degree of interaction among year-classes. These examples suggest the enormous potential economic benefits of experimental management.

Applicability of Experimental Design

Most fisheries situations are amenable to some form of experimental design. Even when there is a large unit stock (e.g., Pacific halibut), temporal replication and interspersion of large and small quotas, for example, could help managers discriminate more effectively among alternative hypotheses about productivity, compared with constant harvest rate policies (Parma and Deriso 1990). However, more features of good experimental design can be employed as the number of statistically independent units available for experimentation increases. This is especially true where noninteracting fish populations occur in separate ponds, lakes, or rivers. Replication of treatments and hence better designs can also be achieved in studies on unit stocks or assemblages of species when the response variables of interest relate to the operation of the fishing fleet (e.g., fishermen behavior, gear design, and gear effectiveness), instead of to the basic population biology of a stock or assemblage (Pikitch 1988; Bergh et al. 1990).

Constraints on Experimental Management

Applying an experimental approach to management, however, is not without severe constraints. For example, an experimental harvest strategy becomes less acceptable as it decreases fishing effort, increases inconvenience to fishermen, or leads to significant displacement of fishermen from their traditional fishing grounds or from the fishery altogether. In contrast, a strategy that permits temporarily increased fishing effort to test for a higher sustainable yield than previously estimated may be readily accepted because of the potential short-term gains for industry. If an experiment is to be made acceptable to fishermen, the expected payoffs from experimentation must be apparent to them. Under some circumstances, managers may make experimental harvest plans acceptable to fishermen by providing some form of compensation (Tyler et al. 1982) or by showing them that the experiment is in their best interest (e.g., the British Columbia groundfish fleet was willing to increase costs of monitoring because they believed that the experiment would lead to increased quotas) (Walters and Collie 1989). Experimental regimes with high fishing effort on some replicate stocks and low effort on others (Holt 1977) could generate useful contrasts in data while at the same time satisfying widely different positions taken by those favoring a moratorium on harvest and those favoring unrestrained effort.

The acceptability and chance of success of an experimental harvest strategy clearly depend on the cooperation of fishermen. A strategy may have the largest expected payoff over other alternatives yet may be opposed by skeptical fishermen. Fishermen may find experimentation more agreeable (1) if they have experienced significant declines in stock abundance, (2) if they engage in debate with managers over stock abundance and the appropriate level of fishing effort, or (3) if managers actively seek cooperation with them, admit to uncertainty, and inform them of management rationale (Pringle 1985; Peyton 1987; Walters and Collie 1989). The experimental idea is novel to most fisheries and may gain acceptance only if introduced on a small scale or with a set of small stocks (Tyler et al. 1982; Bergh et al. 1990).

Conclusions

This review of various management situations demonstrates that proper experimental design and good statistical practice are important means to generate information essential for improved management of fisheries. With a rigorous experimental approach, managers can learn whether stocks are being overharvested or underharvested, whether regulations are having desired effects, and how to modify harvest regimes as a result of improved information (Smith and Walters 1981; Walters 1986; Sainsbury 1988; Peterman 1990). Reduction of uncertainty about biological processes will also reduce controversy over conditions of exploited stocks and effectiveness of methods of stock enhancement or management.

It is unlikely that a perfect experimental design will ever be feasible in fisheries management, given all of the real-world constraints, so analysts must explicitly identify trade-offs among alternative designs. They should examine two or more experimental designs (with different numbers of replicate stocks or different periods of treatment, for instance) and quantify several measures for each design, including (1) the probability of correctly discriminating between the alternative biological models that are behind key areas of uncertainty, (2) the expected economic value of that design, and (3) risks. Such quantitative simulation analyses of alternative experimental designs (e.g., Sainsbury 1988; McAllister 1990) will permit managers to make informed choices of experimental management options.

Despite the constraints noted in the previous section, there is growing interest in the application of experimental design in fisheries management. The experimental designs of Sainsbury (1988), Walters and Collie (1989), and Bergh et al. (1990) for demersal fisheries demonstrate that fishermen, as well as managers, are willing to cooperate. Also, the gap between theory and practice could soon narrow; for example, managers of Fraser River sockeye salmon in British Columbia (Collie et al. 1990) and members of the Northwest Power Planning Council in Portland, Oregon, who are attempting to rehabilitate Columbia River basin salmonids (Lee and Lawrence 1986; Orians 1986) recently have expressed interest in experimental management. Programs that combine good experimental design with comprehensive, quantitative assessments of biological uncertainties, policy options, and potentials for learning will undoubtedly uncover many other situations in which managers and the fishing industry are likely to benefit from experimental management. We hope this review will stimulate more genuine interest in experimental management and help erode existing barriers to its implementation.

Acknowledgments

We are grateful to Kim D. Hyatt, Michael A. Henderson, and Ken Wilson for information concerning various fisheries. We thank Jeremy Collie, Robert Crittenden, Scott Forbes, Michael Henderson, Michael Lapointe, Richard Lockhart, Michael Link, Shane Frederick, Ray Hilborn, Carl Walters, and an associate editor of this journal for their helpful comments.

References

Allen, K. R. 1980. Conservation and management of whales. University of Washington Press, Seattle.

Beacham, T. D., and C. B. Murray. 1988. A genetic analysis of body size in pink salmon (*Oncorhynchus gorbuscha*). Genome 30:31–35.

Berger, J. O. 1985. Statistical decision theory and Bayesian analysis. Springer-Verlag, New York.

Berger, J. O., and D. A. Berry. 1988. Statistical analysis and the illusion of objectivity. American Scientist 76:159–165.

Bergh, M. O., E. K. Pikitch, J. R. Skalski, and J. R. Wallace. 1990. Statistical design of comparative fishing experiments. Fisheries Research 9:143–163.

Bilton, H. T., D. F. Alderdice, and J. T. Schnute. 1982. Influence of time and size at release of juvenile coho salmon (*Oncorhynchus kisutch*) on returns at maturity. Canadian Journal of Fisheries and Aquatic Sciences 39:426–427.

Buckley, R. M. 1989. Habitat alterations as a basis for enhancing marine fisheries. California Cooperative Oceanic Fisheries Investigations Reports 30:40–45.

Burkenroad, M. D. 1948. Fluctuations in abundance of Pacific halibut. Bulletin of the Bingham Oceanography Collection, Yale University 11(4):81–129.

Carpenter, S. R. 1990. Large-scale perturbations: opportunities for innovation. Ecology 71:2038–2043.

Cohen, J. 1988. Statistical power analysis for the behavioral sciences, 2nd edition. L. Erlbaum Associates, Hillsdale, New Jersey.

Colby, P. J., P. A. Ryan, D. H. Schupp, and S. L. Serns. 1987. Interactions in north-temperate lake fish communities. Canadian Journal of Fisheries and Aquatic Sciences 44(supplement 2):104–128.

Collie, J. S., R. M. Peterman, and C. J. Walters. 1990. Experimental harvest policies for a mixed stock fishery: Fraser River sockeye salmon (*Oncorhynchus nerka*). Canadian Journal of Fisheries and Aquatic Sciences 47:145–155.

Dixon, W. J., and F. J. Massey, Jr. 1969. Introduction to statistical analysis, 3rd edition. McGraw-Hill, New York.

Edwards, W., H. Lindman, and L. J. Savage. 1963. Bayesian statistical inference for psychological research. Psychological Review 70:193–242.

Falconer, D. S. 1981. Introduction to quantitative genetics, 2nd edition. Longman, New York.

Foerster, R. E. 1938. An investigation of the relative efficiencies for natural and artificial propagation of sockeye salmon (*Oncorhychus nerka*). Journal of the Fisheries Research Board of Canada 4:151–161.

Foerster, R. E. 1968. Sockeye salmon (*Oncorhynchus nerka*). Fisheries Research Board of Canada Bulletin 162:1–422.

Fukuda, Y. 1962. On the stocks of halibut and their fisheries in the northeastern Pacific. Bulletin of the

International North Pacific Fisheries Commission 7:39–50.

Gerrodette, T. 1987. A power analysis for detecting trends. Ecology 68:1364–1372.

Glantz, M. H., and J. D. Thompson, editors. 1981. Resource management and environmental uncertainty: lessons from coastal upwelling fisheries. Wiley, New York.

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

Green, R. H. 1984. Statistical and nonstatistical considerations for environmental monitoring studies. Environmental Monitoring and Assessment 4:293–301.

Green, R. H. 1987. Statistical and mathematical aspects: distinction between natural and induced variation. Pages 335–354 *in* V. B. Vouk, G. C. Butler, A. C. Upton, D. V. Parke, and S. L. Asher, editors. Methods for assessing the effects of mixtures of chemicals. Wiley, Chichester, UK.

Green, R. H. 1989. Power analysis and practical strategies for environmental monitoring. Environmental Research 50:195–205.

Guthrie, I. C., and R. M. Peterman. 1988. Economic evaluation of lake enrichment strategies for British Columbia sockeye salmon. North American Journal of Fisheries Management 8:442–454.

Healey, M. C. 1986. Optimum size and age at maturity in Pacific salmon and effects of size-selective fisheries. Canadian Special Publication of Fisheries and Aquatic Sciences 89:39–52.

Holling, C. S., editor. 1978. Adaptive environmental assessment and management. Wiley, Chichester, UK.

Holt, R. S., T. Gerrodette, and J. B. Cologne. 1987. Research vessel survey design for monitoring dolphin abundance in the eastern tropical Pacific. U.S. National Marine Fisheries Service Fisheries Bulletin 85:435–446.

Holt, S. J. 1977. Whale management policy. Report of the International Whaling Commission 27:133–137.

Howson, C., and P. Urbach. 1989. Scientific reasoning: the Bayesian approach. Open Court, La Salle, Illinois.

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54:184–211.

Hyatt, K. D., and J. E. Stockner. 1985. Responses of sockeye salmon (*Oncorhynchus nerka*) to fertilization of British Columbia lakes. Canadian Journal of Fisheries and Aquatic Sciences 42:320–331.

Keeney, R. L., and H. Raiffa. 1976. Decisions with multiple objectives. Wiley, New York.

Ketchen, K. S. 1956. Climatic trends and fluctuations in yield of marine fishes in the northeast Pacific. Journal of the Fisheries Research Board of Canada 13:357–374.

Larkin, P. A. 1972. A confidential memorandum on fisheries science. Pages 189–197 *in* B. J. Rothschild, editor. World fisheries policy: multi-disciplinary views. University of Washington Press, Seattle.

LeBrasseur, R. J., and six coauthors. 1978. Enhancement of sockeye salmon (*Oncorhynchus nerka*) by lake fertilization in Great Central Lake: summary and report. Journal of the Fisheries Research Board of Canada 35:1580–1596.

Lee, K. N., and J. Lawrence. 1986. Adaptive management: learning from the Columbia River Basin Fish and Wildlife Program. Environmental Law 16:431–460.

Lindley, D. V. 1985. Making decisions. Wiley, New York.

Loftus, K. H. 1976. Science for Canada's fisheries rehabilitation needs. Journal of the Fisheries Research Board of Canada 33:1822–1857.

Ludwig, D., and R. Hilborn. 1983. Adaptive probing strategies for age-structured fish stocks. Canadian Journal of Fisheries and Aquatic Sciences 40:559–569.

McAllister, M. 1990. Decision and power analyses of large-scale fishing experiments on British Columbia pink salmon (*Oncorhynchus gorbuscha*). Master's thesis. Simon Fraser University, Burnaby, British Columbia. [Also see: McAllister et al. and McAllister and Peterman. In press. Canadian Journal of Fisheries and Aquatic Sciences 49(7).]

Mercer, M. C., editor. 1982. Multispecies approaches to fisheries management advice. Canadian Special Publication of Fisheries and Aquatic Sciences 59.

Millard, S. D., and D. P. Lettenmaier. 1986. Optimal design of biological sampling programs using analysis of variance. Estuarine, Coastal and Shelf Science 22:637–657.

Orians, G. H. 1986. The place of science in environmental problem solving. Environment 28:12–41.

Parma, A. M., and R. B. Deriso. 1990. Experimental harvesting of cyclic stocks in the face of alternative recruitment hypotheses. Canadian Journal of Fisheries and Aquatic Sciences 47:595–610.

Pearson, E. S., and H. O. Hartley, editors. 1976. Biometrika tables for statisticians, volume 2. Charles Griffin, High Wycombe, UK.

Peterman, R. M. 1987. Review of the components of recruitment of Pacific salmon. American Fisheries Society Symposium 1:417–429.

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. 1990. Statistical power analysis can improve fisheries research and management. Canadian Journal of Fisheries and Aquatic Sciences 47:2–15.

Peterman, R. M. 1991. Density-dependent marine processes in North Pacific salmonids: lessons for experimental design of large-scale manipulations of fish stocks. Pages 69–77 *in* S. Lockwood, editor. The ecology and management aspects of extensive aquaculture. International Council for the Exploration of the Sea, Marine Science Symposium 192, Copenhagen.

Peterman, R. M., and M. Bradford. 1987. Statistical power of trends in fish abundance. Canadian Journal of Fisheries and Aquatic Sciences 44:1879–1889.

Peterman, R. M., and R. D. Routledge. 1983. Experimental management of Oregon coho salmon: designing for yield of information. Canadian Journal of Fisheries and Aquatic Sciences 40:1212–1223.

Peyton, R. B. 1987. Mechanisms affecting public acceptance of resource management policies and strategies. Canadian Journal of Fisheries and Aquatic Sciences 44:306–312.

Pikitch, E. K. 1988. Objectives for biologically and technically interrelated fisheries. Pages 107–136 *in* W. S. Wooster, editor. Fisheries science and management: objectives and limitations. Springer-Verlag, New York.

Pringle, J. D. 1985. The human factor in resource management. Canadian Journal of Fisheries and Aquatic Sciences 42:389–392.

Raiffa, H. 1968. Decision analysis. Introductory lectures on choices under uncertainty. Addison-Wesley, Reading, Massachusetts.

Reckhow, K. H. 1990. Bayesian inference in non-replicated ecological studies. Ecology 71:2053–2059.

Ricker, W. E., H. T. Bilton, and K. V. Aro. 1978. Causes of the decline in size of pink salmon (*Oncorhynchus gorbuscha*). Canada Fisheries and Marine Service Technical Report 820.

Roff, D. A. 1984. The evolution of life history parameters in teleosts. Canadian Journal of Fisheries and Aquatic Sciences 41:989–1000.

Sainsbury, K. J. 1988. The ecological basis of multispecies fisheries and management of a demersal fishery in tropical Australia. Pages 349–382 *in* J. A. Gulland, editor. Fish population dynamics, 2nd edition. Wiley, New York.

Silvert, W. 1978. Price of knowledge: fisheries management as a research tool. Journal of the Fisheries Research Board of Canada 35:208–212.

Skud, B. E. 1985. The history and evaluation of closure regulations in the Pacific halibut fishery. FAO (Food and Agriculture Organization) Fisheries Report 289(3):449–456.

Smith, A. D. M., and C. J. Walters. 1981. Adaptive management of stock–recruitment systems. Canadian Journal of Fisheries and Aquatic Sciences 38: 690–703.

Stewart-Oaten, A., W. W. Murdoch, and K. R. Parker. 1986. Environmental impact assessment: "pseudoreplication" in time? Ecology 64:929–940.

Toft, C. A., and P. J. Shea. 1983. Detecting community-wide patterns: estimating power strengthens statistical inference. American Naturalist 122:618–625.

Tyler, A. V., W. L. Gabriel, and W. J. Overholtz. 1982. Adaptive management based on structure of fish assemblages of northern shelves. Canadian Special Publication of Fisheries and Aquatic Sciences 59: 149–156.

Vaughan, D. S., and W. Van Winkle. 1982. Corrected analysis of the ability to detect reductions in year-class strength of the Hudson River white perch (*Morone americana*) population. Canadian Journal of Fisheries and Aquatic Sciences 39:782–785.

Walters, C. J. 1977. Management under uncertainty. Pages 261–297 *in* D. V. Ellis, editor. Pacific salmon management for people. University of Victoria Press, Victoria, British Columbia.

Walters, C. J. 1986. Adaptive management of renewable resources. Macmillan, New York.

Walters, C. J., and J. S. Collie. 1989. An experimental strategy for groundfish management in the face of large uncertainty about stock size and production. Canadian Special Publication of Fisheries and Aquatic Sciences 108:13–25.

Walters, C. J., J. S. Collie, and T. Webb. 1988. Experimental design for estimating transient responses to management disturbances. Canadian Journal of Fisheries and Aquatic Sciences 45:530–538.

Walters, C. J., J. S. Collie, and T. Webb. 1989. Experimental design for estimating transient responses to habitat alterations: is it practical to control for environmental interactions? Canadian Special Publication of Fisheries and Aquatic Sciences 105:13–20.

Walters, C. J., and R. Hilborn. 1976. Adaptive control of fishing systems. Journal of the Fisheries Research Board of Canada 33:145–159.

Walters, C. J., and R. Hilborn. 1978. Ecological optimization and adaptive management. Annual Review of Ecology and Systematics 9:157–188.

Walters, C. J., and C. S. Holling. 1990. Large-scale management experiments and learning by doing. Ecology 71:2060–2068.

Walters, C. J., and B. Riddell. 1986. Multiple objectives in salmon management: the chinook sport fishery in the Strait of Georgia, B.C. Northwest Environmental Journal 2:1–15.

Welch, D. W., and D. J. Noakes. 1990. Cyclic dominance and optimal escapement of Adams River sockeye salmon (*Oncorhynchus nerka*). Canadian Journal of Fisheries and Aquatic Sciences 47:838–849.

White, B. A. 1988. Benefit–cost evaluation of sockeye salmon (*Oncorhynchus nerka*) hatchery on the Nahmint River, British Columbia. Master's thesis. Simon Fraser University, Burnaby, British Columbia.