

WHAT LEVEL OF EFFECT IS A NO OBSERVED EFFECT?

MARK CRANE*† and MICHAEL C. NEWMAN‡

†School of Biological Sciences, Royal Holloway, University of London, Egham, Surrey TW20 OEX, United Kingdom

‡Department of Environmental Sciences, The College of William and Mary, Virginia Institute of Marine Science, Gloucester Point, Virginia 23062-1346, USA

(Received 9 December 1998; Accepted 6 June 1999)

Abstract—The no observed effect concentration (NOEC) is currently a major summary statistic in ecotoxicity testing. Its use is heavily criticized, partly because it is a poor estimator of “safe” chemical concentrations. In this short communication, we review the limited information available on the percentage effect that corresponds with the NOEC, a value designated the EC_{NOEC} , and calculate EC_{NOEC} values for fish growth data. The average EC_{NOEC} for different test protocols was found to vary widely, with values for chronic studies ranging from 10 to 34%. Individual results varied even more widely. This analysis suggests that the NOEC is neither a consistent summary statistic nor an indicator of safe concentrations of toxic chemicals.

Keywords—No observed effect concentration Median effective concentration Median lethal concentration Risk assessment

INTRODUCTION

Statistical analysis is essential for effective interpretation of ecotoxicity test results. It allows estimation of summary statistics such as the no observed effect concentration (NOEC) and the median effective concentration (EC50) for continuous or quantal data. Currently, two general approaches are used to analyze such data [1]: (1) hypothesis tests that determine the NOEC and lowest observed effect concentration (LOEC) and (2) regression analyses that estimate an effect concentration, for example, a 96-h LC50 (median lethal concentration) or EC50.

The NOEC is the highest concentration in a test with a mean response not differing significantly from the mean response of the control if compared statistically. The LOEC is the lowest test concentration having a mean response that does differ significantly from that of the control. Both are generated by post-analysis of variance (ANOVA) multiple comparison tests, such as Dunnett's test [2].

The perceived advantage of the NOEC relative to regression-derived estimates, such as the LC50 or EC50, is that it is easy to calculate, easy to understand, and is an important component of current chemical and effluent risk assessment procedures [3–5]. Its derivation does not require the assumption of a specific model, such as a lognormal concentration-effect model. However, its application does carry the assumption of a toxic threshold: that there is no effect below some threshold concentration. This threshold concentration is often assumed to be between the NOEC and LOEC.

There are major disadvantages to the use of the NOEC that are increasingly considered to be insurmountable [6,7] and provide impetus for rejection of NOEC-associated methods. Four of the most important disadvantages follow.

First, the current use of post-ANOVA testing to identify a concentration at which there is no effect is a fundamentally invalid interpretation of hypothesis testing. With a hypothesis

test, you can ascertain whether there was an effect at a particular concentration with a specified chance of being wrong (the Type I error), but when you do not detect a difference, you cannot state that an effect was not present. Failure to reject the null hypothesis of no difference does not mean that there was no effect. The NOEC may not be a “safe,” no-effect concentration because a large, biologically significant effect may occur at a test concentration yet not be detected by a hypothesis test. The ability to detect a statistically significant effect depends not only on biological response but also on the experimental design (e.g., number of replicates at each concentration and spacing of concentrations), variation within treatments, and the power ($1 - \beta$) of the applied statistical test. The correct interpretation of error rates for hypothesis testing is inconsistent with current application of the NOEC in risk assessment. “The use of conventional hypothesis testing procedures (with $\alpha = 0.05$ and β , the Type II/false negative error rate, unconstrained) implies that it is very important to avoid declaring that a concentration is toxic when it is not, but it is not so important to avoid declaring that a concentration is not toxic when it is” [8]. The emphasis on Type I error and cursory consideration of Type II error are inconsistent with risk assessment, in which the probability of not identifying a problem is more of a concern than the possibility of falsely noting a problem when one does not exist. Dixon [9,10] and Erickson and MacDonald [11] identify a way of shifting the focus back to an appropriate and statistically valid context. They advocate the replacement of the conventional ANOVA-derived NOEC approach with bioequivalence testing methods.

Second, the generally accepted α of 0.05 used to identify statistical significance is a reasonable but arbitrary number. A biologically adverse effect with an associated $p = 0.06$ might be ignored as insignificant if the assessor inappropriately applied the test α of 0.05 to hypothesis test results in a thoughtless fashion. The term “significant” suggests a significant deviation from the null hypothesis, not biological significance.

Third, the NOEC is strongly affected by the specific con-

* To whom correspondence may be addressed (m.crane@rhbc.ac.uk).

Table 1. Percentage reduction in response corresponding to the maximum acceptable toxicant concentration (MATC)^a

Response	n ^b	Mean (%)	Range (%) ^c
Parental generation survival after spawning	13	20	0–61
Hatching success (hatch of normal larvae)	16	12	0–69
Fecundity (eggs produced per female)	27	42	0–87
Larval survival to juvenile stage	40	19	0–71
Early juvenile weight	52	20	0–44
Weight/egg ^d	24	35	0–71
Lowest observed effect	77	28	0–84

^a The MATC was estimated to be the geometric mean of the no observed and lowest observed effects concentrations (modified from table 2 of Suter et al. [12]).

^b n = number of studies included in the analysis.

^c 0 was substituted for any values ≤0.1%.

^d Weight/egg = “an integrative parameter [that was] created by multiplying the number of eggs exposed at each concentration by the proportion hatching, the proportion of the larvae surviving, and the weight of the survivors” [12].

concentrations chosen for the test. The NOEC and LOEC can take the values only of a tested concentration.

Finally, because the NOEC must be selected from one of the test concentrations, there is no way of producing confidence limits for it.

These and other shortcomings have led many ecotoxicologists to advocate gradual replacement of the NOEC approach with a regression approach [6,7]. However, a temporary lack of information needed to make this transition currently results in a hesitancy to adopt the regression approach. As one important example of this lack of knowledge, the general response level (EC_x) equivalent to the NOEC values from ecotoxicity tests has been addressed in only a few cases. Statements such as that of Stephan and Rogers [6] that “10 to 50% adverse effect can occur at concentrations that are reported to be no different from the control” underscore the possibility that the level of effect at the NOEC might be quite high. Also, the question emerges whether a statistic with such high effect levels and wide variation is a valuable tool or simply another source of uncertainty.

NO-EFFECT LEVELS IN THE LITERATURE

Several studies suggest, but do not directly provide, an answer to the question of what level of effect is occurring at the NOEC for conventional ecotoxicity tests. In one such study [12], the geometric mean of the NOEC and LOEC was used as a measure of the maximum acceptable toxicant concentration (MATC) of a large and diverse data set for chronic effects on fish. These data were also used in regression analysis and the percentage effect level corresponding to the MATC estimated with a regression model (Table 1). The MATC values averaged from 12% (hatching success) to 42% (fecundity). The average across all response scales was 28% with a wide range from less than 0.1% to 84%. In other words, the effect level equivalent to the MATC was roughly 28% with very wide variation around this mean. On the basis of this and additional analyses, Suter et al. [13] suggested that one should not generally expect to detect an effect that was less than about 20 to 25%. Obviously the effect level for the NOEC would be lower than that of the MATC.

Other studies also imply the general level of effect corre-

Table 2. Summary of regression-derived percentages corresponding to the no observed effect concentration (NOEC) from an OECD ring test for *Daphnia magna* reproduction [14]

Toxicant	Mean (%)	Range (%)
3,4-Dichloroaniline	8.8	0–37.1
Cadmium	9.6	1.4–19.3
Phenol	9.0	0.1–31.9

sponding to a NOEC. In an analysis of *Daphnia magna* chronic toxicity ring test data, the calculated least significant difference for the post-ANOVA testing (Dunnett's test) suggested that a difference of 20 to 25% or higher could be seen as significantly different from the control mean [14]. The least significant difference indicates how large a difference must be between the control mean and any treatment mean in order to be judged significant, based on a 95% confidence limit (i.e., 95% of the times when such an experiment is done, a difference of this magnitude would be detected with the hypothesis test). Focusing on the statistical power of the conventional *Ceriodaphnia dubia* reproductive effects test, Oris and Bailer [15] calculated that the least significant difference expected with 10 animals per treatment ranged from 31 to 100%. (Table 2 in Norberg-King [16] suggested that the NOEC for this test corresponds to about the 25% effect level.) Similarly, van der Hoeven [17] performed power analyses on standard ecotoxicity tests using *Daphnia magna*, *Eisenia fetida*, and *Folsomia candida*. For *D. magna*, a drop in reproduction of 25% or greater would have been detected in 95% of tests. However, for the other two species, a reduction of 20% would have been detected in only one of five tests. A decrease as large as 44% was required to be detected in 8 of 10 tests with these two species.

Few studies appear to go further by using regression to estimate the effect levels corresponding to the NOEC: We found only one study that answered this question directly. Chronic toxicity to *Daphnia magna* was measured during a ring test and regression methods used to estimate the level of effect corresponding to the NOEC [14]. The NOEC values (Table 2) corresponded to about a 10% level of effect; however, there was wide variation (0–37%) among NOEC values.

We performed two additional analyses of existing data because directly relevant information was so limited in the literature.

ANALYSIS OF NO-EFFECT LEVELS IN FISH CHRONIC TOXICITY TESTS

Chronic fish growth data

The first data set came from ring tests with the trout *Oncorhynchus mykiss*. These data were from 28-d juvenile growth tests in which fish were exposed to 3,4 dichloroaniline (DCA) or linear alkylbenzene sulfonate (LAS) [18,19]. These data were analyzed in the following way.

An NOEC was calculated for each test result by using one-way ANOVA and an appropriate multiple comparison test. If responses increased monotonically with increasing concentration, Williams's test was used, as this is generally considered to be among the most powerful under these circumstances [2]. Dunnett's test was used if responses did not change monotonically with concentration [2].

The fish growth data were fish weights and growth rates (i.e., continuous responses). The 28-d specific growth rate

Table 3. EC_{NOEC} values from the OECD fish growth ring test [18]

Test substance	Laboratory	NOEC (µg/L)	Control mean specific growth rate (SGR)	Regression intercept	Regression slope	Regression r ²	EC _{NOEC} (µg/L)
DCA	1	55	3.12	3.21	-0.00595	0.97	10.49
DCA	2	158	2.29	2.41	-0.00332	0.95	22.92
DCA	3	26	2.99	2.82	-0.00391	0.88	3.40
DCA	4	54	2.56	2.58	-0.00252	0.90	5.33
DCA	5	49	3.35	3.33	-0.00311	0.90	4.55
DCA	6	147	3.22	3.24	-0.00240	0.98	10.89
DCA	7	146	1.00	1.02	-0.00263	0.96	38.35
DCA	8	126	2.14	1.90	-0.00148	0.78	8.72
DCA	9	54	2.49	2.60	-0.00360	0.95	7.84
LAS	1	826	3.66	3.69	-0.00058	0.73	13.07
LAS	3	1,130	2.77	2.88	-0.00059	0.95	24.13
LAS	6	390	2.74	2.60	-0.00023	0.72	3.30
LAS	6	880	3.16	3.17	-0.00053	0.70	14.76

(SGR) data from the OECD fish growth ring test were fitted to a linear model [18]:

$$y = 100[a(1 - bx)]$$

where $y = EC_{NOEC}$ (the percentage effect at the NOEC), $a =$ intercept (the estimated control specific growth rate), $b =$ slope of mean response, and $x = NOEC$.

Data from the ring test were included only if a NOEC could be calculated using Williams's test or Dunnett's multiple comparison test [19]. Nine sets of data were used in which DCA was the test chemical and four in which LAS was tested.

Table 3 shows EC_{NOEC} values estimated from the fish growth ring test. The median level of effect for the EC_{NOEC} was 10.5%, but values ranged from 3.4 to 38.4% for the DCA data and from 3.3 to 24.1% for the LAS data.

Chronic lethality of fathead minnow to illustrate the test dependence of the NOEC

The widely used data for fathead minnow (*Pimephales promelas*) chronic toxicity of sodium pentachlorophenol from Weber et al. [20] were analyzed by maximum likelihood estimation using a lognormal regression model, that is, the conventional probit analysis, except that the normal equivalent deviate was used instead of the probit (a "normit" analysis). The effect level was estimated from the normit model: Normal equivalent deviate = 1.457(log₁₀ concentration) - 3.920. With this equation, the percentage effect level corresponding to a NOEC can be estimated. However, as shown here, the NOEC is strongly influenced by the power of the selected post-ANOVA test. Using Dunnett's test, a NOEC of 256 µg/L was estimated. The predicted normal equivalent deviate of -0.411 for a NOEC of 256 µg/L corresponds to a 34% effect level. If the more powerful Williams's test is used, the NOEC becomes 128 µg/L, which corresponds to an effect level of 19%. Applying the less powerful Wilcoxon rank sum test results in a NOEC of >512 µg/L, corresponding to an effect level of >51%. Thus, the effect level can vary by more than a factor of 2.5, depending only on the test applied to a data set.

CONCLUSIONS

Average effect levels for the NOEC

Studies of the MATC suggest upper limits for the effect level at the NOEC. The MATC for effects on fish (Table 1) suggests an upper limit that ranges widely. The average for

the most sensitive effect was about 28%. Suter [12] concludes that an effect in the range of 20 to 25% is required for detection in a conventional hypothesis test. Studies focusing on hypothesis test power or least significant difference also suggest a similar effect level corresponding with the NOEC. The statistical power of the standard *D. magna* and *C. dubia* tests would allow one to detect a 25% and a 31 to 100% effect level, respectively. The few direct estimates of the average effect level associated with a chronic NOEC range from about 10% (*D. magna* ring test and fish growth ring test) to 34% (chronic fish mortality). It should be noted that ranges tend to increase with the number of samples and that the ranges reported for these small sample sizes would probably increase if there were more samples.

Variation in effect levels

Regardless of the specific test species, effect, or toxicant, there was very wide variation in the effect level corresponding to the NOEC. Some was associated with the system under study.

Test design and computational methods created significant variation. Assignment of the NOEC to a discrete treatment concentration was an inherent source of variation. If one could conduct an identical pair of toxicity tests, except that one test would have many treatment levels and the second would have only a few, the NOEC would usually be lower for the second than for the first test. Occasionally, they would be the same if the one highest level with no significant difference from the mean happened to be the same for both studies. If two tests were performed that varied only in the number of replicates per treatment, the test with the most replicates would be more likely to have a lower NOEC than the other. Finally, if identical tests were done but results were analyzed with tests differing in power (e.g., Williams's test vs the Wilcoxon rank sum test), the calculated NOEC would tend to be higher for the test analyzed by the statistical method with the lowest power. All these sources of variation exist in the analyzed data sets.

Biological relevance of the NOEC

The biological relevance of the effect levels is extremely important to consider when assessing the adequacy of the NOEC for protecting species. On the basis of the minimal data available for estimating chronic effect levels at the NOEC, average effect levels ranged from 10 to 34%. More important,

effect levels from individual tests ranged from nearly 0% to nearly 100%. An undefined amount of this variation has its source in the methods used to calculate the NOEC. Because test power is rarely reported in primary literature or compilations of NOEC data, it is usually impossible to estimate how much variation is associated with test design and statistics. However, the studies cited here imply that the variation contributed by test design and statistics was significant. Thus, a high NOEC could be entirely due to high variability in the measured endpoint leading to low power to detect significant differences between concentrations.

Despite the uncertainties inherent in extrapolating from laboratory ecotoxicity test results to effects on natural populations, the population fate at a NOEC corresponding to an effect level approaching 100% mortality, reproductive failure, or growth inhibition is fairly easy to predict. Populations would rapidly go extinct. Some NOEC values reported in the cited studies approached 100% effect [15]. Similarly, several power analyses suggested that the NOEC method would detect some effects only at extremely high effect levels. In light of these findings, it would be unreasonable to suggest that such NOECs provide reliable information on safe concentrations of toxic chemicals.

The analyses in this report were based on only a few studies while summarizing many others but clearly show the theoretical and practical inadequacies of the NOEC. Summarizing ecotoxicity data as a NOEC provides a compromised picture of chemical safety in the environment. In most cases, a risk assessor using NOEC values will have no way of knowing whether these values are indicative of low, medium, or high effects on the endpoint of interest, but the NOEC is rarely if ever an indicator of no effect. There are many uncertainties in risk assessment, some due to inherent variability in natural systems and thus intractable. The NOEC is an artificial construct, and the uncertainties associated with its use can be eliminated simply by replacing it with statistical summaries of transparent consistency, accuracy, and precision.

Acknowledgement—We thank Nicky Grandy, Peter F. Chapman, Claus Hansen, and two anonymous reviewers for commenting on earlier drafts. This work was funded by the Organization for Economic Cooperation and Development. This is contribution 2224 of the College of William and Mary's Virginia Institute of Marine Science and School of Marine Science.

REFERENCES

- Crane M, Grosso A, Janssen C. 2000. Statistical techniques for the ecological risk assessment of chemicals in freshwaters. In Sparks T, ed, *Statistics in Ecotoxicology*. John Wiley & Sons, Chichester, UK (in press).
- Newman MC. 1995. *Quantitative Methods in Aquatic Ecotoxicology*. Lewis, Boca Raton, FL, USA.
- Campbell PJ, Hoy SP. 1996. ED points and NOELs: How they are used by UK pesticide regulators. *Ecotoxicology* 5:139–144.
- de Bruijn JHM, Hof M. 1997. How to measure no effect. IV: How acceptable is the EC_x from an environmental policy point of view? *Environmetrics* 8:263–267.
- Organization for Economic Cooperation and Development. 1996. Report of the OECD workshop on statistical analysis of aquatic toxicity data, Braunschweig, Germany, October 15–17, 1996. Organization for Economic Cooperation and Development, Paris, France.
- Stephan CE, Rogers JW. 1985. Advantages of using regression to calculate results of chronic toxicity tests. In Bahner RC, Hansen DJ, eds, *Aquatic Toxicology and Hazard Assessment, 8th Symposium*. STP 891. American Society for Testing and Materials, Philadelphia, PA, pp 328–338.
- Chapman PF, Crane M, Wiles J, Noppert F, McIndoe E. 1996. Improving the quality of statistics in regulatory ecotoxicity tests. *Ecotoxicology* 5:169–186.
- Suter GW II. 1993. *Ecological Risk Assessment*. Lewis, Chelsea, MI, USA.
- Dixon PM. 1998. Assessing effect and no effect with equivalence tests. In Newman MC, Strojjan CL, eds, *Risk Assessment: Logic and Measurement*. Ann Arbor Press, Chelsea, MI, USA, pp. 275–301.
- Dixon PM, Garrett KA. 1994. Statistical issues for field experimenters. In Kendall RJ, Lacher TE, eds, *Wildlife and Population Modeling: Integrated Studies of Agroecosystems*. CRC, Boca Raton, FL, USA, pp 439–450.
- Erickson WP, MacDonald WP. 1995. Tests of bioequivalence of control media and test media in studies of toxicity. *Environ Toxicol Chem* 14:1247–1256.
- Suter GW II, Rosen AE, Linder E, Parkhurst DF. 1987. Endpoints for responses of fish to chronic toxic exposures. *Environ Toxicol Chem* 6:793–809.
- Suter GW II, Cornaby BW, Hadden CT, Hull RN, Stack M, Zafran FA. 1995. An approach for balancing health and ecological risks at hazardous waste sites. *Risk Anal* 15:221–231.
- Organization for Economic Cooperation and Development. 1997. *Report of the Final Ring Test of the Daphnia magna Reproduction Test*. OECD Environmental Health and Safety Publications Series on Testing and Assessment 6. Paris, France.
- Oris JT, Bailer AJ. 1993. Statistical analysis of the *Ceriodaphnia* toxicity test: Sample size determination for reproductive effects. *Environ Toxicol Chem* 12:85–90.
- Norberg-King TJ. 1993. *A Linear Interpolation Method for Sublethal Toxicity: The Inhibition Concentration (IC_p) Approach, Version 2.0*. National Effluent Toxicity Assessment Center Technical Report 03-93. U.S. Environmental Protection Agency, Duluth, MN.
- van der Hoeven N. 1998. Power analysis for the NOEC: What is the probability of detecting small toxic effects on three different species using the appropriate standardized test protocols? *Ecotoxicology* 7:355–361.
- Mallett MJ, Grandy NJ, Lacey RF. 1997. Interlaboratory comparison of a method to evaluate the effects of chemicals on fish growth. *Environ Toxicol Chem* 16:528–533.
- Organization for Economic Cooperation and Development. 1994. *Draft Guideline for Testing of Chemicals. Fish, Juvenile Growth Test—28 Days*. Paris, France.
- Weber CI, et al. 1989. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms. EPA/600/4-89/001. U.S. Environmental Protection Agency, Cincinnati, OH.