Environmental Toxicology and Chemistry, Vol. 35, No. 6, pp. 1337–1339, 2016

Printed in the USA

Critical Perspectives

COMMENT ON *ET&C* PERSPECTIVES, NOVEMBER 2015—A HOLISTIC VIEW

DAVID R. FOX*†‡ and WAYNE G. LANDIS§
†Environmetrics Australia, Victoria, Australia
‡University of Melbourne, Melbourne, Victoria, Australia
§Institute of Environmental Toxicology, Huxley College of the Environment, Western Washington University, Bellingham, Washington, USA

(Submitted 8 December 2015; Returned for Revision 15 January 2016; Accepted 21 January 2016)

Abstract: In response to a recent collection of perspectives published in *Environmental Toxicology and Chemistry*, the authors argue that there is little value in revisiting and rehashing the well-documented issues around toxicity metrics, competing statistical paradigms, legitimacy of theoretical constructs for species sensitivity distributions, and a number of other unresolved (and perhaps unresolvable) attendant statistical issues that have occupied journal space for more than 30 yr. This is not to say that these matters are unimportant—they are; however, the discussion on these topics is mature, with very few new insights being offered. To move forward on some of these seemingly intractable issues, the authors suggest the ecotoxicological community would be better served by the formation of a subdiscipline of "statistical ecotoxicology," where professional statisticians and ecotoxicologists work in unison. As it currently stands, statistical developments in ecotoxicology are not necessarily undertaken or peer-reviewed by professional statisticians, a situation that has no doubt contributed to the lack of real progress on important recommendations such as the phasing out of no-observed-effect concentrations. *Environ Toxicol Chem* 2016;35:1337–1339. © 2016 SETAC

Keywords: Statistical ecotoxicology Future direction

> I have no idea who was the first to invent the concepts of No Observed Effect Concentration (NOEC) and Lowest Observed Effect Concentration (LOEC) in ecotoxicology. But I do believe that the introduction of these terms was the most serious misfortune to happen to ecotoxicology.
>
> —Laskowski [1]

The recent set of articles [2–7] discussing the statistical challenges in ecotoxicology is a timely and welcome addition to an ongoing dialogue on this most important topic. Although many good points are discussed in these articles, the selection of perspectives presented in that series [2–7] is very narrow and in our view lacks balance and a critical evaluation of the state of current practice.

The opening article by Green [2] repeats previous cautions from this author warning of the potential folly of essentially "throwing the NOEC baby out with the bathwater." We, together with others, have been responsible for some of the "many calls within the ecotoxicological community to replace hypothesis testing methods to determine a no-observed-effect concentration (NOEC) with regression models to estimate an effects concentration (EC*x*)" [2]. Although we welcome the diversity of views and wish to be part of a full and open debate on such matters, we believe it is now time that these be conducted within an overarching framework that is both inclusive and representative. To date, opposing views on this most important topic have been presented in isolation through vehicles such as single journal articles, invitation-only symposia, and specialist training courses at major society-sponsored conferences—although, to be fair, Green's short course "Statistical Issues in the Design and Analysis of Ecotox Experiments," which has been running for a number of years at SETAC annual meetings, does cover both model-based effect concentration (EC*x*) estimation and NOECs.

So what would this overarching framework look like? As one of us (D.R. Fox) has long advocated [8], a critical first step is to cement and formalize the relationship between statistics and ecotoxicology through the formation of a well-identified subdiscipline of statistical ecotoxicology. Examples of similar couplings can be found in medicine, biology, chemistry, ecology, and the social sciences. Pockets of support for this concept already exist in Europe, North America, and Australia. For example, E. Szöcs at the University of Koblenz (Germany) has worked up an impressive library of customized R functions specifically developed for ecotoxicology, and L. Hothorn at Leibniz University (Germany) has recently published a new textbook on statistics in toxicology using R [9]. The R package drc by C. Ritz at the University of Copenhagen (Denmark) provides a well-established framework for fitting models to concentration–response data [10]; and, more recently, King et al. from the University of Lyon (France) have released an online tool called Mosaic to fit species sensitivity distributions to truncated toxicity data [11].

Returning to the collection of *ET&C* perspectives [2–7], we note 2 dominant themes: NOECs as a credible alternative to regression-based estimates when the concentration–response relationship is poorly defined, and choice of *x* to harmonize EC*x* and NOEC values. Newman [3] suggests Bayes factors can be used as an alternative when regression fails, whereas Green [2] argues that "the NOEC approach must be retained" under such circumstances. This advice fails to ask the question of why the regression approach has failed, We can think of 3 reasons: 1) the experimental design was inadequate, so an exposure response could not be elucidated; 2) the wrong model was used; and/or 3) within the range of concentrations used, there was no discernible exposure response. The situation is rectified in the second case by finding a better-fitting model, either by increasing the number of parameters or by using a different functional form (or both) and in the third case either by

* Address correspondence to david.fox@environmetrics.net.au
Published online in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/etc.3378

abandoning the pursuit of obtaining a meaningful toxicity metric or by increasing the range of test concentrations to see if a response is evident. Interestingly, the lack of a discernible or "good" response relationship appears to be viewed by these authors as a failure of the model-based approach for deriving a toxicity estimate and not as a failure of the experimental design—or an observation on the true state of nature. The fact is, if there is either no response or only a weak response within the range of tested concentrations, then the correct model is the mean response over all concentrations used—that is a horizontal line! Estimation of an NOEC in such a case is nonsensical because, first, the evidence is that there is no relationship and, second, the value of the NOEC is more a function of the variance of the response at each concentration and has relatively little to do with the size of the dose.

With respect to the second theme, again this is nonsensical since the NOEC is an estimate of a no-effect concentration for which $x = 0$ is axiomatic, a point noted in Mebane's response [4]. A number of articles attempting to identify a suitable value of x to harmonize NOEC and ECx data have appeared since Crane and Newman [12] asked the somewhat rhetorical question, "What level of effect is a no observed effect?" This endeavor fails to make clear what the true objective of the exposure–response experiment is; and we believe, for the reasons just stated together with the litany of documented shortcomings of NOECs, that this is a bankrupt approach. It is our contention that this is a binary decision: the goal is to estimate the concentration for which the proportion of organisms (adversely) affected is either 0% or x%, but not both.

In the context of ecological risk assessment, the defining of the x in an ECx as a single number and one that holds for all cases is fundamentally incorrect. To accurately calculate risk and the associated uncertainty, it is preferred that an exposure–response relationship is clearly documented along with the uncertainty in that estimate. Risk-assessment techniques such as the use of Markov Chain Monte Carlo sampling and Bayesian networks can employ all of the information about the exposure–response relationship. Point estimates, no matter how derived, cannot provide the information to do that. Probabilistic risk assessments can incorporate those features of the entire exposure response into their estimates of risk when informing decisions and when evaluating different management outcomes. Point estimates limit the information and prevent the entire risk picture from being evaluated, although we acknowledge that the use of more sophisticated tools such as Bayesian networks and Markov Chain Monte Carlo methods poses greater challenges for regulators and practicing risk assessors. On this latter point, participants at a workshop on advanced methods for dose–response assessment agreed that the wider use of Bayesian methods can improve human health risk-assessment practices [13]. To this end, it was recommended that regulatory agencies such as the US Environmental Protection Agency take steps to foster the use of Bayesian approaches and that professional societies such as the Society for Risk Analysis and the Society of Toxicology be encouraged "to seek out and recognize meritorious analyses that use Bayesian approaches" [13].

We found it interesting that although Green [2] cautions against the use of regression methods when using "severity scores," because the latter are essentially labels, he overlooks the converse situation in which numeric data in the form of concentrations are treated as labels for the purpose of obtaining an NOEC. As noted by Fox et al. [14], "the NOEC/NOEL is effectively a label which is why it has no statement of precision or uncertainty."

Mebane [4] notes that replication in concentration–response experiments is a legacy of analysis of variance–type procedures and suggests, along with ourselves and others, that this is both wasteful and unnecessary for model-based inference.

The information provided in Aldenberg's response [5] concerning the impact of data error on fitted species sensitivity distributions and derived metrics (such as the x% hazardous concentration [HCx]) was based on earlier work by Aldenberg and Rorije [15], which has recently been shown to be flawed [16] and accordingly should not be relied upon. In particular, the claim that "more data error leads to less conservative estimation" [15] (that is, a larger HCx) is not only counterintuitive but plain wrong. Unfortunately, this flawed conclusion appears to have underpinned the recommendation of aggregating toxicity data prior to further statistical analysis. We believe this practice is unsound because of the masking of the true level of stochastic variation, which in the case of species sensitivity distributions leads to less conservative estimates of an HCx [13]. This would seem to be a view shared by Green [2], who noted that one of the main drawbacks associated with the analysis of severity scores was "losing information by replacing the individual subject scores by a single replicate summary value, such as the median." That 2 experts offer apparently conflicting advice demonstrates the need for greater rigor in statistical ecotoxicology.

On reflection, we find it quite astonishing that 30 yr have elapsed since Stephan and Rogers [17] made 12 cogent arguments in support of regression-based toxicity estimates as an alternative to hypothesis testing. Yet here we are in 2016 still debating whether a *model* of the concentration–response relationship is preferable to a claim of no significant difference from a control. It seems that old habits die hard in ecotoxicology, and Stephan and Rogers would no doubt be disappointed with the slow progress toward the use of regression analysis to "encourage aquatic toxicologists to think of chronic toxicity in terms of a concentration–effect relationship" [17].

In conclusion, we believe the number and complexity of unresolved statistical issues in ecotoxicology demands a coordinated and concerted response. It is time to move beyond documenting the shortcomings of NOECs and the species sensitivity distribution methodology and to encourage and nurture a more collaborative research environment that sees statisticians working more closely with biologists, chemists, toxicologists, ecologists, and others. This is a model that has been successfully applied in other disciplines such as the biological sciences (biometrics), the environment (environmetrics), and chemistry (chemometrics). The risks of "going it alone" have been well documented. For example, Richard Horton, editor of the respected medical journal the *Lancet*, highlighted the problem of lack of statistical collaboration in medicine:

> Yet still today, too much of medicine takes medical statistics for granted. Time and again, we see research that has clearly not been within a hundred miles of a statistical brain. Physicians usually make poor scientists, and physicians and scientists together too often play the part of amateur statistician—with appalling consequences. The future of a successful biomedical research enterprise depends on the flourishing of the discipline we call medical statistics. It is not at all clear to me that those who so depend on medical statistics appreciate either that dependence or the fragility of its foundation [18].

Interestingly, toxicology was an active area of research for statisticians of the caliber of Ronald Fisher, Chester Bliss, and

Joseph Berkson during the 1920s, 1930s, and 1940s. However, it appears that the statistical profession's interest in toxicology waned after the 1970s, which may in part be attributable to the demise of many university statistics departments around the world since that time. However, the rise of big data coupled with proclamations by Google Chief Economist H. Varian and the *Harvard Business Review* that a career in data science is "the sexiest job of the 21st century" [19] have led to a resurgence in interest in the statistical sciences. According to the American Statistical Association, statistics was the fastest-growing STEM (science, technology, engineering, and mathematics) major for 2010 to 2013, with an almost doubling of the number of degrees granted during the period [20]. The challenge as we see it is to capitalize on this renewed interest in statistical science by refocusing the big data spotlight on the "little data" problems that characterize much of ecotoxicology. The increased emphasis on collaboration by universities and research establishments around the world should facilitate this objective. By way of example, the University of Melbourne has recently launched an aggressive media campaign to highlight its "collision of ideas" agenda with examples on its website including engineering colliding with environments to tackle global water shortage problems and accounting colliding with botany to help plan cities of the future [21]. For us, we would like to see statistics collide with ecotoxicology to improve the effectiveness of what we do by reducing the uncertainty in our predictions and increasing the confidence in our actions.

In a recent review, Hothorn concluded, "statistics in toxicology is not at the end—it is in the middle" [22]. We think statistical ecotoxicology is front and center.

Acknowledgment—We are grateful to the anonymous reviewer whose suggestions helped improve the clarity of the final manuscript.

REFERENCES

1. Laskowski R. 1995. Some good reasons to ban the use of NOEC, LOEC and related concepts in ecotoxicology. *Oikos* 73:140–144.
2. Green JW. 2015. In response: Challenges for statistical evaluation of ecotoxicological experiments—An industry perspective. *Environ Toxicol Chem* 34:2437–2439.
3. Newman MC. 2015. In response: Regression or significance tests: What other choice is there?—An academic perspective. *Environ Toxicol Chem* 34:2439–2440.
4. Mebane CA. 2015. In response: Biological arguments for selecting effect sizes in ecological testing—A governmental perspective. *Environ Toxicol Chem* 34:2440–2442.
5. Aldenberg T. 2015. In response: Some species sensitivity distribution statistics revisited—A governmental perspective. *Environ Toxicol Chem* 34:2442–2444.
6. Crane M. 2015. In response: Challenges when weighing evidence about environmental risks—An industry perspective. *Environ Toxicol Chem* 34:2444–2445.
7. Hutchinson H, Burden N. 2015. In response: Benefits of the ARRIVE guidelines for improving scientific reporting in ecotoxicology—An academic perspective. *Environ Toxicol Chem* 34:2446–2448.
8. Fox DR. 2010. Statistics and ecotoxicology: Shotgun marriage or enduring partnership? *Integr Environ Assess Manage* 6:501–502.
9. Hothorn LA. 2016. *Statistics in Toxicology using R*. Chapman and Hall/CRC, Boca Raton, FL, USA.
10. Ritz C, Streibig JC. 2005. Bioassay analysis using R. *J Stat Softw* 12:1–22.
11. King GKK, Veber P, Charles S, Delignette-Muller ML. 2014. MOSAIC_SSD: A new web tool for species sensitivity distribution to include censored data by maximum likelihood. *Environ Toxicol Chem* 33:2133–2139.
12. Crane M, Newman MC. 2000. What level of effect is a no observed effect? *Environ Toxicol Chem* 19:516–519.
13. Wilson JD. 2001. Advanced methods for dose–response assessment: Bayesian approaches—Final report. Discussion paper 01–15. Resources for the Future, Washington, DC.
14. Fox DR, Billoir E, Charles S, Delignette-Muller ML, Lopez C. 2012. What to do with NOECS/NOELS—Prohibition or innovation? *Integr Environ Assess Manage* 8:764–766.
15. Aldenberg T, Rorije E. 2013. Species sensitivity distribution estimation from uncertain (QSAR-based) effects data. *Altern Lab Anim* 41:19–31.
16. Fox DR. 2015. More noise does not mean more precision: A review of Aldenberg and Rorije (2013). *Altern Lab Anim* 43:241–249.
17. Stephan CE, Rogers JW. 1985. Advantages of using regression analysis to calculate results of chronic toxicity tests. In Bahner RC, Hansen DJ, eds, *Aquatic Toxicology and Hazard Assessment: Eighth Symposium*. ASTM STP 891. ASTM International, Philadelphia, PA, pp 328–338.
18. Horton R. 2011. Foreword. In Everitt BS, Palmer CR, eds. *Encyclopaedic Companion to Medical Statistics*, 2nd ed. Wiley, Chichester, UK.
19. Davenport TH, Patil DJ. 2012. Data scientist: The sexiest job of the 21st century. *Harvard Business Review*. 2012 October. [cited 2016 January 19]. Available from: https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century
20. American Statistical Association. 2015. A peek into the largest, fastest-growing undergraduate statistics departments. *AMSTAT News*. 2015 February 1. [cited 2016 January 19]. Available from: http://magazine.amstat.org/blog/2015/02/01/undergraduatedepts_feb2015/
21. University of Melbourne. 2016. Our research. [cited 2016 January 19]. Available from: http://collision.unimelb.edu.au/#layer-research
22. Hothorn LA. 2014. Statistical evaluation of toxicological bioassays—A review. *Toxicol Res* 3:418–432.