Vol. 231: 309–314, 2002 | MARINE ECOLOGY PROGRESS SERIES Mar Ecol Prog Ser | Published April 22

REPLY COMMENT

The joint consequences of multiple components of statistical sampling designs

Charles H. Peterson[1,*], **Lyman L. McDonald**[2], **Roger H. Green**[3], **Wallace P. Erickson**[2]

[1]Institute of Marine Sciences, University of North Carolina at Chapel Hill, Morehead City, North Carolina 28557, USA
[2]WEST, Inc., 2003 Central Avenue, Cheyenne, Wyoming 82001, USA
[3]Department of Zoology, University of Western Ontario, London, Ontario N6A 5B7, Canada

We appreciate the opportunity to express the message of our earlier Review (Peterson et al. 2001) in a compressed format that clarifies how multiple design decisions combine to influence the outcomes and interpretations of statistical tests in marine ecology. We achieve this by responding to the selective and misleading arguments raised in the Comment by Gilfillan et al. (2002) and show how our overall conclusions based upon the comparison of 18 components of sampling design in 4 separate impact studies hold true. Despite the contrary assertion of Gilfillan et al. (2002), the objectives of the studies were similar, especially those of the Coastal Habitat Injury Assessment (CHIA) study (McDonald et al. 1995, Sundberg et al. 1996) and the Stratified Random Sampling (SRS) portion of the Exxon-supported Shoreline Ecology Program (SEP; Gilfillan et al. 1995a, Page et al. 1995). The similarity is acknowledged in Gilfillan et al. (1999). The mutual goals of these studies were to assess biological impacts and recovery for the benthic communities of the intertidal shorelines that may have been impacted by the Exxon Valdez oil spill. However, the basic philosophical strategy of approach differed between studies in a fashion that led to contrasting designs for field sampling and statistical analysis, and thereby to differing results and conclusions (Peterson et al. 2001). Our paper compares the designs of these studies for the purpose of extracting generic principles for future field sampling programs.

It is not necessary for us to repeat here the complete discussion of how the sampling and analytical designs differ among the 4 studies because the details are already available in Peterson et al. (2001). Thus, we restrict our treatment of issues raised by Gilfillan et al. (2002) to those points that most require rebuttal because they are misleading, selective, unresponsive, inaccurate, or deserving of modification of our original assessment. First, we must clarify misrepresentations of our general conclusions. The paired design of the CHIA study had greater ability to detect and characterize true oiling impacts than the SEP, not because fully stratified random designs are necessarily inferior but because of the specific SEP stratified random study design. Many of our criticisms of the SEP design were also applied to the CHIA study. We identified aspects of the SEP that were superior to the CHIA study. To represent our paper as a critique of one study is misleading. We made additional criticisms of the fixed site designs of the NOAA-sponsored Hazardous Materials (HAZMAT) shoreline study (Driskell et al. 1996, Houghton et al. 1996) and the Exxon-sponsored Gulf of Alaska (GOA) study (Gilfillan et al. 1995b). We did not include detailed analysis of the design of the fixed site portion of the SEP for reasons that we summarized in Peterson et al. (2001) and expand upon here.

(1) Area covered per sample. The Comment on this component fails to address the issue of small sample area and its influence on among-sample variability. Instead, Gilfillan et al. (2002) misleadingly discuss the number of sites, a separate component (Point 3) of design. The fact remains that the SEP collected samples of a much smaller area than the CHIA and the other 2 studies. For a fixed sampling effort (defined as total area sampled), more small samples will deal with patchiness better than a smaller number of large samples. However, if total area sampled is not retained as individual sample area is reduced, small samples in a patchy system enhance the among-sample error variance, as compared to larger samples that better characterize the community composition. This is especially important in analyzing such community aspects as average species richness, diversity, and evenness. The SEP fell well short of the CHIA, GOA, and HAZMAT studies on this basic component of sampling effort.

(2) Sample replication. The Comment on this component of design again confounds the relevant issue,

*E-mail: cpeters@email.unc.edu

the number of replicate subsamples per site, with additional aspects of design. Meta-analyses are indeed sensitive to p-values of the separate comparisons that comprise them: combining p-values into a synthesized probability is what meta-analyses do. Gilfillan et al.'s (2002) contentions that the p-values in the CHIA study are not well determined and that the meta-analysis conclusions are misleading remain unsupported. The CHIA study employed both the more liberal Fisher's test and the more conservative Liptak's generalization of Souffer's procedure as a means of fairly bracketing the range of meta-analysis conclusions possible (Highsmith et al. 1996). Peterson et al. (2001) characterize the SEP's computations of power as overestimates because they are based on pseudoreplication, rather than the true number of replicate sites. The fact remains that the SEP and the Exxon-sponsored GOA studies used 3 replicate transects to subsample each site, while the CHIA study used 6, and HAZMAT 5 to 10. Greater replication at this level of subsampling is an important component of study design that provides better mean site estimates of key biological variables.

(3) Numbers of study sites per category. The Comment on this component of design fails to acknowledge that we did report higher numbers of study sites for some categories of oiling treatment and habitat type for Gilfillan et al.'s (1995a) SEP than for the other studies, and higher total numbers of study sites (64 total sites for the SRS component of the SEP vs 56 for CHIA, 43 for GOA, and 22 to 30 for HAZMAT). Our reporting on the CHIA study sites does break these totals down by habitat type and shows explicitly a minimum of 2 pairs of sites for the estuarine habitat. We did not include the 11 or 12 additional sites in our analysis of the SEP design because those fixed sites were analyzed separately and not sampled in a consistent fashion that allows quantification of biological recovery (Page et al. 1995).

(4) Numbers of sampling dates. Peterson et al. (2001) acknowledge existence of the fixed site component of the SEP and explain why it is of little use in quantifying the degree of impact and pattern of recovery of shoreline species and communities after the Exxon Valdez oil spill. To quote from Page et al.'s (1995) description of the fixed site component, '…the fixed site data were not used to estimate shoreline recovery, as in the case of the SRS data…' (p. 276) and '…the fixed site biological data were not used to quantitatively project recovery beyond 1990…' (p. 267). Furthermore, 'All the fixed site biological data taken in 1990 and 1991 were from sediment samples. Because the 1989 studies tended to focus on epibiota, the biological data for comparisons with 1990 and 1991 were very limited.' (p. 276). This methodological change prevents any rigorous inference on epibiota, the group that characterizes rocky shores and constitutes the focus of the studies reviewed in Peterson et al. (2001). The fact stands that the SRS component of the SEP and GOA each involved 1 sampling date, whereas the CHIA study had 4 sampling dates over 2 yr and HAZMAT had 5 or more sampling dates over several years. The numbers of dates on which sampling is conducted and the time period encompassed by the sampling represent important design considerations in estimating the duration of impacts and the length of the recovery period after any environmental perturbation.

(5) Total area sampled. The fact remains that the total area of shoreline habitat sampled per year in any given stratum of habitat type and shoreline elevation varied dramatically across study designs from about 1.36 m^2 for the SEP to 9.6 m^2 for GOA, to 35 m^2 for HAZMAT to 60 m^2 for CHIA. This metric implies not only a varying level of sampling effort but also differing ability to characterize the biological communities. Large differences among studies in total area sampled per stratum show that variations in sampling effort per stratum did not compensate for differences among studies in area covered per individual sample.

(6) Philosophical support for targeting putative affected areas. Peterson et al. (2001) explore at length the underlying similarities and differences in approach among the 4 studies. The most basic difference exists between the stratified random sampling that underlies both the SEP and CHIA studies and the fixed site design of GOA and HAZMAT. The objectives of the SEP and CHIA studies are very similar; both were conducted to quantify impacts of the Exxon Valdez oil spill on shoreline communities. This fundamental similarity in objectives is not eliminated by arguing that the SEP was designed to assess the extent of recovery and the CHIA study to quantify injury (Gilfillan et al. 2002). Logically, the CHIA study with its 4 sampling dates informs more on the recovery process than the single sampling of the SRS component of the SEP, given that recovery estimation requires time series information. Despite similar objectives of the SEP and CHIA studies, the philosophical basis underlying the 2 studies differed fundamentally in ways very significant to the results and their interpretation, and to design of future field sampling studies. Our contrast of designs is not invalidated by differences in statistical philosophy, as asserted by Gilfillan et al. (2002). Indeed, the philosophy underlying designs represents important grounds for distinguishing among them. For example, the use of lightly oiled areas as controls in the CHIA study represents a conservative practice that is commonly adopted in environmental science, yet it is perfectly appropriate to question whether impacts over large areas are underestimated by such a design. The decision of SEP to allocate a large fraction of total sampling effort to places with low likelihood of large injury (lightly and very lightly oiled shores) inevitably led to reduced ability to detect injury as compared to the CHIA-study decision to concentrate sampling effort on moderately and heavily oiled shores. The CHIA study sacrificed the opportunity to assess oiling effects on that large fraction of all oiled shoreline so as to better test the areas most likely to exhibit impacts. This reflects a differing philosophy of assessing impacts that represents one of the most basic decisions to consider in any assessment design.

(7) Random site selection versus matched-pair design. Gilfillan et al. (2002) incorrectly assert that the matching of control sites with randomly chosen oiled sites in the CHIA study was done subjectively. Comprehensive site information was available and used for each matching decision, including orientation, wave exposure, spatial proximity, beach slope, substrate composition, nearshore bathymetry, and proximity to sources of freshwater (McDonald et al. 1995). Peterson et al. (2001) admit that such matching cannot be perfect and was not in the CHIA study. Use of paired designs is common and statistically well-justified in environmental field assessments (e.g. Skalski & Robson 1992). We do not disagree that a stratified random selection of both putative impact sites and control sites can represent an appropriate approach to making quantitative field assessments, provided that important statistical design criteria are met. The stratification in the SEP design did not control for, and covariates did not capture, a strong gradient in summer/fall salinity and turbidity across Prince William Sound (PWS), which was driven by proximity to glacial ice and snow melt in mainland drainages. Oil grounded predominantly upon the islands in the sound, where salinity was relatively high, so selecting control sites at random over the entire sound creates a serious bias in the estimation of oiling impacts. This same problem was encountered by the CHIA investigators in 1989 and led to abandonment of 1989 data and implementation of a matched-pairs design from 1990 on. Use of a design that matches control sites against randomly selected impact sites with geographic proximity as one basis for the matching (McDonald et al. 1995) controls against many errors of ignorance about important but unrecognized covariates and environmental gradients that otherwise bias the outcomes. The fully randomized design of the SEP resulted in biased estimates of oil-spill impacts because of intrinsically lower densities and diversities of shoreline epibiota on those control sites near glacial and terrestrial sources of freshwater and turbidity. Furthermore, the use of covariates in an attempt to control for extraneous sources of variation is not advisable if the covariates are functionally or empirically related to the treatment factor or if they are observed variables with substantial error variation. Our review (Peterson et al. 2001) of this dichotomous choice of sampling designs provides important guidance for future oil spill assessments and counsels caution in choosing randomized selection of both oiled and reference sites and in using organic carbon as a covariate. Oil spills occur in unexpected places, so lack of complete understanding of how to stratify field sampling is likely. Organic content of sediments is likely to be increased by the organic carbon in oil, so using organic carbon as a covariate in ANCOVA is likely to violate an important assumption of this analysis.

(8) Sampling frame. Here the Comment fails to deal with the point at issue, namely recognition that oiling did not occur at random across PWS but instead occurred preferentially on island shores where salinity is higher, turbidity is lower, and biological communities are naturally richer and more abundant. Selection of control sites at random throughout any geographic region should incorporate a sampling frame that would allow oiled sites to match control sites in the absence of any impacts of oiling. Such a selection protocol enables spatial contrasts to produce reliable estimates of impact in the absence of 'before' data. Upon recognition of any unexpected and important bias created by non-random application of the environmental perturburbation, a random identification of reference sites should be done in a fashion that faithfully reproduces the site selectivity of the perturbation. The alternative choice of a paired design, selecting carefully matched control sites to compare against randomly selected impact sites, reduces the problems posed by unrecognized selectivity of the actual oiling process largely by using spatial proximity to protect against a suite of geographically varying environmental factors.

(9) Treatment of habitat heterogeneity within sites. Page et al. (1995) do not clearly identify procedures for excluding secondary habitats when encountered within a site, so we were unaware of the application of protocols to exclude them. Given the clarification about procedures in the Comment, there is no longer a difference between the SEP and CHIA studies in this characteristic of the sampling design. Neither GOA nor HAZMAT identified a procedure for eliminating this type of heterogeneity.

(10) Interspersion of sites. Here Gilfillan et al. (2002) argue that the length of shoreline along 1 island, Perry Island, chosen as an example, invalidates our point about poor interspersion among control and oiled sites in the SEP sampling design. Spatial interspersion is an important aspect of both experimental and sampling designs, helping to insure against bias caused by dependence and autocorrelation among adjacent sites (Hurlbert 1984). In fact, interspersion in the SEP was poor, with 3 of the 4 control sites on 1 of several possible island and mainland shores for 1 habitat, and with other less serious departures for other habitat strata. Peterson et al. (2001) mention the possibility, but do not necessarily advocate, re-randomizing after observing that random site selection produced a serious lack of site interspersion. The CHIA study did not re-randomize: it adopted a paired design to eliminate bias caused by a random selection of control sites that could not reproduce the selective oiling of island shores. The paired design reduces, does not introduce, bias and simultaneously insures excellent interspersion of oiled and control sites (McDonald et al. 1995). Peterson et al. (2001) critically discus problems associated with the matched-pairs design and with systematic sampling designs as means of achieving interspersion.

(11) Controls for shoreline treatment and oiling intensity. The Comment on this point misrepresents the presentation in Peterson et al. (2001) to create the appearance of 1-sided criticism of the SEP design. In fact, the only study of the 4 to which the criticism of confounding oiling with shoreline treatments was not applied is the HAZMAT study design. Furthermore, because the SEP design included multiple levels of oiling as separate treatments, it was judged superior to both the CHIA and GOA studies and given a higher rank for this design criterion.

(12) ANCOVA with covariate affected by treatment. In the absence of presentation of the complete set of statistical test results, we may have misinterpreted the information given in Page et al. (1995) on this point. We interpreted their report that 'the physical variables were not strongly correlated with oiling level' (p. 287) to imply that weak correlations did exist. However, because we accepted Page et al.'s (1995) statement that this was not a serious problem, we concluded (Peterson et al. 2001) that this factor 'probably had relatively small influence on the overall conclusions of the tests and studies' (p. 280). Our discussion of this issue serves to guide future studies using covariates around this pitfall. One would expect organic content of sediments to be correlated with oiling in a system where the organic content of oil would contribute significantly to the low background of organic matter in PWS shoreline sediments. It is this organic enrichment that drives the documented enhancement of bacterial production in oiled areas. Thus, use of organic content of sediments as a covariate in assessment of oil impacts, a variable typically correlated with oiling in numerous past studies, would not normally be a wise practice. If in SEP analyses covariates and oiling are indeed uncorrelated, then both the SEP and CHIA studies do a better job of controlling for natural variation than GOA and HAZMAT. ANCOVA (whether a linear additive normal error model or not is irrelevant) is seriously biased in some direction when the covariate is measured with error and the groups being compared possess different distributions on the covariate.

(13) Pseudoreplication. The cloning of data points by treating transects within sites as if they were separate sites increases apparent replication and thus gives the appearance of increasing power to detect true differences (Hurlbert 1984). However, this form of pseudoreplication misleadingly inflates estimates of power because it leads to increased Type I error in both directions; positive and negative responses to the treatment. The CHIA study followed the conservative procedure of using site as the independent sampling unit. Although we disagree with pseudoreplication by treating subsamples within sites as separate sites, we assigned the SEP the highest ranking among the 4 studies on this design criterion. However, even the practice of pooling variances from multiple sources of variation to achieve a more highly replicated test of another factor in ANOVA and ANCOVA is questioned by many statisticians. Those who advocate it require exceptional evidence that 2 variances are equal before pooling, with $\alpha = 0.20$ to 0.30 to avoid the error of pooling variances when they are not equal (Winer 1971). The $\alpha = 0.05$ used by SEP to justify their pooling practices is inappropriate.

(14) Inferring degree of recovery. The Comment on this design component again fails to acknowledge that Peterson et al. (2001) criticized all 4 studies for their poor ability to assess degree of recovery. Furthermore, any study with a single sampling period is incapable of estimating recovery rate: at least 2 points in time are logically required. Only the HAZMAT study involved quantitative sampling of species abundances and community parameters over several years, thus achieving the temporal scope required for recovery analysis. The 'detrending' applied by SEP to canonical correspondence analysis has been criticized by a number of statisticians (e.g. Wartenberg et al. 1987) for its instability and artificiality in eliminating the 'horseshoe effect': consequently, this method is increasingly avoided.

(15) Power analysis. The Comment is correct that average statistical power reported for the SEP by Gilfillan et al. (1995a) was based upon observations, although then used primarily to support simulations. However, statistical power in SEP was presented only for community parameters, not for tests of individual species abundances, whereas the CHIA study computed power for individual species tests. We continue to recommend that, whenever possible, statistical power be estimated in advance of establishment of any actual study design, so those results can be used to plan designs and allocate available sampling effort to provide ability to detect the effects of most environmental significance. Apparently, none of the 4 studies had sufficient data in advance to help optimize power in study design decisions, and sample sizes and allocations of sampling effort were determined largely by logistics and costs. Peterson et al. (2001) discussed power analyses but did not rank studies on the basis of this criterion because attempts to compute power after completion of a study do not represent design decisions.

(16) Taxonomic level used for analysis. Gilfillan et al. (1995a) report results by species for only 2 species, the rockweed *Fucus gardneri* in 2 habitat strata and the mussel *Mytilus edulis* in only 1 stratum. For 2 other species, both littorine snails, p-values are provided without reporting mean abundances by treatment. Thus, we commented that the SEP rarely reports results by species (Peterson et al. 2001). Gilfillan et al. (1995a) include only a summary table of how many single-species statistical tests were significant in the SEP, without providing mean abundances or identifying the species. For the GOA study, Gilfillan et al. (1995b) provide neither mean abundances nor statistical test results for individual species. Reporting abundance data for component species in multiple habitats permits the level of resolution of differences in community composition that is needed to achieve insights into the mechanisms of community response, especially critical in a system well known for the importance of interspecific interactions (Menge 1995). Species-level abundance information is necessary for inferences on competitive replacements, trophic cascades and other indirect effects that delay recovery of rocky intertidal communities.

(17) Pooling of disparate communities. Peterson et al. (2001) do not criticize the use of a 1 mm mesh for soft-sediment sampling. We criticize the failure to exclude meiofaunal taxa from the resulting macrofaunal community because meiofauna are not quantitatively sampled on 1 mm mesh. While larger nematodes are retained, smaller individuals readily pass through and are discarded. Consequently, separate sampling techniques have been developed and are standard practice in meiofaunal research (e.g. Heip et al. 1985). Nematodes and copepods can provide important insight into the nature of community stress (Peterson et al. 1996), but use of 1 mm mesh to sample them produces incomplete and size-biased enumeration with enhanced error variance when size distributions vary.

(18) Scope of communities and habitats examined. Clearly, the analysis of this aspect of sampling design related to shoreline benthic communities not to the birds, mammals, and other organisms evaluated in 22 studies of other communities. Unless the vertebrate consumers were linked to their roles in driving intertidal community dynamics, such additional studies are irrelevant to our contrasts of designs of shoreline benthic communities. Those 22 studies were not linked to the benthic community dynamics and thus were not applicable to our paper. Of all the shoreline benthic studies, only HAZMAT provided sufficient attention to soft-sediment communities to allow assessment of spill impacts on this pollution-sensitive system and to guide design of impacts of future oil spills.

Discussion. Comparisons of study designs among separate assessments of how an oil spill or any other significant perturbation has affected a particular community of organisms are clearly an appropriate part of ecology, providing great insight into the relationships between statistical sampling design and conclusions. The 4 separate field studies of how the Exxon Valdez oil spill affected benthic communities along intertidal shorelines and how they recovered provide an unprecedented opportunity for such comparisons. Peterson et al. (2001) describe in detail the fundamental differences in philosophy and study designs that led to contrasting results. Perhaps the most important difference exists between the fixed-site designs and the stratified random sampling designs. The results of all 4 studies differed. To argue, as in the Comment, for similarity between the results of 2 of them, the SEP and CHIA studies, based upon the proportion of individual tests showing significance is unfounded. Such a misleading metric ignores direction and magnitude of the documented changes and further misses the central point of our paper: that differences in ability to detect and accurately estimate effects vary among these tests and must be considered in order to reach rigorous conclusions. For example, Gilfillan et al. (1995a) report that the significant tests for oil effects reflected positive far more often than negative directions of differences, whereas the CHIA study reported (Stekoll et al. 1996) that 79% of the significant tests showed negative directions of differences (i.e. reductions in abundance). This is a rather critical distinction, ignored in Gilfillan et al.'s (2002) use of an inappropriate metric of comparison. We concur with Gilfillan et al.'s (1995a) earlier view that 'Disagreement about the state of recovery in 1990 probably stems from differences in site selection and statistical analysis' (p. 435).

The basis on which we ordered the 4 studies by detection power, in a generic non-mathematical sense, for each of their statistical design components is clearly developed in Peterson et al. (2001). Our review of the Exxon Valdez study designs was conducted with care and in good faith, based on extensive but in some aspects incomplete information. We often criticized all the studies for failures in design, a point never acknowledged in Gilfillan et al. (2002). The new information presented in the Comment on exclusion of secondary habitat during field sampling in SEP (Point 9) and on lack of any correlations between oiling and organic content or wave intensity in SEP (Point 12) allows us to revise our Table 4. Gilfillan et al.'s (2002) arguments about ranking sample replication (Point 2), number of sampling dates (Point 4), and inferring degree of recovery (Point 14) have little merit, as we explained above. The revised rankings for Point 9 would be 1.5 for both GOA and HAZMAT and 3.5 for both CHIA and SEP (where the higher number indicates more power). The revised rankings for Point 12 would be 1.5 for both GOA and HAZMAT, 3.5 for the CHIA and SEP studies. The consequences of these revisions to our synthesis table would produce the following sums of all ranks by study design: CHIA at 47.5; HAZMAT at 44.5; SEP at 30; and GOA at 28. Consequently, the overall ordering of studies is modified slightly by now ranking SEP above GOA in decisions that affect general detection power. Furthermore, these revisions would rank the SEP higher than the CHIA study in 2 of 15 criteria, lower in 10, and tied in 3. SEP would rank higher than HAZMAT in 3 of 15 criteria, lower in 11, and tied in 1. GOA would rank higher than the CHIA study in 2 of 15 criteria, lower in 12, and tied in 1. Nevertheless, we repeat our caution that this simplified sum of ranks fails to weight the importance of each component to the overall ability of the sampling design to detect and estimate real effects of the spill. The factors that most strongly influenced the ability of these studies to detect and accurately quantify injury and recovery are probably the appropriateness of the sampling frame relative to salinity/turbidity variation, the numbers of dates of quantitative sampling, the allocation of effort relative to likelihood of impact, and the components affecting the total area sampled per sampling stratum (Peterson et al. 2001). While several previous studies have emphasized the importance of obtaining adequate power in environmental assessment designs (e.g. Peterman & M'Gonicle 1992), the principal new contribution of our paper is to explore how the synthesis of many separate design decisions contributes to overall conclusions of a field sampling study. Our review serves as a statistics primer in identifying the range of separate design decisions that enter into the establishment of a field assessment in marine community ecology and the considerations important to making effective design choices.

LITERATURE CITED

Driskell WB, Fukuyama AK, Houghton JP, Lees DC, Mearns AJ, Shigenaka G (1996) Recovery of Prince William Sound infauna from Exxon Valdez oiling and shoreline treatments, 1989 through 1992. Am Fish Soc Symp 18:362–378

Gilfillan ES, Page DS, Boehm PD , Harner EJ (1995a) Shoreline ecology program for Prince William Sound, Alaska, following the Exxon Valdez oil spill. Part 3: biology. In: Wells PG, Butler JN, Hughes JS (eds) Exxon Valdez oil spill: fate and effects in Alaskan waters. American Society for Testing and Materials Special Technical Publication No. 1219. ASTM, Philadelphia, PA, p 398–443

Gilfillan ES, Suchanek TH, Sloan NA, Page DS, Boehm PD (1995b) Shoreline impacts in the Gulf of Alaska region following the Exxon Valdez oil spill. In: Wells PG, Butler JN, Hughes JS (eds)Exxon Valdez oil spill: fate and effects in Alaskan waters. American Society for Testing and Materials Special Technical Publication No. 1219. ASTM, Philadelphia, PA, p 444–481

Gilfillan E, Harner EJ, O'Reilly JE, Page DS, Burns WA (1999) A Comparison of shoreline assessment study designs used for the Exxon Valdez oil spill. Mar Pollut Bull 38:380–388

Gilfillan ES, Harner EJ, Page DS (2002) Comment on Peterson et al. (2001) 'Sampling design begets conclusions'. Mar Ecol Prog Ser 131:303–308

Heip C, Vincx M, Vranken G (1985) The ecology of marine nematodes. Oceanogr Mar Biol Ann Rev 23:399–489

Highsmith RC, Rucker TL, Stekoll MS, Saupe SM, Lindeberg MR, Jenne RN, Erickson WP (1996) Impact of the Exxon Valdez oil spill on intertidal biota. Am Fish Soc Symp 18:212–237

Houghton JP, Lees DC, Driskell WB, Lindstrom SC, Mearns AJ (1996) Recovery of Prince William Sound epibiota from Exxon Valdez oiling and shoreline treatments, 1989 through 1992. Am Fish Soc Symp 18:379–411

Hurlbert SH (1984) Pseudoreplication and the design of ecological field experiments. Ecol Monogr 54:187–211

McDonald LL, Erickson WP, Strickland MD (1995) Survey design, statistical analysis, and basis for statistical inferences in Coastal Habitat Injury Assessment. In: Wells PG, Butler JN, Hughes JS (eds) Exxon Valdez oil spill; fate and effects in Alaskan waters. American Society for Testing and Materials Special Technical Publication No. 1219. ASTM, Philadelphia, p 296–311

Menge BA (1995) Indirect effects in marine rocky intertidal interaction webs: patterns and importance. Ecol Monogr 65:21–74

Page DS, Gilfillan ES, Boehm PD, Harner EJ (1995) Shoreline ecology program for Prince William Sound, Alaska, following the Exxon Valdez oil spill: Part 1: study design and methods. In: Wells PG, Butler JN, Hughes JS (eds) Exxon Valdez oil spill: fate and effects in Alaskan waters. American Society for Testing and Materials Special Technical Publication No. 1219, ASTM, Philadelphia, PA, p 263–295

Peterman RM, M'Gonigle M (1992) Statistical power analysis and the precautionary principle. Mar Pollut Bull 24:231–234

Peterson CH, Kennicutt MC II, Green RH, Montagna P, Harper DE Jr, Powell EN, Roscigno PF (1996) Ecological consequences of environmental perturbations associated with offshore hydrocarbon production: a perspective on long-term exposures in the Gulf of Mexico. Can J Fish Aquat Sci 53:2637–2654

Peterson CH, McDonald LL, Green RH, Erickson WP (2001) Sampling design begets conclusions: the statistical basis for detection of injury to and recovery of shoreline communities after the Exxon Valdez oil spill. Mar Ecol Prog Ser 210:255–283

Skalski JR, Robson DS (1992) Techniques for wildlife investigations: design and analysis of capture data. Academic Press, New York

Stekoll MS, Deysher L, Highsmith RC, Saupe SM, Guo Z, Erickson WP, McDonald L, Strickland D (1996) Coastal habitat injury assessment: intertidal communities and the Exxon Valdez oil spill. Am Fish Soc Symp 18:177–192

Sundberg K, Deysher L, McDonald L (1996) Intertidal and supratidal site selection using a geographical information system. Am Fish Soc Symp 18:167–176

Wartenberg D, Ferson S, Rohlf FJ (1987) Putting things in order: a critique of detrended correspondence analysis. Am Nat 129:434–448

Winer BJ (1971) Statistical principles in experimental design, 2nd edn. McGraw-Hill, New York

Editorial responsibility: Otto Kinne (Editor), Oldendorf/Luhe, Germany

Submitted: April 2, 2002; Accepted: April 3, 2002
Proofs received from author(s): April 16, 2002