Integrated Environmental Assessment and Management — Volume 9999, Number 00—pp. 1–2
© 2010 SETAC

STATISTICS AND ECOTOXICOLOGY: SHOTGUN MARRIAGE OR ENDURING PARTNERSHIP?

David R. Fox
University of Melbourne, Melbourne, Victoria, Australia
david.fox@unimelb.edu.au
DOI: 10.1002/ieam.81

There are cultures in which people believe that some objects have magical powers; anthropologists call these objects fetishes. In our society, statistics are a sort of fetish. . Statistics direct our concern; they show us what we ought to worry about and how much we ought to worry. In a sense, the social problem becomes the statistic and, because we treat statistics as true and incontrovertible, they achieve a kind of fetishlike, magical control over how we view social problems. We think of statistics as facts that we discover, not numbers we create. (Best 2001)

It might be argued that ecotoxicologists have had a fetish with no-observed-effect concentrations (NOECs) and the equally problematic concept of a species sensitivity distribution (SSD). Despite the many calls to ban NOECs and the documented difficulties with setting safe concentrations using empirical SSDs, the methodology continues to be widely used throughout the world. In Australia, the results from this highly statistical approach have underpinned the assessment of marine impacts from the many desalination plants that have been or are in the process of being built around the country. The software tools Toxcalc^{TM} (Tidepool Scientific) and BurrliOz (http://www.cmis.csiro.au/envir/burrlioz/) have been used extensively to produce toxicity estimates and safe dilutions, respectively. Having pored over many of the ecotox reports for a number of these desalination plants, I suspect that a disturbingly high proportion of the recommended safe dilutions are of dubious value at best and fatally flawed at worst. There are many factors operating here, but at the heart of the problem we see evidence of: 1) a lack of proper statistical QA/QC to ensure the integrity of the input data; 2) an uncritical and mechanistic approach to the statistical analysis of the laboratory ecotoxicity data; 3) unstructured and possibly unsound (statistical) modelling approaches such as the exchangeable use of NOEC, IC_x, and EC_x data together with arbitrary assessment factors in order to achieve the best-fitting SSD distribution; and finally 4) the subjugation of disciplinary expertise by computational complexity and automation.

The trouble with our infatuation with the statistical approach is, as Best (2001) suggests, that the outputs become axiomatic. Thus, in the case of the waste stream from a desalination plant, millions of dollars have been or will be spent on outfalls that have been designed to achieve a $1{:}x$ dilution where x has been determined from SSD modelling. I have not seen or heard any discussion about the (statistical) uncertainty in x let alone an evaluation of the overall uncertainty when hydrodynamic models are used to predict the depth at which x is achieved.

I must at this point plead mea culpa. I was responsible for introducing the Burr family of distributions into the ecotoxicologist's lexicon, around which the BurrliOz software was written. More recently, I proposed a Bayesian methodology for estimating the no-effect concentration (NEC; Fox 2010) as an alternative to the pathologically flawed frequentist approach based on NOECs. My enthusiasm for the Bayesian paradigm in ecotoxicology may at first seem contrary to the tenor of this article. However, the apparent inconsistency evaporates when one appreciates that expert opinion and the elicitation of subjective assessments are hallmarks of the Bayesian approach. In a sense, the Bayesian paradigm places the ecotoxicologist back in the driver's seat, no longer consigned to be a mute, backseat observer to some adaptation of Neyman–Pearson hypothesis testing.

Much has been written on the role of statistics in ecotoxicology, and there have been many good suggestions for raising the bar with respect to data collection, processing, and analysis, including Newman's recent pitch for an increased emphasis on Bayesian statistical methods at an undergraduate level (Newman 2008). Perhaps one of the more comprehensive roadmaps for improving the quality of statistics in ecotoxicology was provided by Chapman et al. (1996), which summarized the deliberations of an international workshop held in London in April, 1995. As noted by Chapman et al. (1996) ''twenty-four invited participants from the US, Canada, the UK, The Netherlands, Denmark, Germany and Italy were asked to consider key questions about the current description and use of statistics in toxicity test guidelines.'' Key findings included:

- Inconsistent and ambiguous use of statistical methods
- Vague or nonexistent objectives of ecotoxicity testing
- Lack of specificity in statistical design considerations
- Imprecise, vague, or nonexistent recommendations on modes of analysis and
- Confusing terminology

Chapman et al. (1996) devoted considerable space to reiterating the failings of the NOEC and proffered the EC_x as an alternative measure that ''generally overcome[s] all of the. . criticisms.'' Uptake of this suggestion in the intervening 14 years appears to have been patchy and, as I recently argued, may in fact be counterproductive (Fox 2009).

The aforementioned desalination projects provide examples of instances in which SSDs have been fitted to mixtures of NOEC, EC_x, and IC_x data. Arguments in support of this approach vary and usually appeal to one or more of the following assertions:

1. NOECs are integral to SSD modelling
2. EC_x values can be estimated more reliably than NOECs and
3. Low-effect IC_x values are robust (see, e.g., van der Hoeven 1997; OECD 1998).

If arguments 1–3 above were not applicable or not seen as sufficiently compelling, the fall-back position is that the resulting mixture of inputs in some way optimized the fit of the SSD. In this case, the end justified the means, no matter what conceptual and linguistic difficulties were generated by such an obfuscation of measures and methods.

Returning to the rhetorical question posed in the title of this article, the answer for me is clear: this must be an enduring, robust, and reciprocal relationship to ensure the credibility of ecological risk assessments. The integration of statistical methods into ecotoxicological practice thus far seems to have occurred more by osmosis than by design. Although there is nothing inherently wrong with this, I suspect that it has resulted in some practices that have escaped rigorous scrutiny and debate, by both statisticians and ecotoxicologists. It took a good 10 y or more for the cracks to emerge in NOEC-based procedures, although in hindsight most, if not all, the criticisms could have been identified in advance. Likewise and perhaps with the exception of Suter (1996), there appears to have been little challenging of statistical conventional norms when applied to ectoxicology. Why, for example do we perform Dunnett's test using the sacred $\alpha = 0.05$ level of significance for testing treatments with a control? What are the implications for the ultimate HC_x if a different α is used in Dunnett's test, and how do we interpret the differences? Why do we use replicates in a concentration–response experiment? On this last question, the only reason I can see why this is done is because the ANOVA procedure (which itself is a precursor to the bankrupt NOEC) demands replication! To me this is a no-brainer: scarce experimental resources are spent replicating for a small number of concentrations simply because the statistical procedure used to generate a wholly unsatisfactory toxicity estimate requires it. As noted by Suter (1996), Chapman et al. (1996), Fox (2010), and others, parameter estimates from a *model* describing the concentration–response relationship would seem to be a more profitable approach.

The (statistical) issues and problems in ecotoxicology are many, the challenges great, and the implications of getting it wrong profound. The de facto marriage between statistics and ecotoxicology has to be sanctified and accorded the same status as other statistical couplings such as statistics and biology (biometrics), statistics and chemistry (chemometrics), and statistics and the environment (environmetrics). At a recent meeting of the Australasian Society for Ecotoxicology, I suggested that the nexus between statistics and ecotoxicology would be enhanced via the formation of a subdiscipline having a well-identified name, such as statistical ecotoxicology. Membership currently stands at one and is open to anyone with an interest in advancing this important area!

REFERENCES

Best J. 2001. Damned lies and statistics: Untangling the numbers from the media, politicians, and activists. Berkeley (CA): University of California.

Chapman PF, Crane M, Wiles J, Noppert F, McIndoe E. 1996. Improving the quality of statistics in regulatory ecotoxicity tests. *Ecotoxicology* 5:169–186.

Fox DR. 2009. Is the EC_x a legitimate surrogate for a NOEC? *Integr Environ Assess Manage* 5:351–353.

Fox DR. 2010. A Bayesian approach for determining the no effect concentration and hazardous concentration in ecotoxicology. *Ecotoxicol Environ Saf* 73:123–131.

Newman MC. 2008. ''What exactly are you inferring?'' A closer look at hypothesis testing. *Environ Toxicol Chem* 27:1013–1019.

OECD [Organization for Economic Co-operation and Development]. 1998. Report of the OECD Workshop on Statistical Analysis of Aquatic Toxicity Data. OECD Environmental Health and Safety Publication, Series on Testing and Assessment, No 10, Environment Directorate, ENV/MC/CHEM(98)18, Paris, France: OECD.

Suter GW. 1996. Abuse of hypothesis testing statistics in ecological risk assessment. *Hum Ecol Risk Assess* 2:331–347.

van der Hoeven N. 1997. How to measure no effect. Part III: Statistical aspects of NOEC, EC_x and NEC estimates. *Environmetrics* 8:255–261.

Author Proof