to tobacco smoke versus nonsmokers' hands (34 ng/hand vs 13 ng/hand, respectively). From this preliminary study, we conclude smoking cigarettes significantly increases PAH residue on smokers' hands by approximately 3 times that of nonsmokers.

This study attempted to be as realistic as possible when evaluating third-hand PAH residues resulting from 1 cigarette. Smokers were not asked to change their smoking habits, except to continuously hold the cigarette in 1 hand during the entire duration of the cigarette's burning. Hand size (that is, the adsorptive surface area), duration of smoking, and environmental conditions such as wind, temperature, and humidity, in addition to other factors, may potentially influence PAH concentration. We conducted our third-hand smoke studies outdoors under environmental conditions, and therefore hypothesize that a similar study conducted in the more stable conditions of an indoor environment may reveal higher levels of contaminant residues on surfaces and smokers' bodies.

Moir et al. (2008) quantified PAH concentrations in second-hand tobacco smoke, defined as environmental tobacco smoke that is inhaled involuntarily or passively by someone who is not smoking. Using their study and our data set, we carried out a ''back of the envelope'' calculation to estimate the percentage of sidestream smoke (i.e., second-hand smoke) that becomes third-hand smoke. We conclude that the PAH inventory on 1 hand of a smoker represents 0.1% to 6% of that emitted from sidestream smoke.

Third-hand PAH residues on a smoker's hand represent only a fraction of the total PAH reservoir for a smoker (compared to residues on all exposed skin and clothing). We have begun to quantify this load of chemicals as the first step in assessing the potential for smokers to act as vectors for impairment of indoor air quality. To completely capture the health risk posed by third-hand smoke, further studies from our research group and others need to address the off-gassing or desorption potential of these compounds and more fully evaluate the significance of third-hand smoke residues in impairing indoor air quality and/or increasing PAH exposure to subpopulations such as children. A thorough ranking of the importance of this exposure route compared to other exposures modes (e.g., release of PAHs from cooking methods such as open fires, incense burning, indoor tobacco smoking, etc.) also remain to be quantified.

REFERENCES

Matt G, Quintana P, Larson S, Hovell M, Bernert J, Song S, Novianti N, Juarez T, Floro J, Gehrman C., et al. 2004. Households contaminated by environmental tobacco smoke: Sources of infant exposure. *Tobacco Control* 13:29–37.

Menzie C, Potocki B, Santodonato J. 1992. Exposure to carcinogenic PAHs in the environment. *Environ Sci Technol* 26:1278–1282.

Moir D, Rickert W, Levasseur G, Larose Y, Maertens R, White P, Desjardins S. 2008. A comparison of mainstream and sidestream marijuana and tobacco cigarette smoke produced under two machine smoking conditions. *Chem Res Toxicol* 21:494–502.

Sleiman M, Gundel L, Pankow J, Jacob P, Singer B, Destaillats H. 2010. Formation of carcinogens indoors by surface-mediated reactions of nicotine with nitrous acid, leading to potential thirdhand smoke hazards. *PNAS* 107:6576–6581.

Winickoff J, Friebely J, Tanski S, Sherrod G, Matt G, Hovell M, McMillen R. 2009. Beliefs about the health effects of ''Thirdhand'' smoke and home smoking bans. *Pediatrics* 123:74–79.

WHAT TO DO WITH NOECS/NOELS—PROHIBITION OR INNOVATION?

David R Fox,*† Elise Billoir,‡ Sandrine Charles,§
Marie Laure Delignette-Muller,§ and Christelle Lopes§
†Environmetrics Australia and University of Melbourne, Victoria, Australia
‡Plateforme de Recherche ROVALTAIN en Toxicologie Environnementale et Ecotoxicologie, Valence, France
§Université de Lyon, Lyon, France
*david.fox@environmetrics.net.au
DOI: 10.1002/ieam.1350

Now more than 30 years old, ecotoxicology's ''problem child''—the no observed effect concentration/no observable effect level (NOEC/NOEL), continues to cause us grief. Recent articles by Landis and Chapman (2011), van Dam et al. (2012), Fox (2012), and Jager (2011) have reiterated long-known concerns with NOECs and NOELs as toxicity measures, and each has adopted a particular view on how to best move forward. Those views fall broadly into 2 camps: the prohibition camp that argues for a ban on the use and reporting of NOECs/NOELs, and the education camp that argues for skills enhancement and tool development. We place ourselves firmly in the latter category. We understand a third group, the status quo camp, which argues for the retention of NOECs and NOELs (perhaps with added caveats) is soon to emerge. The purpose of this Learned Discourse is to provide a clearer articulation of ''the problem'' as we see it and, importantly, to provide some constructive suggestions for an orderly transition to model-based toxicity estimation.

In a nutshell, the NOEC/NOEL ''controversy'' stems from inappropriate use of hypothesis testing in general and multiple comparison techniques in particular. We are not the first to raise this issue. For example, Newman and Clements (2008) note problems with the use of hypothesis testing to estimate toxicity measures and state ''the virtues and shortcomings of these 2 approaches have been and still are debated. . .this debate extends back to the origins of these approaches.''

So what does it take to effect change? Landis and Chapman (2011) argue that a ban is required. We agree that the time for action is now, but we remain unconvinced that prohibition will work or that it is even appropriate. There may be a few instances (such as a screening level risk assessment) where one simply wants to ascertain whether there is a toxic effect and hypothesis testing is an appropriate paradigm. Or, we might want to get a sense of the concentration ''neighborhood'' where that toxic effect ''kicks in'' and use this to inform the design of a follow-up concentration-response experiment—Dunnett's test is an appropriate tool. What is not appropriate is the use of hypothesis tests to generate point estimates of toxicity.

To be clear, statistical inference is concerned with 2 related but distinct problems: 1) hypothesis testing, and 2) estimation—it is ''the process of making conclusions on the basis of data that are governed by probability laws'' (Zelen 2005). Our conceptual model for statistical inference (Figure 1) suggests that, in the context of ecotoxicology, hypothesis tests should be used to assess the presence of a toxic effect whereas estimation techniques should be used to derive a measure of a toxic effect. If one accepts this

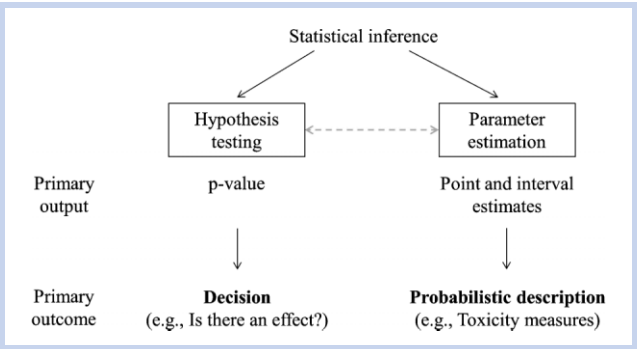

Figure 1. Conceptual representation of statistical inference in ecotoxicology.

taxonomy, then the root cause of the litany of issues with NOECs/NOELs becomes apparent: hypothesis testing has been used not to make a binary decision (accept or reject a null hypothesis about the presence of an effect) but to inappropriately estimate an unknown toxicity measure. Indeed, the NOEC/NOEL is effectively a label which is why it has no statement of precision or uncertainty!

Statisticians do not use hypothesis testing methods for estimation—they use models and statistical estimation theory. However, there is a subtle duality between hypothesis testing and statistical estimation that explains the connecting arrow in Figure 1. For example, consider a simple 1-way ANOVA model to test the hypothesis of equality of k ''treatment'' means (in our case, k concentrations):

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k.$$

Observations belonging to the i^{th} treatment group are represented by the model $Y_i = \mu_i + \varepsilon_i$ where ε_i is a random error term assumed to have a normal distribution with zero mean and variance σ^2. An alternative formulation is to replace the μ_i by the quantity $\mu + \alpha_i$ where μ is the overall mean and α_i the discrepancy between the mean of the i^{th} treatment group and μ with $\sum_{i=1}^{k} \alpha_i = 0$. Clearly, a true null hypothesis requires all α_i to be zero. Thus, the original hypothesis written in terms of means can be rewritten in terms of the treatment effects:

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_k = 0.$$

The estimation of the α_i terms is simply a means to an end in hypothesis testing—we use the estimates to test the hypothesis but not to predict a response at some previously unmeasured concentration, because that is not possible within this framework. In contrast, we use statistical techniques to estimate the parameters of a function describing the concentration–response relationship and then infer either a concentration for a prescribed response or the reverse.

Graduate students in statistics will be familiar with the classic texts by Erich Lehmann: *Testing Statistical Hypotheses* and *Theory of Point Estimation*. That there are 2 separate texts and not a single volume highlights the distinction we believe is fundamental to the misguided application of ANOVA and the companion methods of multiple comparison techniques: measures of toxicity ''live'' in the parameter estimation box of Figure 1 not the hypothesis testing box. So for us, the past complaints and current debate about the NOECs/NOELs are a non sequitur.

So where does this leave us? For us, the answer is clear—we move into the estimation box of Figure 1, which means fitting plausible models to concentration–response data and estimating parameters from the model that have ecotoxicological relevance—such as the no effect concentration (NEC), LC_x, EC_x, or other toxicity measures.

Our session (also cochaired by Wayne Landis), Advanced Statistical Methods in Ecotoxicology, at the recent SETAC Berlin meeting generated much interest and provided us with valuable feedback about this and other suggestions. A recurring concern expressed about the suggested move to model-based inference is that the derived toxicity measure will be influenced by a host of subjective decisions such as the choice of the mathematical model, its parameterization, and the estimation strategy. In response, we ask how is this any different to the raft of subjective (and often-times concealed) decisions used to generate a NOEC/NOEL such as: the set of concentrations to use; which multiple comparison procedure to use (e.g., Dunnett's test, Jonckheere–Terpstra test, Tamhane–Dunnett's test, Cochran–Armitage test, William's test, Mann–Whitney rank sum test with Bonferroni adjustment)?; the level of significance; and, response scale (e.g., original, log-transformed, arc sine transformed)?

We believe there has been a perpetuation of untested assumptions and myths about the role of model-based inference in ecotoxicology that has hindered education and tool development. For example, van der Hoeven (1997) contemplated the use of model-based alternatives to the NOEC/NOEL such as the NEC and EC_x but was dismissive of their use in ecotoxicology because ''data will seldom be sufficient for model verification.'' Although data paucity is an issue, van der Hoeven's argument is equally applicable for NOECs which, as has been repeatedly observed, relies on a process that rewards low-powered experiments. Newman and Clements (2008) reinforce this position with (in our view, flawed) advice that ''hypothesis testing is preferable to modeling if one has no understanding of the relationship between the effect and the toxicant concentration'' although they later ask the rhetorical question ''is it not better to confront these uncertainties at the onset of an investigation?'' Our advice is unequivocal—undertake a pilot study to fill the knowledge gaps and/or use nonparametric smoothers such as generalized additive models (Hastie and Tibshirani 1990), which is a robust statistical technique developed with precisely this type of application in mind.

Embedded in our call to fit plausible models for the generation of toxicity measures is the issue of competency in statistical modeling and estimation. We acknowledge the subjective element of model identification and estimation; however, this is not confined to problems in ecotoxicology—it is omnipotent and a characteristic of any modeling exercise. What we should be more concerned about are the modeler's credentials. To be blunt, anyone can use Excel or some other software with ordinary least-squares to fit a function to data. We have seen gross abuses of this process—for example, fitting a 5-parameter model to 5 data points with attendant claims of success on the basis of an R^2 of unity! For us, we would rather see toxicity measures derived through the use of appropriate tools of estimation by a suitably qualified practitioner than the rote use of statistically inappropriate methods. To this end we propose that quantitative ecotoxicologists be accredited by SETAC as having the necessary training and skills in statistics appropriate to the task of

concentration-response modeling, estimation, and inference. This may seem like a radical step, but in reality it is commonplace in other professions. For example, in Australia professional accountants can hold the Certified Practicing Accountant (CPA) title whereas the statistics profession has various accreditations including CStat (UK), PStat (USA), and AStat (Australia). The accreditation process does not have to be onerous, and we recognize that it is not fail-safe, but it would assist in the screening out of underqualified practitioners.

In conclusion, we find ourselves in agreement with sentiments expressed by van Dam et al. (2012) and Landis and Chapman (2011) about the need to elevate the discussion about NOECs/NOELs to considerations of affirmative action rather than cataloguing the deficiencies of these measures. We may differ in our views on the path but not on the destination.

REFERENCES

Fox DR. 2012. Response to Landis and Chapman (2011). *Integr Environ Assess Manag* 8:4.

Jager T. 2011. Bad habits die hard: The NOEC's persistence reflects poorly on ecotoxicology. *Environ Toxicol Chem* 31:228–229.

Landis WG, Chapman PM. 2011. Well past time to stop using NOELs and LOELs. *Integr Environ Assess Manag* 7:vi–viii.

Newman MC, Clements WH. 2008. Ecotoxicology: A comprehensive treatment. Boca Raton (FL): Taylor and Francis. 852 p.

Hastie TJ, Tibshirani RJ. 1990. Generalized additive models. New York (NY): Chapman and Hall. 352 p.

van Dam RA, Harford AJ, Warne MSJ. 2012. Time to get off the fence: The need for definitive international guidance on statistical analysis of ecotoxicity data. *Integr Environ Assess Manag* 8:242–245.

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

Zelen M. 2005. Inference. In: Armitage P, Colton T, editors Encyclopedia of biostatistics 2nd ed. Chichester, UK: John Wiley & Sons. pp 2533–2545.

COMMENT ON FOX et al. (2012): WHAT TO DO WITH NOECs/NOELs

Wayne G Landis*
Western Washington University, Bellingham, Washington, USA
*Wayne.Landis@wwu.edu
DOI: 10.1002/ieam.1347

I received a copy of the Fox et al. (this issue) Learned Discourse (LD) as it was being proposed for submission. David Fox, Elise Billoir, and I were co-chairs of the Statistical Methods session at the recent SETAC Berlin World Congress; this brief commentary is part of a larger conversation, discussion, and collaboration on the topic.

The Fox et al. (this issue) LD is one of the best short summaries of the statistical issues that I have read. As soon as it is published I will require my students to read it as part of our toxicology program. In several ways I find this proposal more restrictive than what Peter Chapman and I have proposed (Landis and Chapman 2011). Although we called for a ban of the use of hypothesis testing, we left the modeling segment open. Fox et al. have actually set specific criteria for modeling and curve fitting. These criteria could become a part of a reviewer's checklist when reviewing toxicity tests. I think that these criteria are great, but the education curve is going to be steep among the general community.

The way Fox et al. (this issue) addressed common questions and concerns regarding curve fitting was also appropriate. Overall I think the ecotoxicological community has little experience surrounding modeling of almost any sort and are new to many of the fundamental concepts. I also dismiss the issues raised by van der Hoeven (1997) and Newman and Clements (2008) regarding data paucity. Let us do the experiments to get the data and, as Peter Chapman and I pointed out (Landis and Chapman 2011), there are great examples of studies that do exactly that.

However, in screening level risk assessments I rarely see newly generated data. The information is usually from some published source and is often a no observed effect concentration/no observable effect level (NOEC/NOEL) or a toxicity reference value generated from hypothesis testing, probably approximately a 10% to 15% effect level (in other words, not a no-effect level). Years ago in a guideline document written for British Columbia (Canada) our team, which included Peter Chapman's group, suggested that an EC20 or lower be the cutoff even for what were pretty much screening level assessments. So the question is not, Is there toxicity?, but What is the amount of toxicity at each exposure?

In range finding tests, my students use a range of concentrations and we plot the data. The n (sample size) at each concentration is often not high so we can test more concentrations for the same number of organisms and we understand the loss of power. I found plotting the data and plotting a curve more useful for setting up the next set of tests than calculating a NOEC/LOEC in a range-finding exercise. After all, we are most interested in describing (modeling) toxicity at lower concentrations of the toxicant, levels likely to be seen in the environment.

One item that I did try to bring up in my talk at Society of Environmental Toxicology and Chemistry Berlin was that this switch means that the design of toxicity tests will have to be optimized for model construction, not hypothesis testing. This means more concentrations at lower doses (Olmstead and LeBlanc 2005; Rider and LeBlanc 2005).

Finally, my experience is that the ecotoxicological community in general is highly conservative. For instance, although the paradigm change regarding not using NOELs/NOECs occurred decades ago, the approach remains in widespread use. In the year since Landis and Chapman (2011), (39) and (14) additional articles have been published in *Environmental Toxicology and Chemistry* and *Integrated Environmental Assessment and Management*, respectively (search term NOEC, May 14, 2012). Without a clear reason to change I see little change occurring, hence the call for a ban. Bleaney (2012) concurs that a ban may be the only reasonable impetus for change. I am for an orderly transition to such a ban, but eventually the use of inappropriate hypothesis testing has to stop.

REFERENCES

Bleaney A. 2012. Letter to the editor. *Int J Environ Studies* 69:217–219.

Fox DR, Billoir E, Charles S, Delignette-Muller ML, Lopes C. 2012. What to do with NOECs/NOELs—Prohibition or innovation? *Integr Environ Assess Manag* 8:764–766.