

## COMMENT

### Chance and randomness in design versus model-based approaches to impact assessment: comments on Bulleri *et al.* (2007)

Bulleri *et al.* (2007) describe local environmental impacts as ‘assessed by comparing the disturbed site with one or more [unaffected] reference sites . . . [which] can be thought of as a random sample from a population of sites as in design-based approaches (Underwood 1992), or as a set of covariates correlated with the disturbed site as in model-based analyses (Stewart-Oaten & Bence 2001)’ (brackets added). This is misleading. These two approaches are not alternatives. The latter paper shows that reference sites are not essential and that the former paper’s use of them is invalid.

An assessment aims to compare the disturbed site with what it would have been like without the disturbance. The latter, known as a ‘counterfactual’, is hypothetical and cannot be observed. The design-based and model-based approaches can have incompatible ways of estimating parameters of the counterfactual (such as its mean abundance over time after the disturbance) and will have incompatible ways of estimating the estimates’ errors.

#### The design-based approach

In Underwood (1992), estimates of parameters of the counterfactual assume the disturbed site and the observed reference sites are randomly chosen from the same population of sites. If this were true, the counterfactual’s value of a given variable could be estimated by the mean of the reference site values and the distribution of these values could be used for inference (for example for estimating variance). However, this assumption is not true, even approximately, and not plausible as a model. It is a subjective guess by the assessor, and thus invalid for inference.

The ‘random choice’ assumption is taken literally in statistical design-based methods for analysing experiments. ‘It is of the essence that randomization means the use of an objective randomizing device; it does not mean that allocation is vaguely haphazard or even that it is done in a way that looks effectively random to the investigator’ (Cox & Reid 2000, p. 19). The randomizing device is essential because design-based analysis is based on a model of the device’s performance. For most devices (computer programs, tables), this model can be thoroughly checked by running it arbitrarily often, independent of the study.

Disturbed sites in assessments are never chosen by a randomizing device, either by humans selecting a site for a planned disturbance or by ‘Nature’ causing an unplanned one.

A counter-argument is that assessment is not an experiment, with random ‘allocation’ of individual units (sites) to treatments, but a comparison of two populations using random selection. The ‘undisturbed’ population is the set of sites from which the reference sites are chosen (plus the counterfactual). The ‘disturbed’ population consists of the disturbed versions of these sites: what each would be like if it were the disturbed site. Only one of its members is sampled, but this affects only power, not validity, as long as the two populations have the same variance. The samples are not literally random, but ‘random choice’ is often approximate in field biology. Inference about the leaves of trees, or the fish in lakes, is based on assumed random sampling, although the possible samples do not have exactly equal chances.

This counter-argument is flawed because the disturbed ‘sample’ is not approximately random in any objective sense. The leaf and fish samplers have populations: all the leaves on the tree or fish in the lake. Each uses literal random choice. The leaf sampler can divide the tree into roughly equal sections, each section into subsections, etc., and make a chain of random selections until a leaf is selected. The fish sampler can randomly sample strata, such as different parts of the lake or heights in the water column, whose fraction of the whole lake is approximately known. These approximations can be checked, in principle, by counting the leaves in the tree sections, or measuring the area or volume of the lake strata. None of these, a defined population, literal randomization or checkable approximation, applies to the ‘random’ selection of the disturbance site.

Another counter-argument is that the disturbance site is ‘effectively random’. It might be one of several similar bays, or a site on a long uniform coastline: part of a homogeneous environment where all sites, if undisturbed, seem equally likely to yield any given variable value. It is selected for reasons which seem unrelated to its ecology, like ownership for a planned disturbance, or the causes of an accident for an unplanned one. Such ‘as if’ randomness is often accepted. Samples may be opportunistic in field biology, because some areas are inaccessible or animals are cryptic or mobile. A great ‘natural experiment’ in epidemiology is Snow’s (1855) study of cholera deaths in an area where the ‘allocation’ of houses to one of two water companies was made much earlier, without the consent of current occupants, and showed no patterns (Freedman 2005).

This argument is also flawed. ‘Natural’ randomness has no randomizing device. Its inferences are model-based, and credible if the model is plausible. The Snow (1855)

and animal examples involve repeated selections from pre-defined populations by mechanisms that can be plausibly seen as 'chance' and independent of the variables being studied. Snow's (1855) 66 153 houses were allocated to water companies years earlier, so the occupants were allocated by their house choices. These seemed unrelated to cholera vulnerability: few people knew which company they had, and Snow found no patterns of age, occupation, wealth or geography (the companies were mixed even in the same streets). Animal samples result from choices by the animals and the sampler. The sampler can choose accessible areas randomly. The animals' choices can be argued to be unrelated to the study variables. This argument often deserves scepticism: the chance of being seen or caught may depend on an animal's age, size or sex. Inferences become more credible if such factors are assessed, for example by analysing other variables with known distributions, comparing samples from different areas or chosen in different ways, and laboratory studies of behaviour. In contrast, an assessment's disturbance site is not chosen from a population. Defining one afterwards requires subjective decisions. Sites are not really homogeneous. Bays vary in size, substratum, depth and exposure: which ones are included? Weather and currents change along coastlines: how far does the population extend? Sites on the boundary of any population interact directly with sites outside it, but interior sites do not. A plausible model of site selection as a random process applied to a defined population requires an account of the imaginary process and arguments that it approximates the real one. The checks used by Snow or by animal samplers are mostly unavailable. The exceptions, like comparing disturbed and reference site values before the disturbance, are weak because there is only one disturbed site.

A further objection can be noted in passing: the real assessment target is not the mean effect the disturbance would have over a population, but its actual effect at the site chosen.

### The model-based approach

Intervention analysis estimates the counterfactual from information on the disturbed site before the disturbance. It compares before and after time series (for example estimates of their long-term mean abundances) and requires models of the processes producing them. These include chance elements, which need assumptions so that statistical methods based ultimately on repeated independent observations can be applied to complicated (for example correlated) data for inference. Such assumptions and their interpretation are discussed by Stewart-Oaten and Bence (2001, especially pp. 312–317).

In principle, no reference sites are needed: 'BA' is the essential part of BACI (before-after/control-impact). The foundation paper in this area is Box and Tiao (1975), whose examples did not use reference sites. A recent example is the analyses of time series for assessing global warming: these are not invalidated by the lack of reference sites.

Intervention analysis allows for covariates. They can (1) reduce unexplained variance and (2) improve its estimation. Reference sites can provide either benefit in some assessments, but not all. The idea is that much of the natural variation at the disturbed site will be due to factors that also affect the reference sites. If so, these sites might achieve aim (1) by accounting partly for sources of variation which may be unknown, or be hard to measure, or have biological effects which are hard to model (such as invasions by disease or consumers, natural variation in resources or aspects of weather). Aim (2) could be important if the observed time series are short and there is occasional very large natural variation, with long-lasting effects. If such variation occurs between 'before' and 'after' (for example during construction), it may mimic or conceal a disturbance effect; and if it does not occur during the observation periods, our variance estimates may be too small to allow for it. Reference sites might reduce this problem.

However, reference sites may be unavailable or unhelpful. Poor covariates can be worse than none, by introducing more error than they resolve. If the reference sites do not 'track' the disturbance site well, or their values (such as abundances at given times) are poorly estimated, they may miss aim (1) and increase the unexplained variance. We could then be better off without them if the conditions making aim (2) important are lacking. Alternatively, they could be used informally, with the main analysis being a comparison of the before and after time series at the disturbed site, and the reference site(s) being used as a subsidiary check that no region-wide natural change occurred around the time of the disturbance.

### Concluding remarks

The critique of inference based on randomness does not imply that differences among sites are uninformative, or can not be used for inference. Reference sites could be used in a spatial model (for example Cressie 1991) especially if there are no before data at all, or there are none from the disturbed site but some from nearby sites that can be argued to predict it. This approach would need careful statement of its assumptions and their justification.

The methods discussed here are for the assessment of a disturbance at a single site. They are not for an experiment to assess the effects of a given type of disturbance by deliberately disturbing several sites, perhaps leaving others undisturbed. Such studies seek different kinds of conclusions and inferences.

This note does not address Bulleri *et al.*'s (2007) main purpose: to defend two opinions about assessment. These are not invalidated by their misstatement of assessment inference.

### Acknowledgements

Bill Murdoch, Steve Schroeter, David Hinkley and three reviewers provided very helpful comments on earlier drafts.

**References**

- Box, G.E.P. & Tiao, G.C. (1975) Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association* **70**: 70–79.
- Bulleri, F., Underwood, A.J. & Benedetti-Cecchi, L. (2007) The interpretation of ecological impacts in human-dominated environments. *Environmental Conservation* **34**: 181–182.
- Cox, D.R. & Reid, N. (2000) *The Theory of the Design of Experiments*. Boca Raton, FL, USA: Chapman & Hall.
- Cressie, N.A.C. (1991) *Statistics for Spatial Data*. New York, NY, USA: Wiley.
- Freedman, D.A. (2005) *Statistical Models: Theory and Practice*. Cambridge, UK: Cambridge University Press.
- Snow, J. (1855) *On the Mode of Communication of Cholera*. London: John Churchill. (Also at URL <http://www.ph.ucla.edu/epi/snow/snowbook.html>; the natural experiment is in Part 3).
- Stewart-Oaten, A. & Bence, J.R. (2001) Temporal and spatial variation in environmental impact assessment. *Ecological Monographs* **71**: 305–339.
- Underwood, A.J. (1992) Beyond BACI: the detection of environmental impacts on populations in the real, but variable, world. *Journal of Experimental Marine Biology and Ecology* **161**: 145–178.

---

ALLAN STEWART-OATEN\*

*Department of Ecology, Evolution and Marine Biology*  
*University of California, Santa Barbara, CA 93110-9610*  
*USA*

---

\* Correspondence: Dr Allen Stewart-Oaten e-mail: [stewart@lifesci.ucsb.edu](mailto:stewart@lifesci.ucsb.edu)