Figure 1. (A) Relationship between *Lemna gibba* laboratory toxicity data (EC50) and *Lemma gibba*/other macrophyte field toxicity data (EC50) for a variety of plant protection products and other chemicals. **(B)** Relationship between *Lemna minor* and a variety of other macrophyte toxicity data for four commonly used herbicides.

Recommendations

Clearly, more testing with *Lemna* spp. in the field is needed in order to better understand the utility of this laboratory assay, since so few data were available for use in this investigation. With the laboratory and field results consistently falling within an order of magnitude of each other, the application of a 10-fold uncertainty factor, much like the EU approach, is more than adequate to account for variability between the laboratory and field for duckweed and also to account for interspecies differences. Currently, Canada applies an uncertainty factor of 2 and the United States does not apply any uncertainty factor. This may result in unintended toxicity in the field at times and may require revision as more data are collected, but even with these lower uncertainty factors, it is unlikely that significant impacts on macrophytes are being missed. Overall, the *Lemna* spp. assay remains an efficient test system. Duckweed tests are short in duration (typically, 7 d as opposed to the minimum 14 d proposed for rooted submersed macrophytes); the plants are relatively small, meaning less space is required and less waste is generated by the assay; and they are easily maintained in the laboratory allowing testing to occur year round from a consistent stock, reducing variation within and between tests. The duration of duckweed tests can be increased to 14 d or longer, in order to provide more conservative predictions of toxicity to other macrophytes that would be exposed for similar durations. Also, while root measures may be significantly more sensitive in submersed macrophytes relative to other growth measures in *Lemna* spp., perhaps the solution is to find ways to accurately assess the rooting structure of duckweed to see if it too is as sensitive. If work on new macrophyte test species is conducted, there should be evidence that the proposed plant is consistently and significantly more sensitive than *Lemna* spp., otherwise the effort will likely not result in any substantial increase in our understanding of toxicity to macrophytes at the lower tiers of risk assessment.

REFERENCES

Breitholtz M, Ruden C, Hansson SO, Bengtsson B. 2006. Ten challenges for improved ecotoxicological testing in environmental risk assessment. *Ecotoxicol Environ Saf* 63:324–335.

Davy M, Petrie R, Smrchek J, Kuchnicki T, Francois D. 2001. Proposal to update non-target plant toxicity testing under NAFTA. www.epa.gov/scipoly/sap/meetings/2001/june/sap14.pdf. Accessed 16 January 2008.

Fairchild JF, Ruessler DS, Carlson AR. 1998. Comparative sensitivity of five species of macrophytes and six species of algae to atrazine, metribuzin, alachlor, and metolachlor. *Environ Toxicol Chem* 17:1830–1834.

Hanson ML, Arts GHP. 2007. Improving regulatory risk assessment as it relates to aquatic macrophytes. *Integr Environ Assess Manag* 3:466–467.

Van den Brink PJ, Blake N, Brock TCM, Maltby L. 2006. Predictive value of species sensitivity distributions for effects of herbicides in freshwater ecosystems. *Hum Ecol Risk Assess* 12:645–674.

IS THE EC$_x$ A LEGITIMATE SURROGATE FOR A NOEC?

David R Fox*
University of Melbourne, Melbourne, Australia
*david.fox@unimelb.edu.au

The NOEC has been a pivotal quantity in ecotoxicology and continues to be used despite a plethora of papers highlighting its many and severe shortcomings; for instance, Kooijman (2006) stated that "the NOEC should be banned." In response to such criticisms, there is increasing usage of EC$_x$

other ($R = 0.61$, $p < 0.01$, $n = 17$), although field data were consistently the most sensitive (Figure 1A). Although field data showed higher sensitivity, if a 10-fold uncertainty factor is applied, which is used in the European Union, all EC$_{50}$ values from laboratory tests were captured relative to field data. When compared to other macrophytes, *L. gibba* was consistently more sensitive in both field and laboratory settings ($R = 0.83$, $p < 0.01$, $n = 49$) (Figure 1A).

While conducting this small study, we noted that most studies that compare the relative sensitivities between duckweed and other macrophytes do so without considering the actual durations of the assays themselves. If the standard 7-d duckweed assay was increased to 14 d, much like that of a rooted submersed macrophyte test, an increased exposure duration would show higher sensitivity in duckweed. A study conducted by Fairchild et al. (1998) exposed *L. minor* and 4 submersed macrophytes to 4 commonly used herbicides in the laboratory to assess toxicity for 4-d and 14-d durations, respectively. If the *L. minor* data were adjusted with Haber's rule to a 14-d duration, the duckweed data were more or as sensitive as the other species tested, including rooted submersed dicots ($R = 0.63$, $p < 0.01$, $n = 16$) (Figure 1B). Therefore, by applying a 10-fold uncertainty factor and/or a 14-d exposure duration, duckweed would appear to do a reasonable job in protecting other macrophytes from toxicity in the field.

either as a surrogate or alternative to the NOEC. However, I believe that this practice is flawed, and suggest, along with Kooijman (2006), Jager et al. (2006), and others that NOECs be replaced by empirical estimates of model-based no-effect concentrations (NECs).

The NOEC is a surrogate for the NEC and is routinely determined as the concentration in a series of dilution experiments for which the mean response is statistically indistinguishable from the mean response of a "control" group. The most often cited shortcomings of the NOEC include:

- The NOEC is one of the test concentrations;
- The procedure by which a NOEC is determined "rewards bad experiments";
- Statements of precision/uncertainty are not possible;
- NOECs cannot always be determined;
- The size of the NOEC is a function of the choice of statistical test and level of significance.

The statistical procedure most often used to assess the significance of differences between the control response and responses at all other concentrations is Dunnett's test, which is a special case of a more general class of procedures referred to as multiple comparison techniques. Multiple comparison techniques are a companion tool for analysis of variance tests and should only be used after the null hypothesis of the equality of several means has been rejected by the ANOVA procedure—because the multiple comparison tests help identify the source of the significant ANOVA result. The analysis of variance technique examines components of variance but is a blunt instrument that can only conclude that a group of k means are either all the same (the null hypothesis) or at least 2 means are different (the alternative hypothesis). Rejection of the null hypothesis tells us nothing about which means are different.

Multiple comparison procedures such as Fisher's test, Tukey's test, Student Newman Keuls test, Hsu's test, and Dunnett's test are all based on pairwise comparisons of means to explore these differences. The test procedures differ primarily in terms of which error rate is being controlled for (e.g., individual Type I error rate, overall "experimentwise" error rate). Dunnett's test focuses specifically on the $k-1$ comparisons of treatment means with the designated "control" group mean and not on comparisons among pairs of noncontrol responses. Many dose–response experiments employ a geometric progression of dilutions (e.g., a halving of successive concentrations); thus, the NOEC could be in error by up to the same constant of proportionality (e.g., by a factor of 2). Notwithstanding the other important (but often neglected) ANOVA assumptions of normally distributed responses, independence, and constant error variance, why is this the preferred way of estimating the NEC? Why, when we have the opportunity of modeling the dose–response data from which we can directly estimate the NEC, do we elect to use a less efficient and dubious multiple comparison test procedure with all its acknowledged faults? Perhaps it's because there's only one way of performing Dunnett's test whereas there are a multitude of dose–response models, thereby imparting an assumed element of standardization. I don't find this a compelling argument.

Multiple comparison procedures are wasteful of information, are not predicated on any understanding of the system/experiment, and accordingly represent a dumbing-down of ecotoxicology. Nelder (1999) noted that "multiple-comparison methods have no place at all in the interpretation of data," attributing the widespread use of such "non-scientific statistics" to an obsession with p values leading to "the cult of the single study and the proliferation of multiple-comparison tests." He makes a convincing argument for increased focus on modeling, claiming the basis for a "good" model is one that is 1) a priori reasonable, 2) parsimonious, and 3) internally consistent. I believe that dose–response modeling should focus on these 3 qualities rather than the routine application of unstructured statistical tests. This view is consistent with the calls for the use of biologically based models such as those based on Dynamic Energy Budget theory (Kooijman 2006; OECD 2006).

Unfortunately, the problems with the NOEC and its associated statistical methodology are neither overcome nor ameliorated by using a different measure, such as an EC_x. Estimation of the EC_x may be more reliable because we are attempting to estimate something less extreme than the NEC, but there is a fundamental schism that renders the derived value difficult to interpret, for instance when used to develop species sensitivity distributions (SSDs). The schism is explained by the terms themselves: one relates to an effect, the other relates to no effect. While there is nothing to prevent us from taking a collection of EC_x values, fitting a log-logistic distribution, and then using a low-order percentile from this fitted distribution as a threshold concentration, it does generate both a philosophical and operational dilemma. If one accepts the definition of an SSD as being the probability distribution of some measure of toxicity, then this procedure generates an SSD from which an HC_p could be determined. More generally, we can talk of the HC_p determined from the SSD based on EC_x data as being the concentration having an effect of no more than x% on at least $(100 - p)$% of all species. However, except for $x = 0$, this is an awkward and convoluted concept. What fraction of the population is protected by keeping environmental concentrations below an HC_p that has been determined from the SSD fitted to EC_x data, and who decides on the value of x? The questions become even more difficult to answer if, as is often done, we introduce the notion of a $(1 - \alpha)100$% confidence limit on the estimated HC_p. Based on an assumption of a log-normally distributed SSD, van der Hoeven (2004) showed that, when the variance between species is larger than the variance within a species, a relatively large portion of affected species is affected severely.

Further, a no-effect does not correspond to an effect of any magnitude—except zero! While I concede that a confidence interval for the NEC may include concentrations at which an effect is possible, this does not establish a correspondence between an NEC and an EC_x and certainly does not provide a reason for substituting one for the other; to do so only results in further obfuscation.

The NOEC is the largest concentration at which the observed mean response is not statistically different from the mean response of the control group. Whether or not the control response represents an effect or no effect is not considered. It is not uncommon, for example, to see some mortality at a control dose representing a zero concentration of the test chemical. This might be due to factors totally unrelated to the experiment, such as natural mortality or attrition. For the statistician, such outcomes represent "noise" around a true response. Perhaps the confusion could be removed by talking about the control response rather than no effect. The outcome from Dunnett's test (if it must be used) is the largest

concentration at which there is no statistical difference from the control response. Thus the NOEC could be replaced by the IFCR (indistinguishable from control response) or NDFCR (not different from control response)—but I suspect we don't need any more acronyms!

So, where does this leave us? For me, the answer is clear: Invoke principles 1 to 3 above as advocated by Nelder (1999) and adopt a model-based approach to describe the fundamental dose–response mechanism, and the rest follows. The NEC and EC_x are, respectively, a parameter estimate and a model prediction from one and the same model. Uncertainty in these values is handled with confidence and prediction intervals. In an article currently under review, I describe a Bayesian approach for setting credibility intervals using posterior and predictive distributions (Fox, in review). The old method of using an unstructured, uninformed, and insensitive multiple comparison procedure is a bankrupt approach that deserves to be buried. Only then can we move forward and focus on more interesting modeling and estimation issues rather than trying to find ways to prop up the thoroughly flogged, dead NOEC horse.

Acknowledgment

REFERENCES

Jager T, Heugens EHW, Kooijman SALM. 2006. Making sense of ecotoxicological test results: towards application of process-based models. *Ecotoxicology* 15:305–314.

Kooijman SALM. 2006. An alternative for NOEC exists, but the standard model has to be abandoned first. *Oikos* 75:310–316.

Nelder JA. 1999. Statistics for the millennium: From statistics to statistical science. *The Statistician* 48:257–269.

[OECD] Organisation for Economic Co-operation and Development. 2006. Current approaches in the statistical analysis of ecotoxicology data: A guidance to application. Series on testing and assessment, 54. Environmental Health and Safety Publications, Series on testing and assessment. ENV/JM/MONO(2006)18.

van der Hoeven N. 2004. Current issues in statistics and models for ecotoxicological risk assessment. *Acta Biotheoret* 52:201–217.

IS SELENIUM A GLOBAL CONTAMINANT OF POTENTIAL CONCERN?

Peter M Chapman*
Golder Associates
Vancouver, BC, Canada
*pmchapman@golder.com

Selenium background

Selenium (Se) is a metal-like element discovered in 1818 by the Swedish chemist Berzelius and named after Selene, the Greek goddess of the moon. It is a naturally occurring essential element required for the health of humans, other animals, and some plants.

Selenium enters the aquatic environment as a result of natural weathering, sometimes resulting in concentrations higher than most anthropogenic sources (e.g., $>10^3$ µg/L); mining that accelerates natural release from Se-enriched rock strata (coal, phosphate, uranium, some precious and base metal mines); refining/smelting of Se-enriched rock and crude oil containing Se; combustion of fossil fuels; agricultural drainage; production/usage of animal feed (and human) supplements and other Se-enhanced personal care products (e.g., shampoo such as Selsun Blue™); and recently, nanomaterials (e.g., inner core of quantum dots; Bouldin et al. 2008).

Selenium is unique

Selenium has the smallest range between essentiality and toxicity of any of the essential elements. It has 2 modes of toxicity in the aquatic environment: acute via water exposure at relatively elevated concentrations (e.g., >100 µg/L), and chronic to egg-laying animals (e.g., fish, water birds, and possibly amphibians) via dietary exposure at lower concentrations (e.g., <10 µg/L).

Chronic toxicity can occur when inorganic Se released by the processes noted above is biologically transformed into organic Se. Organic Se can be accumulated through the food chain to fish, water birds, and amphibians, and can result in deaths or deformities in the offspring. Organic Se maternally transferred into eggs can substitute for sulfur during the development of amino acid chains necessary for protein synthesis in developing embryos. Transfer of inorganic Se into dietary organo-selenium occurs via a complex series of interconnected hydrological, biogeochemical, and biological pathways that vary over time, among sites, and among receptor taxa.

Selenium puzzles

Various generic puzzles are associated with Se in the aquatic environment, including:

- Increasing concern exists regarding Se in North America, Australia, and New Zealand, but it is generally ignored in most other areas of the world aside from its importance in nutrition and in ameliorating mercury (Hg) bioaccumulation and toxicity. It is unclear why this difference should exist since Se is naturally present in other parts of the world and its release to the environment is similarly enhanced by anthropogenic activities.
- Water Se concentrations do not necessarily predict bioaccumulation and hence toxicity. It is presently uncertain how to relate water concentrations to food chain uptake for regulatory and management purposes.
- Chronic Se toxicity decimated fish populations in Hyco Reservoir (North Carolina, USA), but there was no apparent effect on adjacent Mayo Reservoir with similar Se inputs. There are more sites for which there are predicted Se effects than sites for which there are demonstrated effects.

The reasons for the final puzzle, above, are most probably related to some combination of 3 factors: biogeochemistry of Se (differences in speciation), differences in biota species sensitivity/tolerance to Se, and differences in the different forms/species of Se in different aquatic food chains.

Different sites also have their own puzzles; for example, the following 3 specific puzzles arising from Se released from coal mining into the aquatic environment of the Elk River Valley (BC, Canada). Chapman and de Bruyn (2007) found no increase in Se fish muscle concentrations over 10 y despite approximately 8% increases in water concentrations yearly. In other words, although Se water concentrations doubled over those 10 y, there was no change to the amount of Se in fish muscle. Selenium in benthic invertebrates and in fish eggs also showed a surprising constancy (EVSTF 2008).

Harding (2007) found that red-winged blackbirds exhibited a nonlinear relationship between water Se and mean egg Se. These birds appear to have a declining ability to accumulate Se in their eggs past about 24 mg/kg dry weight. The beneficial