In: Perspectives on the Marine Environment, Proceedings from a Symposium on the Marine Environment of Southern California, P.M. Grifman and S.E. Yoder, eds. USC Sea Grant Program, Los Angeles, CA, p. 75-89, 1992.

Implications For The Design Of Environmental Assessment Studies

Craig W. Osenberg
Coastal Research Center, Marine Science Institute

Sally J. Holbrook
Coastal Research Center, Marine Science Institute, and
Department of Biological Sciences

Russell J. Schmitt
Coastal Research Center, Marine Science Institute
University of California, Santa Barbara

Abstract. Many types of environmental impact assessment studies aim to detect effects of localized impacts. Most of the assessment (or compliance monitoring) designs used in such studies fail to distinguish effects of anthropogenic versus natural origins, and thus might lead to incorrect interpretations. The Before-After- Control-Impact-Paired (BACIP) design surmounts this and other problems, yet has rarely been used in assessment studies. For the past three years, we have used BACIP to study possible effects of nearshore discharge of produced water, an aqueous waste generated during oil production. Results from power analyses suggest that environmental impacts are more likely to be detected for physical and chemical parameters than for biological measures; within biological parameters, effects on individual-based properties (e.g., growth, fecundity) are more likely to be detected than changes in population densities. However, regulatory agencies and resource managers ultimately are concerned with impacts on populations and communities. Our results emphasize the need to: (i) collect adequate (time-series) data before a localized perturbation begins, (ii) understand mechanisms that lead to population change and (iii) develop comprehensive models of processes leading to environmental impacts.

INTRODUCTION

There continues to be considerable debate regarding localized effects of anthropogenic disturbances on marine biotic resources. Controversy arises, in part, from equivocal data obtained from poorly designed environmental assessment studies. Despite great efforts to obtain reliable information, most assessment designs currently in use fail to provide rigorous and convincing tests of possible effects [1]. Most of these studies fall short because the assessment designs employed do not separate changes in ecological systems caused by the putative impact from changes resulting from natural spatial and/or temporal variation [1,2,3,4]. Consequently, it is often difficult to draw scientifically defensible conclusions about the existence or magnitude of a localized environmental impact.

Even when seriously flawed, many assessment designs may be sufficient to "demonstrate" large, severe impacts. However, such dramatic (qualitative) changes in marine ecosystems are unlikely to be common due to regulatory steps such as strict permitting conditions and relatively stringent monitoring of effluents. Of more concern are long-term chronic effects, which can involve quantitative changes (e.g., reductions in population abundance) and which can accrue slowly. By their nature, chronic impacts are more insidious, less easily isolated from natural variability, and therefore demand more rigorous assessment designs to detect. Few assessment studies have used such rigorous designs, and as a result, regulatory agencies typically lack the sufficient scientific information to make environmentally sound decisions concerning management of marine resources.

In this paper, we begin by discussing the goal of environmental assessment studies, and then illustrate how the three most commonly used assessment designs fail to satisfy these goals. We review an alternate design, which controls for many types of natural spatial and temporal variation and therefore provides a more defensible approach to environmental impact assessment. This alternate design reduces the possibility of wrongly concluding that an impact has occurred. However, a concern generic to all assessment designs — that actual impacts might go undetected — still remains whenever statistical power is low. We consider the issue of power, and suggest that the ability to detect actual effects may vary systematically with the type of parameter measured. These results have fundamental implications for interpreting results of all assessment studies.

We illustrate many of our points with data from our ongoing study of possible environmental effects of nearshore discharge of "produced water" on benthic marine organisms. Produced water is an aqueous waste generated during oil production, and is contaminated with various petroleum hydrocarbons, heavy metals and other inorganic chemicals, as well as additives (including biocides) introduced to increase the separation of produced water from crude oil [5,6]. Since January 1988, we have been conducting a detailed study to assess whether environmental impacts result from the discharge of produced water. The system we focus on is a soft-bottom community occurring near Gaviota, California at a bottom depth of approximately 25 m. Although the produced water study is specific in its primary intent, the message of this paper is relevant to the study of localized environmental impacts in general.

THE GOALS OF ENVIRONMENTAL IMPACT ASSESSMENT

In general, the question to be answered by an assessment study of a localized perturbation (e.g., wastewater discharge) is: "How does the ecosystem at the site of perturbation differ from the ecosystem that would have existed had the perturbation never occurred?" Obviously, the answer cannot be obtained by direct observation, and the goal of an assessment design should be to estimate the state of the system that would have existed in the absence of the perturbation [7]. Further, this estimate should be statistically compared to the observed condition (in the presence of the perturbation) and a probability should be assigned that the estimated effect might have arisen by chance (i.e., due to natural variability in the absence of an impact). If a statistically significant result is not obtained, it is absolutely critical to estimate the "power" of the test, which is the probability that the analysis could have detected an impact had it occurred.

Table 1

Two types of errors committed in environmental assessment studies

Type of Error	Conclusion	Reality
False Implication	"Impact"	None
False Exoneration	"None"	Impact

There are two types of errors that can be made in interpreting results from an assessment study (Table 1). We call the first type of error, "False Implication." It arises when we conclude that a perturbation has resulted in an environmental impact when in reality the effects we see arose for another reason (e.g., due to natural variability in the system). The second type of error is "False Exoneration," in which we conclude there has been no impact, but in fact there has been one. The first error might result in unnecessary regulation of environmentally safe projects, while the second might fail to alert regulators to environmental impacts that require prevention or mitigation. Each type of error can have serious implications and should be minimized within constraints imposed by the study. We now consider the relative merits of several impact assessment designs.

THREE COMMON ASSESSMENT DESIGNS AND THEIR LIMITATIONS

A widely used assessment design, often employed in compliance monitoring in the state of California, is one in which an impact site (or a gradient of impact sites) is sampled and compared to a more distant control site(s) after a perturbation has begun. We refer to this as the "Control-Impact" design. Differences in parameters of interest (e.g., population densities) between the sites are taken to represent effects due to the perturbation. However, ecological systems exhibit considerable spatial variation and it is not possible to reliably interpret any difference between sites as being due to the perturbation: differences might exist for a number of possible reasons. For example, we have estimated densities of a large epifaunal gastropod (*Kelletia kelletii*) at our Gaviota study sites. Figure 1 shows densities at two impact sites (50 m and 250 m downcurrent from the diffusers) and at a control site (1500 m upcurrent from the diffusers). Clearly, gastropods were less abundant at the impact sites, and it might be concluded that produced water discharge negatively affected gastropod density. However, at the time these data were collected, produced water had never been discharged. In fact, the differences in densities between control and impact sites (Figure 1) were simply the result of other processes that led to spatial variation among the sites. Had these data been collected after discharge, the "Control-Impact" design could have led to the false implication of an impact (Table 1).

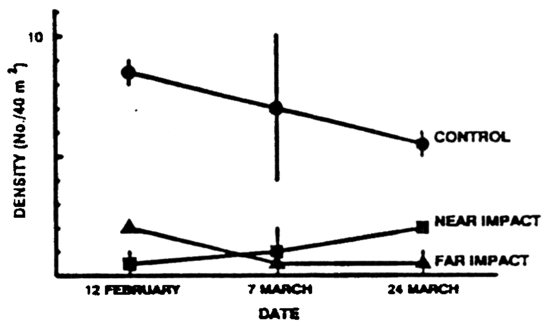

Figure 1 . Density of the gastropod, *Kelletia kelletii*, at three sites over time. The near (square) and far (triangle) impact sites are located 50 m and 250 m downcurrent of a produced water outfall respectively, and the control site (circle) is 1500 m upcurrent. No produced water had ever been discharged at these sites when the data were collected. Shown for each date are the mean and range of gastropod density (N = 2 band transects per site).

A second approach compares the condition of an impact site before the perturbation occurred with the condition of the site after the perturbation. This we call the "Before-After" design. Although this design circumvents problems associated with natural *spatial* variation (as discussed above), it instead ignores natural *temporal* variation, which is also ubiquitous in nature. To illustrate this problem we use data collected by the Marine Review Committee in a study of the San Onofre Nuclear Generating Station, SONGS [8,9]. Densities of pink surfperch were estimated over time before and after new units at SONGS began generating power [8]. The density of pink surfperch declined markedly, with the reduction coinciding with the commencement of power generation by the new units (Figure 2). It is tempting to conclude from these data that the operation of the new units (accompanied by discharge of cooling water) negatively affected the surfperch. However, these data were taken from a control site 18 km from SONGS. In reality, this temporal change in density occurred at all sites (control and impact alike), probably in response to El Niño [10]. Thus, in this case the "Before-After" design would have led to false implication because it failed to separate impacts from temporal variability introduced from natural sources.

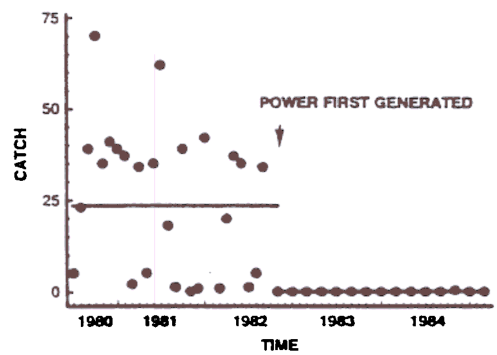

Figure 2 . Density (catch per otter trawl) of pink surfperch, *Zalembius rosaceus*, over time at a location 18 km from the San Onofre Nuclear Generating Station (SONGS). The arrow indicates the first date on which power was generated by two new units of SONGS. Mean densities from the before and after periods are indicated by the solid lines.

One potential solution to the limitations of the "Control-Impact" and "Before-After" designs is to combine them into a single design in which control and impact sites are sampled both before and after a perturbation occurs. In this case, the test for an impact is conducted by asking whether the condition of the impact site relative to the control has changed from the before period to the after period. Green [11] proposed such a design, which he called the "Optimal Impact Assessment" design, but unfortunately recommended an inappropriate statistical test. He suggested using an error term based upon the observed error among all samples collected within a site during a particular period (e.g., replicate samples collected on a single date). For this test to work as designed, it requires the stringent assumption that differences in densities (or other parameters) between the control and impact sites remain exactly the same at all times. However, we know that sites exhibit unique temporal fluctuations under natural conditions. This natural variability, combined with sampling error, comprises the variation from which impacts must be distinguished. Green's design considers only the influence of sampling error. This shortcoming of the "Optimal Impact Assessment" design has been cogently pointed out by Stewart-Oaten et al. [12; see also 13], who noted that the design fails to separate local temporal variability of systems (which arise naturally), from long-term effects indicative of an environmental impact. For example, with sufficiently intensive sampling on two different dates (one in the Before and one in the After periods), use of within-site variation among "replicates" will always yield a significant relative change at the control and impact sites even in the absence of any anthropogenic disturbance.

To illustrate, we conducted such a test for data collected from our study of produced water impacts. The density of the seapen, *Acanthoptilum* sp., was sampled at both the control and impact sites during 1988 and 1990 (Figure 3). These two periods bracketed a projected date on which discharge of produced water was to begin and thus were expected to represent before and after conditions. An "Optimal Impact Assessment" test yielded a significant Site x Period interaction ($F_{1,78}$= 7.26, $P<$ 0.01), suggesting that an impact had occurred at the outfall site. However, commencement of discharge was delayed and did not actually occur during this sampling interval. Therefore, in this instance the "Optimal Impact Assessment" design could have led to false implication.

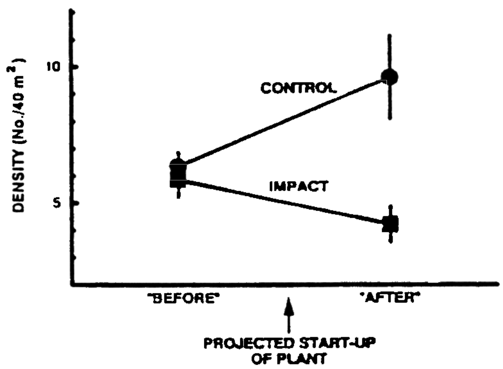

Figure 3. Density of the seapen, *Acanthoptilum* sp., at two sites. The control site is located 1500 m upcurrent, and the impact site 50 m downcurrent, of a produced water outfall. "Before" data are from 1988, and "After" data are from 1990; because of delays, discharge of produced water did not begin as expected between the two surveys. Shown are means (+ SE) using all observations within a period as replicates.

The central problem with the "Optimal Impact Assessment" design is that replicate samples (collected within a date or over a short time span) do not necessarily provide new and independent estimates of the general state of the impact or control sites. Instead, we require estimates obtained on many different dates sufficiently separated in time that data satisfy assumptions of independence. In other words, replication needs to be achieved through time (multiple sampling dates during the before and the after periods) and each replicate observation must be an independent estimate of the average environmental condition [12]. There is a fundamental lack of appreciation for this crucial aspect, yet it distinguishes a proper design from one with superficial similarity.

A PREFERABLE APPROACH
THE BEFORE-AFTER-CONTROL-IMPACT-PAIRED (BACIP) DESIGN

The fourth design we will discuss is the "Before-After-Control- Impact-Paired" (BACIP) design [12,14]. BACIP is somewhat similar in design to the "Optimal Impact Assessment" design, but it explicitly requires that sampling be conducted during several times in the before and after periods at both control and impact sites. For a given parameter (e.g., density), the variate of interest is the *difference* in a parameter value between the control and impact sites on a given date (e.g., population density at the control site minus density at the impact site). The measure of error in the statistical test is the variability of this difference, as assessed through repeated sampling *in time*. A number of assumptions must be satisfied to apply BACIP, and these assumptions (such as independence) have been rigorously elaborated by Stewart-Oaten and co-workers [12,14,15].

BACIP, relative to the other three designs, is most likely to isolate local impacts (e.g., from discharge of produced water) from natural sources of spatial and temporal variation. BACIP controls for the effect of spatial variation by measuring the average difference between the sites during the before period, and uses this difference as an estimate of the expected difference during the after period, assuming no impact. By focusing on differences, BACIP also removes the effect of temporal variation that affects both sites simultaneously (e.g., El Niño, winter storms). In essence, these temporal effects cancel upon subtracting the control and impact values. Finally, BACIP explicitly recognizes that sites fluctuate uniquely through time (i.e., exhibit Site x Time interactions) and therefore uses the variation through time in the difference between the control and impact sites as the estimate of error in the statistical test of an impact. This is the fundamental advantage gained by using a true BACIP design.

Despite its importance in environmental assessment, BACIP is relatively unappreciated as evidenced by its absence in recent discussions of assessment designs sponsored by National Oceanic and Atmospheric Administration and the National Science Foundation [1,13], the Environmental Protection Agency [16], the American Petroleum Institute [5], and the Minerals Management Service, the California State Water Resources Control Board, and the National Academy of Sciences [4]. Even though the development of this assessment design is rather simple and can be traced back to at least 1966 [17], BACIP has rarely been used or even discussed [but see 3,7,18,19,20]. Unfortunately, fundamentally flawed designs such as the "Optimal Impact Assessment" design still motivate large, very expensive assessment programs [e.g., 21] and can lead to erroneous interpretation of environmental impacts.

Both false implication and false exoneration (Table 1) can be costly, and a well designed assessment study should explicitly address the commission of both types of errors. The probability of false implication is greatly reduced using BACIP (relative to the other designs discussed) because the impact is less likely to be confused with natural sources of variability. False exoneration remains a concerns with BACIP (as it does for any design) because there will arise situations in which there is insufficient evidence to statistically reject the null hypothesis of "no impact." In these situations it is tempting to conclude that there was in fact, no impact. In the absence of additional information, this potentially is a dangerous conclusion because there often can be substantial impacts that go undetected. Failure to detect such impacts arises when considerable variability in the system introduces a large error term in the statistical test. The probability of false exoneration is equivalent to the statistical Type II error rate (β). The power of a test is $1 - \beta$, which gives the probability of correctly concluding there has been an impact when an impact of a given size has actually occurred. This explicit specification of the power adds greatly to the interpretation of the test. We now turn to evaluation of power in our BACIP study of produced water effects.

STATISTICAL POWER

The power of a statistical test of an environmental impact (using a BACIP design) is influenced by four statistical attributes: (1) the Type I error rate (we assume here $\alpha = 0.05$), (2) the number of sampling dates (i.e., true replicates — the number of independent estimates of the difference between the control and impact sites), (3) the variability of these estimates (which we term S_Δ), and (4) the size of the impact that we wish to be able to detect. In general, power is high (closer to 1) when the number of survey dates (replicates) is large (Figure 4), the variability of differences (within the before and after periods) is low (Figure 4), and the anticipated impact is large.

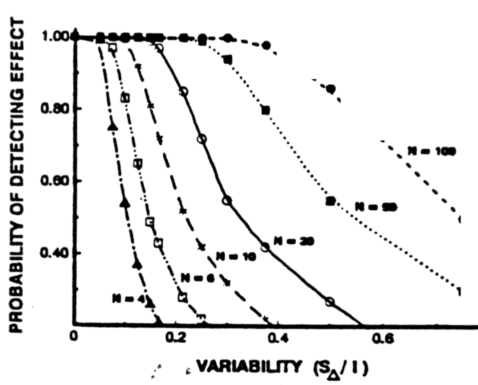

Figure 4. Effect of sample size and variability on statistical power to detect a 30% reduction (relative to control) in population density at the impact site. Given are power curves (after [22]) for 6 sample sizes (total number of sampling dates allocated equally to the before and after periods). Variability is expressed in a standardized form as the standard deviation of the differences between the control and impact sites (S_Δ), divided by the mean density at the impact site (I).

Because power analyses have rarely been applied to BACIP studies [but see 9], and because of continued confusion about the source of variability that is important in tests of impacts (see above), we illustrate the effect of high and low variability on power in Figure 5. In this case, we assumed that an anthropogenic perturbation caused a reduction in density at the impact site of 30%. In the left panel of Figure 5, we assumed that the difference between the control and impact sites varied considerably among sampling dates, while in the right panel we assumed that this variability was considerably less. As expected, it is more difficult to detect the 30% reduction in the case where there is high variability in the difference between control and impact sites (i.e., the power of the test is low). Notice that in both panels, the amount of temporal variability *within* a site is similar; the important distinction between these examples is the amount of variation expressed in the differences between control and impact sites. Further, nothing is assumed about spatial variation that exists within a date at a given site.

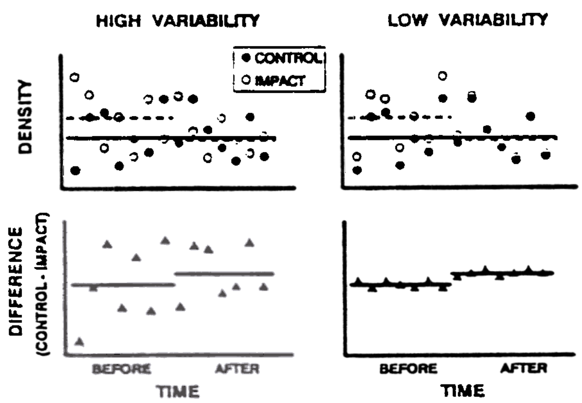

Figure 5. Effect of high and low variability in differences between the control and impact sites on the ability to detect an impact. The two top panels give hypothetical densities at control (solid circle) and impact (open circle) sites during the before and after periods; the bottom panels show the differences. Lines show means in each time period for control sites (solid), impact sites (dashed), and the mean difference (lower graph). Scales for left and right graphs are the same. Variance in densities at each site within a period is identical under the high and low variability scenarios. The degree of temporal consistency in the two sites differs between left and right graphs; under low variability, the control and impact sites are affected more similarly through time than under high variability. Although in each scenario the impacts are of the same size, high variability masks the impact; the impact can be much more easily detected under low variability.

The preceding discussion implies that greater power will arise if temporal changes in the value of a parameter track one another at control and impact sites. If qualitatively different classes of parameters (i.e., physical, chemical, biological) have consistently different patterns of this variability, the power of a test will depend on the particular type of parameter being examined. We have conducted power analyses for a number of parameters estimated at our near impact and control sites near Gaviota. These include population-based parameters (e.g., densities of macroinvertebrates and of infauna, emergence and re-entry rates of demersal zooplankton (estimated using methods of Alldredge and King [23] and Stretch [24]), individual-based parameters (e.g., mean body size, gonadal- somatic index), and physical and chemical parameters (e.g., sedimentation rate, percent organic matter in sediments, grain size of sediments). Here we summarize overall patterns, then illustrate specific conclusions using the white sea urchin, *Lytechinus anamesus.*

Our results indicate that, in general, power to detect impacts on population-level phenomena is relatively weak (compare Table 2 and Figure 4). The average (across species) variability in the difference in population density between sites was particularly large (Table 2), which greatly reduces power. For example, to detect an impact on the density of *Lytechinus* with 80% likelihood (assuming 25 sampling dates in each period), *Lytechinus* densities would have to decline (relative to the control site) by approximately 75%. Although *Lytechinus* provides one of the more extreme examples, power to detect impacts on population densities for most species we have examined also is low. There are, however, some species for which we have relatively high power (80%) to detect effects on density that are comparatively small — on the order of 25% (assuming 25 sampling dates each in the Before and After periods). The average variability for other population-level parameters (i.e., re-entry and emergence rates) was similarly large (Table 2).

As indicated by the lower variability for individual-based measures (Table 2), we have much greater power to detect impacts on such parameters as body size or gonadal-somatic index (GSI). To illustrate using *Lytechinus*, variability in density was approximately 1.5, while variability in mean test diameter was only 0.04. In general, no population-based parameter we investigated has power exceeding that calculated for these individual-based parameters (compare range in variability for various parameters on Table 2). This result — that impacts on individual-based parameters are more likely to be detected than those on population parameters — previously has been suggested [1], but we know of no other data or analyses that explicitly addressed the issue.

Table 2

Relative Variation in Differences Between Control and Near Impact Sites

The standard deviation of differences is standardardized across the various parameters by dividing by the mean parameter value at the Near Impact site. Given is the mean variability (and range) for each type of parameter.

	Index Of Variability (S_Δ / I)	
	Mean	Range
Population-based Biological Parameters		
Population density:	0.69	0.26 - 2.04
Re-entry Rate:	0.57	0.43 - 0.73
Emergence Rate:	0.49	0.44 - 0.58
Individual-based Biological Parameters	0.14	0.04 - 0.25
Physical and Chemical Parameters	0.12	0.05 - 0.19

Because of low levels of variability for physical and chemical measures (Table 2), we have similar, relatively high power to detect changes in these parameters as we have for individual-based parameters. For each of 4 physical/chemical parameters analyzed thus far, we again have greater power to detect a given size impact than for any of the population-based biological parameters

There are at least two explanations for our result that power is greater for impacts on physical/chemical parameters than on biological parameters, and that within biological parameters, individual-based measures have greater power than population-based parameters. In our analyses, power is high when the variability of the difference between control and impact sites over time, $S\Delta$, is low (Table 2, Figure 4). This variability is a function of two underlying sources of variation: within-site sampling error (variability among samples taken from the same site on the same date), and Site x Time interactions (see [25] for a discussion of optimal allocation of resources in BACIP). First, within-site sampling error is probably lower for physical/chemical and individual-based biological parameters than for population parameters because the latter are less efficiently sampled with a given level of effort. Second, physical/chemical parameters might be influenced more by large-scale oceanographic processes (and therefore will show a high degree of synchrony in fluctuations) than are biological parameters. In turn, biological parameters may be more sensitive to local conditions, reducing the degree to which values for different sites track each other through time. We currently are exploring this question by partitioning observed variance to determine the relative contributions of these two sources of error to each parameter type.

The dilemma posed by our results is that tests of impacts on the parameters of greatest interest to resource managers — population densities — have the least power for a given level of effort. One manner by which power can be increased is by increasing the number of sampling dates (true replicates). Figure 4 illustrates how power varies with the number of sampling dates and with variability. For a moderate amount of variability (0.25), increasing the number of sampling dates from 6 to 20 increases the power to detect a 30% reduction in parameter value at the impact site (relative to control) from about 0.2 to > 0.7 (Fig. 4). Unlike many factors that influence power, the number of sample surveys made is under the control of the investigator. However, it is critical that independence be maintained, and this may constrain how frequently sites can be sampled [12].

IMPLICATIONS

Throughout this paper, we have attempted to highlight problems and limitations associated with commonly used environmental impact assessment designs. It is important that such limitations be understood so that better and more effective assessment strategies can be developed and implemented. It is also critical that scientists, policy makers and regulators understand the limitations of each design to better interpret data that arise from each. We began the paper by stressing that many commonly used assessment designs often can lead to erroneous conclusions. These revelations are not new. Indeed they are well appreciated by many members of the scientific community.

However, some of the more subtle distinctions, such as between "Optimal Impact Assessment" and BACIP designs, are not widely appreciated by regulators or organizations conducting assessment studies, with the consequence that flawed designs still are commonly used (e.g., [21]). As a result, most attempts to provide the most rigorous scientific information concerning effects of a localized perturbation fail (for a still relevant review, see [3]). The practice of collecting equivocal data using inadequate assessment designs serves little interest, is unquestionably wasteful, and fails to ensure that the project or development in question is environmentally sound.

Although not widely utilized, the BACIP assessment design [12] has been employed successfully in a comprehensive study of the ecological impacts on the marine environment from the operation of a coastal power generating station [9]. BACIP avoids many of the interpretation errors associated with more limited designs. As such, BACIP is one of the most powerful (and therefore preferred) designs for the assessment of localized environmental impacts from point-source disturbances. This entails the explicit recognition that a time series of "baseline" data is needed before the commencement of the perturbation. Further, our results indicate that BACIP may lack sufficient statistical power to detect many impacts on parameters of most interest to regulators (e.g., population densities). Power of a BACIP test can be improved by increasing the number of sampling surveys (i.e., true replicates), and by increasing the number of samples taken at a site within a survey (i.e., the precision of each replicate). Increasing the number of samples (within a survey) will increase power if a large part of the variation in Control-Impact differences is due to sampling error. On the other hand, increasing the number of surveys will be helpful in most situations due to the influence of natural temporal variation in the Control-Impact differences. However, while a large number of surveys might be necessary, surveys must be spaced sufficiently in time to ensure independence. Thus, the application of BACIP requires extensive planning and foresight. To do so may require a fundamental change in the regulatory process. Regulators and policy makers must allow for a sufficient period of study prior to the perturbation if the goal is to obtain rigorous scientific evidence concerning localized effects.

Of course it will not be possible to conduct an appropriate BACIP assessment study for every new point-source development. Whenever a BACIP approach is employed, additional research should be undertaken to generalize results and thereby provide insight into other situations. This can be accomplished by examining the mechanisms by which environmental perturbations affect marine resources. Indeed, the resolution of environmental impacts ultimately requires this level of comprehension, and mechanistic approaches should be an integral part of any assessment study (be it a BACIP design or not). We need to understand the processes by which changes in the chemical and physical attributes of the environment alter the physiology of individuals (e.g., metabolic rates, energy allocation), how this altered physiology influences vital rates (e.g., birth, death, migration and growth rates), and finally, how these altered vital rates influence population characteristics (e.g., age-structure, density, production). This approach requires mechanistic studies at the toxicological, developmental, physiological and ecological levels, and which are integrated via dynamic (mathematical) models that are rigorously tested under field conditions. This will lead to better understanding of underlying processes, and thereby enhance our ability to predict ecological effects.

There is another compelling reason for an emphasis on mechanistic studies, either in concert with a BACIP assessment or as a "stand-alone" approach. Regulators and resource managers ultimately are interested in protecting marine resources from adverse impacts. The ability to mitigate or ameliorate adverse ecological effects will be greatly strengthened by knowing which attribute(s) of the perturbation are responsible, and how the effect(s) are generated. These issues can be addressed only through the type of mechanistic studies discussed above; environmental assessment designs such as BACIP only can provide information on the existence (and magnitude) of effects, and cannot address the underlying causes. Although environmental agencies have historically been hesitant to fund such "basic" research, there now seems to be a growing appreciation that resolution of critical environmental problems can only be achieved through rigorous development and integration of basic scientific tools within an applied context.

ACKNOWLEDGMENTS

We thank Jim Bence, Joe Connell and Allan Stewart-Oaten for helpful discussions and comments. Many people, particularly Don Canestro, provided valuable assistance in the field and laboratory; Bonnie Williamson rendered technical assistance. Ed DeMartini graciously supplied the data on pink surfperch, which we present with the permission of the Marine Review Committee. This research was supported in part by the University of California Coastal Toxicology Program and the Minerals Management Service, U.S. Dept. of the Interior, under MMS Agreement No. 14-35-0001-30471 (The Southern California Educational Initiative). The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either express or implied, of the U.S. Government.

REFERENCES

1. Carney, R.S. 1987. A review of study designs for the detection of long-term environmental effects of offshore petroleum activities. Pages 651-696 in D.F. Boesch and N.N. Rabalais, Eds. Long-term Effects of Offshore Oil and Gas Development. Elsevier Applied Science, Essex, England.

2. Eberhardt, L.L. 1976. Quantitative ecology and impact assessment. J. Environmental Management 4:27-70.

3. McKenzie, D.H., Arnold, E.M., Skalski, J.R., Fickeisen, D.H., and K.S. Baker. 1979. Quantitative assessment of aquatic impacts of power plants. Pacific Northwest Laboratory, Richland, Washington.

4. National Research Council. 1990. Managing Troubled Waters. National Academy Press, Washington, D.C.

5. Middleditch, B.S. 1984. Ecological effects of produced water effluents from offshore oil and gas production platforms. Ocean Management 9:191-316.

6. Boesch, D.F., Butler, J.N., Cacchione, D.A., Geraci, J.R., Neff, J.M., Ray, J.P., and J.M. Teal. 1987. An assessment of the long-term environmental effects of U.S. offshore oil and gas development activities: future research needs. Pages 1-53 in D.F. Boesch and N.N. Rabalais, Eds. Long-term Effects of Offshore Oil and Gas Development. Elsevier Applied Science, Essex, England.

7. Bence, J. and J. Kastendiek. 1988. Sampling design and analytical procedures (BACIP). Technical report to the California Coastal Commission, Marine Review Committee. 22 pages.

8. DeMartini, E.E. 1987. The effects of operations of the San Onofre Nuclear Generating Station on fish. Final Report. MRC No. D86- 386 and D86-387.

9. Murdoch, W.W., Fay, R.C. and B.J. Mechalas. 1989. Final report of The Marine Review Committee to the California Coastal Commission. August 1989. MRC Document No. 89-02.

10. Murdoch, W.W., Fay, R.C. and B.J. Mechalas. 1988. Interim Technical Report to the California Coastal Commission. 3. Midwater and Benthic Fish. MRC Document No. 87-037.

11. Green, R.H. 1979. Sampling Design and Statistical Methods for Environmental Biologists. John Wiley & Sons, New York.

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

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

14. Stewart-Oaten, A. 1986. The before-after/control-impact pairs design for environmental impact assessment. Marine Review Committee, Encinitas, California.

15. Stewart-Oaten, A., Bence, J.R., and C.W. Osenberg. 1991. Assessing effects of unreplicated perturbations: no simple solutions. Ecology (in press).

16. Tetra Tech, Inc. 1982. Design of 301(h) monitoring programs for municipal wastewater discharges to marine waters. National Technical Information Service, Springfield, Virginia.

17. Campbell, D.T. and J.C. Stanley. 1966. Experimental and quasi- experimental designs for research. Rand McNally, Chicago, Illinois.

18. Skalski, J.R. and D.H. McKenzie. 1982. A design for aquatic monitoring programs. J. Environmental Management 14:237-251.

19. Mar, B.W., Lettenmaier, D.P., Horner, R.R., Richey, J.S., Palmer, R.N., Millard, S.P., MacKenzie, M.C., Vega-Gonzalez, S., and J.R. Lund. 1985. Sampling design for aquatic ecological monitoring, Volume 1: Summary Report. Electric Power Research Institute, Palo Alto, California.

20. Carpenter, S.R., Frost, T.M., Heisey, D. and T.K. Kratz. 1989. Randomized intervention analysis and the interpretation of whole-ecosystem experiments. Ecology 70:1142-1152.

21. Steinhauer, M. and E. Imamura. 1990. California OCS Phase II Monitoring Program. Year-three Annual Report. Volume I. MMS, Pacific OCS Region, Contract No. 14-12-0001-30262.

22. Gill, J.L. 1978. Design and Analysis of Experiments in the Animal and Medical Sciences. Vols. 1-3. Iowa State University Press, Ames, Iowa.

23. Alldredge, A.L. and J.M. King. 1980. Effects of moonlight on the temporal migration patterns of demersal zooplankton. J. Experimental Marine Biology Ecology 44:133-156.

24. Stretch, J.J. 1983. Habitat selection and vertical migration of sand-dwelling demersal gammarid amphipods. Ph.D. dissertation, University of California, Santa Barbara, California.25.

25. Bernstein, B.B. and J. Zalinski. 1983. An optimum sampling design and power tests for environmental biologists. Journal of Environmental Management 16:35-43.