

PRACTICAL IMPLICATIONS OF NONLINEAR EFFECTS IN RISK-ASSESSMENT HARMONIZATION

John A. Bukowski and R. Jeffrey Lewis □ ExxonMobil Biomedical Sciences, Inc.,
Annandale, NJ

□ *Cancer and noncancer health effects have traditionally been handled differently in quantitative risk assessment. A threshold (i.e., safe exposure) has been assumed for noncancer health effects, and low-dose linearity without a threshold has been assumed for cancer. “Harmonization” attempts to reconcile these contrasting assumptions under one paradigm. Recent regulatory initiatives suggest that the U.S. Environmental Protection Agency may be leaning toward a harmonized, probabilistic/linear approach for noncancer health effects. Proponents of this approach cite variability in human susceptibility as an argument against thresholds (i.e., some individuals may be exquisitely sensitive at exposures well below threshold levels). They also cite the results of epidemiological models that suggest low-dose linearity for noncancer health effects. We will discuss the implications of these arguments and compare them to what is known about human biological variability in general. We will also touch on the regulatory implications of hormesis within this framework.*

Keywords. nonlinearity, harmonization, risk assessment

INTRODUCTION

Cancer and noncancer health effects have traditionally been handled differently in quantitative risk assessment (QRA). For noncancer effects, there is the assumption of a “safe” exposure threshold, below which no effects are seen (i.e., the no-adverse-effect level). This is in keeping with the historical toxicological paradigm that “the dose makes the poison.” Cancer risk assessment has used a linear, no-threshold assumption, because cancer can be produced through a genetic mechanism, suggesting that even a single genetic error, if perpetuated, could lead to tumor formation. There is regulatory interest in harmonizing these two approaches under a single set of principles/paradigms (Bogdanffy *et al.*, 2001). Harmonization is very attractive from a regulatory perspective, because it simplifies the process by permitting the use of standardized methodology for all QRA. However, any harmonized QRA approach still needs to reflect the biology of the various diseases involved.

Recent regulatory initiatives suggest that the U.S. Environmental Protection Agency (USEPA) may be leaning toward a probabilistic, linear (i.e.,

Address correspondence to John A. Bukowski, ExxonMobil Biomedical Sciences, Inc., 1545 Route 22 East, Room LF278, Annandale, NJ 08801. Tel: 908-730-1105. Fax: 908-730-1192. E-mail: john.a.bukowski@exxonmobil.com

no-threshold) approach for noncancer health effects. For example, the QRA approach for particulate matter (PM) assumes linear, no-threshold effects for daily morbidity and mortality, (USEPA, 2001, 2002), and the QRA approach for lead accepts that there are linear decreases in children's IQ from even very low blood lead levels (USEPA, 2003). There is also evidence of movement toward increasingly stringent regulatory conservatism, so that even when a threshold is assumed, safe levels would be so low as to represent a *de facto* no-threshold approach. For example, the heightened conservatism associated with regulations such as the Food Quality Protection Act (FQPA) produces a level of uncertainty (often 1000-fold or more) that results in extremely low "safe" levels, consistent with "practical" low-dose linearity.

The current article explores the practical/logical implications of some of the assumptions inherent in low-dose linearity for noncancer effects. The logic behind these assumptions is examined and compared to real-world (e.g., clinical) examples to see if they are coherent with our current state of knowledge. Although we recognize that no-threshold extrapolations are not always linear and linear extrapolations are not always without a threshold, these concepts are similar, from a practical standpoint, and will be used somewhat synonymously in the current discussion.

CONSIDERATION REGARDING LINEAR (I.E., NO-THRESHOLD) LOW-DOSE EXTRAPOLATION

The linear, no-threshold assumption for cancer risk has a theoretical basis, even though it does not address genetic repair (especially from low doses) or carcinogens that act through nongenetic mechanisms. However, linear, no-threshold extrapolation for noncancer health effects runs contrary to the historical toxicological principle that all things are hazardous at some upper dose and nonhazardous at some lower dose. Instead, arguments in favor of linear low-dose extrapolation substitute theoretical assumptions of extreme variability in the range of human sensitivity and/or the results of complex mathematical models. These arguments also tend to ignore or marginalize the growing body of evidence that low-level exposures to many potentially hazardous agents may actually have positive health benefits (i.e., hormesis). These issues are discussed in more detail in the following sections.

Extreme Variation in Sensitivity

Arguments in favor of linear and/or no-threshold low-dose extrapolation for noncancer QRA are often predicated on the assumption of extreme variability in human sensitivity to environmental exposures, such that at least one individual is harmed by even minute exposures to a given agent. For example, in a draft of the PM Criteria Document, the USEPA suggested that a no-threshold linear model could approximate acute morbidity and mortality

from PM. This was partly based on an assumption that “since individual thresholds would vary from person to person due to individual differences in genetic level susceptibility and pre-existing disease condition, it would be almost mathematically impossible for a threshold to exist in the population” (USEPA, 2001). However, there are practical and logical limitations to the preceding theoretical argument.

On a fundamental level, this extreme sensitivity argument ignores the fact that individuals respond to stresses biologically, not “mathematically.” It provides no reasonable clinical/biological mechanism by which trivial exposure to an environmental agent would cause serious harm to even compromised individuals. It is tantamount to saying that one can cook an egg in three minutes at 212°F, so there should be some egg somewhere that can cook at 40°F, given enough time.

The extreme sensitivity argument ignores the fact that the range of susceptibility to physical/biological stress is generally finite (truncated), rather than some theoretical, infinite continuum. In fact, there are many clinical examples where extreme biological variation is incompatible with existence. Table 1 presents the range of normal values for several commonly measured blood parameters. Although it is recognized that some individuals may normally lie outside this range and that others may have abnormal values due to pre-existing medical conditions, there is usually a practical limit to these distributions. For example, random blood glucose levels exist in a range that brackets a mean of approximately 100 mg/dl (Table 1). There is arguably no individual not experiencing an extreme health emergency who has a blood glucose of either 10 or 1000 mg/dl, as such extremes are incompatible with continuing existence. Similarly, resting adult heart rates vary between perhaps 40 and 100 beats per minute, with values of 4 and 1000 being incompatible with long-term existence under any reasonable situation.

Variability arguments also often assume default extreme sensitivity for particular groups, such as children or the elderly. While it is true that these groups may have physiological differences that influence sensitivity to some agents, this is typically neither universal nor extreme. For example, Table 2 shows that dosage rates are usually quite similar between children and adults for drugs commonly used in both groups. This is true for relatively benign

TABLE 1 Range of Normal Laboratory Values for Selected Blood Parameters (Merck, 1999)

Albumin	3.5–5 g/dL
Hemoglobin	13.8–17.2 g/dL (males) 12–15.6 g/dL (females)
White blood cell count	3800–10,800 cells/ μ L
Random glucose	70–125 mg/dL
Blood urea nitrogen	7–30 mg/dL

TABLE 2 Estimated Dosages for Selected Pharmaceuticals (PDR, 1999)

Drug	Type	Adult dose	Pediatric dose	Other considerations
Azithromycin	Antibiotic	7 mg/kg d1 3.5 mg/kg d2–5	10 mg/kg d1 5 mg/kg d2–5	Similar pharmacokinetics in young adults, and children
Hydrocodone	Narcotic	5–15 mg (0.07–0.2 mg/kg)*	2.5–5 mg/d (6–12 yr.) (0.08–0.16 mg/kg)*	
Zyrtec	Antihistamine	5–10 mg/d (0.07–0.14 mg/kg)*	2.5 mg/d (2–5 yr.) (0.15 mg/kg)*	50% longer half-life in hepatic disease
Cytoxan	Cancer chemotherapy	40–50 mg/kg	Same	Reduce if low WBC or if combined with other agents
Vinblastine	Cancer chemotherapy	3.7–11.1 mg/m ²	3–6.5 mg/m ²	50% reduction in hepatic disease

*Assumes 70-kg adult and child weight appropriate for midrange age (USEPA, 1997).

antihistamines and antibiotics, as well as for more toxic cancer chemotherapeutic agents. Of course, there are drugs that have not been approved for use in children and some that are contraindicated because of adverse reactions, such as tooth mottling from tetracycline. However, even these drugs are unlikely to exert adverse effects at levels 100 times below the therapeutic range, as suggested by the FQPA. A possible exception to this might be rare instances of anaphylactic allergy, but these have not been demonstrated for trace exposures to environmental chemicals.

Extreme sensitivity arguments sometimes exhort the need to protect debilitated individuals who suffer from organ impairment, such as those with pulmonary disease, liver disease, or immune suppression. Although these arguments have some merit in general, here again the matter is one of degree. Drug warning labels provide a case in point. These have cautionary statements about the need to reduce dosages with liver or kidney impairment, but suggest reductions of perhaps 2-fold, not 10–100-fold (Table 2). One also needs to consider if populations of extremely debilitated individuals are even at risk from ambient environmental exposures. Those with severe impairment are likely to already live in protected environments or utilize protective equipment, such as oxygen supplies for pulmonary impairment or heightened air/water filtration for severe immune suppression. In such situations, slight changes in ambient pollutant levels are likely to be irrelevant.

Most risks addressed by conservative regulations such as the FQPA are derived from laboratory animal studies. In this situation, the QRA typically assumes a 10-fold uncertainty for animal-to-human extrapolation, an additional 10-fold uncertainty to address within human sensitivity, and an additional 10-fold uncertainty for extreme sensitivity (Kimmel, 2001). This rationale presupposes extrapolation from an average rodent to an average person, so that both routine and extreme human variation still need to be addressed.

However, QRA already begins with results in the most sensitive, rather than average, species of laboratory animal. Furthermore, these laboratory animals are highly inbred strains that lack the outbred vigor of wild animals and are often selected specifically for enhanced sensitivity to disease and environmental insult. In fact, researchers have noted that problems related to genetics and artificial laboratory environments have resulted in highly unthrifty laboratory animals with reduced life spans (Palazzolo, 1995). These factors suggest that laboratory animals are already highly sensitive individuals, so that extrapolation from animals to people might better be viewed as “sensitive to sensitive” rather than “average to average.” The conservatism of regulations such as the FQPA would therefore seem to be excessive.

Reliance on Statistical Models

Support for low-dose linear extrapolation increasingly rests on weak, but statistically significant, epidemiologic results produced with complex statistical models. For example, risk of acute mortality from fine (<10–2.5 μm in diameter) PM is based not on biological models of PM toxicity, but on complex time-series models suggesting a linear association between daily mortality and ambient PM levels (USEPA, 2001, 2002). These PM models generally produce relative risks (RRs) in the range approximately 1.005–1.05 per 10 $\mu\text{g}/\text{m}^3$ fine PM (USEPA, 2001, 2002). Traditionally, such extremely weak associations have been considered to be “outside the resolving power of the epidemiologic microscope” (Shapiro, 1994), especially when not accompanied by strong mechanistic support (Angell, 1990). There are also serious logical/practical limitations that argue against reliance on these results for decisions related to public health and resource allocation.

For one, statistical models are only as accurate as their assumptions. These models assume that data follow particular distributional patterns (e.g., normal, Poisson, etc.) and behave in predictable ways. Modeling assumptions are not always tested and, even when they are, traditional statistical tests (e.g., goodness-of-fit tests) can only detect relatively large departures from the assumptions. Small departures are difficult to detect but may still influence results, especially if these results suggest only tiny increases in risks. One common model assumption is linearity, which is assumed in virtually all of the models used to predict either acute/chronic illness/death associated with PM or decreased IQ associated with low-level lead exposure. Therefore, it should not be a surprise that these models suggest linear effects at low levels of exposure, given that they begin with that assumption.

Another unresolved limitation is the potential for residual bias. Epidemiology studies represent observational data, which have a much higher potential for bias than do experimental data. Multivariate statistical models attempt to correct for one type of bias by adjusting for potentially confounding risk factors. Historically, such adjustment has been adequate, because adjustment

removes enough confounding to allow moderately strong associations (e.g., $RR > 2$) to stand out. However, modern statistical techniques and enhanced computing power have permitted ever more complex statistical models that can measure weaker and weaker mathematical associations. Unfortunately, many researchers and regulators have continued to rely on routine statistical adjustment, even when interpreting increasingly weak associations.

A major problem with using traditional statistical adjustment when assessing vanishingly low risks is that most of the confounders included in multivariate models are only crude surrogates for more complex variables. For example, a simple variable such as years of maternal education may be used as a surrogate for the complexities associated with socioeconomic status. Similarly, crude weather variables, such as average daily ambient temperature measured at a single, remote location (e.g., an airport), have been used as estimates of the heat/humidity stress experienced by individuals in diverse urban microenvironments. Crudely measured covariates permit only imperfect adjustment for major risk factors and result in a certain degree of residual confounding. Furthermore, many minor risk factors are unmeasured and/or not included as covariates in complex statistical models. Only a relatively small amount of residual confounding would be needed to influence the very weak associations suggested by many recent investigations of low-level environmental exposures. In fact, as risks become smaller and smaller, it becomes increasingly difficult to tease out true associations from those related to the influence of other factors (Lumley and Sheppard, 2003). This implies the potential for a practical threshold, due to statistical limitations, in addition to a biological one.

On a more fundamental level, reliance on statistical adjustment to elucidate the true risk posed by low-dose exposures assumes that we can both know all the factors that influence subtle, multifactorial health effects, and that these factors can be accurately measured. For example, reliance on statistical coefficients to determine the subtle, adverse neurodevelopmental (e.g., IQ) impacts caused by low-level chemical exposures suggests that we can sufficiently explain complex human behavior and can predict lowered intelligence and future life success using only a crude estimate of environmental exposure and a few crudely measured covariates. If this were true, the disciplines of psychiatry, psychology, and sociology would be superfluous.

Another argument against undue reliance on statistics as an arbiter of low-dose health effects is the issue of multiple statistical comparisons. The power of modern computing and statistical modeling have allowed smaller and smaller associations to be detected, but have not changed the fundamental limitations inherent in the significance testing used to detect those associations. This testing assumes that one, and only one, test is being conducted, so that the p -value is a reflection of the probability of a chance

event. In actuality, all epidemiological investigations perform scores to hundreds of statistical tests, so that individual p -values are meaningless and do not reflect the true probability of finding the result by chance alone. This multiple-comparison problem is compounded when investigators rely on the data to guide them to the most important models (i.e., those that are most statistically significant) (Faraway, 1992; Chatfield, 1995).

Finally, statistical significance says nothing about the clinical significance of the results. Statistical models may indicate that each $\mu\text{g}/\text{dl}$ of lead exposure is significantly associated with a loss of 0.3 IQ points, but such a change has no measurable clinical impact on intelligence or social functioning (Kaufman, 2001). Similar concerns surround minor pulmonary function fluctuations, subtle changes in heart rate variability, and other trivial “effects” that are well within normal clinical variability for a population. This again suggests a practical threshold, even when a statistical one is not readily apparent.

Hormesis

The most fundamental assumption of low-dose linearity is that environmental exposures cannot be inherently beneficial. This ignores the growing body of evidence in support of hormesis, the notion of beneficial effects from small exposures to agents that are toxic at higher doses. This theory is gaining credibility as a possibly universal phenomenon. There are a considerable number of laboratory examples demonstrating chemical and radiation hormesis among microbial organisms, plants, and mammals (Calabrese, 2003a, 2003b). There are also clinical/practical examples in people, including the hormetic effects of alcohol, micronutrients, exercise, and calorie/fat intake. All are beneficial at low to moderate levels, but harmful in the extreme (Bukowski, 2000). Hormesis suggests that conservative regulatory decisions based on low-dose linearity may not just be incorrect, but may actually be detrimental to public health.

CONCLUSIONS

Complex statistical modeling results and speculation on extreme biological sensitivity do not provide sufficient evidence of low-dose linearity to overturn the traditional threshold paradigm for noncancer QRA. In fact, the empirical evidence cited earlier supports the traditional paradigm and suggests that nonlinear exposure–response relationships (i.e., thresholds and hormesis) may well be the norm at low doses. Therefore, an assumption of low-dose linearity does not appear to be appropriate for noncancer QRA. One could even argue that a realistic range of both human sensitivity to carcinogens and background repair of genetic damage go against low-dose linearity for most cancer risk assessment as well.

REFERENCES

- Angell M. The interpretation of epidemiologic studies. *New Eng J Med*. 1990;323:823–825.
- Bogdanffy MS, Daston G, Faustman EM et al. Harmonization of cancer and noncancer risk assessment: Proceedings of a consensus-building workshop. *Toxicological Sciences*. 2001;61:18–31.
- Bukowski JA, Lewis RJ. Hormesis and health: A little of what you fancy may be good for you. *Southern Medical Journal*. 2000;93:371–374.
- Calabrese EJ, Baldwin LA. Hormesis: The dose-response revolution. *Annu Rev Pharmacol Toxicol*. 2003a;43:175–197.
- Calabrese EJ, Baldwin LA. Toxicology rethinks its central belief. *Nature*. 2003b;421:691–692.
- Chatfield C. Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society (part A)*. 1995;158:419–466.
- Faraway JJ. The cost of data analysis. *Journal of Computational and Graphical Statistics*. 1992;1:213–229.
- Kaufman AS. 2001. Do low levels of lead produce IQ loss in children? A careful examination of the literature. *Archives of Clinical Neuropsychology*. 2001;16:303–341.
- Kimmel CA. 1999 Warkany Lecture: Improving the science for predicting risks to children's health. *Teratology*. 2001;63:202–209.
- Lumley T, Sheppard L. Time series analyses of air pollution and health: Straining at gnats and swallowing camels? *Epidemiology*. 2003;14:13–14.
- Merck Manual*, 17th ed. Rahway, NJ: Merck Research Laboratories; 1999.
- Palazzolo MJ. Decreasing life span of rats poses problems in labs. *Emphasis* (Corning Hazleton). 1995;6(3):1–5 (<http://my.execpc.com/~jwolf/ratlong.pdf>).
- Physicians Desk Reference* (PDR). 53rd ed. Montvale, NJ: Medical Economics Company; 1999.
- Shapiro S. Meta-analysis/Shmeta-analysis. *Am J Epidemiol*. 1994;140:771–778.
- USEPA (U.S. Environmental Protection Agency). *Exposure Factors Handbook*. Vol. 1, EPA/600/P-95/002Fa. Washington, DC: Office of Research and Development; 1997.
- USEPA (U.S. Environmental Protection Agency). *Second External Review Draft of the Particulate Matter Air Quality Criteria Document*. EPA/600/P-99/002bB. Washington, DC: Office of Research and Development; 2001.
- USEPA (U.S. Environmental Protection Agency). *Third External Review Draft of the Particulate Matter Air Quality Criteria Document*. EPA/600/P-99/002aC. Washington, DC: Office of Research and Development; 2002.
- USEPA (U.S. Environmental Protection Agency). *Integrated Risk Information System*. Lead and compounds (inorganic) (CASRN 7439-92-1). www.epa.gov/iris/subst/0277.htm, as referenced on April 28, 2003.